



UNIVERSITÀ
DEGLI STUDI
DI PADOVA

UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Studi Linguistici e Letterari

Dipartimento di Ingegneria dell'Informazione

Corso di Laurea Magistrale in Strategie di Comunicazione
Classe LM-92

**Sentiment analysis per il monitoraggio della reputazione
su Twitter delle compagnie aeree:
il periodo di crisi dell'estate 2022**

Relatore
Prof. Gianmaria Silvello

Correlatore
Prof. Giampietro Vecchiato

Laureanda
Giulia Pasqualini
n° matr. 2027420

Anno accademico 2022/2023

ABSTRACT

This experimental paper proposes a sentiment analysis of data from Twitter, with the aim of monitoring the corporate reputation of five airlines during the crisis period of summer 2022. The sentiment of users who express themselves on four airline services is detected through the analysis of a corpus containing the tweets users published. In addition, a corpus of tweets published by official airline accounts is created to verify if they treated certain topics with the same attention as users.

After providing a theoretical and methodological basis for the definition and measurement of reputation, the procedure used to analyze the data is explained, which involved the use of the Twitter API, the Python programming language, and the VADER (Valence Aware Dictionary and sEntiment Reasoner) module.

SOMMARIO

Il presente elaborato sperimentale propone una sentiment analysis dei dati provenienti da Twitter, con l'obiettivo di monitorare la reputazione aziendale di cinque compagnie aeree nel periodo di crisi dell'estate 2022. Il sentiment degli utenti che si esprimono su quattro servizi delle compagnie aeree viene rilevato attraverso l'analisi di un corpus contenente i tweet da essi pubblicati. Inoltre, viene costituito anche un corpus relativo ai tweet pubblicati dagli account ufficiali delle compagnie aeree per verificare se esse trattano determinati argomenti con la stessa attenzione degli utenti.

Dopo aver fornito una base teorica e metodologica per la definizione e la misurazione della reputazione, si illustra la procedura utilizzata per analizzare i dati, che ha previsto l'utilizzo delle Twitter API, del linguaggio di programmazione Python e del modulo VADER (Valence Aware Dictionary and sEntiment Reasoner).

INDICE

INDICE DELLE FIGURE.....	xi
INDICE DELLE TABELLE.....	xv
CAPITOLO 1 – INTRODUZIONE	1
CAPITOLO 2 – LA REPUTAZIONE AZIENDALE.....	7
2.1 La reputazione nella relazione con gli stakeholder.....	9
2.2 La reputazione nel Web 2.0.....	14
2.2.1 Una definizione di Web 2.0.....	15
2.2.2 Conversazioni e opinioni nel Web 2.0	18
2.2.3 Twitter: microblogging social network	23
2.2.4 Linee guida per preservare la reputazione online	26
2.3 Le crisi reputazionali	29
CAPITOLO 3 – SENTIMENT ANALYSIS PER IL MONITORAGGIO DELLA REPUTAZIONE SU TWITTER	35
3.1 Individuare un percorso di ricerca per monitorare la reputazione	35
3.1.1 Modelli per la misurazione della reputazione.....	39
3.1.2 Strumenti per il monitoraggio della reputazione online	46
3.2 Twitter come fonte di dati sull’opinione dei clienti	47
3.2.1 Monitorare l’attività del servizio clienti	49
3.2.2 Le API di Twitter e la piattaforma Twitter Developer	52

3.3 La sentiment analysis.....	56
3.3.1 Cos'è la sentiment analysis.....	57
3.3.2 Il corpus su cui effettuare la sentiment analysis	59
3.3.4 Punti di forza e di debolezza della sentiment analysis.....	62
3.3.5 Ricerche che utilizzano la sentiment analysis.....	64
3.4 Sentiment analysis con Python	65
3.4.1 Libreria standard di Python.....	69
3.4.2 Librerie open source per l'analisi dei dati	70
3.4.3 VADER	72
CAPITOLO 4 – CASO STUDIO: LE COMPAGNIE AEREE NEL PERIODO DI CRISI DELL'ESTATE 2022.....	77
4.1 L'oggetto della ricerca	77
4.1.1 Caratteristiche principali del settore del trasporto aereo	77
4.1.2 Il settore aereo tra pandemia di COVID-19 e ripresa nell'estate 2022.....	90
4.1.3 Crisi strisciante nell'estate 2022	94
4.1.4 Compagnie aeree selezionate.....	96
4.1.5 Periodo considerato.....	98
4.2 Le domande di ricerca.....	100
4.3 Le fasi della ricerca.....	102
4.3.1 Raccolta, salvataggio e organizzazione dei tweet.....	102
4.3.2 Visualizzazione dei dati non riguardanti la sentiment analysis	111

4.3.3 Pre-processing dei tweet (prima parte)	113
4.3.4 Sentiment analysis con VADER	114
4.3.5 Visualizzazione dei dati provenienti dalla sentiment analysis	116
4.3.6 Pre-processing dei tweet (seconda parte).....	118
4.3.7 Visualizzazione e analisi dei dati testuali.....	121
CAPITOLO 5 – RISULTATI DELL’ANALISI.....	123
5.1 Numero di tweet.....	123
5.1.1 Numero di tweet totale: compagnie vs utenti	123
5.1.2 Numero di tweet diviso per servizio: compagnie aeree vs utenti	126
5.1.3 Correlazione.....	128
5.2 Sentiment rispetto a tutto il periodo e a tutte le compagnie aeree	130
5.2.1 Percentuale di tweet negativi, positivi e neutri	130
5.2.2 Numero di tweet positivi, negativi e neutri diviso per servizio	132
5.2.3 Compound medio.....	133
5.3 Mese più critico	136
5.3.1 Compound medio per ogni mese	136
5.3.2 Risultati specifici per ogni compagnia.....	137
5.4 Elementi comuni tra le varie compagnie	161
CAPITOLO 6 – CONCLUSIONI	169
BIBLIOGRAFIA	177

RINGRAZIAMENTI 197

INDICE DELLE FIGURE

- Figura 1: grafico che mostra il numero di pubblicazioni sulla reputazione nel periodo considerato dagli autori (Veh et al. 2019: 316). 8
- Figura 2: differenza tra Web 1.0 e Web 2.0 (Dincer 2020). 18
- Figura 3: screenshot della prima parte dei testi presenti nel sito ufficiale di Twitter alla pagina About. 24
- Figura 4: screenshot della seconda parte di testi presenti nel sito ufficiale di Twitter alla pagina About. Tra questi si noti il tweet di Jack Dorsey, uno dei fondatori di Twitter. 24
- Figura 5: processo circolare della valutazione e misurazione della reputazione (Cuomo et al. 2012: 28). 36
- Figura 6: rappresentazione grafica del processo di traduzione empirica di un concetto complesso (Corbetta 2014: 111). 39
- Figura 7: esempio di traduzione empirica del concetto di reputazione. I riquadri colorati evidenziano uno dei possibili percorsi della ricerca. 39
- Figura 8: il Reputation Pulse Model (Unipol). 44
- Figura 9: comunicazione da parte di Twitter Dev sull'introduzione dei piani a pagamento. 54
- Figura 10: descrizione nel profilo Twitter di KLM. 89
- Figura 11: screenshot del testo di presentazione dei profili ufficiali delle compagnie aeree su Twitter. 89
- Figura 12: grafico che rappresenta l'evoluzione del numero di passeggeri nel corso del tempo. Traduzione del grafico pubblicato dall'Organizzazione internazionale dell'aviazione civile (ICAO 2023). 91
- Figura 13: grafico che rappresenta le compagnie aeree leader in Europa per numero di passeggeri. Titolo tradotto e compagnie aeree scelte evidenziate. (Statista Research Department 2022). 97
- Figura 14: volume delle ricerche relative agli scioperi delle compagnie aeree. 99
- Figura 15: volume delle ricerche relative ai ritardi delle compagnie aeree. 99
- Figura 16: volume delle ricerche relative alle cancellazioni delle compagnie aeree. 99

▪ Figura 17: scomposizione del concetto di reputazione in variabili empiriche rilevabili.....	101
▪ Figura 18: rappresentazione grafica delle fasi della ricerca.	102
▪ Figura 19: esempio di dataframe dei tweet degli utenti che si esprimono riguardo il servizio “baggage” di British Airways.	109
▪ Figura 20: organizzazione del corpus dei tweet delle compagnie aeree.....	109
▪ Figura 21: organizzazione del corpus dei tweet degli utenti.	110
▪ Figura 22: numero di tweet pubblicati dagli account ufficiali di ogni compagnia aerea.....	124
▪ Figura 23: numero di tweet pubblicati dagli utenti relativamente a ogni compagnia aerea.	124
▪ Figura 24: esempio di descrizione in cui la compagnia aerea invita a scrivere messaggi diretti (DM).....	125
▪ Figura 25: numero di tweet pubblicati dagli account ufficiali delle compagnie aeree relativamente ai propri servizi.....	127
▪ Figura 26: numero di tweet pubblicati dagli utenti relativamente ai servizi di ogni compagnia aerea.	127
▪ Figura 27: percentuale di tweet positivi, negativi e neutri presenti nel corpus 2.	130
▪ Figura 28: percentuale di tweet positivi, negativi e neutri presenti nel sub-corpus di tweet relativi a British Airways.....	130
▪ Figura 29: percentuale di tweet positivi, negativi e neutri presenti nel sub-corpus di tweet relativi a EasyJet.	131
▪ Figura 30: percentuale di tweet positivi, negativi e neutri presenti nel sub-corpus di tweet relativi a KLM.	131
▪ Figura 31: percentuale di tweet positivi, negativi e neutri presenti nel sub-corpus di tweet relativi a Lufthansa.	131
▪ Figura 32: percentuale di tweet positivi, negativi e neutri presenti nel sub-corpus di tweet relativi a Ryanair.....	131
▪ Figura 33: numero di tweet con sentiment positivo, negativo e neutro presenti nel corpus 2 e raggruppati per servizio.....	133

- Figura 34: compound medio per i tweet di ogni compagnia aerea nell'intero arco temporale considerato..... 134
- Figura 35: variazione del compound medio dei tweet degli utenti relativi a British Airways pubblicati dal 23 maggio 2022 al 30 giugno 2022..... 138
- Figura 36: valore medio del compound diviso per servizio, relativo ai tweet pubblicati dal 23 maggio al 30 giugno 2022 riferiti a British Airways. 139
- Figura 37: grafico a barre orizzontali che mostra i bi-gram più frequenti presenti nei tweet negativi riferiti al servizio “baggage” di British Airways e pubblicati tra il 23 maggio e il 30 giugno 2022..... 139
- Figura 38: word cloud relativa ai tweet negativi del servizio “baggage” di British Airways nel periodo che va dal 23 maggio al 30 giugno 2022. 140
- Figura 39: word cloud relativa ai soli tweet negativi riferiti a British Airways pubblicati nella giornata del 31 maggio 2022. 141
- Figura 40: variazione del compound medio dei tweet degli utenti relativi a EasyJet pubblicati dal 23 maggio 2022 al 30 giugno 2022. 143
- Figura 41: valore medio del compound diviso per servizio, relativo ai tweet pubblicati dal 23 maggio al 30 giugno 2022 riferiti a EasyJet. 144
- Figura 42: grafico a barre orizzontali che mostra i bi-gram più frequenti presenti nei tweet negativi riferiti al servizio “flight” di EasyJet, pubblicati tra il 23 maggio e il 30 giugno 2022. 145
- Figura 43: word cloud relativa ai tweet negativi sul servizio “flight” di EasyJet nel periodo che va dal 23 maggio al 30 giugno 2022..... 145
- Figura 44: word cloud relativa ai soli tweet negativi riferiti a EasyJet pubblicati nella giornata del 27 giugno 2022. 146
- Figura 45: variazione del compound medio dei tweet degli utenti relativi a KLM pubblicati dal primo luglio al 31 luglio 2022. 148
- Figura 46: valore medio del compound diviso per servizio, relativo ai tweet pubblicati dal primo luglio al 31 luglio 2022. 149
- Figura 47: grafico a barre orizzontali che mostra i bi-gram più frequenti presenti nei tweet negativi riferiti al servizio “baggage” di KLM pubblicati tra il primo luglio e il 31 luglio 2022. 149

- Figura 48: word cloud relativa ai tweet negativi sul servizio “baggage” di KLM nel mese di luglio 2022..... 150
- Figura 49: word cloud relativa ai soli tweet negativi riferiti a KLM, pubblicati il 4 luglio 2022..... 151
- Figura 50: variazione del compound medio dei tweet degli utenti relativi a Lufthansa pubblicati dal primo luglio al 31 luglio 2022. 152
- Figura 51: valore medio del compound diviso per servizio, relativo ai tweet pubblicati su Lufthansa dal primo luglio al 31 luglio 2022..... 153
- Figura 52: grafico a barre orizzontali che mostra i bi-gram più frequenti presenti nei tweet negativi riferiti al servizio “baggage” di Lufthansa, pubblicati tra il primo luglio e il 31 luglio 2022..... 154
- Figura 53: word cloud relativa ai tweet negativi sul servizio “baggage” di Lufthansa nel mese di luglio 2022..... 154
- Figura 54: word cloud relativa ai soli tweet negativi riferiti a Lufthansa, pubblicati il 10 luglio 2022..... 155
- Figura 55: variazione del compound medio giornaliero dei tweet degli utenti relativi a Ryanair pubblicati dal 1° settembre al 1° ottobre 2022..... 157
- Figura 56: valore medio del compound diviso per servizio, relativo ai tweet pubblicati su Ryanair dal primo settembre al 30 settembre 2022..... 158
- Figura 57: grafico a barre orizzontali che mostra i bi-gram più frequenti presenti nei tweet negativi riferiti al servizio “flight” di Ryanair pubblicati tra il 1° settembre e il 30 settembre 2022. 159
- Figura 58: word cloud relativi ai tweet negativi sul servizio “flight” di Ryanair nel mese di settembre. 160
- Figura 59: word cloud relativa ai soli tweet negativi riferiti a Ryanair, pubblicati il 27 settembre 2022. 160
- Figura 60: tweet di British Airways. 164
- Figura 61: parole più frequenti in comune ai tweet negativi di British Airways, KLM e Lufthansa relativi al servizio “baggage”..... 166
- Figura 62: parole più frequenti in comune ai tweet negativi di EasyJet e Ryanair riguardanti il servizio “flight”..... 166

INDICE DELLE TABELLE

- Tabella 1: selezione di alcune definizioni di reputazione aziendale (Cuomo et al. 2012; Fombrun & Van Riel 1997; corsivo mio). 9
- Tabella 2: differenza tra WOM ed E-WOM. Tabella tradotta e adattata da Huete-Alcocer (2017). 20
- Tabella 3: strategie di risposta alla crisi secondo la *Image Repair Theory* (Benoit 1997 adattato da Splendiani 2022). 32
- Tabella 4: dimensioni individuate da Fombrun et al. (2000), rielaborate da Cuomo et al. (2012). 42
- Tabella 5: dimensioni e indicatori del modello Rep Trak Pulse di Rep Track (Reputation Institute 2009 adattato da Cuomo et al. 2012: 41). 44
- Tabella 6: differenze tra le tre modalità di accesso alle API di Twitter, descritte nel sito ufficiale Twitter Developer. Vengono qui riportate le differenze rilevanti per questa ricerca. 54
- Tabella 7: descrizione dei moduli della libreria standard di Python utili per l'analisi dei tweet. 70
- Tabella 8: pacchetti di NLTK. 72
- Tabella 9: soglie del compound. 73
- Tabella 10: esempio del cambiamento di polarità in due frasi con e senza punteggiatura processate nel Jupyter Notebook. 75
- Tabella 11: esempio del cambiamento di polarità in due frasi con e senza caratteri maiuscoli processate nel Jupyter Notebook. 75
- Tabella 12: esempio del cambiamento di polarità in due frasi con e senza avverbio processate nel Jupyter Notebook. 75
- Tabella 13: esempio del cambiamento di polarità in due frasi con e senza congiunzione e negazione processate nel Jupyter Notebook. 75
- Tabella 14: differenze tra compagnie aeree di bandiera e compagnie aeree low cost (Whyte & Lohmann 2017: 110-112; Wittmer et al. 2011: 72-73; traduzione mia; Bertolin et al. 2018). 82
- Tabella 15: esempio di input inserito per una singola richiesta alle API. 104

- Tabella 16: parole chiave inserite per scaricare i tweet pubblicati dalle compagnie aeree relative ai vari servizi. 105
- Tabella 17: keyword inserite per scaricare i tweet degli utenti relativi alle varie compagnie aeree e ai relativi servizi..... 106
- Tabella 18: sottoperiodi (di 4 giorni ciascuno) considerati per lo scaricamento dei tweet..... 107
- Tabella 19: numero di tweet pubblicati dagli account ufficiali delle compagnie aeree diviso per servizio (Corpus 1). 111
- Tabella 20: numero di tweet pubblicati dagli utenti relativamente a ogni compagnia aerea e ai vari servizi (Corpus 2). 111
- Tabella 21: codice per realizzare un grafico a barre che mostra i tweet raggruppati per compagnia. 112
- Tabella 22: codice per realizzare un grafico a barre che mostra i tweet raggruppati per servizio e divisi per compagnia..... 112
- Tabella 23: codice utilizzato per effettuare una correlazione..... 112
- Tabella 24: codice utilizzato per trovare le menzioni all'interno dei tweet. 113
- Tabella 25: esempio di codice utilizzato per rimuovere gli hashtag. 113
- Tabella 26: esempio di codice utilizzato per rimuovere le menzioni. 113
- Tabella 27: esempio di codice utilizzato per rimuovere i doppi spazi. 113
- Tabella 28: esempio di codice utilizzato per rimuovere gli URL..... 114
- Tabella 29: esempio di tweet sottoposto alla sentiment analysis e relativi risultati del compound e delle singole percentuali di positività, neutralità e negatività. 115
- Tabella 30: esempio di codice utilizzato per mostrare i dati sul sentiment con un grafico a torta..... 116
- Tabella 31: esempio di codice utilizzato per mostrare i dati sul sentiment con un grafico a barre..... 117
- Tabella 32: esempio di codice utilizzato confrontare i valori del compound delle cinque compagnie aeree..... 117
- Tabella 33: esempio di codice utilizzato per individuare il valore medio del compound di ogni mese..... 117
- Tabella 34: esempio di codice utilizzato per visualizzare la variazione del compound nel corso di un intervallo di tempo specifico..... 118

- Tabella 35: esempio di codice utilizzato per individuare i 10 tweet con il
compound più basso in un dataframe. 118
- Tabella 36: codice utilizzato per rimuovere la punteggiatura dal testo. 119
- Tabella 37: codice utilizzato per rimuovere i caratteri maiuscoli dal testo. 119
- Tabella 38: esempio di applicazione della funzione a un tweet in cui sono
presenti delle emoji..... 120
- Tabella 39: codice utilizzato per rimuovere le stopwords dal testo. 120
- Tabella 40: codice utilizzato per dividere il testo dei tweet in token. 120
- Tabella 41: codice utilizzato per la lemmatizzazione dei verbi..... 121
- Tabella 42: codice utilizzato per la creazione di una word cloud. 122
- Tabella 43: dettaglio del corpus 2 che mostra il numero di tweet totali e il
numero di tweet contenenti una menzione della compagnia aerea. 126
- Tabella 44: valore della correlazione tra il numero di tweet degli utenti e il
numero di tweet degli account ufficiali per ogni compagnia aerea. 129
- Tabella 45: valore medio del compound di ogni compagnia relativo ad ogni
mese. 136
- Tabella 46: sintesi della sentiment analysis effettuata sulle cinque compagnie
aeree..... 162

CAPITOLO 1 – INTRODUZIONE

Non ci sono segreti.

Il mercato online conosce i propri prodotti meglio delle aziende stesse.

E diffonde a tutti la propria opinione, buona o cattiva che sia.

(Tesi 12 del *Cluetrain Manifesto*)

Il Web 2.0¹ è caratterizzato da servizi di comunicazione bidirezionale, dall'*open culture* e dai social network che creano una struttura in cui gli utenti (tra cui i consumatori) diventano partecipanti attivi, liberi di dialogare, creare e diffondere contenuti. Il *Cluetrain Manifesto*, pubblicato nel 1999 da Levine et al., proponeva 95 tesi che invitavano le aziende a cambiare il proprio modello comunicativo unidirezionale, in favore di uno più adatto al nuovo mercato interconnesso.

Un'azienda, infatti, secondo la prospettiva delle relazioni pubbliche e della comunicazione d'impresa, non è un'entità che agisce in un contesto isolato, bensì è inserita in un macrosistema in cui sviluppa una relazione interdipendente con una molteplicità di interlocutori (pubblici influenti e stakeholder).

In tale contesto, assume un ruolo di estrema importanza la reputazione, un costrutto che in letteratura viene definito di volta in volta in maniera differente a seconda dei contesti d'applicazione. Questa ricerca si situa in un ambito interdisciplinare per cui risultano utili le definizioni fornite dagli studi di sociologia, comunicazione d'impresa e relazioni pubbliche e che, ad esempio, definiscono la reputazione come:

una rappresentazione collettiva delle azioni e dei risultati passati di un'azienda che descrive la capacità dell'azienda di fornire risultati di valore a una varietà di stakeholder, sia nel suo ambiente competitivo che in quello istituzionale (Fombrun & Rindova 1996, citati in Fombrun & Van Riel 1997).

Una buona reputazione rientra tra gli asset intangibili di un'azienda: pur non essendo un elemento fisico, è capace di generare valore perché permette di attrarre nuovi consumatori, praticare prezzi più alti, aumentare il valore azionario, beneficiare di una

¹ Termine reso popolare da Tim O'Reilly nel 2007.

copertura mediatica positiva, instaurare una relazione proficua con gli interlocutori e, nel caso di situazioni di crisi, essere perdonati più facilmente.

Per coordinare le attività relazionali con gli stakeholder e i pubblici influenti, un ruolo fondamentale è ricoperto dalle attività di comunicazione e, in particolare, dai social network. Le aziende dovrebbero ambire ad instaurare un processo comunicativo di tipo simmetrico bidirezionale, in cui si ascolta in modo sistematico e organizzato l'opinione dei clienti. Ciò è di fondamentale importanza per individuare le loro esigenze, raggiungere una comprensione reciproca e monitorare la reputazione aziendale: tramite il passaparola reso possibile dal Web 2.0 (conosciuto in letteratura come *electronic word of mouth*), qualsiasi opinione e dichiarazione fatta dai clienti può raggiungere una moltitudine di persone e influenzare la reputazione dell'azienda, che si trova così in una situazione più vulnerabile rispetto al passato. È necessaria quindi la presenza di figure aziendali che si occupino quotidianamente di monitorare le conversazioni online, sia su media *owned* che *not owned*, con l'obiettivo di mettere in atto azioni correttive prima che le notizie negative arrivino a un pubblico più ampio.

Un luogo virtuale che dà spazio alla libera circolazione di idee e opinioni su scala globale è Twitter, social network nato nel 2006, che consente agli utenti di pubblicare messaggi brevi e immediati (i *tweet*) e che porta alla generazione di una grande quantità di dati testuali. Questi ultimi possono costituire del materiale empirico che è possibile scaricare tramite le API (*application programming interface*) di Twitter e analizzare nell'ambito del *Natural Language Processing* (NLP), un insieme di tecniche computazionali per l'analisi automatica e la rappresentazione del linguaggio naturale. Nello specifico, il testo presente nei tweet può essere analizzato tramite la *sentiment analysis* (chiamata anche *opinion mining*), una tecnica automatica che permette di rilevare le opinioni contenute in un testo e classificarle come positive, negative o neutre.

Questo elaborato sperimentale presenta un caso studio in cui la sentiment analysis viene utilizzata per monitorare la reputazione su Twitter di cinque compagnie aeree² (British Airways, EasyJet, KLM, Lufthansa, Ryanair) nell'estate del 2022, periodo che

² Per avere un'idea del contesto competitivo, si sono considerate alcune delle maggiori compagnie aeree per numero di passeggeri trasportati. Avere una panoramica della reputazione dell'intero settore è di fondamentale importanza: la reputazione collettiva di un settore può infatti influenzare la reputazione di una singola azienda operante in tale settore. Pertanto, il mantenimento e il miglioramento della reputazione diventa una questione che coinvolge tutte le compagnie aeree nel loro complesso.

può essere considerato una *crisi strisciante*. Dopo due anni dall'inizio della pandemia, nell'estate del 2022 il settore del trasporto aereo ha visto un notevole aumento della domanda, con un livello di traffico aereo pari all'83% di quello toccato nel 2019, secondo i dati forniti da Eurocontrol. Nonostante ci fossero stati dei *segnali deboli* che preannunciavano la ripresa dei viaggi estivi, le compagnie aeree non sono state in grado di organizzarsi in tempo per reintegrare il personale licenziato dallo scoppio della pandemia e i dipendenti rimasti in servizio hanno dovuto accettare condizioni lavorative inadeguate. Queste ultime sono state il motivo scatenante dei numerosi scioperi che hanno colpito le compagnie aeree nel corso dell'estate. La combinazione della mancanza di personale, dell'aumento dei costi (derivante dalla guerra tra Russia e Ucraina) e degli scioperi dei dipendenti ha causato la cancellazione, il ritardo e la riprogrammazione dei voli delle compagnie aeree. Questa situazione ha rappresentato una significativa perdita economica per le compagnie aeree e ha impattato negativamente sull'esperienza dei clienti finali.

L'interesse per questo caso studio specifico è nato dopo aver notato l'attenzione mediatica che stavano ricevendo i problemi del settore del trasporto aereo nell'estate del 2022 e aver intercettato il passaparola offline di viaggiatori che avevano avuto un'esperienza negativa dei servizi delle compagnie aeree. Si sono volute quindi analizzare le opinioni nel contesto online per capire in che modo queste influenzassero la reputazione delle singole compagnie aeree e dell'intero settore.

La ricerca parte dall'ipotesi che su Twitter sia possibile rilevare l'opinione degli utenti che si esprimono sulle compagnie aeree e quindi poter monitorare la reputazione aziendale: le compagnie aeree utilizzano la piattaforma per il proprio servizio clienti e i viaggiatori, a loro volta, hanno la possibilità di fare delle segnalazioni in tempo reale e, nel migliore dei casi, ricevere un aiuto tempestivo.

La reputazione è un concetto che si può sviluppare su più piani di analisi: in questa ricerca, si considera una dimensione specifica del costrutto reputazionale (opinione su determinati servizi delle compagnie aeree) e un gruppo potenzialmente eterogeneo di utenti (su Twitter gli utenti possono appartenere a diverse classi di stakeholder). Una buona reputazione rappresenta un vantaggio economico rispetto alle aziende concorrenti ed è quindi fondamentale riuscire a misurarne l'impatto.

L'obiettivo è monitorare la reputazione su Twitter di cinque compagnie aeree, identificando le eventuali opinioni negative degli utenti riguardo quattro servizi specifici (prenotazione, check-in, bagaglio, volo) e valutando come queste possano influire sulla reputazione aziendale. Il periodo temporale considerato è quello compreso tra il 23 maggio 2022 e il 2 ottobre 2022 che, secondo quanto mostrato da Google Trends, ha visto un alto volume di ricerche riguardante i disagi che stavano caratterizzando il settore del trasporto aereo (scioperi, cancellazioni, ritardi).

Sono stati creati due differenti corpora, contenenti dei tweet in lingua inglese:

- Corpus 1: insieme dei tweet pubblicati dagli account ufficiali delle compagnie aeree tra il 23 maggio e il 2 ottobre 2022 relativamente ai quattro servizi considerati.
- Corpus 2: insieme dei tweet pubblicati dagli utenti tra il 23 maggio e il 2 ottobre 2022 relativamente ai quattro servizi delle compagnie aeree.

Il focus della ricerca è il monitoraggio della reputazione, effettuato tramite una sentiment analysis del secondo corpus, ma vengono presi in considerazione anche i tweet pubblicati dalle compagnie aeree nei propri account ufficiali per verificare se esse trattano determinati argomenti con la stessa attenzione degli utenti.

Nello specifico, le domande che hanno guidato la ricerca sono:

- D1: quali dati emergono dai tweet pubblicati dagli account ufficiali delle compagnie aeree (corpus 1)?
- D2: qual è la reputazione delle compagnie aeree su Twitter? O meglio, quali dati emergono dai tweet che gli utenti pubblicano sulle compagnie aeree (corpus 2)?
- D3: può essere individuata una relazione tra i dati provenienti dai tweet pubblicati dalle compagnie aeree (corpus 1) e quelli provenienti dai tweet pubblicati dagli utenti (corpus 2)

L'elaborato si articola come segue:

Il capitolo 2 fornisce una base teorica e presenta il concetto di reputazione, evidenziando le difficoltà nella sua definizione e circoscrivendolo al contesto della relazione dell'azienda con gli stakeholder, del Web 2.0 e delle situazioni di crisi.

Il capitolo 3 presenta la ricerca dal punto di vista metodologico e mostra come il concetto astratto di reputazione possa essere trasformato in una variabile empiricamente rilevabile, fornendo degli esempi di modelli presenti in letteratura. Si evidenzia come, non esistendo tra gli studiosi una definizione universalmente condivisa di reputazione, non ci sia, di conseguenza, un metodo univoco per la sua misurazione. Viene poi illustrato l'utilizzo delle Twitter API e della piattaforma Twitter Developer per accedere all'intero archivio di tweet pubblicati sulla piattaforma e scaricare i dati desiderati per l'analisi. Infine, viene descritta la tecnica della sentiment analysis, i criteri per la costituzione di un corpus, il linguaggio di programmazione Python, le relative librerie e il modulo VADER (Valence Aware Dictionary and sEntiment Reasoner) utilizzati per svolgere l'analisi.

Il capitolo 4 descrive il caso studio specifico delle compagnie aeree nel periodo di crisi dell'estate 2022. Vengono quindi presentati nel dettaglio l'oggetto della ricerca, le domande di ricerca e le diverse fasi in cui si è sviluppata la raccolta, l'analisi e la presentazione grafica dei dati provenienti da Twitter.

Il capitolo 5, attraverso l'ausilio di grafici e tabelle, mostra i risultati dell'analisi dei due corpora. Viene fornita una panoramica della situazione delle compagnie aeree sia dal punto di vista degli account Twitter ufficiali delle stesse, sia dal punto di vista degli utenti. I risultati relativi al sentiment degli utenti sono presentati considerando sia l'arena competitiva che comprende tutte le compagnie aeree nell'intero arco temporale dei tweet scaricati, sia il contesto caratteristico di ogni singola compagnia aerea in un arco temporale più ristretto. La ricerca si conclude con il capitolo 6, dove vengono commentati i risultati dell'analisi e fornite le risposte alle domande di ricerca.

CAPITOLO 2 – LA REPUTAZIONE AZIENDALE

Il modo per ottenere una buona reputazione sta nell'agire per essere ciò che desideri apparire.

Socrate

Dare una definizione puntuale e univoca di reputazione aziendale (chiamata anche con il termine inglese *corporate reputation*) può risultare complicato e ciò è testimoniato dall'operato di riviste specializzate (Corporate Reputation Review³, Reputation Review⁴), centri di ricerca (Oxford University Centre for Corporate Reputation⁵, Ipsos⁶), società (RepTrak⁷, Reputation Science⁸) e autori che, a seconda dei casi, hanno cercato di individuare e sintetizzare le definizioni esistenti (Barnett et al. 2006; Cuomo et al. 2012) o riflettere su tale costrutto in un'ottica storica, prendendo in considerazione contesti specifici di applicazione come, ad esempio, quello commerciale, bancario, economico, legislativo e professionale (Olegario & McKenna 2013).

Una panoramica generale sugli studi in ambito di reputazione aziendale viene fornita da Veh et al. (2019) che hanno revisionato, tramite un'analisi bibliometrica⁹, più di cinquemila pubblicazioni sulla reputazione aziendale pubblicate tra il 1969 e il 2015. Dalla loro ricerca è emerso che, a partire dagli anni Novanta, c'è stata una rapida crescita nella produzione scientifica riguardo questa tematica (figura 1) e che, a seconda del campo disciplinare in cui si inseriva una ricerca, la reputazione aziendale trovava una concettualizzazione di volta in volta differente.

³ <http://www.palgrave.com/gp/journal/41299> [ultima consultazione: 15/02/23]

⁴ <https://www.reputationreview.org/> [ultima consultazione: 15/02/23]

⁵ <https://www.sbs.ox.ac.uk/research/centres-and-initiatives/oxford-university-centre-corporate-reputation> [ultima consultazione: 15/02/23]

⁶ <https://www.ipsos.com/en/corporate-reputation>

⁷ <https://www.reptrak.com/> [ultima consultazione: 20/04/23]

⁸ <https://www.reputationscience.it/> [ultima consultazione: 20/04/23]

⁹ Un'analisi bibliometrica consiste “nell'applicare metodi statistici per determinare i cambiamenti qualitativi e quantitativi in un determinato argomento di ricerca scientifica, stabilire il profilo delle pubblicazioni sull'argomento e rilevare le tendenze all'interno di una disciplina.” (Rey-Martí et al. 2016, traduzione mia)

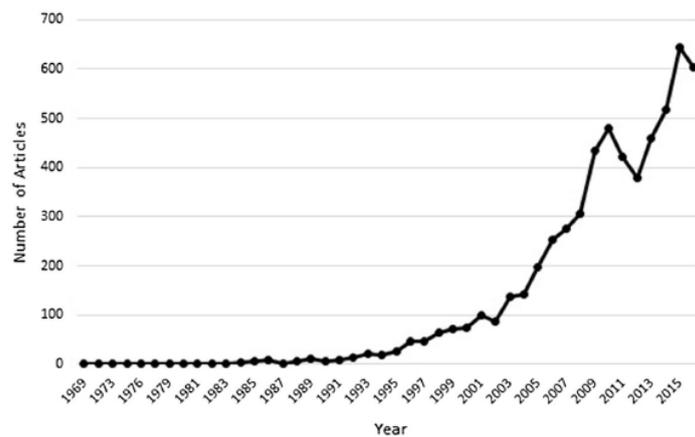


Figura 1: grafico che mostra il numero di pubblicazioni sulla reputazione nel periodo considerato dagli autori (Veh et al. 2019: 316).

Diversi autori (Fombrun & Van Riel 1997; Cuomo et al. 2012) confermano l'esistenza di una molteplicità di definizioni di reputazione aziendale che variano a seconda delle discipline. Questa ricerca si situa in un ambito interdisciplinare per cui risultano utili le seguenti definizioni:

DISCIPLINA	DEFINIZIONE La reputazione aziendale è:
SOCIOLOGIA	Un indicatore di legittimità che racconta dell'operato di un'organizzazione in <i>relazione</i> alle aspettative e alle norme della società. (Cuomo et al. 2012: 18)
CORPORATE COMMUNICATION	L'insieme delle immagini, delle <i>opinioni</i> e delle percezioni sviluppate nel tempo da clienti, fornitori e dal pubblico in generale nei confronti dell'organizzazione. (ibid.)
RELAZIONI PUBBLICHE	Una rappresentazione <i>collettiva</i> delle azioni e dei risultati passati di un'azienda che descrive la capacità dell'azienda di fornire risultati di valore a una <i>varietà di stakeholder</i> , sia nel suo ambiente <i>competitivo</i> che in quello istituzionale. (Fombrun & Rindova 1996 citati in Fombrun & Van Riel 1997, traduzione mia)

MARKETING	<p>L'insieme dei significati cognitivi e <i>affettivi</i> attribuiti dai consumatori ai <i>prodotti</i> con cui sono entrati in contatto.</p> <p>(Cuomo et al. 2012: 18)</p>
-----------	--

Tabella 1: selezione di alcune definizioni di reputazione aziendale (Cuomo et al. 2012; Fombrun & Van Riel 1997; corsivo mio).

Secondo le definizioni sopracitate, la reputazione risulta strettamente connessa alle relazioni sociali e alle opinioni¹⁰ che una varietà di stakeholder ha nei confronti di un'organizzazione in un contesto competitivo. In questo elaborato il concetto di reputazione viene circoscritto al contesto della relazione dell'azienda con gli stakeholder, del web 2.0 e delle situazioni di crisi.

2.1 La reputazione nella relazione con gli stakeholder

Come afferma John W. Hunt (citato in Citarella et al. 2010: 12):

Un'organizzazione è un'entità sociale identificabile che persegue obiettivi multipli attraverso attività coordinate e relazioni fra partecipanti ed oggetti. Un tale sistema sociale è aperto e dipende per la sua sopravvivenza da altri individui e sottosistemi all'interno di un'entità più ampia: la società.

Secondo la prospettiva delle relazioni pubbliche e della comunicazione d'impresa, un'azienda non è un'entità che agisce in un contesto isolato, bensì fa parte di quello che Vecchiato (2015: 26) definisce “sistema aperto”: essa è inserita in un macrosistema e sviluppa una relazione interdipendente con diversi interlocutori, dai quali viene influenzata e sui quali, a sua volta, esercita un'influenza.

Gli interlocutori con cui un'azienda si relaziona possono essere dei *pubblici influenti* o dei *portatori di interesse* (o *stakeholder*). Nel primo caso, si tratta di “soggetti dotati di poteri decisionali ritenuti rilevanti [dalla stessa azienda] per il raggiungimento degli obiettivi oppure perché ritenuti influenti sui primi” (Vecchiato 2004: 27). I *portatori di*

¹⁰ Anche considerando l'etimologia della parola, si trova una connessione tra reputazione e opinione: il vocabolario etimologico di Pianigiani (1907) indica che “reputare” deriva dal latino “reputare” che ha il tema in “pūtare”, ovvero tagliare a una pianta i rami inutili e, metaforicamente, “opinare, pensare”. Il prefisso re- indica ripetizione e quindi abitudine.

interesse (a cui ci si riferisce anche con il termine inglese *stakeholder*¹¹), invece, si attribuiscono autonomamente un titolo per relazionarsi con l'azienda (*ivi*, p.28) e, secondo la definizione di Pastore e Vernuccio (2008), possono essere suddivisi in stakeholder primari e secondari. Gli stakeholder primari influenzano direttamente l'azienda e sono fondamentali per far progredire le attività aziendali. Esempi di stakeholder primari sono (*ivi*, p. 7):

- i proprietari (soci e azionisti);
- i dirigenti;
- i finanziatori (creditori e obbligazionisti);
- i dipendenti;
- i fornitori (di materie e servizi);
- i concorrenti;
- i business partner (distributori o partner orizzontali¹²);
- i clienti (intermedi e finali).

Gli stakeholder secondari, invece, non influenzano direttamente l'azienda e non sono indispensabili per la sua sopravvivenza (*ibid.*). Tra gli stakeholder secondari si trovano:

- le istituzioni (che siano politiche economiche, finanziarie, nazionali, locali);
- la comunità locale;
- il pubblico generico (media);
- i gruppi di rappresentanza e di pressione (tra cui i sindacati, le associazioni ambientaliste, di categoria o dei consumatori, le lobby).

In questo contesto, in cui l'azienda si trova a relazionarsi con una molteplicità di interlocutori:

La reputazione aziendale è l'insieme di tutte le immagini, le aspettative, le percezioni e le opinioni di un'organizzazione sviluppate nel tempo dai suoi clienti, dipendenti, fornitori, investitori e dal pubblico in generale in relazione alle qualità, alle caratteristiche e al

¹¹ *Stakeholder* è una parola composta formata da *to hold*, “tenere”, e *stake*, “interesse”. La traduzione che viene data dal dizionario è “persona che ha un interesse in una ditta, un'organizzazione o simili” (Collins Dictionary).

¹² Una partnership orizzontale è una “partnership fra aziende dello stesso livello che, operando in settori diversi, mettono a disposizione competenze e segmenti di clientela” (Inside Marketing).

comportamento dell'organizzazione, sulla base dell'esperienza personale, dicerie o azioni passate osservate dall'organizzazione. (Bennet & Kottasz 2000: 234, traduzione mia).

Il costrutto “reputazione” sintetizza il risultato dell'efficacia nella gestione delle relazioni con i diversi portatori di interesse. (Cuomo et al. 2012: 10)

È utile chiarire la differenza tra i concetti di *reputazione*, *identità* e *immagine*. Poma e Grandoni (2021: 33-34) spiegano il significato di questi termini che spesso vengono confusi o usati come sinonimi. L'*identità* è l'insieme degli elementi materiali e immateriali che caratterizzano l'azienda nel profondo, ovvero “la sua personalità, la vision, la mission, i valori guida e i comportamenti dei membri”; l'*immagine*¹³, invece, è “il riflesso dell'identità dell'organizzazione così come appare agli occhi dei suoi pubblici”. La *reputazione* si posiziona a metà strada fra questi due concetti e corrisponde al “grado di allineamento tra immagine e identità”. Il livello di reputazione, quindi, è tanto più alto quanto più si avvicina all'identità aziendale: se l'immagine trasmessa ai pubblici è costruita, poco trasparente e non aderente alla realtà si perde di autenticità e si rischia di incorrere in una crisi reputazionale (*ibid.*).

Inoltre, è bene ricordare che la reputazione di un'azienda è influenzata anche dall'operato dei concorrenti che, non a caso, rientrano tra gli stakeholder primari. La reputazione di un intero settore merceologico ha un “effetto deformante” sul giudizio e sulle aspettative che un nuovo cliente ha rispetto a un'azienda. Questo effetto si può manifestare in due forme (Cuomo et al. 2012: 11-12):

- *amplificante*: se un settore gode, in generale, di una buona reputazione, questa influenzerà positivamente anche la reputazione della singola azienda che fa parte di quel settore.
- *riducente*: se un settore registra risultati e performance negativi, questi influenzeranno negativamente anche la reputazione di un'azienda di quel settore che, a differenza della maggioranza, registra risultati e performance positivi.

¹³ Invernizzi e Romenti (2020) precisano che, soprattutto a partire dagli anni Novanta, molti professionisti di relazioni pubbliche hanno invitato ad abbandonare il termine *immagine* per riferirsi agli obiettivi della comunicazione d'impresa. Questo termine, infatti, nel tempo ha assunto una connotazione negativa perché viene associato a qualcosa di superficiale, illusorio e opposto alla realtà, facilmente manipolabile nel breve termine.

Quanto più un interlocutore è estraneo a un'azienda tanto più egli si farà guidare dalla reputazione che caratterizza il settore a cui appartiene data azienda (*ibid.*): un esempio (riportato da Poma & Grandoni 2021) può essere il settore dell'estrazione e della lavorazione del marmo di Carrara in Italia che con la sua reputazione influenza positivamente il giudizio e le aspettative dello stakeholder che si avvicina per la prima volta a un'azienda che fa parte di quel settore. L'effetto deformante ha perciò un'influenza limitata alle prime interazioni tra stakeholder e azienda.

Avere una buona reputazione rientra tra gli asset¹⁴ intangibili di un'azienda: pur non essendo un elemento fisico (come gli asset tangibili) è capace di generare valore in quanto, come evidenziano Poma e Grandoni (2021: 36), permette di:

- attrarre nuovi consumatori, personale qualificato e investitori;
- praticare prezzi più alti;
- aumentare il valore azionario;
- registrare meno pressioni da parte dei gruppi di rappresentanza;
- beneficiare di una copertura mediatica positiva;
- instaurare una relazione proficua con gli interlocutori e con la comunità in cui si opera;
- nel caso di crisi o eventi critici, essere perdonati più facilmente dagli stakeholder.

La generazione di valore da parte dell'azienda è talmente connessa al governo della relazione con gli stakeholder che Costabile (2001 citato in Vecchiato 2015: 36) parla di *capitale relazionale*, espressione che indica “lo stock di fiducia, fedeltà e lealtà che deve essere accumulato dall'impresa per poter accrescere la propria capacità competitiva e il proprio valore di mercato nel tempo”. A questo proposito, Vecchiato (2015) menziona una ricerca condotta dalla rivista Fortune nel 2001 (Y. Kim, *Measuring the economic value of Public Relations*) che ha evidenziato lo stretto legame causale tra la spesa nelle attività di relazioni pubbliche e il livello di reputazione e tra il livello di reputazione e i risultati economici, concludendo quindi che “le relazioni pubbliche non sono un costo ma

¹⁴ Un *asset* è un bene che l'azienda possiede: esso può essere *tangibile*, quindi materiale, come nel caso di materie prime, macchinari, edifici, terreni; o *intangibile*, quindi immateriale, ad esempio nel caso di marchi, brevetti e reputazione. (Rudy 2021)

un investimento che contribuisce a creare reddito e valore per l'organizzazione" (Vecchiato 2015: 34).

Le attività di comunicazione rappresentano un "collante" e uno strumento fondamentale per coordinare le attività relazionali (Pastore & Vernuccio 2008: 17). In particolare, il processo comunicativo a cui un'azienda dovrebbe ambire è quello *simmetrico bidirezionale* che, secondo Grunig e Hunt¹⁵ (1984 citati in Vecchiato 2004) meglio si addice alle attività di relazioni pubbliche. A differenza della *comunicazione unidirezionale*, che Grunig individua nelle attività di propaganda e informazione, il modello *simmetrico bidirezionale* mira a raggiungere una comprensione reciproca tra le parti, tramite l'ascolto e il dialogo. La simmetria viene raggiunta solo se "esiste un bilanciamento tra gli interessi delle parti in gioco e l'organizzazione è aperta all'ascolto delle esigenze dei pubblici al punto di arrivare a modificare in parte la propria posizione" (Invernizzi & Romenti 2020: 7). Nella comunicazione con gli interlocutori questa propensione al cambiamento da parte dell'azienda non è detto che sia presente e si è quindi in presenza di un modello che Grunig definisce *asimmetrico bidirezionale* (1984 citato in Vecchiato 2004): in questo caso il feedback del ricevente viene ascoltato ma ha come obiettivo la persuasione dell'interlocutore.

Un esempio chiarificatore della differenza tra modello bidirezionale simmetrico e asimmetrico è quello proposto da Invernizzi e Romenti (2020): si ipotizzi che un'azienda pubblichi un post su Facebook e i clienti inizino ad esprimere delle critiche attraverso dei commenti. A questo punto, l'azienda potrebbe mettere in atto una comunicazione bidirezionale asimmetrica cercando di far prevalere la propria posizione e far cambiare opinione ai clienti; diversamente, l'azienda potrebbe cogliere i feedback degli interlocutori e modificare la propria posizione, praticando quindi una comunicazione bidirezionale simmetrica, in cui emittente e ricevente mettono in atto una negoziazione e

¹⁵ James E. Grunig è uno dei più grandi esperti di Relazioni Pubbliche. Nel 1984, insieme a Todd Hunt, pubblica il volume *Managing Public Relations*, in cui propone quattro modelli di relazioni pubbliche a ognuno dei quali corrisponde una precisa fase storica:

- il modello *press-agentry publicity* (monodirezionale) che si afferma intorno a inizio Ottocento con la figura del *press agent* ed è tipico dell'attività di propaganda;
- il modello *public information* (monodirezionale), impiegato da Ivy L. Lee a inizio Novecento che mira a informare il pubblico;
- il modello *two-way asymmetric*, incarnato da Edward L. Bernays che aveva come scopo quello di persuadere i riceventi;
- il modello *two-way symmetric*, di cui, secondo Grunig, i massimi esponenti sono Cutlip e Center (Grunig & Hunt, 1984).

si trovano in una posizione di pari potere. Tramite questo esempio gli autori mettono in luce anche come il modello di comunicazione adottato prescindendo dallo strumento impiegato: i social network permettono un dialogo maggiore con i clienti, ma non sono una garanzia che questo avvenga in una modalità simmetrica.

Mettere in pratica una comunicazione bidirezionale simmetrica, ascoltando in modo sistematico e organizzato i clienti, permette all'azienda di migliorarsi: le aziende possono agire sulla base di quanto ascoltano, cogliere dei suggerimenti per migliorare i prodotti e i servizi offerti o sviluppare nuove idee suggerite dai clienti (Rappaport 2012). Secondo Vecchiato (2015), in una situazione di questo tipo in cui viene dato valore al *capitale relazionale*, le attività delle relazioni pubbliche si integrano e sovrappongono a quelle del marketing, che da "transazionale" diventa "relazionale" (Vecchiato 2015: 35). Questa propensione verso la relazione e la generazione di valore, è stata messa in luce anche da Kotler e Armstrong (2011: 5, traduzione e corsivi miei), che hanno definito il marketing come "il processo mediante il quale le aziende creano *valore* per i clienti e costruiscono solide *relazioni* al fine di acquisire altrettanto valore in cambio." In letteratura, tale approccio aziendale è definito anche come *customer centricity*: le aziende concentrano la propria attenzione sul cliente e sull'instaurazione di una relazione con esso che sia soddisfacente per entrambe le parti (Day 2003 citato in Lamberti 2013). Nel momento in cui i clienti esprimono dei bisogni, le aziende si attivano per sviluppare delle soluzioni atte a soddisfarli (Lamberti 2013).

2.2 La reputazione nel Web 2.0

Uno dei contesti in cui può avvenire il dialogo con gli stakeholder e l'ascolto delle loro opinioni è quello del Web 2.0, in cui la reputazione può essere definita come "il processo di posizionamento, monitoraggio, misurazione, dialogo e ascolto mentre l'organizzazione si impegna in un dialogo trasparente ed etico con i vari stakeholder" (Jones et al. 2009). Lampignano (2016: xxv) nel suo *Reputation Manifest*, spiega in dieci punti come si articola la reputazione in rete:

1. Socialmente siamo quello che si dice di noi.
2. Ciò che si dice di noi e ciò che siamo, non sempre coincidono.
3. Ciò che coincide non è quello che deve essere socialmente curato.

4. Quello che deve essere curato può seguire due direzioni: assomigliare a quello che vogliamo essere; assomigliare a quello che la società crede e dice di noi.
5. La società non conosce il nostro essere, capisce solo il nostro fare.
6. Il nostro fare è proiezione sociale.
7. Il sociale è un'arena dove si incontrano e scontrano interessi, i quali non sono neutri ma ideologici.
8. L'ideologia è uno dei prodotti di una specifica cultura.
9. La cultura è il luogo dove si sviluppa, fin dall'infanzia, il singolo individuo.
10. L'individuo lasciato solo, nello scontro con il sociale, soccombe.

I dieci punti richiamano nuovamente la riflessione sul rapporto tra identità e reputazione: ciò che un'azienda pensa di essere spesso non è ciò che gli utenti percepiscono. È su questo disallineamento che l'azienda deve lavorare, decidendo se “assomigliare a ciò che vuole essere” basandosi quindi su un'identità predefinita, o se “assomigliare a quello che la società crede e dice”, integrando quindi anche l'opinione degli stakeholder. Tradurre in azioni concrete ciò che l'azienda afferma di essere è fondamentale per risultare autentici agli occhi del pubblico, che grazie alle opportunità offerte dal Web 2.0 può esprimere facilmente la propria opinione e difendere i propri interessi.

2.2.1 Una definizione di Web 2.0

Tim O'Reilly¹⁶ è stato il primo studioso ad aver reso popolare il termine Web 2.0¹⁷ e ad averlo definito come (O'Reilly 2007: 17, corsivo mio):

...the *network as platform*, spanning all connected devices; web 2.0 applications are those that make the most of the intrinsic advantages of that platform: delivering software as a continually-updated service that gets better the more people use it, consuming and remixing data from multiple sources, including individual users, while providing their own data and services in a form that allows remixing by others, creating network effects

¹⁶ L'origine del termine *Web 2.0* viene indagata da Aced Toledano (2013) tramite una rassegna della letteratura: il termine è stato utilizzato per la prima volta dalla designer Darcy DiNucci nel 1999 nell'articolo *Fragmented Future* pubblicato nel *Print Magazine*; gran parte degli studiosi però ne attribuisce l'origine a Tim O'Reilly, che lo utilizza per la prima volta nell'articolo *What is Web 2.0. Design Patterns and Business Models for the Next Generation of Software*. In questo articolo O'Reilly spiega che il termine è nato nel 2003 durante una sessione di brainstorming tra le compagnie O'Reilly Media e MediaLive International; poco dopo è stato utilizzato per la prima volta di fronte al grande pubblico nel *Web 2.0 Summit*, organizzato ogni anno a San Francisco dal 2004 al 2011 da Tim O'Reilly e Dale Dougherty. Aced Toledano (2013) conclude affermando che Tim O'Reilly non è stato il primo ad utilizzare il termine ma è stato il primo ad averlo reso popolare.

¹⁷ La notazione “2.0” riprende quella utilizzata per numerare le varie versioni di un software e segnala un cambiamento rispetto al Web 1.0 (Wikipedia).

through an “*architecture of participation*”, and going beyond the page metaphor of Web 1.0 to deliver *rich user experiences*.

Foiaia (2007) individua le caratteristiche principali che differenziano il Web 2.0 dal Web 1.0¹⁸:

- Ampia diffusione di tecnologie che supportano l’utente e che gli permettono di plasmare e modificare servizi di comunicazione di tipo bidirezionale. Nel Web 1.0 invece si trovavano siti con una comunicazione monodirezionale (dal sito all’utente), in cui i contenuti erano già predisposti e meno modificabili (ibid.);
- Rich User Experience. Le applicazioni e i siti su Internet offrono un’esperienza utente paragonabile a quella che si ha nell’utilizzo di applicazioni utilizzate nel computer in locale. Alcune tecnologie software (RSS¹⁹, AJAX²⁰) rendono il web una piattaforma alternativa in cui poter sviluppare applicazioni ricche e *user friendly*, ovvero di semplice utilizzo e grande intuitività. Il web diventa quindi concorrente rispetto ai sistemi operativi come Windows o MacOS (ibid.);
- Open Culture (Source, Application, Data, Content). Diventa centrale il concetto di *open culture* per cui gli utenti possono creare, diffondere, modificare e riutilizzare i contenuti in rete. Si forma così una *intelligenza collettiva*²¹, ovvero un contesto in cui i contenuti sono il risultato “del lavoro, diretto o implicito, condiviso e collaborativo di una molteplicità di individui” (Foiaia 2007: 124). Questa apertura ha diverse sfaccettature (*ivi*, p. 92):
 - o Open source: i programmatori sviluppano e distribuiscono i propri codici e applicazioni in modo da ottenere collaborazione da parte di altri utenti e progredire nello sviluppo (*ibid.*).

¹⁸ Il web 1.0 corrisponde al World Wide Web, la cui data di nascita viene comunemente indicata nel 6 agosto 1991, giorno in cui l'informatico inglese Tim Berners-Lee pubblicò il primo sito web. (Wikipedia).

¹⁹ RSS (Really Simple Syndication) è un formato per la distribuzione di contenuti Web. Esso permette di rendere le informazioni accessibili all'utente “aggregandole” nella stessa pagina o finestra, senza dover ogni volta ritornare al sito principale. Esempi di aggregatori sono Reddit e Google News. (Wikipedia).

²⁰ AJAX (Asynchronous JavaScript and XML) è una tecnica di sviluppo software per realizzare applicazioni web interattive (Rich Internet Application). Essa permette di scambiare dati in background fra web browser e server, consentendo così di aggiornare una pagina web senza esplicito ricaricamento da parte dell'utente. (Wikipedia).

²¹ Il concetto di *intelligenza collettiva* è stato definito per la prima volta nel 1994 da Pierre Lévy nel libro *L'intelligenza collettiva. Per un'antropologia del cyberspazio*. Essa è definita come “un’intelligenza distribuita ovunque, continuamente valorizzata, coordinata in tempo reale, che porta a una mobilitazione effettiva delle competenze” (Levy 1996).

- Open application: gli utenti possono usufruire delle API (Application Programming Interface), cioè delle “interfacce che permettono alle applicazioni di interagire con altre applicazioni.” (Azione Digitale 2021).
 - Open data: i dati (di qualsiasi tipo, come immagini, audio, video, testo) sono spesso fruibili gratuitamente e separatamente rispetto alle pagine Web native grazie, ad esempio, ai RSS che permettono di distribuire i dati liberamente attraverso la rete. (Foiaia 2007)
 - Open content (*User generated content*): tutti gli utenti, a prescindere dalla propria professione e competenza, hanno la possibilità di generare dei contenuti sulle proprie pagine personali (ad esempio un blog WordPress o un profilo Instagram). In alcuni casi, in assenza dei contenuti generati dalla partecipazione degli utenti, le piattaforme e le community non avrebbero ragione d’essere (Foiaia 2007). Degli esempi sono Twitter, Youtube, Facebook, social network che basano il loro funzionamento sullo *user generated content*.
- Partecipazione ai social network. I social network sono stati definiti da Boyd ed Ellison (2007: 211) come un servizio web che permette agli individui di: costruire un profilo pubblico o semi-pubblico all’interno di un sistema delimitato; creare una lista di utenti con cui connettersi; visualizzare la propria e l’altrui lista di contatti. Inoltre, Beer (2008 citato in Tisconi 2014) aggiunge che i social network, grazie all’azione degli utenti che li riempiono di contenuti, possono essere considerati come un vasto archivio di informazioni relativi alla vita quotidiana dei consumatori.

In questo stadio evolutivo del Web, “al centro di tutto ci sono le persone e i fenomeni culturali e sociali sviluppati a partire dalla loro interazione sul Web, e che difficilmente possono essere controllati dalle aziende” (Foiaia 2007: 14). Il Web ha avuto un’evoluzione simile a quella dei modelli comunicativi descritti da Grunig:

- Web 1.0 – comunicazione monodirezionale: il web è diretto da chi gestisce il software, le applicazioni o i contenuti. Gli utenti hanno accesso a funzioni limitate di lettura e scrittura, non hanno ampio margine di personalizzazione e usano Internet come mera fonte di informazioni (Foiaia 2007).

- Web 2.0 – comunicazione bidirezionale: gli utenti da audience diventano partecipanti attivi che danno il proprio contributo (ibid.). Il web 2.0 può essere considerato uno spazio *democratico* (Jones et al. 2009), poiché, grazie alla sua struttura, lascia libertà di espressione e partecipazione agli utenti e si sviluppa grazie ai contenuti da essi prodotti. Come afferma O’Reilly (2007) nella definizione di Web 2.0, tale spazio è dunque caratterizzato da una “architecture of participation” che favorisce una comunicazione bidirezionale tra emittente e ricevente e in cui i messaggi vengono trasmessi in una struttura da uno a molti.

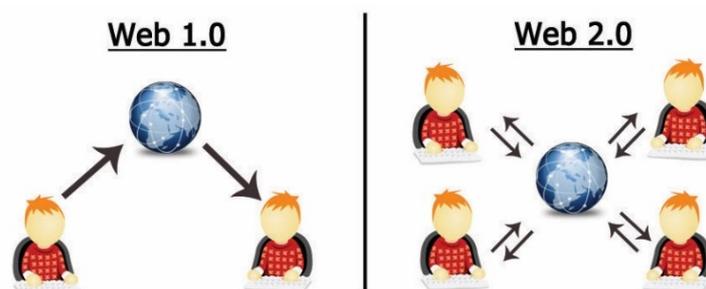


Figura 2: differenza tra Web 1.0 e Web 2.0 (Dincer 2020).

Ora che tutti gli utenti hanno la possibilità di pubblicare contenuti online, l’elemento distintivo di un’azienda non risiede più nella pubblicazione di contenuti, bensì nell’abilità di instaurare e mantenere una relazione con l’audience (Davis 2012 citato in Aced Toledano 2013). La comunicazione bidirezionale porta aziende e stakeholder in una situazione di “co-protagonismo” (Poma & Grandoni 2021: 69): la reputazione aziendale non è più plasmata dalla sola comunicazione unidirezionale che parte dall’azienda, bensì è fortemente influenzata dalle azioni e dalle opinioni degli stakeholder online (Jolly 2001 citato in Jones et al. 2009). In un contesto sociale così connesso, la reputazione aziendale può essere danneggiata più velocemente, poiché l’utente “si informa, legge, domanda, critica, partecipa e contribuisce alla costruzione della marca o dello specifico prodotto” (Lampignano 2016: 32).

2.2.2 Conversazioni e opinioni nel Web 2.0

Grazie alla sua struttura ubiqua e partecipativa, il Web 2.0 permette di strutturare un dialogo, collegare periferie e confrontarsi con opinioni diverse (Poma & Grandoni 2021: 70). Ancor prima del passaggio al Web 2.0, nel 1999 il *Cluetrain Manifesto* scritto da

Rick Levine, Christopher Locke, Doc Searls e David Weinberger, proponeva 95 tesi che invitavano le aziende a cambiare il proprio modello comunicativo unidirezionale, in favore di uno più adatto al nuovo mercato interconnesso. Alcune di queste tesi sottolineavano la natura conversazionale che stava assumendo il mercato dei consumatori e a cui avrebbero dovuto aderire anche le aziende:

- Tesi 1: i mercati sono conversazioni.
- Tesi 6: Internet dà agli esseri umani la possibilità di conversare in un modo che era semplicemente impossibile nell'era dei mass media.
- Tesi 9: Queste conversazioni in Rete stanno facendo nascere nuove forme di organizzazione sociale e nuovi scambi di conoscenze.
- Tesi 10: ...i consumatori stanno diventando più intelligenti, più informati, più organizzati.
- Tesi 12: Non ci sono segreti. Il mercato online conosce i propri prodotti meglio delle aziende stesse. E diffonde a tutti la propria opinione, buona o cattiva che sia.
- Tesi 38: Le comunità umane sono basate sulla comunicazione – discorsi umani su problemi umani.
- Tesi 40: Le aziende che non appartengono a una comunità basata sulla comunicazione sono destinate a morire.
- Tesi 53: Ci sono due conversazioni in corso. Una all'interno dell'azienda, l'altra con il mercato.
- Tesi 60: ... I mercati vogliono parlare con le aziende.
- Tesi 76: Abbiamo anche noi [i consumatori] qualche idea da proporvi. [...] Avete un minuto?

Il *Chuetrain Manifesto*, antesignano delle riflessioni che ci sarebbero state riguardo le caratteristiche del Web 2.0, si faceva portavoce delle nuove esigenze del mercato: i consumatori chiedono all'azienda di essere reperibile, aperta a conversare con loro in modo autentico, umano e onesto e di prendere realmente in considerazione la loro opinione. Come evidenzia la tesi 53, “ci sono due conversazioni in corso”, una all'interno dell'azienda e una con il mercato. In particolare, nel contesto online, le fonti di conversazione sono quelle che Rappaport (2012: 16) definisce fonti *brand backyard* e fonti *consumer backyard*. La prima fonte consiste negli *owned media*, ovvero tutti quegli

spazi ufficiali di proprietà dell'azienda (siti web, account social aziendali, customer service, community ufficiali), mentre le fonti *consumer backyard* sono gli spazi in cui il pubblico può esprimersi liberamente (account social personali, siti di valutazione e recensione, forum).

I consumatori esprimendo la loro opinione plasmano la reputazione dell'azienda e determinano la fiducia che i consumatori le concedono. Ciò accadeva già in passato, in assenza di Internet, (Olegario & McKenna 2013) quando la fonte di informazione più affidabile era l'opinione di familiari, amici e conoscenti, ovvero il passaparola, conosciuto in letteratura anche come *word of mouth*. Con Internet e la relativa "architecture of participation" (O'Reilly 2007) aumentano le possibilità di diffondere la propria opinione anche tra individui sconosciuti e a grande distanza fra loro. Nasce così l'*electronic word of mouth* (E-WOM), definito da Hennig-Thurau et al. (2004: 39, traduzione mia) come "qualsiasi dichiarazione positiva o negativa fatta da clienti potenziali, effettivi o precedenti su un prodotto o un'azienda, che viene messa a disposizione di una moltitudine di persone e istituzioni tramite Internet". Tra il passaparola tradizionale (WOM) e quello online (E-WOM) ci sono alcune differenze che vengono qui sintetizzate:

	WOM	E-WOM
Credibilità	Alta credibilità: il ricevente conosce di persona l'emittente dell'informazione.	Credibilità variabile: emittente e ricevente non è detto che si conoscano e ciò potrebbe influire negativamente sulla credibilità.
Privacy	La conversazione tra emittente e ricevente è privata, interpersonale e avviene in tempo reale.	L'informazione trasmessa di solito è pubblica, viene trasmessa in forma scritta e può essere letta da chiunque in qualsiasi momento.
Velocità di diffusione	L'informazione si diffonde lentamente e in presenza dei due interlocutori.	L'informazione si trasmette più velocemente attraverso Internet in qualsiasi momento.

Tabella 2: differenza tra WOM ed E-WOM. Tabella tradotta e adattata da Huete-Alcocer (2017).

I consumatori hanno un grande potere nei confronti dell'azienda, perché la fiducia che un consumatore ripone in un altro è maggiore rispetto a quella riposta nelle comunicazioni provenienti dall'azienda: "il consumatore considera scontato che la pubblicità provi a convincerlo della bontà del prodotto e che il produttore farebbe di tutto per invogliarlo a seguire i suoi consigli" (Lampignano 2016: 30); al contrario, l'opinione proveniente da un altro utente è percepita come disinteressata e non commerciale. Questa tendenza è rilevata anche da una ricerca del 2022 di ReferralCandy (Matson 2022) che indica, in ordine decrescente, le fonti di informazione verso cui i consumatori mostrano più fiducia, ovvero il *word of mouth*, le opinioni online e, infine, la pubblicità.

Hennig-Thurau et al. (2004), attraverso una revisione della letteratura esistente, individuano undici motivi²² che potrebbero spingere un utente a pubblicare la propria opinione nelle piattaforme online:

- (1) *Preoccupazione per gli altri consumatori*: l'utente mette in atto un comportamento prosociale²³ per aiutare gli altri consumatori a prevenire un'esperienza negativa.
- (2) *Desiderio di aiutare l'azienda*: dopo un'esperienza positiva con l'azienda, il consumatore sente di voler dare qualcosa "in cambio" e contribuire positivamente al suo successo nel mercato. Anche in questo caso, viene messo in atto un comportamento prosociale che può essere spiegato dalla *equity theory* (Oliver & Swan 1989 citati in Hennig-Thurau et al. 2004): se il consumatore sente di aver ricevuto più risorse rispetto all'azienda, tenderà a sanare questo disequilibrio tramite una recensione positiva.
- (3) *Benefit sociale*: l'utente pubblica un commento per sentirsi parte di una community online.
- (4) *Esercizio del potere (collettivo) sulle aziende*: dato che le opinioni negative possono influenzare la reputazione di un'azienda, le recensioni negative possono essere utilizzate come strumento di potere da parte degli utenti, specialmente quando vengono pubblicate più recensioni negative contemporaneamente.

²² Tutti i motivi esposti nel saggio di Hennig-Thurau et al. (2004) sono stati tradotti in italiano.

²³ Un comportamento prosociale è un insieme di "azioni di individui o gruppi finalizzate a realizzare o a migliorare il benessere di un'altra persona" (Oldini 2002: 283).

- (5) *Ricerca di consigli post-acquisto*: il consumatore chiede informazioni su un prodotto o un servizio dopo averlo acquistato ad altri consumatori che si sono trovati nella stessa situazione.
- (6) *Auto-miglioramento*: l'utente vuole ottenere un riconoscimento positivo da altre persone ed essere considerato, ad esempio, un consumatore esperto che ha conoscenze approfondite riguardo l'oggetto o il servizio in questione.
- (7) *Compenso economico*: in alcune piattaforme si può ottenere un compenso economico nel momento in cui si lascia una recensione, motivo per cui l'utente è spinto a pubblicarla.
- (8) *Comodità* + (9) *Supporto nel risolvere un problema*: nel caso si fosse su una piattaforma con dei moderatori, si potrebbe chiedere aiuto all'operatore per comodità o perché impossibilitati a trovare una soluzione.
- (10) *Espressione di emozioni positive* + (11) *Sfogo di emozioni negative*: secondo Zajonc (1971 citato in *ibid.*), gli individui sentono un desiderio di equilibrio nella propria vita. Nel momento in cui questo equilibrio viene a mancare, gli individui cercano di ripristinarlo tramite, ad esempio, la pubblicazione di un commento online. Esprimere un'emozione positiva permette di soddisfare il desiderio dell'individuo di condividere la propria gioia; al contrario, esprimendo un'emozione negativa permette all'individuo di ridurre il proprio senso di frustrazione. In entrambi i casi, quindi, l'individuo ritrova un equilibrio e riduce una tensione interna.

I motivi identificati da Hennig-Thurau (2004) possono essere utilizzati come guida interpretativa durante l'analisi delle opinioni dei consumatori. Un'azienda che registra un grande numero di recensioni da parte dei clienti potrebbe percepirle come una minaccia alla reputazione oppure sfruttarle a proprio favore e considerarle come dei dati preziosi, spesso ottenuti gratuitamente²⁴, su cui effettuare un'analisi volta ad individuare i punti forti e i punti deboli dell'azienda.

²⁴ I dati vanno sempre ottenuti legalmente rispettando le regole della privacy.

2.2.3 Twitter: microblogging social network

L'E-WOM può aver luogo nei *microblogging social network*, ovvero “social network che consentono, attraverso l'utilizzo di messaggi molto brevi [chiamati ‘microblog’], una forma di comunicazione in tempo reale” (Tissoni 2014:102). Come altri tipi di contenuti pubblicati online, anche i microblog permettono agli utenti di esprimere i propri pensieri e le proprie opinioni da qualsiasi parte del globo, attraverso una molteplicità di dispositivi (fissi o mobili), su vasta scala e ad alta velocità. Come sottolineano Jansen et al. (2009) un microblog, però, a differenza di altri contenuti, è caratterizzato dalla brevità: un messaggio di questo tipo ha un limite fissato di caratteri e ciò lo rende facile da produrre e fruire, immediato e non invasivo. I motivi²⁵ per cui un utente è portato a pubblicare un microblog, secondo Nardi et al. (2004 citati in ibid.) sono la possibilità di documentare la propria vita, fornire commenti e opinioni, esprimere emozioni o idee e partecipare a una community.

Uno dei *microblogging social network* più popolari è Twitter, che nel 2022 conta 368 milioni di utenti attivi mensilmente in tutto il mondo (Bagadiya 2023) e dal 27 ottobre 2022 vede come CEO Elon Musk. Questo social network nacque come prototipo nel marzo 2006 a San Francisco grazie a Jack Dorsey, Biz Stone ed Evan Williams che lavoravano per il sito internet Odeo.com. Inizialmente Twitter veniva utilizzato per permettere ai dipendenti di Odeo di comunicare in modo rapido e simultaneo. Il primo tweet risale al 21 marzo 2006 e fu pubblicato da Jack Dorsey che scriveva “just setting up my twttr”. Nel luglio 2006 venne lanciata la versione aperta al pubblico e a ottobre dello stesso anno venne fondata la società Obvious Corporation dai tre creatori di Twitter e da alcuni dipendenti di Odeo. Dal 2013 la Obvious Corporation è diventata una società quotata in borsa.

I contenuti di Twitter consistono in messaggi di testo (i *tweet*) che dal 2017 hanno un limite di 280 battute²⁶. Il nome “Twitter” tradotto letteralmente significa “cinguettare” e, metaforicamente, “ciarlare”. Tale significato viene trasmesso sia attraverso il logo che rappresenta un uccellino sia attraverso la pagina About²⁷ del sito ufficiale dove si leggono

²⁵ Alcuni di questi motivi sono gli stessi individuati Hennig-Thurau (2004) per quanto riguarda l'E-WOM.

²⁶ Inizialmente il limite era di 140 battute (Tissoni 2014).

²⁷ <https://about.twitter.com/en> [ultima consultazione: 22/02/23]

i seguenti testi (Figure 3 e 4), che si possono considerare espressione della *mission* dell'azienda:



Figura 3: screenshot della prima parte dei testi presenti nel sito ufficiale di Twitter alla pagina About.



Figura 4: screenshot della seconda parte di testi presenti nel sito ufficiale di Twitter alla pagina About. Tra questi si noti il tweet di Jack Dorsey, uno dei fondatori di Twitter.

La *mission* ufficiale, che si trova nella sezione del sito dedicata alla relazione con gli investitori²⁸, sintetizza gli obiettivi dell'azienda:

The mission we serve as Twitter, Inc. is to give everyone the power to create and share ideas and information instantly without barriers. Our business and revenue will always follow that mission in ways that improve – and do not detract from – a free and global conversation.

Dai testi sopra riportati e dalla *mission* ufficiale si comprende come la priorità di Twitter sia quella di dare spazio alla circolazione libera di idee e opinioni, su scala globale e senza barriere. Ciò che contraddistingue questo social network è la facilità con cui gli utenti (tra cui le aziende) possono reperire i tweet su un determinato argomento, capire immediatamente quali siano i temi caldi e le opinioni a riguardo ed eventualmente inserirsi nella conversazione. Ciò è consentito da:

²⁸ <https://investor.twitterinc.com/contact/faq/default.aspx> [ultima consultazione: 22/02/23]

- Natura pubblica dei tweet: quando ci si iscrive, i messaggi vengono postati automaticamente in modalità pubblica (possono essere resi privati solo cambiando manualmente le impostazioni). Come riporta Dardi (2011: 8), il carattere pubblico e aperto dei messaggi inizialmente aveva destato qualche preoccupazione, ma è proprio grazie a questo che i messaggi possono propagarsi velocemente.
- Funzione di Retweet: questa funzione permette di ripubblicare il tweet di un altro utente, contribuendo così alla sua diffusione.
- Hashtag²⁹: i tweet possono essere categorizzati tramite gli hashtag, preceduti dal simbolo del cancelletto (#). Essi consistono in parole chiave che permettono di rintracciare i tweet inerenti a uno stesso argomento. L’hashtag venne usato per la prima volta nel 2007 da Chris Messina che, come spiega Sorbo (2015), voleva migliorare il tracciamento dei contenuti che si riferivano a uno stesso argomento.
- Possibilità di rispondere e menzionare: è possibile scrivere messaggi di risposta agli utenti. Se due utenti conversano, i loro follower e chi ha partecipato alla conversazione vedranno i tweet nella cronologia del proprio profilo. Se nella cronologia si clicca su una risposta, la schermata si espande per mostrare il tweet che ha dato inizio alla conversazione. Se ci si vuole rivolgere a qualcuno direttamente, lo si può menzionare inserendo il simbolo “@” prima del nome utente.
- Ricerca avanzata: essa permette di ottenere risultati personalizzati, impostando diversi filtri di ricerca dei tweet, come parole chiave, lingua, data o intervallo di date, account da cui proviene il tweet o account a cui è diretto, menzioni presenti.
- Scheda “Esplora”: in questa scheda vengono indicati gli ultimi argomenti di tendenza³⁰, individuati tramite un algoritmo che considera il luogo, gli interessi e i profili seguiti dall’utente.

²⁹ <https://help.twitter.com/it/using-twitter/how-to-use-hashtags> [ultima consultazione: 20/02/23]

³⁰ <https://help.twitter.com/it/using-twitter/twitter-trending-faqs> [ultima consultazione: 20/02/23]

2.2.4 Linee guida per preservare la reputazione online

Data la possibilità degli utenti di conversare online in qualsiasi momento, le aziende si trovano in una posizione vulnerabile e per continuare a svolgere le proprie attività serenamente non devono subire le opinioni degli utenti, sperando che siano positive, bensì gestire in modo sistematico e attivo la propria reputazione.

In generale, gestire la reputazione è un'attività complessa perché, come spiegano Poma e Grandoni (2021), consiste nell'ascoltare in modo organizzato tutti gli stakeholder, interni ed esterni all'azienda, e individuare il canale comunicativo e gli strumenti di controllo più adatti per farlo. Lo strumento di controllo e monitoraggio della reputazione utilizzato in questa ricerca verrà descritto nel capitolo 3; in questo paragrafo vengono elencate delle linee guida che, a prescindere dal metodo scelto, possono aiutare a gestire la reputazione online.

Rappaport (2012), grazie all'analisi di diversi casi studio inerenti all'attività di piccole e grandi aziende, ha individuato sei approcci vincenti per gestire la reputazione online:

1. *Anticipare i problemi e le preoccupazioni degli stakeholder monitorando costantemente le conversazioni online.*
2. *Coinvolgere i settori aziendali che possono influire sulla reputazione e sulla relazione con il cliente.*
3. *Sviluppare un atteggiamento lungimirante per reagire a un'eventuale crisi.*

Un esempio positivo dell'applicazione di questi tre approcci è quello di Hasbro³¹ (Rappaport 2012; Shin 2006; Reuters 2006): nel 2006 un dipendente dell'azienda, che stava controllando le recensioni di un prodotto su Amazon, scoprì che un bambino era morto giocando con “un banco da lavoro” in miniatura che conteneva utensili, viti e chiodi di plastica. L'azienda contattò immediatamente la Consumer Product Safety Commission che dichiarò che i chiodi erano adatti solo a bambini dai tre anni in su. Sulla scatola del prodotto Hasbro aveva inserito l'etichetta con indicata l'età corretta, mentre il bambino vittima dell'incidente aveva un'età inferiore e ciò sollevava Hasbro dalla responsabilità dell'accaduto. L'azienda continuò a monitorare le conversazioni online e rilevò che le

³¹ Casa editrice statunitense di giochi e giocattoli (Wikipedia).

preoccupazioni dei clienti riguardavano quel prodotto in particolare e non l'azienda Hasbro nel suo complesso. I dati provenienti dall'ascolto dei social media sono stati condivisi anche con l'ufficio legale e il reparto produttivo che, in quel caso, erano fondamentali nel determinare una gestione efficace della reputazione. Nonostante Hasbro non fosse responsabile dell'incidente, l'azienda decise di far fronte alle preoccupazioni dei genitori e di ritirare il giocattolo dal mercato. L'azienda non ha quindi aspettato che ci fosse un alto numero di opinioni negative sulla pericolosità del proprio prodotto ed è riuscita ad anticipare un'eventuale crisi e dei danni alla reputazione.

4. Capire chi sono gli stakeholder coinvolti e quali sono le loro posizioni.

5. Riunire fonti complementari di dati per avere una panoramica della situazione.

Quando un'azienda deve risolvere un determinato problema, di solito gli utenti interessati consistono in diversi tipi di stakeholder. L'azienda, quindi, a seconda del caso, deve riuscire a individuare i diversi utenti coinvolti e le fonti da cui attingere per ascoltare la loro opinione e avere una panoramica della situazione. Un esempio positivo è quello della vicenda che ha coinvolto la Apple dopo l'uscita dell'Iphone4, conosciuta anche come "antenna-gate" (Rappaport 2012). La linea telefonica del nuovo modello dell'iPhone cadeva se il telefono veniva impugnato in un certo modo. Riguardo questa problematica, attraverso un'analisi completa delle opinioni dei clienti, Apple ha compreso che le posizioni degli stakeholder erano due e opposte: i clienti consideravano il problema di poca importanza, mentre i giornalisti e gli esperti del settore avvertivano un disagio maggiore. Apple si è ritrovata quindi a fronteggiare due posizioni opposte e ha applicato delle strategie comunicative ad hoc per gestire la relazione con i diversi gruppi di stakeholder.

6. Coinvolgere gli stakeholder per valorizzare o ristabilire la reputazione.

Ascoltare gli stakeholder permette di capire quali elementi del proprio prodotto e servizio migliorare. Ad esempio, Nestlé nel 2010 (Rappaport 2012; Armstrong 2010) è stata colpita duramente dalle proteste dell'organizzazione ambientalista Greenpeace che accusava l'azienda di essersi rifornita di olio di palma, uno degli ingredienti della barretta KitKat, dalla società Sinar Mas. Quest'ultima piantava alberi di palma distruggendo l'habitat della foresta pluviale indonesiana. Greenpeace mise in atto una vera e propria

campagna di pressione sui social media: diffuse un video³² in cui un impiegato, durante una pausa in ufficio, mangia una barretta che ha lo stesso packaging del KitKat e all'interno vi trova un dito di orango (uno degli abitanti della foresta pluviale). Il video, dalla macabra satira, sensibilizzò i consumatori sull'origine dell'olio di palma e provocò la loro indignazione, espressa online tramite migliaia di commenti. Inizialmente Nestlé non seppe reagire in modo adeguato e fece rimuovere il video accusando Greenpeace di aver violato il copyright. Questa azione portò a un *effetto Streisand*³³ nei confronti del video e i consumatori continuarono ad esprimere opinioni negative che Nestlé poi eliminava dalle proprie pagine. Dato che la situazione non sembrava migliorare, l'azienda capì che per ripristinare la propria reputazione doveva aprire un dialogo con gli stakeholder e modificare il proprio operato: dopo dieci settimane dall'inizio della campagna di Greenpeace, l'azienda si scusò per il suo approccio difensivo e dichiarò di aver interrotto il rapporto con la società che piantava alberi di palma in Indonesia e di essersi rivolta a una fonte di olio di palma sostenibile.

Idealmente un'azienda dovrebbe mettere in atto le azioni individuate dallo studio di Rappaport (2012) ma, come sottolinea Lampignano (2016: 87-88), è necessario che tra i dipendenti siano presenti delle figure che gestiscano la reputazione e affianchino il reparto marketing. Una di queste figure è quella del *Reputation Manager*, definito come colui che sa “contribuire alla proiezione nel mondo dell'essenza di un soggetto (individuo o azienda)” e sa gestirla nel tempo, anche e soprattutto nei momenti di crisi (*ivi*, p. 88). Egli ha diversi compiti, tra cui:

- presidiare le conversazioni rilevanti per la reputazione che avvengono sia su fonti media *owned* (canali ufficiali dell'azienda) che su fonti media *not owned* (siti di news, blog, forum, social network). Poma e Vecchiato (2012) sottolineano come il monitoraggio delle conversazioni su Internet debba essere effettuato quotidianamente come nel caso della tradizionale rassegna stampa, in modo da mettere in atto azioni correttive prima che le notizie negative arrivino a un pubblico più ampio;

³² Il video è un'opera di finzione e satirica e oggi è ancora disponibile su YouTube alla pagina ufficiale di Greenpeace Italia: <https://www.youtube.com/watch?v=iv5jsSkJqbA> [ultima consultazione: 22/02/23].

³³ L'effetto Streisand si ha quando, nel tentativo di censurare un'informazione, se ne provoca, contrariamente alle attese, l'ampia pubblicizzazione (Wikipedia).

- valutare il sentiment online, grazie a un uso corretto della sentiment analysis e dell'opinion mining³⁴;
- ingaggiare gli influencer per azioni mirate;
- gestire i contenuti dannosi per il marchio, con l'obiettivo di prevenire le crisi: il *Reputation Manager* deve essere un portavoce della storia della marca ed essere preparato a gestire la narrazione negativa che potrebbe essere messa in atto dai consumatori. Deve quindi colmare il disallineamento che potrebbe esistere tra l'identità dell'oggetto di analisi e il suo modo di essere percepito.

2.3 Le crisi reputazionali

Durante il proprio ciclo di vita è possibile che un'azienda si trovi ad affrontare una crisi reputazionale. La crisi è “una situazione operativa che, se non affrontata e adeguatamente risolta, potrebbe avere conseguenze negative sul rapporto con uno o più stakeholder (e pubblici influenti) e sulla *business continuity*³⁵”. C'è una crisi ogniqualvolta “un elemento nuovo e inaspettato, endogeno o esogeno [all'azienda], mette a rischio la continuità degli affari” (Poma & Grandoni 2021: 238-239, corsivo mio).

Le **cause** di una crisi sono potenzialmente infinite, ma Poma e Grandoni fanno un elenco delle più frequenti (ivi, p. 239): disastro naturale, errore umano, guasto meccanico, débâcle tecnologica, sabotaggio o estorsione, attacco esterno, pratiche finanziarie discutibili, problemi nei vertici aziendali, riorganizzazioni e licenziamenti, problemi di comunicazione interna, problemi di concorrenti o dell'intero settore, lobbies economiche e politiche, contingenze politiche internazionali, diffusione di notizie false.

Inoltre, le crisi possono essere suddivise in **tipologie differenti**, a seconda del percorso evolutivo, del soggetto ritenuto responsabile e dell'impatto. Seguendo la classificazione proposta da Poma e Grandoni (2021: 241), a seconda del percorso evolutivo, una crisi può essere:

- *Strisciante*: all'inizio la crisi è quasi impercettibile e ci sono solo alcuni *segnali deboli*, ovvero segnali “che spesso precedono e danno avvisaglia dell'incombere di un evento negativo” (Poma & Vecchiato 2012: 66) che permetterebbero di

³⁴ Questi metodi verranno descritti nel capitolo 3 di questa ricerca.

³⁵ Per *business continuity* o continuità operativa si intende la “capacità di un'organizzazione di continuare a erogare prodotti o servizi a livelli predefiniti accettabili a seguito di un incidente” (Wikipedia).

prendere dei provvedimenti meno costosi e più facili da gestire. Col passare del tempo, però, la situazione si aggrava e, se non fronteggiata in tempo, porta allo scoppio di un grave caso di crisi reputazionale. “Più tardi si deciderà di intervenire, più difficile sarà riparare alla situazione” (Poma & Grandoni 2021: 241).

- *Improvvisa*: la crisi scoppia senza aver dato dei segnali premonitori, coglie di sorpresa l'organizzazione e diventa la “notizia del giorno” dei mass media. Questo tipo di crisi è il più traumatico perché avviene senza che l'azienda se lo aspetti (*ibid.*).

Secondo l'Institute of Public Relations (n.d. citato in *ibid.*), la crisi può essere anche classificata in base a chi ne viene ritenuto responsabile dall'opinione pubblica e dai tribunali competenti:

- *Crisi in cui l'azienda è vittima*: è il caso delle crisi provocate da eventi naturali, calunnie, atti violenti o estremi che mirano ad attaccare direttamente l'azienda. La responsabilità dell'azienda è minima o nulla.
- *Crisi accidentali*: si tratta di eventi in cui l'azienda non ha avuto successo (come esperimenti non riusciti o errori tecnici) e quindi viene critica dagli stakeholder per non aver messo in atto una corretta strategia. La responsabilità dell'azienda è bassa.
- *Crisi prevedibili*: in questo caso, l'organizzazione ha deliberatamente deciso di adottare comportamenti scorretti, come nei casi di negligenza, incuria, ricerca di scorciatoie. La responsabilità dell'azienda è alta.

A seconda dei casi, le crisi possono avere un diverso impatto sulla reputazione delle organizzazioni. Secondo Van den Hurk (2013 citato in Splendiani 2022), si possono identificare quattro livelli di impatto:

- *Impatto minimo*: l'azienda potrebbe non essere direttamente coinvolta nella crisi, ad esempio nel caso di crisi di settore o di un'azienda dello stesso settore. In questo caso, potrebbe solo essere necessario monitorare la situazione.
- *Impatto moderato*: la crisi non è grave ma l'azienda si trova comunque a dover dare una risposta ai clienti o ad attivisti e gruppi di pressione.
- *Impatto maggiore*: la crisi è causata da una condotta scorretta del management o dai reati perpetrati dall'azienda.

- *Impatto catastrofico*: in questo caso, diventa difficile rispondere alla crisi nell'immediato, perché essa esplode senza che l'azienda se lo aspetti, come nel caso di vittime, lesioni gravi o disastri naturali o di origine umana.

Una volta che un'organizzazione si trova in una situazione di crisi arriva il momento di render conto del proprio operato e lavorare per proteggere la propria reputazione. Splendiani (2022) nel suo volume menziona diverse teorie proposte dalla letteratura che identificano le **strategie di risposta** alla crisi. Quella che risulta più adatta agli obiettivi di questa ricerca è la *Image Restoration Theory*³⁶, elaborata da William Benoit nel 1997 e pubblicata sul *Public Relations Review*. La teoria viene presentata da Splendiani (2022: 32) nella seguente tabella:

Strategia generale	Variante	Messaggio chiave
Negazione	<i>Simple denial</i>	“Non abbiamo commesso il fatto”
	<i>Shift the blame</i>	“Il fatto è stato commesso da altri”
Elusione di responsabilità	<i>Provocation</i>	“Abbiamo risposto ad un atto provocatorio”
	<i>Defeasability</i>	“Non sapevamo di poter provocare conseguenze negative”
	<i>Accident</i>	“È stato un incidente involontario”
	<i>Good intentions</i>	“Avevamo buone intenzioni”
Riduzione dell'offensività	<i>Bolstering</i>	“La nostra storia dimostra che abbiamo sempre agito per il bene dei nostri clienti”
	<i>Minimization</i>	“Non è poi così grave come sembra”

³⁶ Nota anche come *Image repair theory*.

	<i>Differentiation</i>	“In casi simili i danni provocati sono stati peggiori”
	<i>Transcendence</i>	“Lo abbiamo fatto per un bene superiore”
	<i>Attack accuser</i>	“Il nostro accusatore non è credibile”
	<i>Compensation</i>	“Siamo disposti a rimborsare le parti lese”
Azione correttiva	<i>Corrective action</i>	“Ripristineremo la situazione a prima dell’accaduto e ci assicureremo che non accada più in futuro”
Dichiararsi mortificati	<i>Mortification</i>	“Ci scusiamo per l’accaduto”

Tabella 3: strategie di risposta alla crisi secondo la *Image Repair Theory* (Benoit 1997 adattato da Splendiani 2022).

Come spiega Splendiani (2022:) la teoria indica le strategie di risposta alla crisi e i relativi messaggi chiave da trasmettere:

- La prima strategia di risposta è la *negazione*. L’organizzazione può negare che l’atto sia avvenuto (*Simple denial*) oppure attribuirne la colpa ad altri soggetti (*Shift the blame*).
- La seconda strategia di risposta è l’*elusione di responsabilità*. Ciò consiste nel dichiarare che:
 - o il comportamento è stato una risposta a un atto provocatorio (*Provocation*);
 - o l’organizzazione non aveva informazioni sufficienti per controllare la situazione e questo l’ha portata a compiere l’atto (*Defeasability*);
 - o il fatto è accaduto in modo accidentale e involontario (*Accident*);
 - o il fatto è avvenuto nonostante le buone intenzioni dell’organizzazione (*Good intentions*).
- La terza strategia è la *riduzione dell’offensività* dell’atto e consiste nel:

- rafforzare i sentimenti positivi nei confronti dell'organizzazione, ricordando le azioni lodevoli del passato (*Bolstering*);
 - ridurre la percezione degli effetti negativi (*Minimization*);
 - menzionare atti simili ma più gravi (*Differentiation*);
 - descrivere il contesto in cui è verificato l'evento e fornire delle attenuanti (*Transcendence*);
 - screditare l'accusatore per ridurre l'impatto dell'accusa (*Attack the accuser*);
 - offrire un compenso alla parte lesa (*Compensation*).
- La quarta strategia è l'*azione correttiva*. L'organizzazione dichiara che il problema verrà risolto e che non si ripeterà in futuro.
 - La quinta strategia consiste nel *dichiararsi mortificati*. L'organizzazione si scusa e ammette la propria colpa.

Il modello di Benoit è stato applicato a tre social media (Facebook, Twitter, Instagram) da Triantafyllidou e Yannas (2020 citati in Splendiani 2022), per indagare gli effetti delle risposte sui riceventi. Gli autori affermano che la strategia di risposta che ha più successo è quella dell'*azione correttiva*, a prescindere dalla piattaforma utilizzata.

Il contesto del Web 2.0 rappresenta un'opportunità per stabilire una comunicazione più veloce ed efficace tra azienda e stakeholder, ma porta a un'accelerazione dello sviluppo della crisi (Bridgeman 2008 citato in Arokiasamy et al. 2019). L'"architecture of participation" offerta dal Web 2.0 permette a chiunque, da qualsiasi parte del mondo, di informarsi, fornire testimonianze o esprimere la propria opinione riguardo la crisi tramite la condivisione di contenuti (testo, immagini, video, audio). In particolare, come evidenzia Norsa (2002a citato in Poma e Vecchiato 2012: 180), i contenuti condivisi su Internet sono incontrollabili, veloci, possono raggiungere un pubblico ampissimo ed essere una fonte di informazione anche per i media tradizionali³⁷. Tutto ciò che viene pubblicato nel Web entra a far parte di quello che Sernovitz (2011: 74) chiama "archivio permanente", ovvero un archivio informatico che non ha gli stessi limiti fisici di quelli cartacei e in cui è possibile rintracciare le informazioni anche a distanza di molto tempo. Di conseguenza, come fanno notare Poma e Vecchiato (2012), anche a distanza di anni

³⁷ Stampa, radio e televisione.

da un evento di crisi, un utente avrà sempre a disposizione delle informazioni a esso relative e sarà più difficile dimenticarsene e restaurare completamente la reputazione aziendale.

Poma e Vecchiato (2012: 183) individuano alcune caratteristiche del Web 2.0 che hanno richiesto delle modifiche nell'approccio alle situazioni di crisi. L'azienda deve essere sempre pronta ad affrontare la crisi, "24 ore al giorno 7 giorni su 7". Come fa notare anche Splendiani (2022), nel contesto online, l'avvento di una crisi e la propagazione di notizie e opinioni a riguardo sono simultanei e, allo stesso modo, la risposta da parte dell'azienda dovrebbe essere immediata e finalizzata a instaurare un dialogo con gli interlocutori. Il pubblico durante una crisi ha infatti un "grande appetito di notizie" (Poma & Vecchiato 2012: 185) e l'azienda deve essere pronta a trasmetterle in modo chiaro e preciso.

Ai fini della gestione della crisi, secondo Splendiani (2022), non c'è un social network più adatto di un altro. La scelta su quale social network comunicare dipende dalla familiarità che l'azienda ha con il pubblico su quella piattaforma e, quindi, da una strategia di lungo termine avviata già da prima della crisi e non estemporanea: "il numero di follower e l'intensità dei rapporti con essi andranno coltivati nel tempo, per poi partire da una posizione di vantaggio nel momento in cui la crisi dovesse esplodere" (ivi, p.27).

CAPITOLO 3 – SENTIMENT ANALYSIS PER IL MONITORAGGIO DELLA REPUTAZIONE SU TWITTER

If you cannot measure it, you cannot improve it.

W. Thomson

L'avvento del Web 2.0, e in particolare l'*electronic word of mouth*, ha aumentato le possibilità di interazione tra gli utenti e portato alla generazione di una grande quantità di dati che possono rappresentare del materiale empirico con cui costruire una ricerca. Come fa notare Corbetta (2014: 457), in questi casi il ricercatore non deve più scendere sul campo e può svolgere il proprio studio "senza staccarsi dalla scrivania e dal suo computer".

In questa ricerca sono rilevanti i dati testuali generati dagli utenti di Twitter, piattaforma di microblogging che permette di esprimere la propria opinione e comunicare con altri utenti e con le aziende. In questo capitolo viene mostrato come sia possibile trasformare un concetto astratto come quello di reputazione in una variabile che sia rilevabile tramite la *sentiment analysis*, una tecnica che rientra nell'ambito del *Natural Language Processing* e che, in questo caso, viene applicata ai tweet.

Si adotta un approccio *data-driven*³⁸, simmetrico bi-direzionale e orientato al cliente, in cui la rilevazione dei dati ha come obiettivo primario il monitoraggio della reputazione e l'ascolto autentico dell'opinione degli utenti, ovvero l'individuazione delle loro esigenze e delle aree problematiche e di miglioramento di un'azienda.

3.1 Individuare un percorso di ricerca per monitorare la reputazione

Come sottolineano Poma e Grandoni (2021: 379), una buona reputazione rappresenta un vantaggio economico rispetto alle aziende concorrenti ed è quindi fondamentale riuscire a misurarne l'impatto. Avere una misura del livello reputazionale permette di

³⁸ L'approccio data-driven considera il dato come un mezzo per raggiungere un determinato obiettivo (Brancale 2018). In particolare, l'ascolto della rete e la raccolta dei dati possono avere svariati obiettivi, anche molto diversi fra loro, come la comprensione della mentalità dei clienti, la profilazione dei clienti attuali e potenziali, l'integrazione di altre fonti di dati, l'identificazione dei problemi, la previsione dei cambiamenti del mercato e delle vendite, l'analisi dei concorrenti e il monitoraggio della reputazione (Rappaport 2012: 3).

intraprendere eventuali interventi correttivi, monitorarne i cambiamenti nel corso del tempo e prevenire eventuali crisi.

Nella prospettiva di Cuomo et al. (2012), la gestione della reputazione consiste in un processo circolare che si articola in più momenti e prevede una fase di *valutazione* e una di *misurazione*.

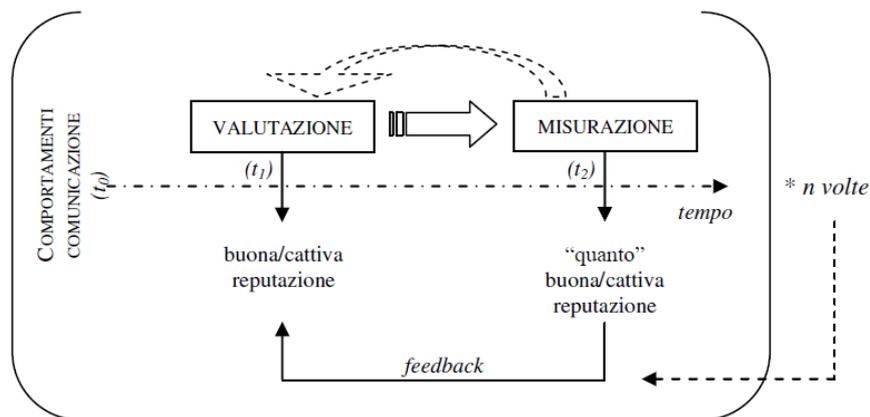


Figura 5: processo circolare della valutazione e misurazione della reputazione (Cuomo et al. 2012: 28).

Nella figura 5, il tempo t_0 rappresenta il momento in cui un'azienda fa ingresso in un nuovo mercato e deve iniziare a instaurare delle relazioni con gli interlocutori. In questo momento non è ancora possibile fare una valutazione della reputazione aziendale, in quanto questo costrutto è dipendente dagli stakeholder e non può essere determinato a priori ancor prima di instaurare delle relazioni con essi. Dopo un po' di tempo che l'azienda è presente nel mercato (al tempo t_1), inizia il processo di costruzione della reputazione che permette di accumulare capitale relazionale ed è quindi possibile iniziare una fase di valutazione. Infine, al tempo t_2 , l'azienda può procedere con la misurazione del livello reputazionale.

Nella fase di valutazione della reputazione viene messo in atto un processo di *operativizzazione*³⁹, in cui vengono individuate le dimensioni che compongono il costrutto reputazionale. Come spiega Corbetta (2014: 88) *operativizzare* un concetto significa trasformarlo in una variabile empiricamente rilevabile che permetta di

³⁹ Corbetta (2014: 89) introduce il termine di “operativizzazione” per distinguerlo da quello di “misurazione”: quando non si dispone di un'unità di misura è improprio parlare di misurazione. A seconda dei casi, il passaggio da una proprietà a una variabile, consisterà in una *classificazione*, un *ordinamento* o un *conteggio*. Il termine di operativizzazione rappresenta quindi un termine unico per definire tutte queste operazioni.

comunicare una realtà complessa in modo immediato e sintetico: “la teoria ha un senso per la scienza empirica solo nella misura in cui riesce a connettersi proficuamente con il mondo empirico; i concetti sono gli strumenti [...] per stabilire una tale connessione” (Blumer 1969 citato in Corbetta 2014: 88).

Una guida generale e metodologicamente corretta per scomporre il concetto di reputazione e individuare delle variabili rilevabili potrebbe essere il processo di traduzione empirica di un concetto complesso proposto da Corbetta (2014). La scomposizione segue diverse fasi che vengono qui elencate e spiegate usando come esempio il concetto di reputazione.

- *Fase 1 - Riflessione teorica sul concetto e individuazione delle dimensioni.*

Come si è esposto nel capitolo 2, si possono individuare diverse definizioni di reputazione. Due di queste considerano la reputazione come:

L'insieme delle immagini, delle opinioni e delle percezioni sviluppate nel tempo da clienti, fornitori e dal pubblico in generale nei confronti dell'organizzazione (Cuomo et al. 2012: 18).

L'insieme dei significati cognitivi e affettivi attribuiti dai consumatori ai prodotti [o servizi] con cui sono entrati in contatto (ibid.).

Così presentato, il concetto di reputazione ha un alto grado di astrazione: se un ricercatore si chiedesse semplicemente quale fosse la reputazione di un'azienda farebbe fatica a trovare immediatamente una risposta. Bisogna quindi rendere il concetto meno astratto identificando delle dimensioni in cui si articola. In questo caso, delle dimensioni potrebbero essere le seguenti:

- Dimensione 1: opinione degli stakeholder sui servizi dell'azienda.
- Dimensione 2: opinione degli stakeholder sulla *vision* e la *mission* dell'azienda.
- Dimensione 3: opinione degli stakeholder sull'ambiente di lavoro.

- *Fase 2: Individuazione degli indicatori*

In questa fase ci si inizia ad avvicinare alla traduzione empirica del concetto di partenza. Le dimensioni individuate possono essere suddivise in degli *indicatori*, ovvero dei concetti più semplici, ognuno dei quali è una rappresentazione solo parziale del concetto

generale di partenza. Come evidenzia Corbetta (2014: 109) “un concetto (generale) non può essere esaurito da un solo indicatore (specifico)”. Inoltre, la scelta tra i possibili indicatori è arbitraria e dipende dal ricercatore che avrà come unico dovere quello di argomentare la propria scelta di fronte alla comunità scientifica (ibid.). Ad esempio, data la dimensione 1, riguardo l’opinione degli stakeholder sui servizi dell’azienda, alcuni degli indicatori possibili possono essere:

- Opinione riguardo il servizio A
- Opinione riguardo il servizio B
- Opinione riguardo il servizio C

- *Fase 3: Operativizzazione degli indicatori*

In questa fase viene individuata una variabile relativa al singolo indicatore. Per il servizio A, ad esempio, la variabile sarà il tipo di opinione degli stakeholder e le relative modalità della variabile saranno positivo, negativo o neutro. Lo stesso ragionamento può essere applicato al servizio B e al servizio C.

- *Fase 4 (non sempre necessaria): costruzione degli indici*

La fase 4 si concentra sulla costruzione di un indice, ovvero “una variabile funzione di altre variabili che sintetizza le informazioni contenute nelle singole variabili” (Corbetta 2014: 531). Un indice permette di ritrovare l’unicità del concetto di partenza dopo che questo è stato scomposto. Questa fase è presente solo nel caso in cui ci siano più indicatori e quindi più variabili che devono essere sintetizzate da un unico indice. Se invece si sta lavorando con un concetto semplice, rappresentato da un’unica variabile, non è possibile (né necessario) trovare un indice. Nel caso qui proposto, un indice potrebbe essere il tipo di opinione su tutti i servizi.

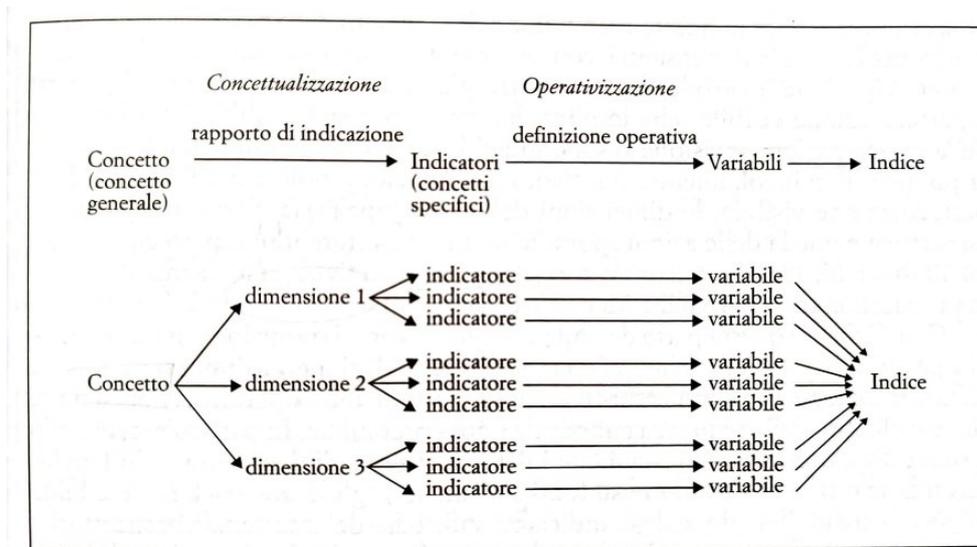


Figura 6: rappresentazione grafica del processo di traduzione empirica di un concetto complesso (Corbetta 2014: 111).

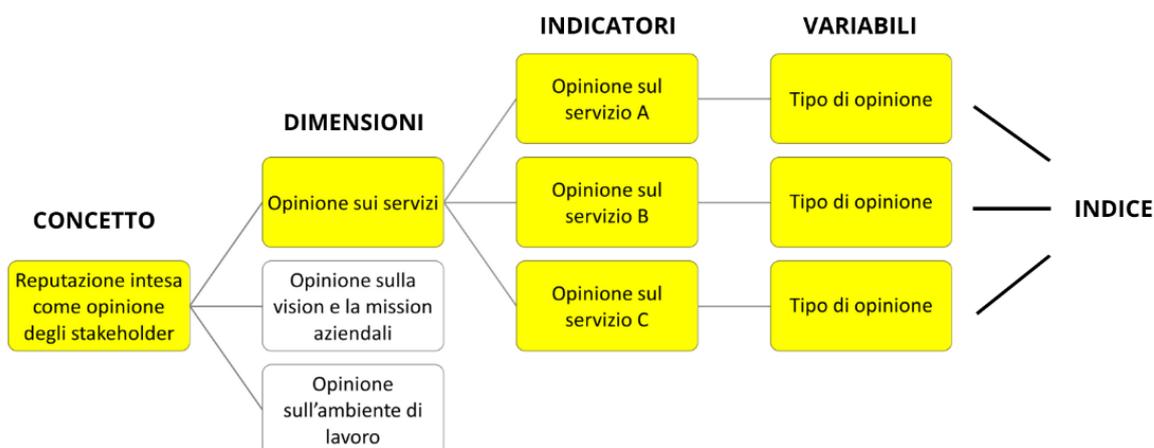


Figura 7: esempio di traduzione empirica del concetto di reputazione. I riquadri colorati evidenziano uno dei possibili percorsi della ricerca.

3.1.1 Modelli per la misurazione della reputazione

Poma e Grandoni (2021: 388) analizzano diversi modelli di misurazione della reputazione presenti in letteratura e affermano che nello stato odierno degli studi sulla reputazione aziendale non esiste una definizione universalmente condivisa di reputazione né, di conseguenza, un metodo univoco per la sua misurazione. È assente quindi una standardizzazione di concetti, modelli e procedure che porti alla costruzione di un sapere condiviso. Le dimensioni concettuali considerate nella misurazione della reputazione aziendale variano notevolmente da una fonte all'altra, soprattutto a causa delle specifiche

esigenze conoscitive delle diverse imprese. Gli autori evidenziano come spesso manchi collaborazione tra gli addetti ai lavori circa la costruzione di un modello di misurazione universalmente riconosciuto:

Emerge limpidamente la riluttanza degli autori dei modelli [analizzati] a svelare i meccanismi che sottostanno alle metodologie di misurazione della reputazione, azione che – se da una parte può essere giustificata dal vantaggio economico di governare “in solitaria” un modello di successo – rappresenta al contempo un forte limite per gli sviluppi che potrebbero derivare da una condivisione e da una riflessione su ampia scala su questi temi (Poma & Grandoni 2021: 425).

Degli esempi di valutazione della reputazione sono le classifiche reputazionali stilate da alcune riviste specializzate⁴⁰ che, a seconda dei casi, hanno individuato diverse dimensioni del costrutto reputazionale. Una delle classifiche più popolari è quella delle *World's Most Admired Companies*⁴¹ stilata ogni anno a partire dal 1997 dalla rivista Fortune, in collaborazione con la società di consulenza Korn Ferry. Per stilare questa classifica è stata svolta un'inchiesta campionaria⁴² in cui è stato chiesto a 3.740 dirigenti, direttori e analisti finanziari di selezionare le dieci società che ammiravano di più. Nel sito web di Korn Ferry⁴³ vengono riportati i nove attributi reputazionali considerati (che corrispondono quindi a diverse dimensioni del costrutto reputazionale) per stilare la classifica:

- (1) *Capacità di attrarre e trattenere persone di talento*
- (2) *Qualità della gestione*
- (3) *Responsabilità sociale nei confronti della comunità e dell'ambiente*
- (4) *Innovatività*
- (5) *Qualità dei prodotti o dei servizi*
- (6) *Uso sapiente dei beni aziendali*

⁴⁰ Fombrun et al. (2000) riportano alcuni esempi di riviste specializzate che, nel corso degli anni, hanno stilato delle classifiche reputazionali. Le riviste menzionate sono Fortune, Manager Magazin, Management Today, Asian Business, Far Eastern Economic Review, Financial Times, Industry Week.

⁴¹ <https://fortune.com/franchise-list-page/methodology-worlds-most-admired-companies-2022/> [ultima consultazione: 25/03/23]

⁴² L'inchiesta campionaria è definita da Corbetta (2014) come la ricerca condotta tramite questionario. Tale termine corrisponde al termine inglese *survey* che spesso non viene tradotto nella letteratura sociologica italiana.

⁴³ <https://www.kornferry.com/insights/this-week-in-leadership/fortune-worlds-most-admired-companies-2022#:~:text=How%20is%20it%20done%3F,their%20industries%20and%20across%20industries.> [ultima consultazione: 20/03/23]

(7) *Solidità finanziaria*

(8) *Valore dell'investimento a lungo termine*

(9) *Efficacia nel fare affari a livello globale*

Fombrun⁴⁴ et al. (2000) esaminano otto classifiche reputazionali (tra cui quella delle *World's Most Admired Companies*) stilate da riviste specializzate (Fortune, Manager Magazin, Management Today, Asian Business, Far Eastern Economic Review, Financial Times, Industry Week) ed evidenziano come le ricerche alla base di tali classifiche siano caratterizzate da una “debolezza concettuale e metodologica”, in quanto spesso coinvolgono un gruppo limitato di stakeholder e si concentrano soprattutto su aspetti finanziari. Secondo gli autori (Fombrun et al. 2000) un metodo per la misurazione della reputazione che possa definirsi efficace deve rispettare due requisiti:

- Considerare la molteplicità di stakeholder da cui dipende il costrutto reputazionale.
- Considerare tutte le dimensioni del costrutto reputazionale (non solo la dimensione finanziaria).

Per questo motivo, a partire dalle dimensioni considerate dalle classifiche delle riviste, gli autori (ibid.) svolgono dei focus group coinvolgendo diversi stakeholder e individuano delle dimensioni del costrutto reputazionale che ritengono essere più bilanciate. Il modello proposto si chiama *Reputation Quotient Model* ed è composto da sei dimensioni:

Dimensione	Significato
(1) Emotional appeal	Percezione della capacità dell'impresa di suscitare negli stakeholder una complessiva sensazione di piacevolezza, ammirazione, rispetto e fiducia.
(2) Prodotti e servizi	Percezione che gli stakeholder hanno dell'innovazione, della qualità, del rapporto

⁴⁴ Charles Fombrun è il fondatore e presidente emerito di The RepTrak Company, azienda e centro di ricerca americano, precedentemente noto come Reputation Institute. Fombrun ha fondato il Reputation Institute nel 1997 e oggi è uno dei centri di ricerca più importanti nell'ambito degli studi sulla reputazione.

	qualità/prezzo rispetto ai prodotti e ai servizi dell'azienda.
(3) Vision e leadership	Percezione che gli stakeholder hanno della vision aziendale, della capacità di sfruttare le opportunità del mercato e della qualità della leadership.
(4) Ambiente di lavoro	Percezione della qualità della vita aziendale e della capacità dell'azienda di attrarre talenti.
(5) Responsabilità sociale e ambientale	Percezione della responsabilità ambientale dell'azienda e della sua capacità di instaurare buoni rapporti con le comunità locali.
(6) Performance reddituale e finanziaria	Percezione relativa alle prospettive di crescita dell'azienda, alla possibilità di registrare performance migliori rispetto ai concorrenti e di gestire gli investimenti.

Tabella 4: dimensioni individuate da Fombrun et al. (2000), rielaborate da Cuomo et al. (2012).

Cuomo et al. (2012) spiegano che le dimensioni proposte da Fombrun et al. (2000) costituiscono diverse aree di valore che possono portare un'azienda a raggiungere un vantaggio competitivo e prendono in considerazione anche il sentiment generato da parte di diversi stakeholder. La reputazione aziendale può quindi essere scomposta in diverse dimensioni e l'azienda, a seconda delle proprie esigenze e della fase del ciclo di vita in cui si trova, può decidere di focalizzarsi su un determinato aspetto. Gli autori (Cuomo et al. 2012: 37) suggeriscono al ricercatore di “guardare sempre alla reputazione come un fenomeno complesso [...] che si dipana su più piani di analisi”.

Un altro modello di misurazione della reputazione è il RepTrak⁴⁵ Pulse, sviluppato dal Reputation Institute (dal 2020 rinominato RepTrak), un'azienda di Boston fondata nel 1997 dallo stesso Charles Fombrun del Reputation Quotient Model e da Cees van Riel. Dal 2023 il CEO è Mark Sonders. Il RepTrak Pulse è stato sviluppato per sostituire il precedente Reputation Quotient Model e fornire un modello di misurazione che fosse

⁴⁵ <https://www.reptrak.com/> [ultima consultazione: 25/03/23]

ancora più sintetico⁴⁶, che agevolasse nella raccolta e nell’analisi dei dati sulla reputazione e che potesse essere utilizzato in più contesti culturali (Fombrun et al. 2011). Secondo Poma e Grandoni (2021), oggi questo modello è quello con maggiori conferme sperimentali di validità, soprattutto per quanto riguarda la validità a livello cross-culturale e l’indipendenza dal gruppo di stakeholder considerato. Il gruppo Unipol⁴⁷, che utilizza questo modello, spiega infatti come esso adotti un approccio integrato che permette di misurare la reputazione su più gruppi di stakeholder. “Pulse” rimanda al cuore pulsante del modello, ovvero alla componente emotiva che permette di creare una relazione tra stakeholder e azienda. La forza di tale legame dipende dagli attributi di stima, fiducia, ammirazione e atteggiamento positivo ed è indice della reputazione aziendale. Il modello individua sette pilastri, spiegati da Cuomo et al. (2012), su cui l’azienda si fonda e può lavorare per migliorare la percezione degli stakeholder:

DIMENSIONE	INDICATORI
Prodotti e servizi (Products & Services)	Qualità di prodotti e servizi.
Innovazione (Innovation)	Livello di innovazione.
Ambiente di lavoro (Workplace)	L’ambiente aziendale è di qualità e i dipendenti vengono trattati in modo adeguato.
Eticità (Governance)	L’azienda ha un comportamento etico.
Responsabilità sociale (Citizenship)	L’azienda supporta delle buone cause e protegge l’ambiente con cui si relaziona.
Leadership	Nell’azienda è presente una forte leadership ed è gestita efficacemente.

⁴⁶ Secondo gli autori (Fombrun et al. 2011) i modelli precedenti proposti dalla letteratura si basavano su troppi attributi reputazionali (dai 10 ai 20) e ponevano domande ridondanti ai partecipanti alle ricerche. Lo stesso Reputation Quotient Model elaborato da Fombrun et al. nel 2000 utilizza venti variabili associate alle varie dimensioni.

⁴⁷ <https://www.unipol.it/it/rep-trak-model> [ultima consultazione: 26/03/23]

Performance	L'azienda ha una performance e dei risultati finanziari di qualità.
-------------	---

Tabella 5: dimensioni e indicatori del modello Rep Trak Pulse di Rep Track (Reputation Institute 2009 adattato da Cuomo et al. 2012: 41).

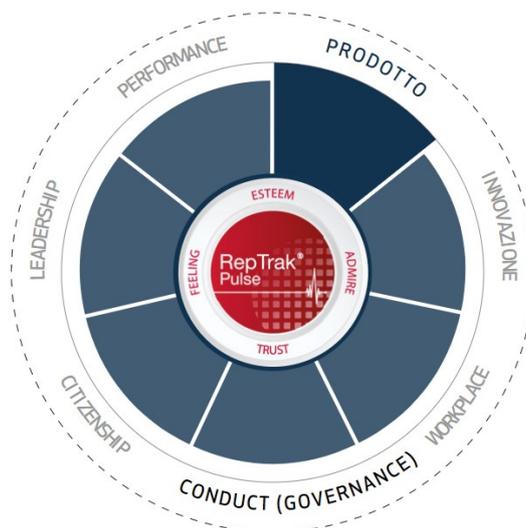


Figura 8: il Reputation Pulse Model (Unipol⁴⁸).

Alcuni modelli si concentrano su stakeholder specifici, come nel caso del *Corporate Character* (Davies et al. 2004), un modello di misurazione della reputazione che si focalizza sulla percezione della reputazione di un'organizzazione da parte dei dipendenti e dei clienti, stakeholder di estrema importanza soprattutto nel settore dei servizi, in cui la prospettiva di un dipendente può influenzare direttamente quella di un consumatore. Gli autori di questo modello utilizzano una personificazione, in quanto l'azienda viene considerata come un'entità dotata di carattere (*character*), ovvero un insieme di caratteristiche umane che la rendono distinguibile agli occhi degli stakeholder. Le dimensioni in base alle quali viene valutata la reputazione sono piacevolezza (*agreeableness*), intraprendenza (*enterprise*), competenza (*competence*), eleganza (*chic*), spietatezza (*ruthlessness*), informalità (*informality*), mascolinità (*machismo*).

In letteratura sono presenti anche dei modelli che si focalizzano in modo specifico sulla misurazione della reputazione nel contesto online. Come fanno notare Cuomo et al. (2012: 46-47) “ogni processo di monitoraggio della corporate reputation condotto

⁴⁸ <https://www.unipol.it/it/rep-trak-model> [ultima consultazione: 26/03/23]

nell'arena competitiva reale – *offline* – non può mancare di integrare le valutazioni raccolte con quanto dell'impresa si riferisce da più parti nell'arena competitiva virtuale – *online*". Cuomo et al. (2012) spiegano che i modelli che analizzano la reputazione online possono:

- prendere in considerazione la presenza dell'azienda online (*brand presence*), contando il numero di contenuti e menzioni presenti su tale azienda;
- analizzare le tematiche più discusse riguardo l'azienda;
- analizzare le fonti da cui provengono le varie opinioni, quindi capire dove si concentrano le opinioni e da quali stakeholder provengono;
- analizzare il *sentiment* riguardo un'azienda, ovvero capire come si parla dell'azienda (se in modo positivo o negativo).

Inoltre, il centro di ricerca Ipsos⁴⁹ ha sviluppato RISE⁵⁰ (*Reputation Intelligence for Strategic Evaluation*), un sistema per combinare i dati provenienti dai sondaggi con quelli provenienti da altre fonti (come i media tradizionali o i social media). Questo sistema parte dai dati provenienti dai sondaggi per organizzare quelli provenienti dai social media ed è anche in grado di filtrare i risultati in base agli stakeholder di interesse (e.g.: opinion leader, politici).

I modelli esistenti sono numerosi e attualmente non si è ancora giunti a un metodo universalmente condiviso per misurare la reputazione di un'azienda nella sua totalità. Un ricercatore, anche nel caso in cui fosse consapevole che il concetto di reputazione è multidimensionale e dipendente da una molteplicità di stakeholder, può compiere delle scelte discrezionali e decidere di circoscrivere la propria ricerca a determinate dimensioni e stakeholder. Nell'ottica della sociologia comprendente (Sbalchiero 2021) le scelte discrezionali del ricercatore non pregiudicano l'oggettività della ricerca. Durante un processo di ricerca scientifica il ricercatore deve eliminare l'influenza del proprio *giudizio di valore*, cioè non deve esprimere una posizione soggettiva rispetto al fenomeno studiato. Tuttavia, la definizione dell'oggetto di studio della ricerca sarà inevitabilmente frutto di una scelta discrezionale, ovvero di una *relazione al valore* che consiste nelle motivazioni

⁴⁹ Ipsos è una società multinazionale con sede a Parigi che si occupa di ricerche di mercato e consulenza (Wikipedia).

⁵⁰ <https://www.ipsos.com/it-it/corporate-reputation> [ultima consultazione: 30/03/23]

personali e negli interessi conoscitivi del ricercatore. Una ricerca può definirsi oggettiva se lo scienziato esplicita le motivazioni (relazione al valore) e i passaggi che lo hanno condotto alla costruzione dell'oggetto di studio e ai risultati, "in modo che tutto sia chiaro, immediato (e riproducibile) anche al lettore di un'altra cultura. E questo, indipendentemente dai giudizi di valore" (Sbalchiero 2021: 48).

3.1.2 Strumenti per il monitoraggio della reputazione online

Online sono presenti diversi strumenti, la maggior parte a pagamento, per il monitoraggio della reputazione. Uno strumento gratuito è Google Alert⁵¹ che permette di monitorare i contenuti pubblicati riguardo una determinata parola chiave inserita dall'utente e di ricevere avvisi via e-mail ogni qualvolta ci siano delle novità. Degli strumenti a pagamento invece sono SEMRush⁵² che offre diversi tipi di servizi, tra cui il monitoraggio delle menzioni online e del sentiment, la visualizzazione degli account raggiunti e la rilevazione degli influencer del settore; Mention⁵³ per tenere sotto controllo le menzioni dell'azienda nei social media; SentiONE⁵⁴ che consente di accedere a dati real-time o in archivio di più social network contemporaneamente, di controllare le menzioni, i profili, le keyword, di visualizzare i dati e creare dei report; Reputology⁵⁵ e ReviewPush⁵⁶ per controllare e gestire le recensioni online sulla propria azienda e i relativi servizi/prodotti; Reputation Rating⁵⁷, invece permette di incrociare le dimensioni reputazionali con gli stakeholder d'interesse e di ottenere dei dati a riguardo.

Per monitorare la reputazione nel Web 2.0, in alternativa agli strumenti online, che offrono soluzioni dal semplice utilizzo ma spesso a pagamento e poco personalizzabili, un ricercatore può creare degli algoritmi ad hoc tramite un linguaggio di programmazione che, ad esempio, può essere utilizzato per elaborare il linguaggio naturale (come quello contenuto nelle recensioni o nei post online). In questa ricerca, come si vedrà, è stato utilizzato il linguaggio di programmazione Python.

⁵¹ <https://www.google.it/alerts> [ultima consultazione: 25/03/23]

⁵² <https://it.semrush.com/> [ultima consultazione: 25/03/23]

⁵³ <https://mention.com/en/> [ultima consultazione: 25/03/23]

⁵⁴ <https://sentione.com/> [ultima consultazione: 25/03/23]

⁵⁵ <https://www.reputology.com/> [ultima consultazione: 25/03/23]

⁵⁶ <https://www.reviewpush.com/> [ultima consultazione: 25/03/23]

⁵⁷ <https://www.reputationrating.it/> [ultima consultazione: 25/03/23]

3.2 Twitter come fonte di dati sull'opinione dei clienti

Twitter è una piattaforma di microblogging dedicata allo sviluppo di conversazioni online. Queste ultime possono essere considerate una fonte preziosa da cui attingere per ottenere dei dati sull'opinione che i clienti hanno riguardo un'azienda e i suoi concorrenti e quindi monitorare la reputazione. Alcune ricerche inerenti all'anno 2022 offrono una panoramica dell'utilizzo di Twitter a livello globale:

- Secondo una ricerca di BusinessDIT (Yakub 2023), sono stati pubblicati in media 867 milioni di tweet al giorno (602.000 tweet al minuto).
- I dati di Socialpilot (Bagadiya 2023) informano che gli utenti attivi mensilmente sono stati 368 milioni.
- La ricerca di Hootsuite (Martin 2023) mostra come Twitter sia il settimo social media preferito, dopo Whatsapp, Facebook, Instagram, WeChat, Douyin, TikTok.

Dato l'ingente numero di utenti e di tweet pubblicati, Twitter rappresenta per le aziende una fonte da cui raccogliere grandi quantità di dati testuali che, nel momento in cui vengono analizzati, possono aiutare a indirizzare o correggere le proprie strategie comunicative e di marketing. Diversi studi evidenziano i vantaggi che derivano dall'utilizzo di Twitter come fonte di dati:

- I caratteri limitati dei tweet portano gli utenti ad essere più concisi e più espressivi rispetto ai contenuti che vengono pubblicati su altri social network e nei blog (Kontopoulos et al. 2013). Twitter risulta quindi essere una piattaforma di microblogging particolarmente adatta alla rilevazione del tipo di opinioni espresse nei testi, ovvero l'*opinion mining*. (Pak & Paroubek 2010).
- È possibile raccogliere un grande numero di tweet e quindi creare un ampio corpus di dati testuali (id).
- Si possono raccogliere tweet scritti in diverse lingue (ibid.).
- Gli argomenti di discussione su Twitter sono svariati (Thelwall et al. 2010): a questo proposito, Twitter pubblica nel suo sito dedicato al marketing⁵⁸ delle informazioni che possono risultare molto utili per i ricercatori e gli addetti ai

⁵⁸ <https://marketing.twitter.com/en> [ultima consultazione: 25/03/23]

lavori. Ad esempio, alla fine di ogni anno viene condiviso con gli utenti iscritti alla newsletter un report chiamato *The Conversation: Twitter Trends*, in cui vengono indicati gli argomenti su cui gli utenti hanno pubblicato più tweet.

- La raccolta delle informazioni tramite Twitter permette di effettuare una ricerca meno intrusiva nei confronti degli utenti e spesso meno dispendiosa per il ricercatore, soprattutto in termini di tempo, in quanto i dati sono già disponibili nel web (Bhatia et al. 2013; Misopoulos et al. 2014). In particolare, Hasson et al. (2019) esaminano l'utilizzo di dati provenienti da Twitter come alternativa ai dati ottenuti dai sondaggi. Nel loro caso specifico, estrarre i dati da Twitter ha avuto i seguenti vantaggi:
 - I dati provenienti da Twitter possono includere l'opinione di un gruppo più diversificato di utenti, inclusi i dipendenti dell'azienda, i quali non sempre accettano di partecipare ai sondaggi. Nell'ottica degli studi sulla reputazione aziendale, ciò significa ottenere dati provenienti da stakeholder molto diversi fra loro.
 - Non viene richiesto alcuno sforzo all'utente: scrivere un tweet richiede meno tempo rispetto a un sondaggio ed è un'azione guidata da un interesse e un'esigenza personale che non ha quindi come obiettivo quello di fornire dei dati al ricercatore. Come evidenziano anche Austin e Wittes Schlack (2010), ciò comporta una partecipazione più autentica da parte degli utenti che esprimono liberamente la propria opinione dai propri dispositivi e in qualsiasi luogo vogliano senza che il ricercatore interferisca nelle loro vite.
 - Raccogliere tweet con metodi automatici è una pratica potenzialmente meno costosa rispetto a un sondaggio condotto su larga scala. Come evidenzia anche Corbetta (2014), grazie alla presenza dei dati online, il ricercatore può raccogliere i dati direttamente da casa propria, azzerando i costi di spostamento per raggiungere la fonte o per trasferirla e i costi di rilevazione e di registrazione dei dati in database informatici.
 - Twitter permette di raccogliere e monitorare i dati in modo costante e in real-time, mentre il sondaggio è svolto con una frequenza molto minore: nel caso analizzato da Hasson et al. (2019) il sondaggio viene usato dall'azienda su base annuale.

È necessario però riflettere sui limiti e sulle sfide che pone la rilevazione dei dati online. Corbetta (2014) afferma che la grande ricchezza di dati provenienti dal web porta anche a una difficoltà di gestione e controllo dei dati, di selezione, archiviazione e, come si vedrà, di generalizzabilità dei risultati. La ricerca tramite i dati provenienti da Internet è un ambito in continua evoluzione e pone dei quesiti di natura metodologica che sono ancora irrisolti.

3.2.1 Monitorare l'attività del servizio clienti

Tra le informazioni che un'azienda può ricavare da Twitter vi sono quelle provenienti dai clienti che scambiano opinioni fra loro o con l'azienda stessa. Le aziende, infatti, possono essere presenti su Twitter con i loro account ufficiali e usare questi ultimi per il servizio clienti, ovvero per l'assistenza fornita "ai clienti prima, durante e dopo l'acquisto di un prodotto/servizio" (Live Agent, n.d.). Tra i casi esemplari negli Stati Uniti, la sezione Business di Twitter (Alton, n.d.) menziona Warby Parker (rivenditore online di occhiali da vista e da sole), Adobe XD (software di progettazione), Evernote (applicazione), Glossier (sito web del settore Beauty). In un articolo del 2016 pubblicato nel blog ufficiale di Twitter (Twitter 2016) si sottolinea come questo social network permetta ai clienti non solo di comunicare più velocemente con le aziende, ma anche di poterlo fare in una struttura da uno a molti, in cui il cliente che chiede assistenza non si interfaccia solo con l'azienda ma anche con gli altri utenti che potrebbero prendere parte alla conversazione.

Blunt e Hill-Wilson (2013) evidenziano come alcune aziende decidano consapevolmente di non partecipare alla conversazione con gli utenti per paura di ricadute reputazionali. La scarsa partecipazione però, come già esposto nel capitolo 2, non è una strategia adeguata a risolvere i problemi e prevenire eventuali crisi. Un *reputation manager* (e, più in generale, un comunicatore) deve presidiare le conversazioni e gestire i contenuti dannosi per l'azienda (Lampignano 2016), come ad esempio un commento negativo di un utente che può amplificarsi e raggiungere un'audience molto ampia.

Inoltre, gli autori (Blunt & Hill-Wilson 2013) fanno notare che l'utilizzo di Twitter come servizio clienti aiuta a risolvere i problemi più semplici e alleggerire il carico di richieste che arriva ad altri canali dell'azienda (e.g. sito web, call center). I caratteri limitati del tweet non permettono sempre di fornire una soluzione definitiva al problema

del cliente, ma possono indicare la strada da percorrere per trovarla. L'operatore del servizio clienti, ad esempio, potrebbe portare l'utente a passare a un canale comunicativo alternativo indicando un numero di telefono, una e-mail o un link grazie ai quali troverà le informazioni più adatte alla risoluzione del proprio problema (ibid.).

Alcune ricerche sul servizio clienti (Twitter 2016; Cover 2021) riportano dei dati che potrebbero tornare utili alle aziende:

- L'83% degli intervistati che hanno utilizzato il servizio clienti è riuscito a risolvere il problema per cui ha chiesto assistenza (Twitter 2016).
- Gli utenti utilizzano il servizio clienti soprattutto per rivolgersi ad aziende del settore della vendita al dettaglio e dei viaggi (Twitter 2016).
- Tra i fattori determinanti per un'esperienza positiva con il servizio clienti vengono menzionati la personalizzazione e il tono amichevole dei tweet pubblicati dall'azienda (Twitter 2016) e la reattività, intesa come tempo di risposta (Cover 2021).

Per monitorare su Twitter l'attività di un'azienda e dei suoi concorrenti e le opinioni espresse dagli utenti può risultare utile seguire il modello di analisi proposto da Rodríguez-Díaz et al. (2018), chiamato *gap model*: esso aiuta a individuare i problemi che un'azienda può avere nella gestione della reputazione online⁵⁹ e tiene conto anche “delle dinamiche competitive dei mercati, dove tutte le aziende sono focalizzate a soddisfare i bisogni dei clienti e fidelizzarli” (ivi, p.11, traduzione mia). Nel modello sono descritte sette possibili carenze (chiamate *gap*) che l'azienda potrebbe dover colmare durante il monitoraggio della propria reputazione online e di quella dei concorrenti:

(1) *GAP 1: Valutazione delle opinioni dei clienti*

Questo gap riguarda la capacità dell'azienda di catturare, analizzare e reagire alle recensioni che i clienti pubblicano online. Un'attività di monitoraggio costante può rappresentare un vantaggio competitivo rispetto ai concorrenti e, in generale, uno strumento per riuscire a comprendere e far fronte alle esigenze dei clienti.

(2) *GAP 2: Coerenza dell'opinione dei clienti*

⁵⁹ Il modello proposto dagli autori non fa riferimento a un social network specifico, bensì al contesto online generico.

La coerenza riguarda il livello di convergenza o divergenza esistente nel tipo di opinioni che i clienti esprimono attraverso i vari canali comunicativi: ad esempio, è possibile che in un determinato sito (es: Twitter) i clienti abbiano espresso mediamente un parere positivo sull'azienda, mentre in un altro sito (es: Trustpilot) il parere è negativo.

(3) *GAP 3: Verità nelle opinioni dei clienti*

La reputazione online non sempre corrisponde alla reputazione reale perché, secondo gli autori (Rodríguez-Díaz et al. 2018), i clienti insoddisfatti tendono a commentare di più rispetto a quelli soddisfatti. È necessario quindi che l'azienda svolga una ricerca (ad esempio tramite un sondaggio) con un campione rappresentativo di clienti, in modo da ottenere una valutazione della reputazione il più possibile vicina al reale, che sarà poi confrontata con le opinioni online dei clienti. Nel caso in cui ci fosse un disallineamento, l'azienda dovrà prendere delle misure correttive e fare in modo che le aspettative dei clienti siano vicine alla realtà.

(4) *GAP 4: Intensità della comunicazione aziendale*

Questo gap si riferisce al livello di visibilità, impatto e reattività di un'azienda nei canali online, ovvero al numero di azioni svolte e alla varietà dei canali utilizzati. Una comunicazione che si traduce in più azioni e che avviene su più canali aumenta le possibilità di un'azienda di ottenere visibilità. L'intensità della comunicazione aziendale può essere misurata, ad esempio, con Google Analytics che fornisce dei dati sul traffico di una determinata pagina web.

(5) *GAP 5: Qualità di prodotti e servizi* + (6) *GAP 6: Valore di prodotti e servizi*

Idealmente, la qualità e il valore percepiti dai clienti relativamente a prodotti e servizi dovrebbero superare le aspettative del cliente. Se ciò accade, l'azienda ha un alto livello di reputazione e un vantaggio competitivo. Monitorare e colmare questi due gap consiste nel determinare quanto valore e qualità attesi differiscano da valore e qualità percepiti.

(7) *GAP 7: Valutazione dei concorrenti*

Questo aspetto indica quanto l'azienda si impegni a ottenere informazioni sull'attività dei propri concorrenti per poi valutarle e prendere delle decisioni mirate.

3.2.2 Le API di Twitter e la piattaforma Twitter Developer

I tweet che gli utenti scelgono di condividere pubblicamente (che in questo caso costituiscono dei dati testuali) possono essere ottenuti da un ricercatore tramite le API di Twitter, in modo da costruire un *dataset*⁶⁰ ed effettuare un'analisi. Un'API (*application programming interface*, traducibile in italiano come “interfaccia di programmazione delle applicazioni”) è un intermediario che permette a due componenti software di scambiare informazioni (Salesforce 2022). Le API creano un collegamento tra un'applicazione che fa una richiesta (client) e un'applicazione che fornisce una risposta (server)⁶¹. Il sito ufficiale di Twitter⁶² spiega che ciò che l'applicazione (client) richiede è un *endpoint*, ovvero un indirizzo univoco (come fosse un numero di telefono) che permette di ottenere determinate informazioni da parte del server. Gli endpoint consentono a un'azienda di integrare i propri dati e le proprie applicazioni con le informazioni provenienti da Twitter. Per accedere alle API di Twitter bisogna registrare un'applicazione che, per impostazione predefinita, riesce ad accedere solo alle informazioni pubbliche di Twitter.

Gli endpoint disponibili su Twitter sono di cinque tipologie:

1. Endpoint per gestire account e utenti: essi permettono di gestire il proprio profilo, le impostazioni e i follower.
2. Endpoint per gestire i tweet e le risposte: con questo tipo di endpoint si possono pubblicare, scaricare e monitorare i tweet. Come esempio, il sito di Twitter cita una ricerca della Northeastern University⁶³ che permette di monitorare e prevedere le epidemie di influenza tramite i dati ottenuti da Twitter.
3. Endpoint per gestire i messaggi diretti: questo tipo di endpoint non viene concesso in maniera predefinita, bensì solo dopo che l'utente ha fornito un'autorizzazione aggiuntiva dal proprio account. Un esempio di questo endpoint citato da Twitter è quello delle aziende che gestiscono le interazioni con i clienti tramite un chatbot⁶⁴.

⁶⁰ Un dataset è una collezione di dati.

⁶¹ <https://www.salesforce.com/it/blog/2022/09/definizione-application-programming-interface-API.html> [ultima consultazione: 25/03/23]

⁶² <https://help.twitter.com/it/rules-and-policies/twitter-api> [ultima consultazione: 25/03/23]

⁶³ <https://help.twitter.com/it/rules-and-policies/twitter-api> [ultima consultazione: 25/03/23]

⁶⁴ Un chatbot è un software che viene progettato per simulare delle conversazioni con gli esseri umani via chat (Wikipedia).

4. Endpoint per gestire gli annunci: grazie a questo endpoint le aziende possono creare e gestire le proprie campagne pubblicitarie su Twitter e raggiungere diversi target.
5. Endpoint che usano SDK (*Software Development Kit*): Twitter fornisce anche degli SDK, cioè degli strumenti per incorporare i contenuti provenienti da Twitter (come pulsanti e cronologie). Un'azienda, ad esempio, può inserire nel proprio sito le conversazioni di Twitter che stanno avvenendo in tempo reale.

Nel sito di Twitter dedicato agli sviluppatori (chiamato Developer Platform⁶⁵) sono contenute le informazioni sulle API e le indicazioni per creare un'applicazione che permetta di "chiamare" un endpoint. L'ultima versione delle API di Twitter è "Twitter API v2"⁶⁶ ed è caratterizzata da tre possibili livelli di accesso: *Essential*, *Elevated* e *Academic Research*. Alcune delle differenze tra i tre tipi di accesso sono mostrate nella seguente tabella:

	Essential	Elevated	Academic Research
Accesso	Da definire	Da definire	Da definire
Prezzo	Da definire	Da definire	Da definire
Numero massimo di progetti	1	3	1
Numero massimo di applicazioni	1	3	1
Numero massimo di tweet scaricabili	500mila tweet al mese	2milioni di tweet al mese	10milioni di tweet al mese

⁶⁵ <https://developer.twitter.com/en> [ultima consultazione: 25/03/23]

⁶⁶ <https://developer.twitter.com/en/docs/twitter-api/getting-started/about-twitter-api>[ultima consultazione: 25/03/23]

Accesso all'intero archivio di tweet	No	No	Sì
---	----	----	----

Tabella 6: differenze tra le tre modalità di accesso alle API di Twitter, descritte nel sito ufficiale Twitter Developer⁶⁷. Vengono qui riportate le differenze rilevanti per questa ricerca.

Fino a gennaio 2023 l'utilizzo delle API è stato gratuito: la modalità *Essential* era aperta a tutti, mentre per usufruire delle modalità *Elevated* e *Academic Research* era necessario mandare una richiesta di approvazione a Twitter, in cui lo sviluppatore⁶⁸ specificava in che modo e con che finalità avrebbe utilizzato il servizio. Da febbraio 2023, invece, le modalità di accesso hanno subito delle modifiche: su Twitter, nella pagina @TwitterDev si legge che verrà proposto un abbonamento al posto delle API gratuite, ma non sono ancora state fornite informazioni precise su come usufruire delle modalità di accesso e a che prezzo (per questo motivo nella tabella 6 alcune informazioni sono “da definire”).



Figura 9: comunicazione da parte di Twitter Dev⁶⁹ sull'introduzione dei piani a pagamento.

Molti sviluppatori e giornalisti (Powers 2023; University of Washington 2023; Arnoldi 2023) hanno criticato la decisione del nuovo CEO di Twitter Elon Musk⁷⁰: l'introduzione di un canone mensile, di cui ancora non si conosce l'entità, potrebbe rappresentare una battuta d'arresto per lo sviluppo dei progetti di studenti, accademici o piccole aziende che non sempre hanno grandi budget a disposizione. Le API a pagamento potrebbero portare

⁶⁷ <https://developer.twitter.com/en/docs/twitter-api/getting-started/about-twitter-api> [ultima consultazione: 25/03/23]

⁶⁸ Uno sviluppatore è un creatore di programmi per calcolatori (Enciclopedia Treccani).

⁶⁹ https://twitter.com/TwitterDev/status/1621026986784337922?ref_src=twsrc%5Etfw%7Ctwcamp%5Etwetembed%7Ctwterm%5E1621153089381343233%7Ctwgr%5Eaac6efb485db47e0e7770474a416823452d6c2d1%7Ctwcon%5Es3_&ref_url=https%3A%2F%2Fwww.grid.news%2Fstory%2Ftechnology%2F2023%2F02%2F02%2Fcharging-for-twitter-api-access-risks-killing-the-things-that-made-twitter-great%2F [ultima consultazione: 25/03/23]

⁷⁰ Elon Musk è diventato il CEO di Twitter il 28 ottobre 2022, acquistando la società per 44 miliardi di dollari (Gaggi 2022).

alla scomparsa di quei bot⁷¹ che, fino adesso, sono riusciti a fornire gratuitamente aggiornamenti importanti per la sicurezza pubblica, come quelli riguardanti i disastri naturali, le pandemie, gli attentati terroristici o le guerre.

Le API usate in questo elaborato sono state richieste prima dell'introduzione dei piani a pagamento e la modalità di accesso utilizzata è l'*Academic Research*, di cui possono usufruire laureandi magistrali, dottorandi o ricercatori che fanno parte di un'istituzione accademica. Rispetto alle altre opzioni, questo tipo di accesso permette di scaricare un numero maggiore di tweet (10 milioni al mese) e, grazie all'endpoint *Full-archive search*⁷², di accedere a un archivio di dati che contiene tutti i tweet pubblicati online a partire da marzo 2006⁷³, data di nascita di Twitter. Selezionando un arco temporale specifico nella query di ricerca, è possibile studiare conversazioni del passato o monitorare quelle del presente. Gli altri livelli di accesso (*Essential* ed *Elevated*), invece, prevedono l'utilizzo dell'endpoint *Recent Search*⁷⁴ che consente di scaricare solamente i tweet pubblicati nell'ultima settimana.

Per monitorare le conversazioni su Twitter è sufficiente creare un'*app only authentication*, ovvero un'applicazione che permette solo di accedere alle informazioni pubbliche su Twitter e non di interagire con i tweet o con gli utenti (come nel caso delle applicazioni che permettono di pubblicare tweet, seguire o bloccare i follower). Gli endpoint vengono chiamati senza che l'utente debba fornire le proprie credenziali di accesso, poiché si accede a informazioni di pubblico dominio. Per accedere alle API di Twitter è necessario seguire i passaggi descritti nella Developer Platform⁷⁵:

1. Creare un account per sviluppatori (*Developer Account*): chiunque abbia un normale account Twitter può accedere alla piattaforma per sviluppatori e creare un progetto e la relativa app. Nel caso di questa ricerca, è stato richiesto l'accesso nella modalità *Academic Research*.

⁷¹ Un bot è un programma automatizzato che accede alla rete attraverso gli stessi canali degli utenti. (Wikipedia).

⁷² <https://developer.twitter.com/en/docs/twitter-api/tweets/search/introduction> [ultima consultazione: 25/03/23]

⁷³ Il primo tweet risale al 21 marzo 2006 e fu pubblicato da Jack Dorsey, uno dei co-fondatori di Twitter: <https://twitter.com/jack/status/20> [ultima consultazione: 20/04/23]

⁷⁴ <https://developer.twitter.com/en/docs/twitter-api/tweets/search/introduction> [ultima consultazione: 25/03/23]

⁷⁵ <https://developer.twitter.com/en/docs/twitter-api/getting-started/getting-access-to-the-twitter-api> [ultima consultazione: 25/03/23]

2. Salvare le credenziali: durante la creazione del progetto vengono generate delle credenziali da utilizzare durante ogni richiesta. Esse sono:
 - *API Key and Secret*: consistono in uno username e in una password.
 - *Access Token and Secret*: permettono di identificare l'utente che sta facendo la richiesta.
 - *App Only Access Token – Bearer Token*⁷⁶: il Bearer Token permette di identificare l'applicazione che sta facendo la richiesta al server di Twitter.

Nel caso del monitoraggio dei tweet tramite un'*app only authentication*, l'unica credenziale di cui si ha bisogno è il bearer token.

3. Utilizzare le credenziali per fare delle richieste⁷⁷ alle API di Twitter. Ciò può essere fatto tramite un linguaggio di programmazione come, ad esempio, Python.

3.3 La sentiment analysis

I tweet a cui è possibile avere accesso tramite le API di Twitter costituiscono dei dati testuali su cui è possibile effettuare delle analisi nell'ambito del *Natural Language Processing*⁷⁸ (NLP, in italiano *Elaborazione del linguaggio naturale*): come spiegato da diversi autori (Chaudhari 2020; Rajput 2020; Sensini 2019), esso consiste in un insieme di tecniche computazionali per l'analisi automatica e la rappresentazione del linguaggio naturale. Si tratta di applicare degli algoritmi in modo tale che i dati provenienti dal linguaggio naturale non strutturato possano essere convertiti in un formato comprensibile a un computer. Per chiarire la definizione di elaborazione del linguaggio naturale è necessario fare alcune precisazioni (Liddy 2001):

- Per linguaggio naturale si intende un testo orale o scritto di qualsiasi tipo e genere (ovvero di qualsiasi varietà linguistica⁷⁹). L'unico requisito è che il testo

⁷⁶ <https://developer.twitter.com/en/docs/authentication/oauth-2-0/bearer-tokens> [ultima consultazione: 25/03/23]

⁷⁷ <https://developer.twitter.com/en/docs/tutorials/step-by-step-guide-to-making-your-first-request-to-the-twitter-api-v2> [ultima consultazione: 25/03/23]

⁷⁸ Nissim e Pannitto (2022) segnalano che, come sinonimo di Natural Language Processing, viene usata anche la denominazione *Computational Linguistics* (CL, in italiano *Linguistica computazionale*).

⁷⁹ Le varietà di una lingua sono le diverse attualizzazioni di tale lingua da parte dei parlanti. Esse dipendono da cinque parametri (Bonomi et al. 2011): *diamesia* che è la variazione della lingua secondo il

sia usato dagli esseri umani per comunicare in un contesto naturale e non sia costruito artificialmente per l'analisi da effettuare (ibid.). In questa ricerca il linguaggio naturale analizzato è quello contenuto nei tweet.

- L'elaborazione del linguaggio naturale si può sviluppare su uno o più livelli di analisi linguistica (fonologico, morfologico, lessicale, sintattico e semantico), proprio come accade quando l'uomo produce e comprende il linguaggio. Ciò si traduce nell'esistenza di diversi metodi e tecniche, ognuna delle quali permette di effettuare una determinata analisi linguistica.
- L'elaborazione del linguaggio naturale è considerata una branca dell'intelligenza artificiale⁸⁰ (AI). L'analisi automatica dei testi mira, infatti, a raggiungere un grado di interpretazione del testo pari a quella che fornirebbe un essere umano, obiettivo che però non viene raggiunto pienamente a causa di alcune limitazioni che verranno menzionate in seguito.

3.3.1 Cos'è la sentiment analysis

I tweet raccolti con le API possono essere analizzati tramite una *sentiment analysis* (chiamata anche *opinion mining*) che, dall'inizio degli anni Duemila, è una delle aree di ricerca più attive nell'ambito del *Natural Language Processing* (Liu 2012). La *sentiment analysis*⁸¹ o *analisi del sentiment* è “lo studio computazionale di opinioni, sentimenti ed emozioni espressi nei testi” (Liu 2010: 3, traduzione mia). In altre parole, essa è una tecnica automatica che permette di rilevare le opinioni contenute in un testo e classificarle come positive, negative o neutre (Oliveira et al. 2017; Zhang et al. 2014).

Liu (2012) individua cinque componenti chiave di un'opinione:

mezzo fisico impiegato; la *diastrotia* che è legata alle condizioni sociali del parlante; la *diafasia* che dipende dalla situazione comunicativa; la *diacronia*, ovvero la trasformazione legata al tempo; la *diatopia* che corrisponde ai mutamenti della lingua nello spazio geografico.

⁸⁰ L'intelligenza artificiale è la disciplina che studia se e in che modo si possano riprodurre le capacità dell'intelligenza umana tramite l'uso di un computer (Enciclopedia Treccani). È lo studio delle modalità che permettono ai “computer di fare delle cose in cui, al momento, gli esseri umani sono migliori” (Rich & Knight 1991 citati in Russel & Norvig 2010)

⁸¹ Secondo Liu (2012) il termine “sentiment analysis” è apparso per la prima volta nel 2003 nello studio “Sentiment analysis: Capturing favorability using natural language processing” di Nasukawa e Yi, mentre il termine “opinion mining” è stato utilizzato per la prima volta sempre nel 2003 nella ricerca “Mining the peanut gallery: Opinion extraction and semantic classification of product reviews” di Dave, Lawrence e Pennock.

- Un'entità (*entity*) che corrisponde all'oggetto che viene valutato tramite l'opinione. Può essere un prodotto, un servizio, un argomento, un evento o un'organizzazione.
- Un aspetto (*aspect*), cioè una componente dell'entità su cui viene espressa l'opinione.
- Un *sentiment* nei confronti dell'aspetto: il sentiment può essere rappresentato da una polarità positiva, negativa o neutra e da un punteggio numerico che ne denota l'intensità.
- Un portatore di opinione (*opinion holder*), ovvero la persona che ha espresso l'opinione.
- Il tempo (*time*), che corrisponde al momento in cui l'opinione è stata espressa.

Questi cinque aspetti sono essenziali, poiché permettono di contestualizzare l'opinione, capire chi l'ha espressa, in che momento e riguardo quale aspetto. Inoltre, le cinque componenti individuate da Liu (2012) fungono da framework concettuale per trasformare il linguaggio naturale non strutturato in dati strutturati e rappresentabili in un database.

Come segnala Liu (2012), la *sentiment analysis* viene applicata a una vasta gamma di contesti: nell'ambito del marketing, essa permette la valutazione di dati testuali non strutturati provenienti dai social media, facilita "il processo decisionale nei confronti del controllo della qualità, della valutazione delle prestazioni e della progettazione di esperienze di prodotti/servizi" (Wright 2009 citato in Misopoulos et al. 2014: 707) e aiuta così a sviluppare nuove idee per migliorare l'esperienza del consumatore. Inoltre, come evidenziano Colleoni et al. (2011: 11), la valutazione del sentiment degli utenti sui social media è una buona approssimazione della reputazione di un'azienda. Grazie alla *sentiment analysis*, le opinioni degli utenti sparse online vengono raggruppate in una forma di reputazione rilevabile e misurabile. In particolare, nel caso di questa ricerca, il concetto di reputazione astratto e generico viene scomposto e ricondotto al tipo di opinione che gli utenti hanno su determinati servizi, ovvero una variabile empirica misurabile tramite la *sentiment analysis*.

3.3.2 Il corpus su cui effettuare la sentiment analysis

L'insieme dei tweet raccolti tramite le API costituisce un *corpus*⁸², cioè “una collezione di dati testuali selezionati e organizzati secondo criteri linguistici espliciti per servire da campioni di linguaggio” (Sinclair 1996 citato in Bolasco 2013: 68). Bolasco (2013: 64) distingue due tipi di collezioni di testi: le raccolte di *testi veri*, cioè dati testuali in cui i discorsi sono pieni e le frasi e i periodi sono ben articolati, e le raccolte di *testi-frammento*, ossia testi corti (come i tweet) in cui non sono presenti veri e propri periodi e in cui la sintassi tipica del linguaggio scritto non viene rispettata o è del tutto assente.

Come sottolinea Cortelazzo (2013) un'analisi di tipo quantitativo (come nel caso della sentiment analysis), “è tanto migliore quanto è migliore il modo con cui viene costruito il dato, cioè quanto è migliore è il modo con cui viene costruito il corpus”. Un corpus, per essere ritenuto tale, deve rispettare alcuni requisiti:

- Omogeneità: i testi presenti nel corpus devono essere prodotti all'interno della stessa varietà linguistica, presentando quindi condizioni di enunciazione simili (Bolasco 2013; Cortelazzo 2013). I tweet considerati per questa ricerca appartengono tutti alla stessa *varietà diamesica*, ovvero una variazione sincronica della lingua relativa a un canale di comunicazione specifico (in questo caso Twitter). Inoltre, per garantire l'omogeneità, il corpus deve essere composto da testi che abbiano una lunghezza comparabile: nel caso dei tweet si tratta di testi di massimo 280 caratteri.
- Possibilità di costruire dei sotto-*corpus*: i testi all'interno di un corpus devono poter essere articolabili in sotto-corpus. Come spiega Cortelazzo (2013), infatti, un'analisi quantitativa è utile se permette di comparare tra loro realtà diverse, quindi, in questo caso, sotto-corpus differenti. Ad esempio, nella sua ricerca, Cortelazzo compara i discorsi di più presidenti della Repubblica. La comparazione permette di individuare le specificità di ogni sotto-corpus e, nell'ottica delle relazioni pubbliche e della comunicazione aziendale, consente di confrontare più realtà concorrenti e giungere a delle valutazioni sul panorama competitivo.

⁸² Il corpus in letteratura viene anche chiamato *dataset* (Biber et al. 1998 citati in Nissim & Pannitto 2022; McEnery & Wilson 2001)

- Grandezza finita: un corpus deve essere composto da un numero limitato e definito di testi (McEnery & Wilson 2001).
- Formato leggibile: il corpus deve essere contenuto in dei file che siano leggibili dal computer e dall'applicazione utilizzata per analizzarlo (ibid.).
- Disponibilità per altri ricercatori: un corpus dovrebbe essere reso disponibile anche ad altri ricercatori, in modo che questi possano sviluppare ulteriormente la ricerca, collaborare e comparare i risultati ottenuti (ibid.). La condivisione di dataset apre la possibilità a ricercatori provenienti da diverse parti del mondo di partecipare alla ricerca. Ad esempio, il sito web Kaggle⁸³, dedicato alla data science⁸⁴, ha un'intera sezione dedicata alla condivisione di dataset sui più svariati argomenti che possono essere scaricati gratuitamente dagli utenti e utilizzati per condurre delle ricerche.
- Rappresentatività: la questione della rappresentatività si pone in tutti quei casi in cui non si disponga di un inventario chiuso ed esaustivo di testi rispetto al fenomeno che si vuole indagare ed è quindi necessario costituire il corpus attraverso una selezione tra i testi esistenti. Il corpus viene creato attraverso un campionamento, ovvero un "procedimento attraverso il quale si estrae, da un insieme di unità (popolazione) costituenti l'oggetto di studio, un numero ridotto di casi (campione) scelti con criteri tali da consentire la generalizzazione all'intera popolazione dei risultati ottenuti studiando il campione" (Corbetta 2014: 317). Il campione di testi, quindi, costituisce un sottoinsieme di tutti i testi esistenti rispetto al fenomeno indagato e il ricercatore deve verificare che il campione di dati sia rappresentativo della popolazione indagata.

I tweet, come già spiegato precedenza, possono essere considerati dei dati alternativi rispetto a quelli ottenuti dai sondaggi d'opinione. Diversi autori (Kaschesky et al. 2013; Oliveira 2017), però, evidenziano che lo svantaggio di utilizzare i tweet come dati risiede nel fatto che essi non possono costituire un campione di dati statisticamente rappresentativo. Il ricercatore non dispone dei dati sociodemografici della popolazione (gli utenti Twitter), né conosce la sua ampiezza (numero di utenti che hanno espresso la

⁸³ <https://www.kaggle.com/datasets>

⁸⁴ La data science è un campo multidisciplinare che consiste nell'utilizzo dei dati per generare delle informazioni di valore (Sas, n.d.).

propria opinione circa il fenomeno indagato). Inoltre, gli utenti presenti online non rappresentano tutti i settori sociali per via del divario digitale⁸⁵ (Corbetta 2014).

Questi ostacoli potrebbero essere superati tramite un campionamento sistematico⁸⁶, in cui ogni unità viene estratta secondo intervalli regolari. Nel caso di Twitter, però, le API⁸⁷ selezionano i tweet (es: 500 tweet) ritenuti più rilevanti⁸⁸ e li forniscono in ordine cronologico inverso⁸⁹, non rispettando quindi un intervallo di estrazione regolare. Il procedimento messo in atto dalle API non ha un vero e proprio disegno di campionamento e può quindi essere considerato un campionamento di convenienza, in cui il gruppo di unità viene scelto solo per il fatto di essere facilmente accessibile⁹⁰ (le API selezionano i tweet più rilevanti). Come sottolinea Corbetta (2014), il ricercatore che utilizza questo tipo di campionamento deve essere consapevole che i risultati non possono essere generalizzabili e rappresentativi di tutta la popolazione. Un campione di convenienza però può essere utilizzato per effettuare delle ricerche esplorative⁹¹ che, nel campo delle ricerche di mercato, permettono di comprendere meglio la natura di un problema e valutare il comportamento di consumatori e competitor. Le ricerche esplorative non hanno l'obiettivo di raccogliere dati che siano rappresentativi e corretti dal punto di vista statistico, bensì mirano a individuare delle aree problematiche o di possibile miglioramento che necessitano di ulteriori ricerche di tipo statistico. Anche Oliveira et al. (2017) evidenziano la scarsa rappresentatività dei dati provenienti da Twitter e sottolineano la necessità di utilizzare i risultati della sentiment analysis come integrazione ai metodi di ricerca tradizionali.

⁸⁵ Il divario digitale indica la disparità di accesso ai mezzi di informazione e comunicazione da parte di determinate aree geografiche o fasce di popolazione (Enciclopedia Treccani).

⁸⁶ Il campionamento sistematico è un tipo di campionamento probabilistico in cui le unità campionarie vengono selezionate scorrendo la lista dei soggetti e selezionandone una ogni dato intervallo. I campioni casuali possono essere ottenuti anche nel caso in cui manchi la lista della popolazione e l'ampiezza N della popolazione sia sconosciuta. Ad esempio, nelle ricerche di mercato si seleziona a intervalli regolari un cliente da intervistare tra quelli che escono dal supermercato a spesa conclusa, senza sapere a priori quanti sono i clienti totali (Corbetta 2014).

⁸⁷ <https://developer.twitter.com/en/docs/twitter-api/tweets/search/integrate/paginate> [ultima consultazione: 27/03/23]

⁸⁸ <https://developer.twitter.com/en/docs/twitter-api/v1/tweets/search/api-reference/get-search-tweets> [ultima consultazione: 27/03/23]

⁸⁹ Nel sito ufficiale di Twitter non vengono forniti dettagli specifici sui criteri di selezione dei tweet.

⁹⁰ Nel sito ufficiale di Twitter si afferma che i tweet che si ottengono dalle ricerche sono quelli ritenuti più pertinenti, affidabili e sicuri: <https://help.twitter.com/it/using-twitter/twitter-search-not-working> [ultima consultazione: 27/03/23]

⁹¹ <https://it.surveymonkey.com/mp/3-types-survey-research/> [ultima consultazione: 27/03/23]

3.3.4 Punti di forza e di debolezza della sentiment analysis

I punti di forza e di debolezza della sentiment analysis possono essere individuati riflettendo sulle differenze tra metodi qualitativi e quantitativi di analisi dei testi che vengono messe in luce da diversi autori (Cortelazzo 2013; McEnery & Wilson 2001). La sentiment analysis fa parte dei metodi quantitativi di analisi dei testi che permettono di analizzare corpora di grandi dimensioni e mirano a identificare dei fenomeni linguistici e la loro frequenza e utilizzare statisticamente le informazioni raccolte. Nelle analisi qualitative, invece, non vi è l'esigenza di assegnare delle frequenze alle caratteristiche linguistiche dei dati testuali e l'obiettivo principale è identificare degli aspetti dell'uso della lingua e fornire degli esempi di particolari fenomeni (McEnery & Wilson 2001).

Gli studiosi (Cortelazzo 2013; McEnery & Wilson 2001) fanno notare che i punti di forza di un tipo di metodo sono i punti di debolezza dell'altro. Nel caso dei metodi quantitativi, vengono individuati i seguenti vantaggi:

- se il corpus viene costruito adeguatamente ed è statisticamente rappresentativo, i risultati che emergono dall'analisi possono essere generalizzabili e confrontati con quelli di altri corpora (McEnery & Wilson 2001);
- grazie alla lettura computerizzata, si riesce ad elaborare una grande quantità di testi che sarebbe difficile da gestire e analizzare con la sola lettura umana (Cortelazzo 2013);
- i metodi quantitativi di analisi dei testi sono un potente strumento di sintesi statistica: ad esempio, è possibile individuare la frequenza di una parola, nella sua morfologia originaria o nel suo lemma, cioè sulle "parole raggruppate neutralizzando la variazione morfologica"⁹² (Cortelazzo 2013: 302) oppure si può individuare la frequenza degli n-grams (sequenze di due o più parole). Una volta ottenuti i risultati dell'analisi, è possibile derivarne una rappresentazione grafica complessiva (ibid.).

⁹² Ciò avviene ad esempio quando si raggruppano il singolare e il plurale di una stessa parola.

L'analisi automatica del contenuto però non è in grado di analizzare accuratamente i vari livelli linguistici come invece è possibile fare grazie ai metodi qualitativi. I punti di forza dell'analisi qualitativa sono quelli di debolezza dell'analisi quantitativa:

- Dal punto di vista del lessico, l'analisi quantitativa riesce a individuare le frequenze delle parole, quindi lavora sul piano del significante, ma non riesce ad esaminare l'aspetto semantico, cioè il significato: non vengono rilevate le informazioni sui sinonimi o sugli omografi (Cortelazzo 2013).
- L'analisi qualitativa, rispetto alla quantitativa, riesce ad analizzare gli aspetti sintattici e retorici di un testo, che spesso sono di estrema importanza, come nel caso dell'analisi del linguaggio politico (ibid.). In particolare, per quanto riguarda le figure retoriche, i metodi quantitativi non sono in grado di rilevare il sarcasmo, “figura retorica in cui il parlante o lo scrivente afferma il contrario di ciò che intende” (Liu 2012: 52, traduzione mia). Liu (2012) sottolinea che, dal punto di vista della sentiment analysis, il sarcasmo consiste nell'espressione di un'opinione di tipo positivo che però ne racchiude una di tipo negativo e viceversa.

L'analisi dei testi è uno di quei campi in cui la via primaria di ricerca è quella qualitativa (Cortelazzo 2013), in quanto permette di mantenere una ricchezza e una precisione linguistica non raggiungibile tramite metodi quantitativi (McEnery & Wilson 2001). L'approccio quantitativo ristruttura il testo ed è utile quando “ha senso sacrificare la ricchezza di un testo per ridurlo a rappresentazioni basate su grafici, indici, tabelle, percentuali” (Cortelazzo 2013: 302). In ogni caso, Cortelazzo (2013) sottolinea che spesso torna utile affiancare i due tipi di metodo: in un'ottica esplorativa, grazie all'analisi quantitativa è possibile individuare orizzonti di ricerca non rilevabili tramite un'analisi qualitativa, in quanto gli indizi che li suggeriscono sono diluiti in testi molto ampi. Il testo analizzato con metodi quantitativi può fornire delle informazioni da indagare ulteriormente con un'analisi qualitativa. In un'ottica confermativa, invece, un'analisi quantitativa può confermare e rendere più solidi i risultati trovati con una precedente analisi qualitativa (ibid.).

3.3.5 Ricerche che utilizzano la sentiment analysis

La tecnica della sentiment analysis può essere applicata in ambiti e con scopi molto diversi. Ad esempio, in alcuni studi presenti in letteratura, la sentiment analysis è utilizzata:

- in ambito politico per prevedere il risultato delle elezioni: Nausheen e Begum (2018) analizzano i tweet degli utenti che si esprimono sui candidati Hillary Clinton, Donald Trump e Bernie Sanders alle elezioni presidenziali statunitensi del 2016.
- per la sicurezza pubblica: Sakaki et al. (2010), ricercatori giapponesi, hanno sviluppato un sistema che, tramite il monitoraggio dei tweet in tempo reale, individua quando e dove avviene un terremoto.
- nel marketing per confrontare l'opinione dei consumatori rispetto a diverse aziende operanti nello stesso settore: si possono confrontare alcune marche di telefoni cellulari (Sharma et al. 2020) o si può indagare il sentiment dei clienti rispetto diverse assicurazioni sanitarie (Van den Broek-Altenburg & Atherly 2019).
- nel contesto del servizio clienti, per indagare l'esperienza con un prodotto o un servizio. Ai fini di questa ricerca, è utile menzionare gli studi presenti in letteratura che hanno rilevato le opinioni dei clienti nel settore del trasporto aereo.
 - o Misopoulos et al. (2014) analizzano l'esperienza degli utenti con diverse compagnie aeree in una data specifica (18 dicembre 2010⁹³) in cui delle condizioni meteorologiche avverse hanno causato la cancellazione e riprogrammazione di numerosi voli. Vengono utilizzate sia tecniche lexicon-based che di machine learning e vengono presi in considerazione i tweet pubblicati in lingua inglese contenenti determinate parole chiave. Inoltre, un limite di questo studio consiste nel fatto che le parole chiave considerate per scaricare i tweet corrispondono a dieci parole negative e dieci positive (come "good"). Tuttavia, tale approccio ha mostrato alcune carenze in quanto non sempre alle parole chiave positive e negative

⁹³ Data in cui è iniziata una serie di eruzioni del vulcano Eyjafjöll in Islanda che ha bloccato il traffico aereo.

corrispondono dei tweet positivi o negativi rispettivamente. Si è infatti riscontrato che in alcuni casi, tweet contenenti parole chiave positive presentavano delle negazioni che ne invertivano il significato positivo (come nel caso di “no good”). Si nota quindi come la scelta delle parole chiave sia di estrema importanza per garantire l’efficacia dell’analisi dei dati e come probabilmente possa essere meglio utilizzare delle parole chiave dalla connotazione neutra.

- Kwon et al. (2021), invece, analizzano le recensioni online pubblicate sul sito web Airline Quality⁹⁴ e scritte da clienti che hanno usato compagnie aeree asiatiche. Viene utilizzato il linguaggio di programmazione R.
- Nella ricerca di Monika et al. (2019) vengono analizzati i tweet degli utenti contenuti in un file di Kaggle riguardante sei compagnie aeree statunitensi. L’analisi viene eseguita tramite tecniche di machine learning.

Dalle ricerche sopracitate si nota che per effettuare una sentiment analysis il ricercatore può scegliere tra diverse fonti di dati (siti di recensioni, social network), diversi linguaggi di programmazione (Python, R) e approcci (lexicon-based e machine learning). Per quanto riguarda la rilevazione del sentiment dei clienti, in letteratura sembrano mancare degli studi recenti⁹⁵ in cui la sentiment analysis abbia come scopo non solo individuare la natura delle opinioni dei clienti ma anche fornire delle riflessioni su come questa possa influenzare la reputazione aziendale online.

3.4 Sentiment analysis con Python

Fondamentali per lo sviluppo di questa ricerca, e in generale per l’elaborazione del linguaggio naturale, sono il linguaggio di programmazione Python e l’applicazione Jupyter Notebook che permette di scrivere ed eseguire codici in tale linguaggio. Uno dei modi per utilizzare Jupyter Notebook è tramite l’installazione di Anaconda Navigator,

⁹⁴ <https://www.airlinequality.com/> [ultima consultazione: 27/03/23]

⁹⁵ Colleoni et al., in una ricerca pubblicata nel 2011, propongono uno strumento (E-Daemon) per visualizzare e misurare la reputazione in tempo reale. Esso utilizza le Twitter API e permette di ricercare contenuti inserendo specifiche parole chiave. Tale strumento non sembra però aver trovato applicazione in altri studi.

l'interfaccia grafica (GUI, *Graphical User Interface*) di Anaconda⁹⁶ con cui è possibile accedere a una vasta gamma di strumenti per l'analisi dei dati. In questa ricerca è stata utilizzata la versione di Python 3.9.12 distribuita da Anaconda.

Le API creano un collegamento tra due applicazioni che, in questo caso, sono Twitter e Jupyter Notebook, un'applicazione web open source lanciata nel 2014 che consente di creare dei documenti ed eseguirli tramite un browser web (Driscoll, n.d.). I documenti notebook sono caratterizzati da una sequenza di celle in cui è possibile inserire codice sorgente⁹⁷ eseguibile, testi, equazioni e contenuti multimediali. Come spiega Shafi (2023) il nome di questa applicazione è un acronimo che deriva dalle iniziali di Ju(lia), Pyt(hon), R, ovvero i primi tre linguaggi di programmazione che erano supportati dall'applicazione. Oggi Jupyter permette di scrivere codici in più di 40 linguaggi, e i kernel⁹⁸ dei codici sono disponibili su Github⁹⁹, un sito web in cui gli sviluppatori possono caricare il codice sorgente dei loro programmi e permettere agli altri utenti di scaricarlo. Paul e Harvey Deitel (2021: 21) affermano che “i notebook Jupyter sono il mezzo preferito per eseguire analisi di dati in Python e comunicare i risultati in maniera riproducibile”: i file notebook, infatti, possono essere condivisi facilmente tra i gruppi di lavoro ed essere visualizzati come pagine web all'interno di GitHub.

Il linguaggio di programmazione Python fu sviluppato nel 1989 dal programmatore olandese Guido Van Rossum del National Research Institute for Mathematics and Computer Science di Amsterdam che durante le vacanze di Natale decise di progettare un linguaggio di programmazione che fosse semplice e intuitivo, adatto anche ai non professionisti, facile da insegnare e open source, aperto alle modifiche e agli sviluppi da parte della comunità di programmatori. Il nome di questo linguaggio rende omaggio al gruppo di comici inglesi Monty Python di cui Van Rossum era un grande fan (Deitel 2021; Tibone 2020).

⁹⁶ Anaconda è una distribuzione open source dei linguaggi di programmazione Python e R che può essere installata nei sistemi operativi di Windows, Linux e MacOS ed è utilizzata soprattutto nell'ambito dell'analisi dei dati. È stata lanciata per la prima volta nel 2012 da Peter Wang e Travis Oliphant: <https://www.anaconda.com/> [ultima consultazione: 15/04/23]

⁹⁷ Il codice sorgente (chiamato anche sorgente o codice) è “il testo di un algoritmo di un programma scritto in un determinato linguaggio di programmazione [...] che definisce il flusso di esecuzione del programma stesso, ovvero la sua codifica software”. (Wikipedia)

⁹⁸ Un kernel notebook è un “motore computazionale” che esegue il codice contenuto in un documento Notebook. (Ingargiola 2015)

⁹⁹ <https://github.com/jupyter/jupyter/wiki/Jupyter-kernels> [ultima consultazione: 27/03/23]

Python è un linguaggio di scripting interpretato¹⁰⁰, cioè “i codici sorgente non vengono compilati ma interpretati ed eseguiti da un apposito software” (Tibone 2020: 5). I file contenenti programmi scritti in Python sono semplici file di testo (chiamati *script*), con estensione “.py” e per essere eseguiti necessitano di un software chiamato interprete che riesce a tradurre il codice sorgente in linguaggio macchina (ibid. p.3). Python è un linguaggio di alto livello “perché delega all’interprete la gestione delle componenti hardware, permettendo al programmatore di concentrarsi unicamente sugli obiettivi specifici del programma che sta scrivendo” (ibid. p.3). È anche un linguaggio multiplatforma perché lo stesso script può essere eseguito su sistemi operativi diversi (Linux, Windows, MacOS) a patto che venga installato l’interprete adatto (ibid.).

Python è tra i linguaggi di programmazione più popolari al mondo, è utilizzato per la scrittura di software di aziende di successo (Google, Youtube, Instagram, Reddit, Spotify, Dropbox, Quora¹⁰¹, Nasa¹⁰²) e ha superato R in popolarità nell’ambito della data science. Ci sono diversi motivi per cui Python è così apprezzato:

- Sintassi semplice e pulita: Python è più facile da imparare rispetto ad altri linguaggi di programmazione (Ruchko n.d.), poiché presenta meno costrutti sintattici¹⁰³. Ciò facilita l’apprendimento anche da parte di utenti non esperti di informatica ma interessati, ad esempio, ad analizzare dei dati (Conant 2017). Come fa notare Beri (2010) la sintassi di Python è così chiara che spesso porta a capire quale comando utilizzare senza dover consultare la documentazione¹⁰⁴. Inoltre, leggendo uno script è probabile che se ne comprenda lo scopo anche in assenza di commenti¹⁰⁵. Una particolarità della sintassi è l’indentazione, ovvero l’inserimento di una rientranza a inizio riga per delimitare i blocchi di istruzioni.

¹⁰⁰ In un linguaggio compilato la macchina svolge direttamente l’operazione di traduzione. Nel linguaggio interpretato, invece, la fonte del codice non viene tradotta direttamente dalla macchina, bensì è un programma diverso, detto interprete, che legge ed esegue il codice (FreeCodeCamp 2022)

¹⁰¹ <https://codeinstitute.net/global/blog/7-popular-software-programs-written-in-python/> [ultima consultazione: 28/03/23]

¹⁰² <https://aulab.it/notizia/300/cose-python-e-perche-e-cosi-popolare> [ultima consultazione: 28/03/23]

¹⁰³ <https://it.wikipedia.org/wiki/Python#Sintassi> [ultima consultazione: 29/03/23]

¹⁰⁴ La documentazione ufficiale di Python è disponibile online e fornisce informazioni dettagliate su come utilizzare i vari strumenti disponibili.

¹⁰⁵ In un linguaggio di programmazione, un commento è una parte del codice sorgente che descrive le righe del codice e non fa parte dell’algoritmo e quindi non viene eseguito (Wikipedia). In Python si può inserire un commento utilizzando il simbolo del cancelletto (#).

- Ciò semplifica la sintassi in quanto non è necessario scrivere delle parole chiave o inserire delle parentesi per delimitare i vari blocchi del programma (Beri 2010).
- Utilizzo per diversi scopi: Python può essere utilizzato per l'analisi dei dati, finanza e trading, calcoli matematici, automazione, machine learning¹⁰⁶, computer grafica, sviluppo di videogiochi (Ruchko n.d.).
 - Community attiva in tutto il mondo: “it’s the people behind the language that make it special” (Ankur Gupta, esperto di Python intervistato da Ruchko, n.d.). La community di Python è molto vasta e formata da utenti disponibili a condividere conoscenze e fornire aiuto a chi si trova in difficoltà durante la programmazione. Online è possibile trovare centinaia di risorse gratuite, come libri, video tutorial e script condivisi su GitHub che fanno progredire la conoscenza e la partecipazione.

Tutti i linguaggi di programmazione hanno delle librerie di funzioni¹⁰⁷ che, al pari delle librerie fisiche che contengono molti libri da poter consultare, presentano al loro interno diversi codici preconfezionati da utilizzare per compiere determinate operazioni informatiche. In questo modo, i codici sono facilmente condivisibili tra gli sviluppatori e non devono essere scritti ogni volta da zero (operazione molto onerosa in termini di tempo).

Uno dei motivi per cui Python è così popolare è l'ampia gamma di librerie di funzioni di cui dispone. Una libreria è una raccolta di moduli o pacchetti¹⁰⁸ correlati, mentre un modulo è un insieme di codici o funzioni (Chiluka 2022). Per accedere alle librerie è necessario installarle nel progetto Python attraverso il comando `pip install`, mentre per utilizzare i moduli di una libreria si digita l'istruzione `import`.

Al momento dell'installazione di Python si ha accesso a una libreria standard¹⁰⁹ (Iglesias Moreno 2020) e non è quindi necessario installarla in un secondo momento. Al suo interno si trovano numerose funzionalità già codificate, pronte per scrivere il

¹⁰⁶ Il machine learning è una branca dell'intelligenza artificiale che mira a sviluppare sistemi che apprendono o migliorano le performance in base ai dati che utilizzano. (Oracle, n.d.)

¹⁰⁷ Una funzione è una parte di programma a cui viene dato un nome, è una sequenza di istruzioni che esegue un compito specifico (Open Education Polito, n.d.). Le funzioni accettano come input zero o più argomenti (o parametri), li elaborano, e restituiscono in output un risultato (Melotti 2017).

¹⁰⁸ Un pacchetto è un insieme di moduli: si tratta di una cartella contenente dei moduli o delle sotto-cartelle che, a loro volta, contengono altri moduli o cartelle (Udacity team, 2021).

¹⁰⁹ <https://docs.python.org/3/library/functions.html> [ultima consultazione: 20/04/23]

programma ed elencate nella documentazione¹¹⁰ ufficiale di Python. Inoltre, su Python è possibile installare anche molte librerie open source rese disponibili dall’ampia comunità di sviluppatori.

La sentiment analysis fa parte di un processo di raccolta ed elaborazione dei dati che si sviluppa in più fasi e prevede l’utilizzo di diverse librerie. Di seguito vengono elencati e descritti alcuni dei moduli e delle librerie utilizzati.

3.4.1 Libreria standard di Python

Per l’elaborazione dei tweet tornano utili alcuni moduli della libreria standard di Python che vengono qui elencati insieme alla loro descrizione.

MODULO	DESCRIZIONE DEL MODULO
csv	Consente l’elaborazione di file csv con dati separati da virgole (Deitel 2021). Il csv (acronimo di <i>Comma Separated Values</i>) è un formato di file basato su testo, utilizzato per l’esportazione e l’importazione di database o fogli elettronici. In questo formato ogni riga di testo corrisponde a una riga di tabella e le virgole all’interno della riga di testo fungono da “separatori” e delimitano le varie colonne del database. Il formato csv è tra i formati di file supportati da Excel ¹¹¹ .
datetime, time	Da utilizzare per la manipolazione di date e orari (Deitel 2021).
json	Modulo che permette l’elaborazione del formato JavaScript Object Notation (JSON), che è quello attraverso cui le Twitter API forniscono i dati ¹¹² .
re (regex)	Il modulo regex (acronimo di <i>regular expressions</i>) consente di lavorare con le espressioni regolari. Queste ultime consistono in sequenze di caratteri che definiscono un pattern di ricerca e sono utili per trovare delle corrispondenze in stringhe di testo (Deitel 2021).

¹¹⁰ <https://docs.python.org/3/library/> [ultima consultazione: 30/03/23]

¹¹¹ https://it.wikipedia.org/wiki/Comma-separated_values [ultima consultazione: 30/03/23]

¹¹² <https://developer.twitter.com/en/docs/twitter-api/v1/data-dictionary/overview> [ultima consultazione: 02/04/23]

Tabella 7: descrizione dei moduli della libreria standard di Python utili per l'analisi dei tweet.

3.4.2 Librerie open source per l'analisi dei dati

Per svolgere l'analisi dei dati raccolti da Twitter è necessario ricorrere anche a delle librerie open source, qui elencate e brevemente descritte.

La libreria *Pandas* è utile per la manipolazione e l'analisi di dati tabulari, come quelli provenienti da file in formato csv o da cartelle di lavoro Excel. Wes McKinney ha creato questa libreria nel 2008 mentre lavorava per AQR Capital Management spinto dall'esigenza di avere a disposizione uno strumento flessibile e potente per effettuare analisi quantitative su dati finanziari. Il nome *Pandas* deriva dal termine "panel data" che si riferisce ai set di dati in cui vengono osservate le variabili in più periodi di tempo¹¹³. La libreria ha due strutture dati fondamentali (Deitel 2021): le *Series* per collezioni di dati monodimensionali, come nel caso di una colonna di un foglio di calcolo e i *DataFrames* per collezioni di dati bidimensionali, come nel caso di una tabella con righe e colonne. La libreria *Pandas* è utile per effettuare calcoli statistici sui dati, pulire e filtrare i dati, memorizzare i dati ripuliti e salvarli in un nuovo file (McIntire et al., n.d.).

Per visualizzare i dati in modo sintetico e più accessibile, invece, possono essere utilizzate le librerie *Matplotlib* e *Seaborn*, che consentono di realizzare grafici di diverso tipo (a dispersione, a barre, a torta, 3D ecc.), statici o animati. *Seaborn* è una libreria costruita a partire da *Matplotlib* e, rispetto a questa, è di livello più alto: i grafici realizzati con *Seaborn* sono più piacevoli esteticamente e vengono creati con meno righe di codice (Deitel 2021).

Infine, una libreria fondamentale per l'elaborazione del linguaggio naturale è *NLTK* (*Natural Language Toolkit*). È stata sviluppata nel 2001 da Steven Bird ed Edward Loper, inizialmente come risorsa da utilizzare nel corso di linguistica computazionale al dipartimento di Computer and Information Science dell'Università della Pennsylvania (Bird et al. 2009). Gli autori spiegano che la libreria nasce con l'obiettivo di insegnare l'elaborazione del linguaggio naturale a studenti di linguistica e di informatica che, tramite uno stesso strumento, possono soddisfare interessi differenti: nel primo caso analizzare il linguaggio per sviluppare nuove teorie e riflessioni linguistiche, mentre nel secondo capire come implementare i contenuti della libreria. Nel corso del tempo la

¹¹³ [https://it.wikipedia.org/wiki/Pandas_\(software\)](https://it.wikipedia.org/wiki/Pandas_(software)) [ultima consultazione: 30/03/23]

libreria si è arricchita grazie ai contributi di decine di sviluppatori ed è stata utilizzata in diversi corsi universitari (Bird et al. 2009 xv; Bird et al. 2008). Come affermano Bird et al. (2009) NLTK è stata progettata seguendo quattro principi:

- *Semplicità*: costruire una struttura intuitiva che permetta di capire come elaborare il linguaggio naturale senza troppi sforzi. I creatori della libreria avevano come obiettivo quello di riuscire a insegnare a studenti appartenenti a diversi corsi di studio come elaborare il linguaggio naturale in un solo semestre universitario. Necessitavano quindi di uno strumento che fosse comprensibile anche per chi non avesse esperienze in programmazione.
- *Coerenza*: denominare i metodi e le varie componenti della libreria in modo coerente e conciso, in modo che un utente sappia cosa aspettarsi dall'utilizzo di una determinata funzione.
- *Estensibilità*: fare in modo che i nuovi moduli sviluppati possano essere inseriti facilmente all'interno della libreria.
- *Modularità*: permettere agli utenti di usare una determinata componente della libreria senza dover per forza conoscere il resto del toolkit.

Per effettuare una sentiment analysis, risultano utili diversi pacchetti di NLTK, di seguito elencati e le cui caratteristiche si trovano nella documentazione sul sito ufficiale¹¹⁴ della libreria:

PACCHETTO	DESCRIZIONE DEL PACCHETTO
nltk.corpus	Viene utilizzato per accedere a diversi corpora inclusi nella libreria o esterni ad essa ¹¹⁵ .
nltk.tokenize	Consente di suddividere il testo in <i>token</i> , ovvero unità linguistiche che possono corrispondere a parole, parti di parole o caratteri (Burchfiel 2022).
nltk.stem	Si utilizza per ridurre le parole alla loro radice (<i>stem</i>), eliminando così le desinenze o gli affissi ¹¹⁶ .

¹¹⁴ <https://www.nltk.org/py-modindex.html> [ultima consultazione: 30/03/23]

¹¹⁵ <https://www.nltk.org/api/nltk.corpus.html#module-nltk.corpus> [ultima consultazione: 30/03/23]

¹¹⁶ <https://www.nltk.org/api/nltk.stem.html#module-nltk.stem> [ultima consultazione: 30/03/23]

nltk.sentiment	Permette di effettuare diversi tipi di sentiment analysis.
----------------	--

Tabella 8: pacchetti di NLTK.

Infine, un altro pacchetto open source che risulta utile per l’elaborazione del linguaggio naturale è *NeatText*¹¹⁷, che include diverse funzioni per pulire il testo dagli elementi non necessari ai fini di una sentiment analysis o di un’analisi testuale.

3.4.3 VADER

Per svolgere la sentiment analysis si può usare VADER (Valence Aware Dictionary and sEntiment Reasoner), un modulo di NLTK sviluppato da Hutto e Gilbert nel 2014. Innanzitutto, bisogna ricordare che per effettuare la sentiment analysis esistono essenzialmente due tipi di approcci, quello lexicon-based e quello che utilizza il machine learning (chiamato anche apprendimento automatico).

Nel primo caso, in cui rientra anche VADER, l’analisi si basa su un *opinion lexicon/sentiment lexicon* (Kontopoulos 2013; Hutto & Gilbert 2014), ovvero una lista di elementi lessicali predefiniti (e.g. parole) etichettati come positivi o negativi da degli esseri umani. Per analizzare un nuovo testo il sistema di sentiment analysis cerca una corrispondenza tra le parole presenti nel testo e quelle presenti nel lessico preso come riferimento e in questo modo determina la polarità delle singole parole e del testo nel suo complesso (Nausheen & Begum 2018).

Nel machine learning, invece, un classificatore (un algoritmo) viene addestrato a rilevare la polarità dei testi in modo automatico. Per l’addestramento del classificatore si parte da un dataset molto ampio in cui i testi sono etichettati come positivi o negativi da un essere umano. Il classificatore, quindi, partendo dai testi già etichettati “impara” a etichettarne di nuovi in autonomia senza dover consultare un lessico di riferimento (Kontopoulos 2013; Bonta & Janardhan 2019).

Entrambi gli approcci hanno dei punti di forza e debolezza: i lessici di parole possono essere facilmente estesi e migliorati ma crearli da zero è molto dispendioso in termini di tempo e per questo spesso si utilizzano lessici già esistenti (Bonta & Janardhan 2019),

¹¹⁷ <https://pypi.org/project/neattext/> [ultima consultazione: 30/03/23]

che però potrebbero non essere sempre adatti (Blinov et al. 2013). Come evidenziato in diversi studi (Blinov et al. 2013; Devika et al. 2016; Rathee et al. 2018) l'approccio del machine learning è generalmente molto accurato, ma solo nell'ambito lessicale in cui il classificatore viene addestrato. Nel momento in cui lo stesso classificatore è usato in un ambito diverso, la sua performance diminuisce drasticamente. Altre criticità di questo approccio dipendono dal fatto che la performance dipende molto anche dalla grandezza e dalla qualità del dataset usato per l'addestramento del classificatore (Hutto & Gilbert 2014; Zhang et al. 2014), l'addestramento del classificatore può durare molto tempo (ciò può essere un ostacolo quando bisogna analizzare dei dati testuali in tempo reale) e l'algoritmo utilizzato richiede una potenza di elaborazione elevata in termini di CPU¹¹⁸ (Hutto & Gilbert 2014). Non c'è quindi un approccio che sia migliore in termini assoluti e la scelta tra uno e l'altro dipende dalle esigenze specifiche della ricerca.

In questa ricerca è stato utilizzato VADER che fa parte degli approcci lexicon-based. Bonta e Janardhan (2019) spiegano nel dettaglio come VADER calcoli il punteggio del sentiment nei testi. Una volta eseguito questo modulo di NLTK, il testo viene analizzato per cercare delle corrispondenze con il lessico di riferimento. La polarità viene individuata tramite la funzione "polarity_scores()". VADER è in grado di indicare se un testo è positivo, negativo o neutro e ne mostra anche una misura tramite il *compound*: esso corrisponde alla somma dei punteggi delle singole parole del testo, che sono valutate come positive, negative o neutre e normalizzate in una scala tra -1 e +1 (dove -1 indica il massimo della negatività e +1 il massimo della positività). Le soglie standardizzate del valore del compound per classificare il testo come positivo, negativo o neutro sono:

Sentiment positivo: punteggio di compound ≥ 0.05
Sentiment neutro: punteggio di compound > -0.05 e < 0.05
Sentiment negativo: punteggio di compound ≤ -0.05

Tabella 9: soglie del compound.

¹¹⁸ La CPU (acronimo di *central processing unit*) è il processore di un computer che "carica le istruzioni dei programmi, le interpreta e manipola i dati di conseguenza" (Enciclopedia Treccani).

Inoltre, per ogni frase analizzata, VADER fornisce tre punteggi percentuali¹¹⁹ della positività, negatività e neutralità della frase. Un testo, ad esempio, potrebbe avere le seguenti percentuali: 'neg': 0.636, 'neu': 0.364, 'pos': 0.0.

In questa ricerca è stato scelto VADER per effettuare la sentiment analysis in quanto esso è un modello particolarmente adatto all'analisi dei testi provenienti da piattaforme di microblogging come Twitter. Hutto e Gilbert (2014) hanno sviluppato questo modulo di NLTK con l'obiettivo di costruire un motore di sentiment analysis che lavori efficacemente con i testi dei social media ma sia anche estendibile ad altri ambiti; che non usi dati per l'addestramento di un motore di machine learning ma si basi su un lessico costruito e classificato accuratamente dall'uomo; che sia abbastanza veloce da poter essere usato per monitorare i dati in tempo reale e che presenti un buon bilanciamento tra velocità e performance.

In particolare, gli autori hanno sviluppato VADER spinti dall'esigenza di colmare le carenze di altri approcci lexicon-based che, durante la misurazione del sentiment, ignorano alcune peculiarità dei testi dei microblog che possono influire sulla polarità. VADER riesce a rilevare tali peculiarità poiché esse sono state inserite nel lessico di riferimento (in lingua inglese). Esse sono:

- le emoticon occidentali¹²⁰ come “:-)”;
- gli acronimi e gli inicialismi come “LOL¹²¹” e “WTF¹²²”;
- espressioni provenienti dallo slang come “nah”, “meh”, “giggly”.

Inoltre, questo modello prende in considerazione anche degli elementi sintattici ignorati da altri approcci che possono provocare un cambiamento nell'intensità della polarità e nel valore del compound:

- la punteggiatura (ad esempio uno o più punti esclamativi)

```
example_sentence = "This is bad"
scores = sia.polarity_scores(example_sentence)
print(scores)

{'neg': 0.636, 'neu': 0.364, 'pos': 0.0, 'compound': -0.5423}
example_sentence = "This is bad!!"
```

¹¹⁹ La somma dei tre punteggi è 1.

¹²⁰ https://en.wikipedia.org/wiki/List_of_emoticons [ultima consultazione: 31/03/23]

¹²¹ Acronimo di “laughing out loud”.

¹²² Acronimo di “what the fuck”.

```
scores = sia.polarity_scores(example_sentence)
print(scores)

{'neg': 0.671, 'neu': 0.329, 'pos': 0.0, 'compound': -0.6229}
```

Tabella 10: esempio del cambiamento di polarità in due frasi con e senza punteggiatura processate nel Jupyter Notebook.

- l'uso delle lettere maiuscole

```
example_sentence = "This is bad"
scores = sia.polarity_scores(example_sentence)
print(scores)

{'neg': 0.636, 'neu': 0.364, 'pos': 0.0, 'compound': -0.5423}
example_sentence = "This is BAD"
scores = sia.polarity_scores(example_sentence)
print(scores)

{'neg': 0.679, 'neu': 0.321, 'pos': 0.0, 'compound': -0.6408}
```

Tabella 11: esempio del cambiamento di polarità in due frasi con e senza caratteri maiuscoli processate nel Jupyter Notebook.

- gli avverbi di quantità (come “extremely”)

```
example_sentence = "This is bad"
scores = sia.polarity_scores(example_sentence)
print(scores)

{'neg': 0.636, 'neu': 0.364, 'pos': 0.0, 'compound': -0.5423}
example_sentence = "This is extremely bad"
scores = sia.polarity_scores(example_sentence)
print(scores)

{'neg': 0.558, 'neu': 0.442, 'pos': 0.0, 'compound': -0.5849}
```

Tabella 12: esempio del cambiamento di polarità in due frasi con e senza avverbio processate nel Jupyter Notebook.

- la presenza di congiunzioni avversative (“but”) e di negazioni (“not”)

```
example_sentence = "This is fine"
scores = sia.polarity_scores(example_sentence)
print(scores)

{'neg': 0.0, 'neu': 0.526, 'pos': 0.474, 'compound': 0.2023}
example_sentence = "This is fine, but it doesn't work that well"
scores = sia.polarity_scores(example_sentence)
print(scores)

{'neg': 0.209, 'neu': 0.659, 'pos': 0.132, 'compound': -0.2074}
```

Tabella 13: esempio del cambiamento di polarità in due frasi con e senza congiunzione e negazione processate nel Jupyter Notebook.

L'elenco completo delle caratteristiche di VADER è disponibile nello script su Github¹²³ (in formato .py) che contiene il codice sorgente che viene eseguito dall'interprete Python e tradotto in istruzioni eseguibili dal pc.

Nello studio presentato da Hutto e Gilbert (2014) viene confrontata la performance dell'analisi svolta con VADER con quella svolta tramite altri approcci lexicon-based: i risultati mostrano che VADER performa meglio degli altri approcci in tutti gli ambiti, tra cui il caso specifico dell'analisi dei tweet. Inoltre, viene fatto un confronto anche con la performance di alcuni algoritmi di machine learning da cui risulta che VADER performi con uguale o maggiore efficacia in tutti gli ambiti rispetto a quanto facciano gli algoritmi di machine learning nell'ambito per cui sono stati addestrati.

In conclusione, quindi, gli autori (Hutto & Gilbert 2014) affermano che l'utilizzo di VADER porta con sé diversi vantaggi: in primo luogo, si tratta di un approccio che, nonostante sia veloce e poco dispendioso per il computer (in termini di utilizzo di CPU), non sacrifica l'accuratezza dell'analisi. In base all'esperienza degli autori, un corpus che viene analizzato da VADER in una frazione di secondo può impiegare decine di minuti o addirittura ore se analizzato con algoritmi di machine learning. In secondo luogo, le regole alla base del funzionamento di VADER e il lessico di riferimento possono essere facilmente accessibili per il ricercatore rispetto ai più complicati modelli matematici utilizzati nell'ambito del machine learning. VADER riesce quindi ad essere utilizzato anche da utenti meno esperti e non appartenenti alla comunità di informatici. Infine, VADER dà risultati efficaci a prescindere dall'ambito di applicazione, mentre gli approcci di machine learning tendono ad essere affidabili solo nell'ambito in cui sono stati addestrati.

¹²³ <https://github.com/cjhutto/vaderSentiment/blob/master/vaderSentiment/vaderSentiment.py> [ultima consultazione: 31/03/23]

CAPITOLO 4 – CASO STUDIO: LE COMPAGNIE AEREE NEL PERIODO DI CRISI DELL’ESTATE 2022

Le riflessioni teoriche e metodologiche dei capitoli 2 e 3 tornano utili in questo capitolo per l’analisi di un caso studio specifico. Di seguito si presentano l’oggetto della ricerca, le domande di ricerca e le fasi in cui si è sviluppata la raccolta, l’analisi e la presentazione grafica dei dati provenienti da Twitter.

4.1 L’oggetto della ricerca

Questa ricerca, in un’ottica esplorativa, vuole offrire una panoramica della reputazione e, in generale, dell’attività su Twitter di cinque compagnie aeree nell’estate del 2022, tramite la tecnica della sentiment analysis. La reputazione è un concetto che si può sviluppare su più piani di analisi: questa ricerca si focalizza su una dimensione specifica del costrutto reputazionale (opinione su determinati servizi) e su un gruppo potenzialmente eterogeneo di utenti (gli utenti di Twitter potrebbero appartenere a diverse classi di stakeholder).

L’interesse per questo caso studio specifico è nato dopo aver letto numerosi articoli che riportavano i problemi che stava vivendo il settore del trasporto aereo nell’estate del 2022 e aver intercettato il passaparola offline di viaggiatori che avevano avuto un’esperienza negativa del servizio delle compagnie aeree. Si è voluto quindi analizzare il passaparola e l’espressione delle opinioni nel contesto online (*electronic word of mouth*) per capire in che modo questi influenzassero la reputazione delle compagnie aeree e dell’intero settore.

4.1.1 Caratteristiche principali del settore del trasporto aereo

Il settore del trasporto aereo è caratterizzato da una molteplicità di attività e stakeholder, tra cui compagnie aeree, aeroporti, governi, organizzazioni internazionali (IATA) e intergovernative (ICAO) e clienti finali. Per avere una panoramica dell’organizzazione di questo settore, vengono qui descritte alcune delle sue caratteristiche raggruppate nelle seguenti aree tematiche:

- *Regolamentazione*

L'ICAO (Organizzazione Internazionale dell'Aviazione Civile) e la IATA (International Air Transport Association) svolgono un ruolo fondamentale nella regolamentazione delle operazioni del settore del trasporto aereo. Come si legge nel sito web ufficiale¹²⁴, l'ICAO è un'agenzia delle Nazioni Unite con sede a Montreal in Canada, fondata nel 1944 durante la Convenzione di Chicago¹²⁵, a cui oggi aderiscono 193 paesi al fine di gestire la diplomazia e la cooperazione nel trasporto aereo. Gli obiettivi primari di questa agenzia sono quelli di garantire la sicurezza, l'efficienza e la capacità della navigazione aerea, sviluppare l'economia del trasporto aereo e ridurre al minimo gli effetti negativi delle attività di questo settore sull'ambiente. La IATA¹²⁶, invece, è un'organizzazione internazionale, sempre con sede a Montreal, fondata nel 1945, di cui fanno parte circa 300 compagnie aeree (ovvero l'83% delle compagnie aeree nel mondo). Essa ha diversi obiettivi, come la digitalizzazione dei servizi offerti dal settore del trasporto aereo, il rafforzamento dei sistemi di sicurezza, l'attenzione alla sostenibilità, all'ambiente e alla responsabilità sociale e la pubblicazione di dati aggiornati riguardanti l'intero settore.

Queste due autorità, a seconda delle situazioni che si manifestano (emergenze sanitarie, disastri naturali, attacchi terroristici), pubblicano dei regolamenti e delle linee guida che influenzano fortemente l'attività delle compagnie aeree e degli aeroporti (Macilree & Duval 2021 citati in Thepchalerm & Ho 2021). Ad esempio, nel caso della pandemia di COVID-19, entrambe le organizzazioni hanno dato indicazioni precise sia agli addetti ai lavori che ai passeggeri e reso disponibile il materiale informativo nei propri siti ufficiali.

- *Domanda derivata e tipo di prodotto*

Il settore del trasporto aereo è caratterizzato da una *domanda derivata*, ovvero una domanda che deriva dalla domanda per un altro prodotto o servizio (Budd & Ison 2017). I passeggeri, le merci o la posta trasportati hanno come obiettivo primario quello di raggiungere una determinata destinazione e non quello di vivere l'esperienza di viaggio in sé offerta dalle compagnie aeree. Da ciò ne consegue che la domanda dei servizi di

¹²⁴ <https://www.icao.int/about-icao/Pages/default.aspx>

¹²⁵ Si tratta di una convenzione che elenca le regole principali alla base dell'aviazione civile e del trasporto aereo mondiale (Wikipedia).

¹²⁶ <https://www.iata.org/en/about/>

trasporto aereo varia a seconda della domanda di altre attività economiche (turismo, commercio, servizi postali) che a sua volta è influenzata da numerosi fattori come il periodo dell'anno, le guerre, le epidemie, gli attacchi terroristici, le condizioni atmosferiche, il prezzo e la qualità dei servizi offerti dalle strutture turistiche.

Inoltre, il prodotto venduto dalle compagnie aeree è un bene di consumo monouso, in quanto, una volta che l'aereo è in volo, i posti vuoti non possono più essere venduti (Budd & Ison 2017). Le compagnie aeree, quindi, pongono molta attenzione al coefficiente di riempimento degli aeromobili (*load factor*), un valore che indica la percentuale di posti occupati dai passeggeri (Beers 2022). Trasportando più passeggeri, le compagnie aeree aumentano le loro entrate e possono raggiungere un pareggio¹²⁷, oltre il quale realizzano un profitto (Pande & Hardiman 2022).

- *Modello di business delle compagnie aeree*

Tutte le linee aeree hanno come obiettivo il trasporto di passeggeri o di merci da una destinazione a un'altra ma differiscono in base al proprio modello di business¹²⁸: esso definisce gli obiettivi e gli scopi specifici delle varie attività e può essere ridefinito al variare delle condizioni del mercato (Whyte & Lohmann 2017). Tra le linee aeree dedicate al trasporto passeggeri si hanno innanzitutto i modelli di business delle compagnie aeree di bandiera¹²⁹/compagnie aeree tradizionali (FSNC - *full service network carriers*) e delle compagnie aeree low cost (LCC - *low cost carriers*). Nelle compagnie aeree low cost si rivolge particolare attenzione alla riduzione e al controllo dei costi: si risparmia spazio a bordo riducendo lo spazio tra le file di sedili; si tende a non usare manicotti d'imbarco¹³⁰ negli aeroporti; si cerca di spingere i clienti ad acquistare direttamente tramite i canali ufficiali della compagnia anziché tramite canali secondari (Whyte & Lohmann 2017). Dagli anni Novanta, da quando sono emerse le

¹²⁷ Il punto di pareggio corrisponde al numero di quantità vendute necessario a coprire i costi. Tra le varie compagnie aeree, il punto di pareggio del coefficiente di riempimento può variare ma in media è intorno al 70% (Pande & Hardiman 2022).

¹²⁸ Un business model è una "struttura concettuale o piano che definisce come una compagnia conduce il proprio business" (Whyte & Lohmann 2017: 108, traduzione mia).

¹²⁹ Il termine "compagnia aerea di bandiera" in origine designava le compagnie aeree pubbliche, di proprietà dello stato, ma oggi viene utilizzato anche nel caso in cui l'originaria compagnia di bandiera sia stata ceduta in parte o venduta dallo stato a dei privati, passando così dall'essere un'impresa pubblica a un'impresa privata (Wikipedia).

¹³⁰ Un manicotto d'imbarco è una struttura chiusa che collega un terminal al velivolo, consentendo ai passeggeri di salire e scendere rapidamente. (Wikipedia).

compagnie aeree low cost (ibid.), anche le compagnie aeree tradizionali, nel caso delle tratte più brevi, hanno cercato di ridurre i costi e presentare un'offerta per cui a differenti tariffe corrispondono diversi servizi e benefit. Degli esempi di compagnie aeree di bandiera sono Air France-KLM, Lufthansa, British Airways e ITA Airways, mentre delle compagnie aeree low cost sono Ryanair, EasyJet, Vueling e Volotea. Le differenze tra i due tipi di compagnie vengono descritte da diversi autori (Whyte & Lohmann 2017; Wittmer et al. 2011; Bertolin et al. 2018) ed evidenziate nella tabella 14.

Caratteristica del business model	Compagnia di bandiera	Compagnia aerea low cost
Interno degli aeromobili	Gli interni degli aeromobili hanno una o più isole di sedili. La distanza minima tra le file di sedili è intorno agli 81-86 cm.	Vengono utilizzati gli aeromobili di tipo Boeing 737 o Airbus A319/320, che all'interno hanno una struttura molto stretta di sedili. La distanza minima tra le file di sedili è intorno ai 71-76 cm.
Utilizzo degli aeromobili	Gli aeromobili vengono utilizzati per lunghe tratte che iniziano e terminano in luoghi che potrebbero avere diversi fusi orari. Per questo, tra un utilizzo e l'altro del velivolo, ci sono dei periodi di pausa per permettere al personale di riposare.	Gli aerei vengono utilizzati per un massimo di 11 ore consecutive al giorno che comprendono voli brevi (durata di 1-2 ore) e periodi di pausa tra un volo e l'altro molto veloci (25-30 minuti).
Modello di rete delle rotte	Le rotte hanno generalmente una struttura <i>hub-and-spoke</i> ¹³¹ che si sviluppa tra grandi aeroporti e su lungo e breve raggio: ogni compagnia ha un "hub" centrale, ovvero un aeroporto in cui concentra la propria flotta aerea. Per essere trasportati da una destinazione all'altra, i passeggeri fanno scalo	Le rotte hanno generalmente una struttura <i>point-to-point</i> : i voli sono diretti da una destinazione all'altra senza scali. La flotta aerea è dislocata in un numero maggiore di aeroporti.

¹³¹ L'espressione *hub-and-spoke* è stata creata per analogia con la ruota della bicicletta: *hub* significa mozzo, mentre *spoke* raggio. Questa organizzazione delle rotte ha quindi una parte centrale da cui si collegano numerose parti periferiche (Wikipedia).

	nell'hub della compagnia e poi ripartono con dei voli "spoke" verso le destinazioni finali.	
Prodotti offerti	Vengono offerte diverse tipologie di biglietto aereo (prima, seconda, terza o quarta classe) e nel prezzo del biglietto sono compresi dei pasti e dei servizi di intrattenimento per il viaggio.	Di solito è disponibile una sola classe che prevede i servizi di base, mentre i servizi aggiuntivi (scelta del posto, pasti, intrattenimento a bordo) necessitano un pagamento extra.
Strategia di prezzo	Vengono proposte diverse categorie di prezzo. Inizialmente venivano offerti solo biglietti andata-ritorno, ma poi, nel corso del tempo, sono stati introdotti i biglietti di sola andata per adattarsi all'offerta delle compagnie low cost.	Di solito la struttura tariffaria è semplice e prevede un'unica tariffa base a cui eventualmente aggiungere dei servizi extra a pagamento. Vengono spesso attivate promozioni per stimolare la domanda.
Target	I consumatori a cui si rivolgono queste compagnie aeree sono generalmente coloro che viaggiano per lavoro, devono compiere medie o lunghe tratte e sono disposti a pagare un prezzo più alto per avere inclusi nel biglietto dei servizi aggiuntivi.	I consumatori a cui si rivolgono le linee aeree low cost sono tutti quei viaggiatori che compiono brevi tratte e che non sono interessati a particolari tipi di servizio (pasti o intrattenimento).
Relazione con il cliente	Per queste compagnie aeree è importante instaurare una relazione duratura con i clienti e portarli a fare degli acquisti ripetuti. Molte compagnie offrono i programmi fedeltà <i>frequent flyer</i> che mirano ad incentivare gli acquisti dei clienti tramite una raccolta punti da utilizzare successivamente per la riscossione di premi.	Le compagnie aeree low cost di solito non offrono programmi fedeltà e non mirano a sviluppare una relazione duratura con il cliente.

Bagaglio	Nella fascia economy il prezzo del biglietto include un bagaglio da stiva da 23 kg.	Per portare un bagaglio da stiva si deve solitamente pagare un prezzo aggiuntivo.
Comunicazione e promozione	La reputazione aziendale è molto importante e la comunicazione aziendale potrebbe includere degli stereotipi che si basano sulla cultura di appartenenza della compagnia aerea (ad esempio, puntualità per le compagnie dei paesi del nord Europa, cibo di qualità e accoglienza per le compagnie italiane).	I messaggi comunicativi spesso sono molto semplici ed esaltano il prezzo e le destinazioni.
Canali di distribuzione	Per gli acquisti dei biglietti viene incoraggiato l'utilizzo delle piattaforme online, ma continuano ad avere un ruolo rilevante anche le agenzie viaggi.	L'obiettivo è portare i clienti ad acquistare dai canali online ufficiali e ad evitare fonti secondarie a cui la compagnia poi dovrebbe pagare una commissione.
Fonte di guadagno e coefficiente di riempimento	Per queste compagnie un buon coefficiente di riempimento si avvicina all'80%. Il guadagno deriva da un prezzo del biglietto più elevato rispetto alle compagnie aeree low cost.	Le compagnie aeree low cost mirano a raggiungere un coefficiente di riempimento prossimo al 100% in modo da poter mantenere i prezzi dei biglietti più bassi.

Tabella 14: differenze tra compagnie aeree di bandiera e compagnie aeree low cost (Whyte & Lohmann 2017: 110-112; Wittmer et al. 2011: 72-73; traduzione mia; Bertolin et al. 2018).

Oltre a questi due modelli di business è presente il modello delle *compagnie aeree charter* che offrono un servizio di trasporto diretto per determinate destinazioni turistiche: di solito, un tour operator stipula un contratto di noleggio (*charter* significa infatti “noleggiare”) con la linea aerea, acquistando tutti i post sull’aereo a un prezzo concordato (Whyte & Lohmann 2017). Un altro modello di business è quello delle *compagnie aeree regionali*, che operano su brevi tratte che hanno una bassa domanda e si trovano in aeree geografiche difficili da raggiungere con mezzi di trasporto di superficie (ibid.). Per trasportare merci o recapitare la posta, alcune compagnie aeree tradizionali hanno una parte di flotta dedicata a voli cargo (e.g. Lufthansa Cargo), mentre in altri casi sono

presenti *compagnie aeree cargo* (ibid.) dedicate esclusivamente al trasporto di merci (come nel caso di FedEx, Polar Air, Cargolux). Infine, esistono anche modelli di business ibridi che caratterizzano alcune compagnie aeree low cost nel momento in cui offrono dei servizi aggiuntivi molto simili a quelli offerti dalle compagnie aeree tradizionali. Ad esempio, Easyjet offre dei servizi aggiuntivi a pagamento come l'imbarco prioritario o l'accesso a un'area lounge¹³² in aeroporto, mentre Jetblue, offre dei posti in cui i sedili possono essere reclinati fino a diventare dei letti (ibid.).

- *Gestione degli aeroporti*

Per far sì che un aeroporto funzioni adeguatamente bisogna considerare i seguenti postulati presentati da De Neufville (2017: 61-78):

1. Nel settore del trasporto aereo è quasi impossibile fare delle previsioni corrette (ivi, p. 62). Più la previsione è a lungo termine e dettagliata, più è probabile che emergano degli eventi che porteranno a una deviazione dalla previsione. Le attività delle compagnie aeree e degli aeroporti sono ampiamente influenzate dai *trend breakers* (ibid.), ovvero tutti quegli eventi o fattori che interrompono i modelli di servizio, domanda e modalità operative consolidati, come nel caso di crisi economiche, eventi geopolitici, crisi sanitarie, catastrofi naturali, nuove rotte aeree e nuove alleanze tra compagnie aeree. Nel momento in cui emerge un evento inaspettato avranno successo quelle compagnie aeree che saranno in grado di adattarsi velocemente al cambiamento.

2. La flessibilità nell'organizzazione e nel design degli spazi è essenziale (ivi, p.66). Data l'impossibilità di avere delle previsioni affidabili, diventa difficile essere preparati e pianificare le attività degli aeroporti in anticipo. È necessario quindi organizzare gli spazi tramite un *design flessibile*¹³³ che consenta agli aeroporti di adattare le proprie infrastrutture (piste, sale d'attesa, lounge, area bagagli, terminal) in modo rapido ed efficiente di fronte a livelli imprevisi di traffico e clienti, a nuove

¹³² Easyjet è stata la prima compagnia aerea low cost ad aprire un'area lounge nell'aeroporto di Londra Gatwick (Strano 2021).

¹³³ Nel design, la flessibilità si riferisce ad un tipo di progettazione che tiene conto della possibilità di adattare gli spazi e le strutture a diverse esigenze o situazioni in modo rapido ed efficiente (Budd & Ison 2017).

procedure e requisiti normativi. In questo modo non bisognerà adattare le strutture in circostanze difficili e non aumenteranno i costi, i tempi di realizzazione e lo spreco di risorse.

3. Ogni aeroporto compete con gli altri in termini di traffico di passeggeri ed entrate economiche e il suo vantaggio competitivo è influenzato da diversi fattori, come la presenza o meno degli hub¹³⁴ delle compagnie aeree, le rotte disponibili e la qualità dei servizi aeroportuali (ivi, p.68). La presenza dell'hub di una compagnia aerea influenza il livello di traffico aereo: ad esempio, quando IAG ha acquisito British Airways e Iberia nel 2011, c'è stato un calo del 30 per cento del traffico aereo e di passeggeri verso Madrid, in quanto gli hub sono stati spostati nel Regno Unito. Inoltre, ogni compagnia aerea può riorganizzare facilmente le proprie rotte grazie alla mobilità dello staff e degli aeromobili causando, anche in questo caso, delle ripercussioni negative sul traffico di passeggeri e sulle entrate economiche degli aeroporti. Infine, i passeggeri che devono effettuare degli scali hanno molte opzioni a disposizione per scegliere sia la compagnia aerea che l'aeroporto. La scelta potrebbe essere guidata non solo dal prezzo del biglietto aereo ma anche dalla qualità dell'aeroporto (e.g. efficienza delle infrastrutture, presenza di negozi) in cui viene effettuata la connessione.

4. La misura della capienza degli spazi di un aeroporto può essere ingannevole (ivi, p. 70). Essa non è un semplice numero che può essere determinato in modo assoluto, ma dipende dalle aspettative dei clienti riguardo la qualità, la gestione e la fornitura dei servizi: una scarsa capienza non significa che l'utente non possa usufruire di un servizio, ma piuttosto che ne usufruisce con un livello di qualità inferiore rispetto a quello desiderato. La capienza degli spazi di un aeroporto dipende dalla velocità dei servizi offerti e dalla gestione del flusso di viaggiatori. L'incremento del numero di unità operative (ad esempio le cabine per il controllo dei documenti o le corsie per i controlli di sicurezza) e il conseguente aumento del personale impiegato, può aumentare la velocità dei servizi offerti. Inoltre, per controllare il flusso dei viaggiatori è possibile adottare strategie mirate: negli aeroporti di Heathrow e Singapore, ad esempio, le aree del gate vengono aperte solo pochi minuti prima della partenza, in modo da imbarcare immediatamente i passeggeri ed evitare sovraffollamenti nelle aree

¹³⁴ Un hub è l'aeroporto in cui una compagnia aerea concentra la propria flotta aerea (Bisignani 2009).

del gate. Per gestire la capacità delle piste di atterraggio, invece, è stato introdotto il sistema degli slot¹³⁵, che consistono in una autorizzazione al decollo o all'atterraggio del volo in un determinato giorno e in una determinata fascia oraria. Come spiegano alcuni studiosi (Budd & Ison 2017; Graham 2017) la International Air Transport Association (IATA) regolamenta l'assegnazione degli slot tramite regole precise. Una di queste è il principio dei diritti acquisiti ("grandfather rights"), per cui una linea aerea che ha utilizzato uno slot nella stagione precedente ha il diritto di utilizzarlo nuovamente, ma solo a patto che abbia rispettato la regola del "use it or lose it", secondo cui la linea aerea deve far volare l'80 per cento dei voli associati a quello slot. L'adesione alla IATA è volontaria e quindi anche il rispetto delle regolamentazioni da essa proposte ma, in ogni caso, nell'Unione Europea vige il regolamento EU/95/93 per l'assegnazione degli slot.

5. Le file d'attesa sono un elemento costantemente presente nelle operazioni aeroportuali (De Neufville 2017: 74): i viaggiatori usufruiscono dei servizi in ordine sequenziale (prima si recano al check-in, poi ai controlli di sicurezza e infine al gate) e si ritrovano ad aspettare il proprio turno in delle code, che si formano per via della frequenza irregolare con cui i essi si recano nei vari punti dell'aeroporto. Le code formate dai viaggiatori vanno gestite intervenendo sulla quantità di personale impiegato, per "prevenire, o almeno per ridurre al minimo, le peggiori condizioni operative" (De Neufville 2017: 77, mia traduzione).

- *La reputazione delle compagnie aeree*

Come fa notare Newton (2014) il settore del trasporto aereo costituisce un esempio di come la reputazione collettiva di un intero settore possa influenzare la reputazione di una singola azienda operante in tale settore. Le scarse prestazioni che si potrebbero registrare all'interno del settore potrebbero avere un "effetto deformante riducente" (Cuomo et al. 2012: 11) sul giudizio e sulle aspettative dei clienti di una compagnia aerea, anche nel caso in cui questa stia performando meglio rispetto alla media. Pertanto, il mantenimento

135

https://www.faa.gov/about/office_org/headquarters_offices/ato/service_units/systemops/perf_analysis/slot_administration/slot_definition#:~:text=In%20the%20context%20of%20airport,control%20clearance%20or%20similar%20authorizations [ultima consultazione: 09/04/23]

e il miglioramento della reputazione diventa una questione collettiva, che coinvolge tutte le compagnie aeree nel loro complesso.

Inoltre, Newton (2014) sottolinea che le scarse prestazioni di questo settore spesso sono causate da fattori non controllabili dalle compagnie aeree (i sopracitati *trend breakers*) come, ad esempio, le condizioni meteorologiche, le regolamentazioni nazionali e internazionali e il prezzo delle materie prime.

Secondo Matthews (2022), nel momento in cui la reputazione di una singola azienda è fortemente influenzata dalla reputazione collettiva e da fattori negativi comuni a tutte le aziende (come nel caso del periodo di crisi e disagi dell'estate del 2022), una compagnia aerea dovrebbe concentrare i propri sforzi sull'implementazione dei servizi di comunicazione e assistenza, in modo da riacquistare la fiducia dei clienti e mettere in risalto i propri punti di forza rispetto ai concorrenti. Ciò significa riuscire a supportare i clienti in ogni fase del viaggio e creare la migliore esperienza possibile.

- *Gestione del personale nel settore del trasporto aereo*

Come evidenziano gli autori (Harvey & Turnbull 2017), una gestione efficace del personale è essenziale per garantire un'erogazione di qualità dei servizi delle compagnie aeree.

Le compagnie aeree tradizionali puntano sull'offerta di un servizio clienti di alto livello e necessitano quindi di personale altamente qualificato, mentre le compagnie aeree low cost, che basano la propria competitività sul prezzo dei biglietti, chiedono ai propri dipendenti di essere efficienti, veloci e produttivi, in modo da ridurre i costi unitari (costo per passeggero per chilometro volato¹³⁶) e trasferire questo risparmio ai consumatori offrendo loro tariffe più basse.

Come fanno notare gli autori (ibid.) il settore del trasporto aereo è caratterizzato da una domanda ciclica e strettamente connessa alla variazione del prodotto interno lordo di un paese. La business class è quella che risente maggiormente di queste fluttuazioni, poiché le aziende che acquistano i biglietti per i propri dipendenti sono meno inclini a utilizzare le proprie risorse economiche per servizi premium nei momenti di recessione. Per fronteggiare un calo della domanda e un conseguente calo del fatturato e dei ricavi è

¹³⁶ L'RPK (revenue passenger kilometer) misura il numero di chilometri volato per passeggero.

necessario che la compagnia aerea tagli dei costi. Una delle principali strategie è quella di tagliare i costi relativi al personale, in quanto sono facilmente controllabili dalle compagnie aeree (rispetto invece ai costi del carburante o dei velivoli) e costituiscono una parte significativa dei costi totali (circa un terzo nel caso delle compagnie aeree tradizionali, secondo quanto esposto da Harvey & Turnbull 2017). In passato, ogni qualvolta ci sia stata una recessione economica (ad esempio dopo l'attacco terroristico alle Torri Gemelle o nel 2008 con la crisi finanziaria) le misure per ridurre i costi si sono tradotte nel congelamento degli stipendi, in tagli salariali, nella sospensione dei benefit, in offerte di licenziamento volontario, congedo volontario o obbligatorio, nel mancato rinnovo di contratti a tempo determinato e nella presenza di personale in prova non trasferito a tempo pieno. Queste misure hanno provocato delle proteste da parte dei lavoratori tramite sindacati e scioperi che hanno bloccato il normale proseguimento dei servizi e hanno causato una perdita di guadagno per le compagnie aeree.

Si comprende quindi che questo tipo di approccio è deleterio, in quanto impatta primariamente sui lavoratori che sono fondamentali per il successo della compagnia aerea e nella gestione della relazione con i passeggeri. Southwest Airlines (citato in Harvey & Turnbull 2017: 317) è un esempio di compagnia aerea che è riuscita a costruire una relazione proficua con i propri dipendenti e, di conseguenza, con i propri clienti. Essa offre al proprio staff stipendi e benefit adeguati, corsi sulla sicurezza, corsi di formazione e possibilità di crescita. I sindacati sono trattati come "business partner" e non come nemici. Il risultato di questo approccio è che in quarant'anni di storia i dipendenti hanno scioperato una sola volta (Gittel et al. 2004 citati in Harvey & Turnbull 2017: 317).

- *Compagnie aeree e servizio clienti su Twitter*

Le compagnie aeree possono svolgere l'attività di servizio clienti online, in particolare attraverso i social media come Twitter, che danno l'opportunità di interagire con gli utenti in modo più rapido e su scala globale. I viaggiatori possono pubblicare vari tipi di contenuti (testo, video, audio), accedere a quelli condivisi da altri utenti e comunicare con l'azienda. Inoltre, in caso di disagi o problemi, i clienti possono fare delle segnalazioni in tempo reale e, nel migliore dei casi, le compagnie aeree sono in grado di fornire un aiuto tempestivo. Inoltre, come evidenziato precedentemente, Twitter consente agli utenti di comunicare in una struttura da uno a molti in cui un emittente trasmette un messaggio a

più riceventi contemporaneamente (Twitter 2016). Come evidenzia Budd (2017) nel contesto online attuale i clienti si aspettano una presenza stabile e attiva delle compagnie aeree e queste ultime riconoscono l'importanza dei social media nelle loro strategie comunicative e di marketing.

I social media portano con sé delle opportunità ma anche delle sfide. Grazie ad essi le compagnie aeree possono dare aggiornamenti in tempo reale sullo stato dei voli, sull'insorgere di particolari problemi (ritardi, cancellazioni) o sull'evoluzione delle situazioni di crisi. Al contempo però, la compagnia aerea deve preoccuparsi di investire in personale specializzato che si dedichi alla gestione di questo tipo di comunicazione, ovvero che si occupi di pubblicare contenuti coerenti tra le varie piattaforme, monitorare le conversazioni, rispondere tempestivamente alle richieste dei clienti e gestire in modo adeguato i feedback negativi con lo scopo di preservare la reputazione (ibid.).

La prima compagnia aerea che ha compreso il potenziale dell'utilizzo dei social media per il servizio clienti è l'olandese KLM, che viene citata da Budd (2017) come un caso esemplare. Nel 2014, durante un'intervista con l'MIT Sloan Management Review (Kane 2014), Martijn van der Zee (che in quel momento era il vicepresidente di Air France-KLM¹³⁷) ha descritto l'evoluzione dell'utilizzo dei social media da parte della compagnia per l'attività di servizio clienti. Tutto è iniziato nella primavera del 2010 quando le eruzioni del vulcano islandese Eyjafjöll interruppero i viaggi aerei nell'Europa occidentale e settentrionale per diversi giorni, influenzando anche la programmazione delle rotte di volo nelle settimane successive. In questa situazione di blocco per il trasporto aereo, i clienti hanno iniziato a contattare la compagnia aerea KLM (presente nei social media dal 2009), che avrebbe dovuto operare nella zona che si trovava in condizioni meteorologiche avverse. Arrivavano moltissime richieste da parte dei clienti che non potevano essere processate solo tramite gli aeroporti, il sito web o il call center della compagnia. Gli utenti iniziarono così ad usare Twitter e Facebook come canali alternativi per comunicare con l'azienda. In questa situazione, KLM è riuscita a individuare una nuova soluzione per l'attività di servizio clienti. Dopo il ritorno alla normalità e il ripristino dei viaggi aerei, la compagnia ha deciso di integrare l'utilizzo dei social media per il servizio clienti nella propria strategia comunicativa anche in situazioni

¹³⁷ Air France-KLM è una holding franco-olandese, risultato della fusione del 2004 tra la compagnia francese Air France e la compagnia olandese KLM (Wikipedia).

normali di business, riconoscendo il valore aggiunto che portava nella gestione dei problemi e nel supporto ai clienti. Grazie all'intuizione avuta in un momento di disagio dei trasporti, oggi KLM è in grado di fornire un servizio clienti attivo 24 ore su 24 e 7 giorni su 7.

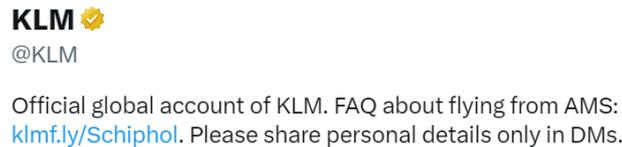


Figura 10: descrizione nel profilo Twitter di KLM¹³⁸.

Col tempo KLM è stata raggiunta anche da molte altre compagnie che sono presenti sui social media (come Twitter) con i propri account ufficiali e tramite questi danno supporto diretto ai clienti 24 ore su 24, 7 giorni su 7 o in determinate fasce orarie.



Figura 11: screenshot del testo di presentazione dei profili ufficiali delle compagnie aeree su Twitter.

Inoltre, durante l'intervista con l'MIT Sloan Management Review, Martijn van der Zee ha sottolineato come KLM ponga molta attenzione all'ascolto dei clienti sui social, raccogliendo dati sul sentiment, sui paesi di provenienza dei messaggi e sulle ore di punta. L'obiettivo è avere una visione precisa di cosa viene detto sulla compagnia e comprendere i punti di forza e i punti di debolezza dei servizi offerti, in modo da sviluppare soluzioni efficaci per i viaggiatori.

6. Aspetti finanziari (Jackson 2017)

¹³⁸ <https://twitter.com/KLM> [ultima consultazione: 15/04/23]

Per garantire la continuità operativa di una compagnia aerea, che necessita di ingenti capitali per poter mantenere la propria flotta e investire in nuove rotte, è fondamentale che le condizioni finanziarie siano solide. Come spiega Jackson (2017), le compagnie aeree che non riescono a finanziare le proprie attività rischiano di essere acquisite da altre compagnie o di cessare di esistere.

Gli aeromobili hanno un costo molto alto e le compagnie aeree possono ricorrere a diverse opzioni per il loro acquisto. Tra queste si hanno i prestiti bancari, i crediti all'esportazione (forniti dai governi per supportare la vendita dei propri velivoli in altre nazioni), l'emissione di nuove azioni per ottenere finanziamenti aggiuntivi e il leasing, ovvero il noleggio degli aeromobili (ibid.).

Jackson (2017) spiega che alla fine di ogni anno, a prescindere dalla composizione della flotta aerea (aeromobili nuovi o usati, acquistati o noleggiati) e dal modello di business, ogni compagnia aerea deve presentare un report finanziario che dia un'indicazione dettagliata della performance finanziaria della compagnia aerea a tutti i suoi stakeholder.

Quando la domanda di voli cala, le compagnie aeree registrano una rapida diminuzione delle proprie entrate e devono mettere in atto delle strategie per tutelare il proprio stato finanziario ed evitare di essere insolventi¹³⁹, dichiarare bancarotta o, nel peggiore dei casi, uscire dal mercato (ibid.). Per ridurre i costi si possono restituire gli aeromobili a chi ha concesso il leasing, rinegoziare i termini del leasing o delle tariffe aeroportuali, cambiare le rotte, licenziare il personale o modificare i contratti d'assunzione.

4.1.2 Il settore aereo tra pandemia di COVID-19 e ripresa nell'estate 2022

Nell'estate del 2022 il settore del trasporto aereo è stato caratterizzato da diversi disagi, tra cui cancellazioni e ritardi dei voli, affollamento degli aeroporti e scioperi del personale. Per capire come si è giunti a questa situazione è necessario riflettere sugli effetti negativi della pandemia di COVID-19 sul settore dell'aviazione.

Lo stato di pandemia, dichiarato dall'Organizzazione Mondiale della sanità l'11 marzo 2020, ha impattato negativamente su molti settori, tra cui quello del trasporto aereo, che ha dovuto affrontare una crisi globale senza precedenti. I dati forniti

¹³⁹ Un'azienda è insolvente quando non riesce a pagare i propri debiti.

dall'Organizzazione internazionale dell'aviazione civile (ICAO 2023) mostrano infatti che l'epidemia di COVID-19, rispetto alle situazioni di crisi del passato (crisi petrolifera, Guerra Iran-Iraq, Guerra del Golfo, crisi economica in Asia, attacco terroristico alle Torri Gemelle, SARS 2003, crisi finanziaria del 2008) ha fatto diminuire con velocità e intensità maggiore il numero di passeggeri che, rispetto al 2019, ha visto un declino del 60% nel 2020, per poi recuperare fino a -28% nel 2022, man mano che la situazione sanitaria è migliorata e si è stabilizzata (si veda la figura 12).

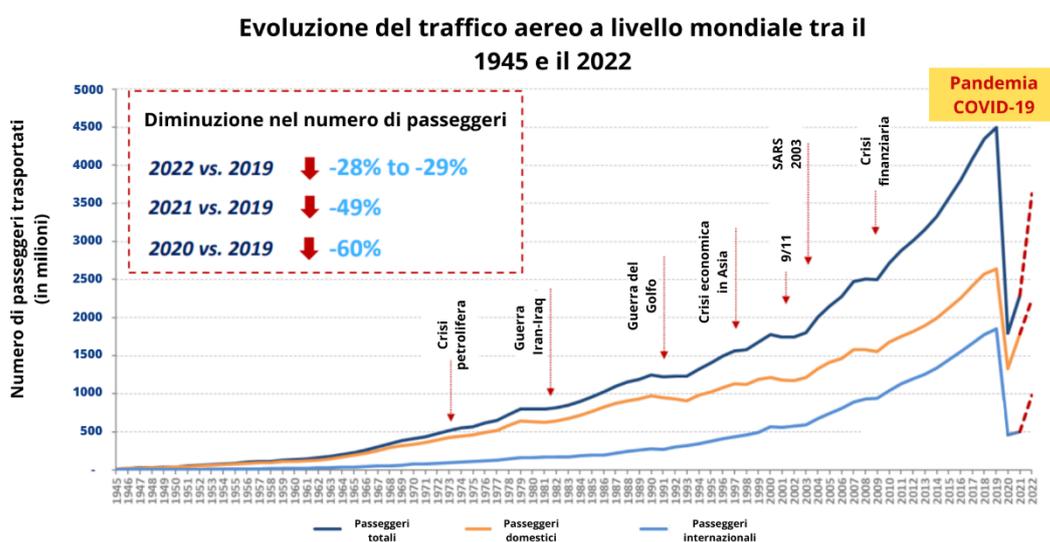


Figura 12: grafico che rappresenta l'evoluzione del numero di passeggeri nel corso del tempo. Traduzione del grafico pubblicato dall'Organizzazione internazionale dell'aviazione civile (ICAO 2023).

Con lo scoppio della pandemia i governi di tutto il mondo hanno adottato delle misure di contenimento del virus (lockdown, chiusura delle frontiere, distanziamento sociale) e il settore del trasporto aereo, come tanti altri, ha dovuto reagire tempestivamente alle nuove disposizioni riorganizzandosi a livello operativo ed economico (Thepchalerm & Ho 2021).

Per poter adattare le proprie procedure di volo e di terra a questa crisi sanitaria improvvisa, le compagnie aeree hanno seguito le linee guida emanate dai governi e da organizzazioni quali ICAO e IATA. Quest'ultima (IATA 2020b, citato in Thepchalerm & Ho 2021), ad esempio, ha indicato come procedure di contenimento del virus il distanziamento sociale e l'utilizzo delle mascherine a bordo degli aeromobili, la frequente sanificazione degli interni degli aeromobili, il controllo della temperatura dei viaggiatori prima e dopo il volo, il tracciamento dei passeggeri e dello staff e la somministrazione di

test rapidi. Oltre a queste linee guida, le compagnie aeree hanno applicato misure aggiuntive, come la creazione di uniformi protettive per il personale e la sospensione del servizio bar durante il viaggio (in modo da ridurre il contatto tra staff e passeggeri). Negli aeroporti, invece, come prevedevano Serrano e Carra (2020), la diminuzione della domanda di voli ha portato a concentrare i velivoli in pochi terminal e, in alcuni casi, per contrastare la diffusione del virus sono stati introdotti sistemi self-service per il check-in, l'imbarco dei bagagli e il controllo dei documenti.

In alcuni casi, però, le disposizioni dei governi hanno previsto la chiusura degli aeroporti e le compagnie aeree si sono trovate a dover sospendere del tutto le proprie attività (IFC 2020).

La riduzione e la sospensione delle attività delle compagnie aeree ha impattato negativamente sull'utilizzo della flotta di aeromobili e sul raggiungimento del numero minimo di voli per mantenere gli slot attivi¹⁴⁰. Per questo motivo a marzo 2020, sia l'Unione Europea (con il regolamento UE 2020/459¹⁴¹) che la IATA (con il *Justified non utilization of slots*¹⁴²) hanno introdotto delle deroghe alla regola del “use it or lose it”, in modo che le compagnie aeree non perdessero gli slot per cui non riuscivano a garantire un numero minimo di voli.

Come spiega Oliviero Baccelli, docente di economia e politica dei trasporti all'Università Bocconi di Milano, intervistato da Rosalba Reggio de Il sole 24 ore (2022) a causa delle restrizioni alla mobilità, c'è stata una diminuzione del numero di voli passeggeri (soprattutto per quanto riguarda le rotte intercontinentali) e, in generale, un calo del coefficiente di riempimento degli aeromobili. Una tendenza opposta, invece, si è registrata nel mercato dei voli cargo: gli acquisti online si sono moltiplicati e quindi anche la domanda per lo spostamento delle merci. Alla luce della diminuzione della domanda di voli passeggeri, per poter continuare a rimanere in una condizione finanziaria solida le compagnie aeree hanno messo in atto diversi espedienti, tra cui la richiesta di aiuti economici al governo (70 miliardi di dollari forniti dai governi alle compagnie aeree

¹⁴⁰ Prima della pandemia, secondo le regole imposte da IATA e dall'Unione Europea, per mantenere attivo uno slot, la compagnia aerea doveva far volare almeno l'80 per cento di voli associati a quello slot.

¹⁴¹ <https://eur-lex.europa.eu/legal-content/IT/TXT/?uri=celex:32020R0459> [ultima consultazione: 10/04/23]

¹⁴² <https://www.iata.org/contentassets/4ede2aabfcc14a55919e468054d714fe/wasg-edition-1-english-version.pdf> [ultima consultazione: 10/04/23]

europee¹⁴³ tra il 2020 e il 2021), la conversione dei voli passeggeri in voli cargo (come spiega Baccelli intervistato da Reggio nel 2022) e il licenziamento del personale (secondo i dati di Eurostat¹⁴⁴, nel primo trimestre del 2022, rispetto al 2019, nel settore del trasporto aereo europeo hanno lavorato 85.000 persone in meno).

Dopo due anni dall'inizio della pandemia, nell'estate del 2022, il settore del trasporto aereo ha visto un notevole aumento della domanda (traffico pari all'83% del livello toccato nel 2019, secondo i dati forniti da Eurocontrol¹⁴⁵) grazie al miglioramento della situazione sanitaria e al desiderio delle persone di tornare a viaggiare. Molti viaggiatori però hanno avuto esperienza di lunghe code in aeroporto, bagagli smarriti, scioperi, voli cancellati o in ritardo. In un articolo sull'Harvard International Review, Chen (2022) analizza la situazione caotica in un cui si è ritrovato il settore del trasporto aereo nell'estate del 2022 e ne individua le cause principali nelle decisioni prese dal settore durante la pandemia e nella guerra in Ucraina.

A causa dell'emergenza sanitaria il settore del trasporto aereo si è dovuto riorganizzare e ha dovuto ridurre le proprie operazioni. In particolare, sono stati fatti dei tagli al personale, riducendo gli stipendi o licenziando i piloti, i tecnici altamente specializzati, il personale di terra o di volo. Nell'estate del 2022, nonostante l'aumento della domanda, le compagnie aeree non sono riuscite a organizzarsi in tempo per reintegrare il personale licenziato durante la pandemia e si sono ritrovate in una condizione inadeguata a soddisfare il volume e le esigenze dei passeggeri. In primo luogo, c'è stata una penuria di piloti: chi era stato licenziato si è trovato un nuovo lavoro, mentre chi è rimasto in servizio ha dovuto accettare orari di lavoro prolungati e livelli salariali più bassi. Queste condizioni di lavoro sono state il motivo scatenante dei numerosi scioperi che hanno colpito le compagnie aeree in più momenti dell'estate 2022.

Inoltre, si è verificata una carenza di tecnici altamente specializzati, la cui assunzione da parte delle compagnie aeree e degli aeroporti prevede un periodo di diversi mesi¹⁴⁶, in

¹⁴³ Secondo quanto riportato da Zandt (2021) sul sito di Statista, le compagnie aeree che hanno ricevuto più aiuti finanziari sono quelle nordamericane (105 miliardi di dollari), seguite da quelle europee (70 miliardi di dollari) e da quelle della regione asiatica (30 miliardi di dollari).

¹⁴⁴ <https://ec.europa.eu/eurostat/en/web/products-eurostat-news/-/ddn-20220810-1> [ultima consultazione: 10/04/23]

¹⁴⁵ <https://www.eurocontrol.int/publication/eurocontrol-analysis-paper-2022-year-european-aviation-bounced-back> [ultima consultazione: 10/04/23]

¹⁴⁶ Il periodo necessario all'assunzione del personale è di 10-12 settimane, secondo quanto riportato da Berberi (2022) sul Corriere della sera.

quanto è richiesto un addestramento specifico e l'ottenimento del nulla osta di sicurezza¹⁴⁷ (Chen 2022). Alcune compagnie aeree hanno cercato di affrontare la situazione adottando strategie, come l'offerta di bonus ai dipendenti che segnalavano all'azienda potenziali lavoratori disponibili o l'assunzione di personale già attivo all'estero, ma in molti casi esse hanno dovuto cancellare¹⁴⁸ i voli per mancanza di personale (ibid.).

Un altro elemento a sfavore della ripresa delle attività delle compagnie aeree è stato lo scoppio della guerra tra Russia e Ucraina, che ha causato un aumento dei costi per le materie prime e il carburante. Inoltre, la cancellazione e la modifica delle rotte (per non sorvolare su determinate aree geografiche) hanno portato anche ad utilizzare una maggiore quantità di carburante¹⁴⁹. Questi fattori hanno fatto sì che le compagnie aeree facessero ulteriori tagli sul personale o non avessero abbastanza fondi per degli stipendi adeguati, portando così, ancora una volta, i dipendenti a scioperare.

La combinazione della mancanza di personale, dell'aumento dei costi e degli scioperi ha causato la cancellazione, il ritardo e la riprogrammazione dei voli delle compagnie aeree. Questa situazione ha rappresentato una significativa perdita economica per le compagnie e ha impattato negativamente sull'esperienza dei clienti finali (Chen 2022) e sulla loro fiducia nel settore del trasporto aereo (Street 2022).

4.1.3 Crisi strisciante nell'estate 2022

La situazione in cui si è ritrovato il settore del trasporto aereo nell'estate del 2022 può essere considerata una crisi, ovvero una situazione operativa che ha messo a rischio la *business continuity* e che, se non adeguatamente risolta, può avere delle conseguenze negative sul rapporto con gli stakeholder (e pubblici influenti). Il periodo di crisi dell'estate del 2022 si inserisce nel contesto di una crisi più ampia e globale, la crisi

¹⁴⁷ Il nulla osta di sicurezza è un'abilitazione al trattamento di informazioni classificate contenute in documenti o altro materiale (Wikipedia).

¹⁴⁸ Le 10 compagnie aeree con la percentuale più alta di voli cancellati durante l'estate del 2022 sono, in ordine decrescente, Lufthansa, SAS, KLM, Eurowings, TAP Air Portugal, British Airways, Wizz Air, EasyJet, AirFrance, Vueling (Mabrian Technologies 2022).

¹⁴⁹ Le compagnie aeree sono sensibili all'aumento del prezzo e della quantità di carburante, in quanto esso costituisce il 35% dei costi di una compagnia aerea (Morrison et al. 2020 citati in Thepchalerm & Ho 2021).

pandemica iniziata nel 2020, che è stata *improvvisa*, poiché non ha dato segnali premonitori e ha colto di sorpresa il settore del trasporto aereo.

Il periodo dell'estate del 2022, invece, può essere definito come una crisi *strisciante* che prima di manifestarsi apertamente è stata anticipata da dei *segnali deboli*, ovvero indizi che forniscono un'avvisaglia di un evento negativo e che, se rilevati in tempo, permettono di prendere dei provvedimenti meno costosi e più facili da gestire. Delle previsioni sull'aumento della domanda di voli per l'estate del 2022, infatti, erano già disponibili a inizio 2022 (come quelle fornite da Eurocontrol¹⁵⁰ e IATA¹⁵¹) e molte prenotazioni erano già state effettuate nei primi mesi dell'anno, come riportato a gennaio 2022 dal World Travel and Tourism Council¹⁵². Inoltre, il Consiglio Europeo, a febbraio 2022, di fronte a una maggiore copertura vaccinale e a un'evoluzione positiva della pandemia, invitava i paesi membri a consentire la libera circolazione a partire dal primo marzo ai viaggiatori vaccinati o guariti. Di fronte a questi cambiamenti, a un aumento della domanda e a un miglioramento della situazione sanitaria, le compagnie aeree non sono riuscite a riorganizzarsi in tempo e far fronte alla mancanza di personale. Come attenuante di questo mancato adattamento, bisogna osservare che la variante Omicron del coronavirus, diffusasi tra fine 2021 e inizio 2022, ha portato ulteriori rallentamenti alle assunzioni (The Economist 2022) e la guerra tra Russia e Ucraina, scoppiata a febbraio 2022, ha causato un aumento dei costi per le compagnie aeree.

Sulla responsabilità dei disguidi che hanno caratterizzato il settore del trasporto aereo nell'estate del 2022 le opinioni sono molteplici e discordanti. Come riportano Philip et al. (2022), da un lato le compagnie aeree hanno lamentato di non essere riuscite a prepararsi adeguatamente a una domanda così alta per via di regolamenti che non favorivano una veloce ripresa (come nel caso dei lunghi tempi di attesa per i nulla osta di sicurezza) e degli aiuti finanziari limitati ricevuti dai governi; dall'altro, invece, i governi affermano che spetta al settore aereo gestire le proprie operazioni dopo aver ricevuto ingenti sussidi. Stephen Cotton (citato in Georgiadis & Chavez 2022, traduzione mia), segretario generale del sindacato International Transport Workers' Federation, afferma che il caos in cui si è

¹⁵⁰ <https://www.eurocontrol.int/press-release/eurocontrol-think-paper-15-charting-european-aviation-recovery-2021-covid-19-impacts>

¹⁵¹ <https://www.iata.org/en/pressroom/2022-releases/2022-03-01-01/>

¹⁵² <https://wtc.org/news-article/wtcc-predicts-a-strong-recovery-of-the-european-travel-and-tourism-sector-in-2022>

ritrovato il settore è “il risultato diretto di una serie di cattive decisioni prese dal governo e dalle compagnie aeree”. I governi avrebbero dovuto fornire più supporto alle compagnie aeree e queste ultime avrebbero dovuto riflettere sulle conseguenze del taglio dei costi e dei posti di lavoro. Willie Walsh (citato da *ibid.*) invece, presidente della IATA, difende le compagnie aeree e afferma che queste non avevano altra scelta se non quella di ridurre il personale per preservare le proprie attività durante la crisi pandemica.

I disagi vissuti dal settore del trasporto aereo sono stati il sintomo di una crisi strisciante ma, secondo Georgiadis e Chavez (2022) sono il presagio di una futura ripresa, del risanamento delle finanze e del ritorno alla redditività del settore.

4.1.4 Compagnie aeree selezionate

In questa ricerca si è scelto di raccogliere e analizzare dei tweet in lingua inglese, in modo da rilevare l'opinione di una maggiore quantità di parlanti e utilizzare efficacemente il modulo VADER di NLTK, che è stato sviluppato per processare i testi scritti in lingua inglese.

Per scegliere le compagnie aeree su cui raccogliere i dati, si è utilizzata la classifica fornita da Statista delle compagnie aeree leader in Europa per numero di passeggeri trasportati nell'anno 2021.

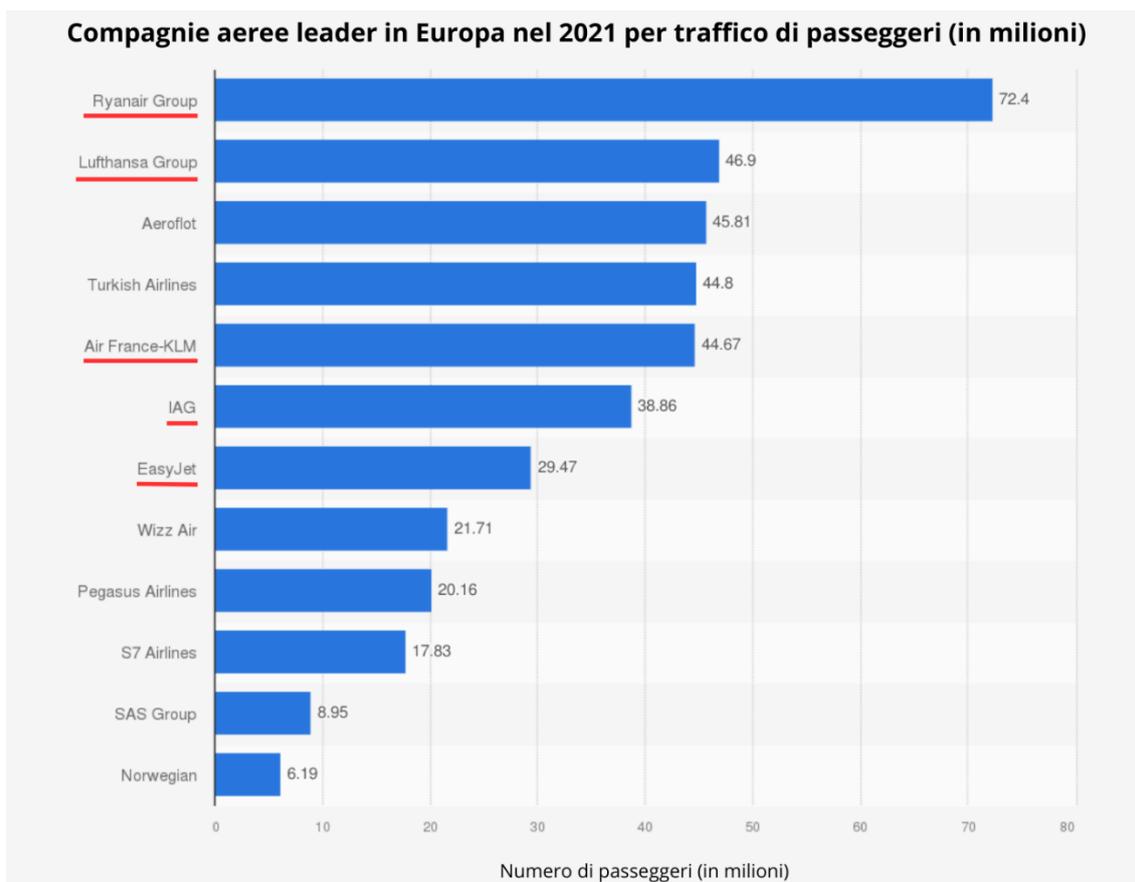


Figura 13: grafico che rappresenta le compagnie aeree leader in Europa per numero di passeggeri. Titolo tradotto e compagnie aeree scelte evidenziate. (Statista Research Department 2022).

Tra le dieci compagnie aeree mostrate ne sono state scelte cinque che avessero un account ufficiale su Twitter in lingua inglese e che utilizzassero quest'ultimo regolarmente per il servizio clienti e per fornire aggiornamenti sulle proprie attività. Le compagnie aeree scelte¹⁵³ sono, in ordine decrescente per numero di passeggeri trasportati:

- Ryanair: compagnia aerea low cost, fondata nel 1984 in Irlanda.
- Lufthansa: compagnia aerea tradizionale, fondata nel 1926 in Germania¹⁵⁴. Del Lufthansa Group¹⁵⁵ fanno parte anche le compagnie aeree Austrian Airlines, Eurowings, Brussels Airlines e Swiss International Airlines.

¹⁵³ Nell'analisi sono state escluse la compagnia aerea russa Aeroflot e la compagnia turca Turkish Airlines, in quanto la prima non ha un account ufficiale su Twitter, mentre la seconda, pur avendo un account ufficiale per il servizio clienti (@TK_HelpDesk) pubblica tweet prevalentemente in lingua turca.

¹⁵⁴ <https://www.lufthansagroup.com/en/chronicle.html> [ultima consultazione: 14/04/23]

¹⁵⁵ <https://www.lufthansa.com/it/it/lufthansa-aviation-group> [ultima consultazione: 14/04/23]

- KLM¹⁵⁶: compagnia aerea tradizionale, fondata nel 1919 nei Paesi Bassi¹⁵⁷. Dal 2004 fa parte del gruppo Air France-KLM.
- British Airways: compagnia aerea tradizionale, che nasce nel Regno Unito inizialmente come Aircraft Transport and Travel Limited (AT&T) nel 1919 e assume la forma attuale nel 1974 dopo la fusione di diverse compagnie aeree britanniche¹⁵⁸. Dal 2011, insieme a Iberia, compagnia aerea tradizionale spagnola, fonda il gruppo IAG (International Airlines Group) che controlla anche Aer Lingus, Vueling e Air Europa.
- EasyJet: compagnia aerea low cost, fondata nel 1995 nel Regno Unito¹⁵⁹.

4.1.5 Periodo considerato

Dopo aver scelto queste compagnie aeree, è stato utilizzato Google Trends per avere una panoramica dell'interesse degli utenti rispetto ai disagi che stavano caratterizzando il settore del trasporto aereo nell'estate del 2022.

Google Trends permette di inserire una query di ricerca, selezionare un periodo temporale e una regione specifica e ottenere dati circa il volume di ricerca di quella query, ovvero il livello di interesse degli utenti. I dati relativi al volume sono indicizzati in una scala da 1 a 100, dove 100 corrisponde al livello di interesse massimo. In questo caso, si è individuato il volume delle ricerche, da inizio gennaio a inizio ottobre 2022 in tutto il mondo, delle espressioni:

1. “[compagnia aerea]¹⁶⁰ strike”
2. “[compagnia aerea] delay”
3. “[compagnia aerea] cancellation”.

¹⁵⁶ KLM è l'acronimo di Koninklijke Luchtvaart Maatschappij, espressione olandese che significa “Compagnia Reale d'Aviazione”.

¹⁵⁷ <https://www.klm.com/information/corporate/history> [ultima consultazione: 14/04/23]

¹⁵⁸ <https://www.britishairways.com/it-it/information/about-ba/history-and-heritage/explore-our-past> [ultima consultazione: 14/04/23]

¹⁵⁹ <https://corporate.easyjet.com/about/our-journey> [ultima consultazione: 14/04/23]

¹⁶⁰ “[compagnia aerea]” corrisponde ai diversi nomi delle compagnie aeree considerate e inserite nelle parole chiave utilizzate per le ricerche su Google Trends.

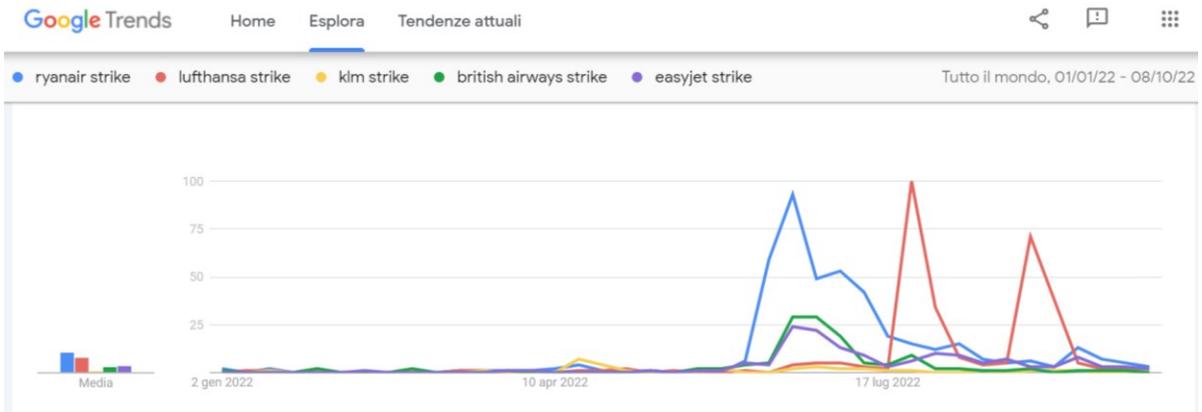


Figura 14: volume delle ricerche relative agli scioperi delle compagnie aeree.

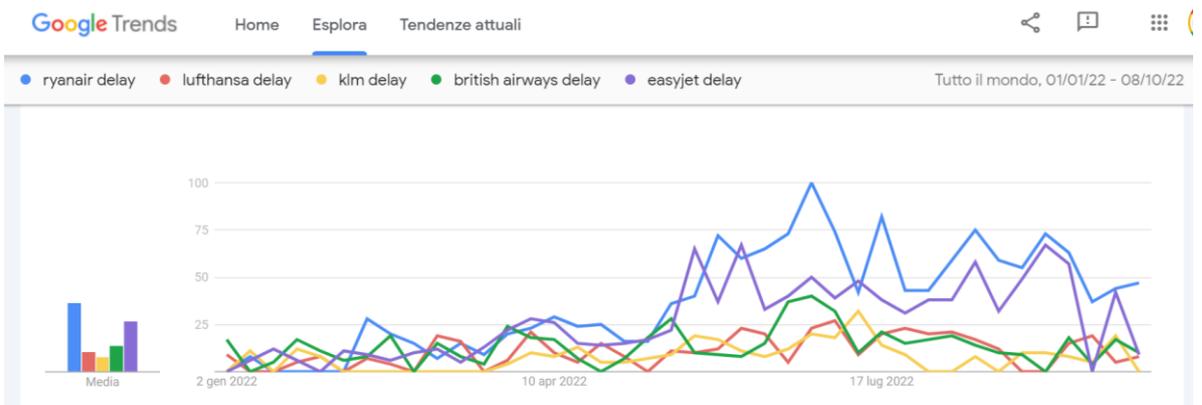


Figura 15: volume delle ricerche relative ai ritardi delle compagnie aeree.

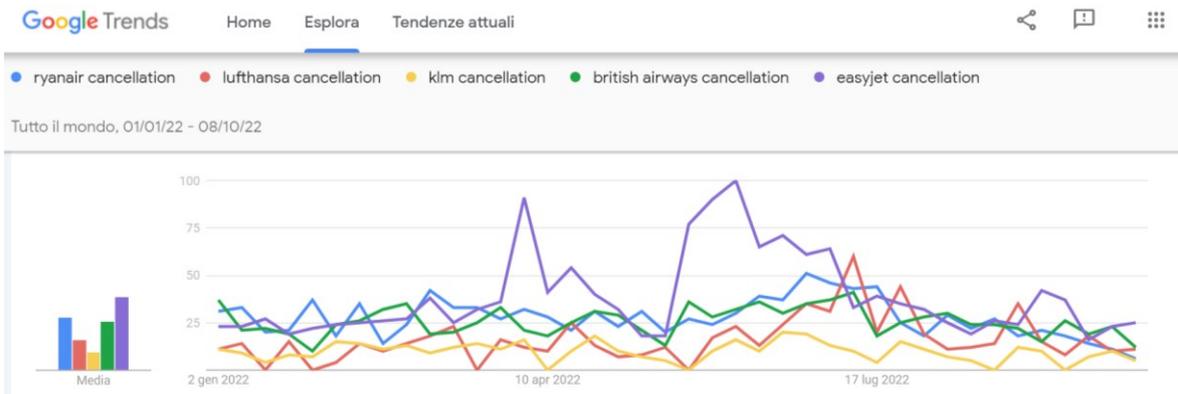


Figura 16: volume delle ricerche relative alle cancellazioni delle compagnie aeree.

Come si nota dal primo grafico (figura 14) l'interesse riguardo agli scioperi è stato quasi nullo fino a maggio 2022 ed è aumentato notevolmente nel periodo estivo nel caso di tutte le compagnie, tranne KLM che ha un volume di ricerca pressoché costante e molto basso. Gli utenti hanno cercato informazioni soprattutto rispetto agli scioperi di

Ryanair e Lufthansa e, in misura minore, rispetto agli scioperi di British Airways e EasyJet.

Nel secondo grafico (Figura 15), invece, relativo ai ritardi, il valore minimo del volume di ricerca è intorno a 25, ma da maggio a ottobre è aumentato per tutte le compagnie. La compagnia che ha raggiunto il picco più alto è stata Ryanair, seguita da EasyJet, British Airways, KLM e Lufthansa.

Nell'ultimo grafico (figura 16), che rappresenta il volume delle ricerche riguardanti le cancellazioni, l'andamento è costante per tutte le compagnie e sotto il valore di 50, tranne nel caso di EasyJet che ha dei picchi nel mese di aprile e a inizio giugno e Lufthansa che ha un picco intorno a metà luglio.

Rispetto a tutti e tre i grafici, si nota che la compagnia aerea KLM è quella che presenta i volumi di ricerca più bassi e ciò potrebbe essere dovuto a una buona performance sui tre aspetti indagati (scioperi, cancellazioni, ritardi) o a una concentrazione dell'interesse in un'area geografica diversa da quella qui presentata (in questo caso si considera l'interesse globale; a livello regionale potrebbero cambiare i volumi delle ricerche).

Dai dati forniti da Google Trends si è notato che il volume delle ricerche, in molti casi, è iniziato a salire già da fine maggio e si è quindi scelto come periodo di riferimento per lo scaricamento dei tweet quello compreso tra il 23 maggio 2022 al 2 ottobre 2022, che include tutti i mesi estivi.

4.2 Le domande di ricerca

Questa ricerca si propone di offrire una panoramica dell'attività del servizio clienti e della reputazione su Twitter di cinque compagnie aeree. Per farlo, vengono creati e analizzati due differenti corpora che contengono i tweet relativi a quattro servizi delle compagnie aeree (prenotazione, check-in, bagaglio, volo). Il primo corpus contiene i tweet pubblicati dalle compagnie aeree nei propri account ufficiali, mentre il secondo corpus contiene i tweet pubblicati dagli utenti.

Il focus della ricerca è la misurazione della reputazione delle cinque compagnie aeree selezionate, effettuata tramite una sentiment analysis del secondo corpus, contenente i tweet degli utenti che si esprimono su determinati servizi. Si vuole prestare particolare attenzione alle eventuali opinioni negative degli utenti e valutare come esse possano influire sulla reputazione aziendale. Tuttavia, vengono presi in considerazione anche i

tweet pubblicati dalle compagnie aeree nei propri account ufficiali per verificare se esse trattano determinati argomenti con la stessa attenzione degli utenti.

L'analisi dei tweet degli utenti parte dall'ipotesi che su Twitter sia possibile rilevare l'opinione di coloro che si esprimono sulle compagnie aeree e che la valutazione del sentiment di tali opinioni possa costituire una buona approssimazione della reputazione aziendale, come viene osservato anche da Colleoni et al. (2011: 11) relativamente al contesto dei social media.

Il concetto astratto di reputazione viene quindi scomposto e ricondotto al tipo di opinione che gli utenti hanno su determinati servizi, ovvero una variabile empirica rilevabile tramite la sentiment analysis.

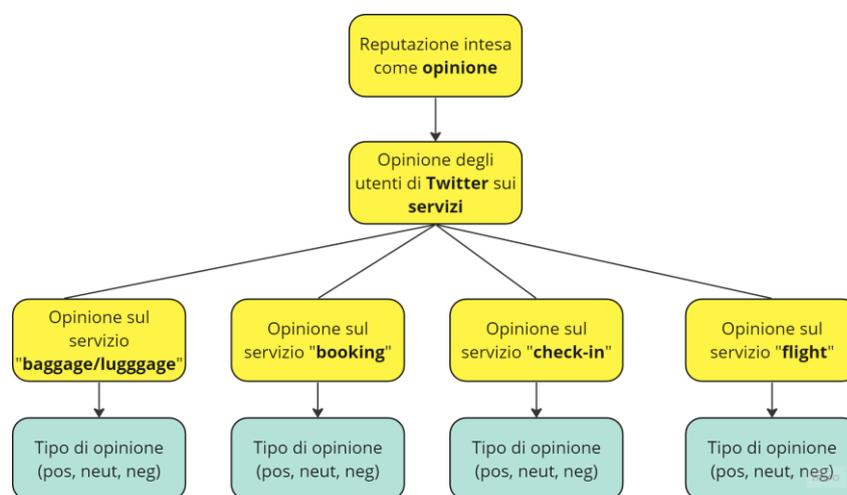


Figura 17: scomposizione del concetto di reputazione in variabili empiriche rilevabili.

Si precisa che il corpus su cui viene effettuata l'analisi non contiene un campione di dati statisticamente rappresentativo, in quanto non sono conosciuti i dati sociodemografici della popolazione indagata (gli utenti di Twitter), né la sua ampiezza (numero totale di utenti che hanno espresso la propria opinione). I risultati, quindi, non sono generalizzabili ma, in un'ottica esplorativa, possono aiutare a individuare delle aree problematiche o di possibile miglioramento che necessitano di ulteriori ricerche di tipo statistico.

Le domande che hanno guidato la ricerca sono:

- D1: quali dati emergono dai tweet pubblicati dagli account ufficiali delle compagnie aeree (corpus 1)?

- D2: qual è la reputazione delle compagnie aeree su Twitter? O meglio, quali dati emergono dai tweet che gli utenti pubblicano sulle compagnie aeree (corpus 2)?
- D3: può essere individuata una relazione tra i dati provenienti dai tweet pubblicati dalle compagnie aeree (corpus 1) e quelli provenienti dai tweet pubblicati dagli utenti (corpus 2)?

4.3 Le fasi della ricerca

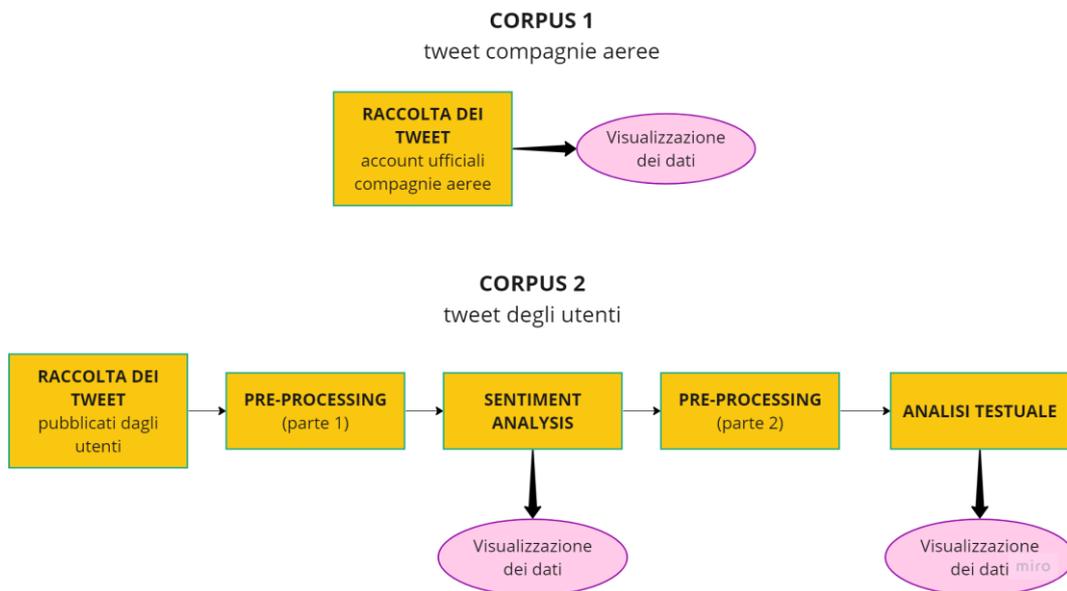


Figura 18: rappresentazione grafica delle fasi della ricerca.

La ricerca si è articolata in più fasi, ognuna delle quali ha previsto l'utilizzo del linguaggio di programmazione Python e del Jupyter Notebook. Nella figura 18, tramite un diagramma di flusso, si mostra la sequenza con cui sono state eseguite le diverse operazioni.

4.3.1 Raccolta, salvataggio e organizzazione dei tweet

Per prima cosa si è provveduto allo scaricamento, al salvataggio e all'organizzazione dei tweet in due corpora distinti. Si sono ricercati i tweet relativi a quattro componenti del servizio delle compagnie aeree: prenotazione, check-in, bagaglio, volo. I due corpus che si sono andati a costituire sono:

- Corpus 1: insieme dei tweet pubblicati dagli account ufficiali delle compagnie aeree tra il 23 maggio e il 2 ottobre 2022 relativi ai servizi di *baggage/luggage*¹⁶¹, *booking*, *check-in* e *flight*¹⁶².
- Corpus 2: insieme dei tweet pubblicati dagli utenti tra il 23 maggio e il 2 ottobre 2022 relativi ai servizi di *baggage/luggage*, *booking*, *check-in* e *flight* delle compagnie aeree.

Le procedure seguite per questa prima fase sono state le stesse per entrambi i corpora, tranne nel caso delle parole chiave inserite per scaricare i tweet che, a seconda del corpus di riferimento, permettono di scaricare i tweet pubblicati dalle compagnie aeree o dagli utenti, relativamente a un servizio specifico.

Edwards (2021) propone nel sito web *Towards Data Science* una procedura per scaricare i tweet tramite le Twitter API e l'accesso *Academic Research*, che è quello di cui si dispone in questa ricerca e che permette di accedere all'intero archivio di tweet pubblicati nel corso del tempo.

In questa prima fase della ricerca sono stati messi in atto i seguenti passaggi:

1. Importazione delle librerie e dei moduli da utilizzare:
 - *Requests*, libreria per mandare richieste alle API.
 - *Os*, modulo per gestire i file presenti nel sistema operativo.
 - *Json*, modulo per gestire le risposte ricevute dalle API¹⁶³.
 - *Pandas*, libreria per manipolare i dati ottenuti.
 - *Csv*, modulo per salvare i dati in un file *csv* (*comma separated values*).
 - I moduli *Datetime* e *unicodedata* e la libreria *dateutil.parser* per modificare il formato delle date di pubblicazione dei tweet.
 - *Time*, modulo per gestire il tempo, come ad esempio far trascorrere alcuni secondi tra le varie richieste fatte alle API.

¹⁶¹ Nelle keyword relative al bagaglio sono state inserite sia la parola *baggage* che *luggage*, in quanto i parlanti inglesi le usano come sinonimi (Collins Dictionary).

¹⁶² I nomi dei servizi in questo caso non vengono tradotti in italiano in quanto si riferiscono al corpus analizzato che è in lingua inglese.

¹⁶³ Il formato JSON (JavaScript Object Notation) è quello tramite cui le API forniscono i dati.

2. Creazione di una funzione `auth()` per inserire il Bearer Token (identificativo dell'applicazione che sta facendo la richiesta alle API) e autenticarsi alle API di Twitter.
3. Definizione della richiesta da fare all'endpoint *Full archive search*¹⁶⁴. Nel codice vengono inseriti tutti i parametri necessari a trovare i tweet d'interesse, ovvero `keyword` (parola chiave), `start_date` (data di inizio ricerca), `end_date` (data di fine ricerca), `max_results` (numero massimo di risultati per ogni ricerca). La richiesta fatta alle API permette non solo di ottenere il testo dei singoli tweet, ma anche la loro data di creazione, lo user ID¹⁶⁵, il numero di retweet, di risposte, di "mi piace" e di menzioni associate al tweet.

```
#input per la richiesta
bearer_token = auth()
headers = create_headers(bearer_token)
keyword = "(baggage OR luggage) (from:askryanair) lang:en"
start_time = "2022-05-23T00:00:00.000Z"
end_time = "2022-05-27T00:00:00.000Z"
max_results = 500
```

Tabella 15: esempio di input inserito per una singola richiesta alle API.

Nella tabella seguente vengono indicati gli account Twitter ufficiali delle compagnie aeree da cui si sono ricavati i tweet e le keyword utilizzate all'interno del codice Python per trovare i tweet d'interesse:

DATI PER LO SCARICAMENTO DEI TWEET DEGLI ACCOUNT UFFICIALI DELLE COMPAGNIE AEREE						
COMPAGNIA AEREA	NOME UTENTE	USER ID	QUERY PER SERVIZIO BAGAGLIO	QUERY PER SERVIZIO VOLO	QUERY PER SERVIZIO PRENOTAZIONE	QUERY PER SERVIZIO CHECK-IN
British Airways	@British_Airways	18332190	(baggage OR luggage) (from:British_Airways) lang:en	flight (from:british_airways) lang:en	booking (from:british_airways) lang:en	\"check in\" (from:british_airways) lang:en

¹⁶⁴ <https://developer.twitter.com/en/docs/twitter-api/tweets/search/quick-start/full-archive-search>
[ultima consultazione: 15/04/23]

¹⁶⁵ Lo user ID è un numero identificativo associato ad ogni account Twitter. Ogni componente su Twitter (tweet, user, messaggio) ha un ID unico. A differenza del nome utente, esso non può essere modificato. Per ricavare lo user ID si possono utilizzare diversi siti web gratuiti.

EasyJet	@easyJet	38676903	(baggage OR luggage) (from:easyJet) lang:en	flight (from:easyJet) lang:en	booking (from:easyJet) lang:en	\\"check in\\" (from:easyJet) lang:en
KLM	@klm	56377143	(baggage OR luggage) (from:klm) lang:en	flight(from:klm) lang:en	booking (from:klm) lang:en	\\"check in\\" (from:klm) lang:en
Lufthansa	@lufthansa	124476322	(baggage OR luggage) (from:lufthansa) lang:en	flight (from:lufthansa) lang:en	booking (from:lufthansa) lang:en	\\"check in\\" (from:lufthansa) lang:en
Ryanair	@askryanair	268942773	(baggage OR luggage) (from:askryanair) lang:en	flight (from:askryanair) lang:en	booking (from:askryanair) lang:en	\\"check in\\" (from:askryanair) lang:en

Tabella 16: parole chiave inserite per scaricare i tweet pubblicati dalle compagnie aeree relative ai vari servizi¹⁶⁶.

Allo stesso modo, vengono presentate nella tabella 17 le keyword utilizzate per ricavare i tweet pubblicati dagli utenti. In questo caso, sono stati scaricati i tweet che contengono il nome della linea aerea e del servizio e non solo quelli in cui la linea aerea fosse “taggata”, in quanto gli utenti potrebbero parlare della linea aerea anche senza taggarla:

DATI PER LO SCARICAMENTO DEI TWEET DEGLI UTENTI						
COMPAGNIA AEREA PRESENTE NEL TWEET	NOME UTENTE DELLA LINEA AEREA	USER ID DELLA LINEA AEREA	QUERY PER SERVIZIO BAGAGLIO	QUERY PER SERVIZIO VOLO	QUERY PER SERVIZIO PRENOTAZIONE	QUERY PER SERVIZIO CHECK-IN
British Airways	@British_Airways	18332190	british airways (baggage OR luggage) lang:en	british airways flight lang:en	british airways booking lang:en	british airways \\"check in\\" lang:en

¹⁶⁶ Per la linea aerea Ryanair, si sono scaricati i tweet dal profilo @askryanair: questo, infatti, è quello dedicato al servizio clienti. Al contrario, l'altro profilo ufficiale @Ryanair si dedica alla pubblicazione di contenuti vari, molti dei quali dal tono scherzoso e sarcastico.

EasyJet	@easyJet	38676903	easyjet (baggage OR luggage) lang:en	easyjet flight lang:en	easyjet booking lang:en	easyjet \"check in\" lang:en
Klm	@klm	56377143	klm(baggag e OR luggage) lang:en	klm flight lang:en	klm booking lang:en	klm \"check in\" lang:en
Lufthansa	@lufthansa	124476322	lufthansa (baggage OR luggage) lang:en	lufthansa flight lang:en	lufthansa booking lang:en	lufthansa \"check in\" lang:en
Ryanair	@askryanair	268942777 3	ryanair (baggage OR luggage) lang:en	ryanair flight lang:en	ryanair booking lang:en	ryanair \"check in\" lang:en

Tabella 17: keyword inserite per scaricare i tweet degli utenti relativi alle varie compagnie aeree e ai relativi servizi.

Per comporre una keyword che permettesse di trovare i tweet desiderati è venuta in aiuto la funzione di ricerca avanzata¹⁶⁷ di Twitter. Essa consente di generare una keyword specifica (da inserire poi nel codice Python) grazie alla compilazione dei seguenti campi: intervallo di tempo, parole presenti nel tweet in qualsiasi ordine, sequenza esatta di parole, account da cui proviene o a cui è diretto il tweet, hashtag, località geografica, lingua del tweet. Ad esempio, se si cercano i tweet in lingua inglese pubblicati dall'account @askryanair che contengano le parole “baggage” o “luggage”, la keyword da inserire nel codice sarà: “(baggage OR luggage) (from:askryanair) lang:en”. Si noti che per comporre la parola chiave vengono inseriti degli operatori logici booleani (come OR e AND), il cui utilizzo è spiegato in modo approfondito sul sito ufficiale di Twitter¹⁶⁸.

4. Scaricamento dei tweet

L'arco temporale considerato per questa ricerca è quello compreso tra lunedì 23 maggio 2022 e domenica 2 ottobre 2022. È stato scelto questo periodo in base alle tendenze rilevate su Google Trends che mostravano come ci fossero dei picchi di interesse tra l'inizio e la fine della stagione estiva. Con il profilo *Academic Research* ogni richiesta fatta alle API relativa a un determinato periodo temporale

¹⁶⁷ <https://help.twitter.com/it/using-twitter/twitter-advanced-search> [ultima consultazione: 15/04/23]

¹⁶⁸ <https://developer.twitter.com/en/docs/twitter-api/tweets/search/integrate/build-a-query#boolean>
[ultima consultazione: 15/04/23]

permette di scaricare al massimo 500 tweet¹⁶⁹ alla volta. I tweet disponibili per una determinata keyword però potrebbero essere più di 500. In questa ricerca, ad esempio, si sono voluti scaricare i tweet pubblicati in un arco temporale di circa quattro mesi, per cui solamente 500 tweet sarebbero stati poco rappresentativi dell'intero periodo. Per questo motivo, l'intero arco temporale è stato suddiviso in sottoperiodi di quattro giorni, per ognuno dei quali è stata fatta una richiesta che permette di scaricare al massimo 500 tweet. Il numero massimo di tweet scaricabili è stato quindi $(500 * n)$, dove n corrisponde al numero di periodi considerati. In alcuni casi sono stati scaricati meno tweet in quanto per la keyword inserita era stato pubblicato un numero inferiore di tweet.

N° di periodi	Data di inizio	Data di fine
1	23/05/2022	27/05/2022
2	27/05/2022	31/05/2022
3	31/05/2022	04/06/2022
4	04/06/2022	08/06/2022
5	08/06/2022	12/06/2022
6	12/06/2022	16/06/2022
7	16/06/2022	20/06/2022
8	20/06/2022	24/06/2022
9	24/06/2022	28/06/2022
10	28/06/2022	02/07/2022
11	02/07/2022	06/07/2022
12	06/07/2022	10/07/2022
13	10/07/2022	14/07/2022
14	14/07/2022	18/07/2022
15	18/07/2022	22/07/2022
16	22/07/2022	26/07/2022
17	26/07/2022	30/07/2022

18	30/07/2022	03/08/2022
19	03/08/2022	07/08/2022
20	07/08/2022	11/08/2022
21	11/08/2022	15/08/2022
22	15/08/2022	19/08/2022
23	19/08/2022	23/08/2022
24	23/08/2022	27/08/2022
25	27/08/2022	31/08/2022
26	31/08/2022	04/09/2022
27	04/09/2022	08/09/2022
28	08/09/2022	12/09/2022
29	12/09/2022	16/09/2022
30	16/09/2022	20/09/2022
31	20/09/2022	24/09/2022
32	24/09/2022	28/09/2022
33	28/09/2022	02/10/2022

Tabella 18: sottoperiodi (di 4 giorni ciascuno) considerati per lo scaricamento dei tweet.

Nello specifico, per poter fare più richieste:

- Si crea un *ciclo for* (for-loop) che permette di fare una richiesta alle API per ogni sottoperiodo.

¹⁶⁹ <https://developer.twitter.com/en/docs/twitter-api/tweets/search/introduction> [ultima consultazione: 15/04/23]

- Si utilizza il `next_token()`, una variabile fornita dall'endpoint *Full archive search* per andare a scaricare il blocco di risultati successivo a quello precedentemente scaricato. Grazie al `next_token()` si può riprendere lo scaricamento dei tweet da dove lo si era terminato nella chiamata precedente e si impaginano (*pagination*¹⁷⁰) i risultati uno dopo l'altro.
 - Si inserisce anche la funzione `time.sleep()` per mettere in pausa l'esecuzione del programma per alcuni secondi (nel codice utilizzato sono stati impostati 5 secondi). Le Twitter API prevedono un numero massimo di richieste¹⁷¹ che possono essere fatte in un determinato periodo di tempo. Se si fanno troppe richieste, quindi, Twitter potrebbe bloccare temporaneamente l'accesso alle API. Con la funzione `time.sleep()` si riesce a far passare del tempo tra una richiesta e l'altra e si evita di ricevere un messaggio di errore.
5. I tweet scaricati, relativi a una specifica keyword, sono stati salvati in dei file csv, in modo da poter essere letti (tramite la funzione `read()` di Pandas) e analizzati in un secondo momento. Le keyword inserite per scaricare i tweet, come illustrato sopra (tabelle 16 e 17) sono 20 per corpus, perciò inizialmente i tweet sono stati salvati in 20 file csv diversi per ogni corpus.
 6. I file csv sono stati visualizzati e manipolati tramite la libreria Pandas, che permette di leggere i file tabulari contenuti nei file csv e visualizzarli in dei dataframe¹⁷². Tramite la manipolazione del dataframe, si è aggiunta una colonna "airline" e una colonna "service", in modo che ad ogni tweet fosse associato la compagnia aerea e il servizio di riferimento. Ciò consente di tenere traccia delle informazioni dei tweet ogni qualvolta si vada a manipolare i dati.

¹⁷⁰ <https://developer.twitter.com/en/docs/twitter-api/tweets/search/integrate/paginate> [ultima consultazione: 15/04/23]

¹⁷¹ <https://developer.twitter.com/en/docs/twitter-api/rate-limits> [ultima consultazione: 15/04/23]

¹⁷² I dataframe sono tabelle bidimensionali formate da righe e colonne che permettono di lavorare sui dati senza andare a modificare il file csv originale. I file modificati poi possono essere salvati in un nuovo file.

Un dataframe, quindi, si presenta come segue:

	author id	created_at	tweet	service	airline
0	33980482	2022-05-26 21:20:41+00:00	@Gabri_EllaConti @British_Airways Basically no...	baggage	britishair
1	878145649	2022-05-26 20:43:11+00:00	.@richardbranson.@VirginAtlantic shocking Cuto...	baggage	britishair
2	4214708495	2022-05-26 20:12:31+00:00	RT @gongimenez: @British_Airways @British_Airw...	baggage	britishair
3	2999796643	2022-05-26 20:08:27+00:00	@British_Airways never again. First flight ca...	baggage	britishair
4	3001143320	2022-05-26 20:04:38+00:00	@British_Airways i have arrived in faro but 2 ...	baggage	britishair

Figura 19: esempio di dataframe dei tweet degli utenti che si esprimono riguardo il servizio “baggage” di British Airways.

Successivamente, per disporre di meno file su cui lavorare e facilitare la gestione dei dati, i 20 file csv singoli sono stati organizzati in 5 file diversi per corpus, ognuno relativo a una compagnia aerea. All’interno di un singolo file sono stati concatenati i dataframe relativi ai tweet riguardanti tutti i servizi presi in considerazione (baggage/luggage, booking, flight, check-in). Per chiarezza, si mostra una rappresentazione grafica dell’organizzazione dei due corpus (figure 20 e 21).

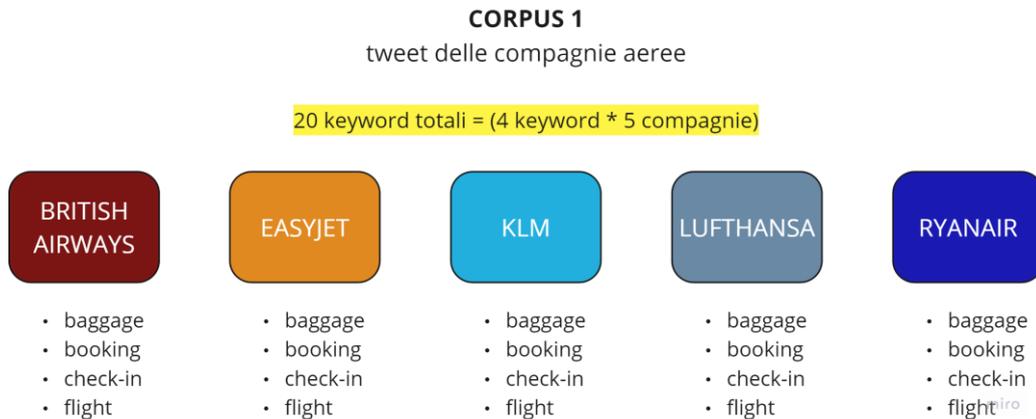


Figura 20: organizzazione del corpus dei tweet delle compagnie aeree.

CORPUS 2

tweet degli utenti che esprimono la propria opinione sulle compagnie aeree

20 keyword totali = (4 keyword * 5 compagnie)



Figura 21: organizzazione del corpus dei tweet degli utenti.

Nel caso del corpus 2, inoltre, cercando per parola chiave il nome della compagnia aerea e del servizio (ad esempio “klm flight lang:en”) è probabile che venissero inclusi anche dei tweet pubblicati dalle stesse compagnie aeree che all’interno contenevano tale parola chiave. Per evitare, quindi, che i tweet pubblicati dalle compagnie aeree si mischiassero con quelli pubblicati dagli utenti, sono stati rimossi dal corpus tutti i tweet provenienti dagli account ufficiali delle compagnie. Ogni account Twitter è associato a uno user ID univoco¹⁷³: grazie alle funzioni della libreria Pandas è stato possibile selezionare ed eliminare tutti i tweet associati allo user ID degli account ufficiali delle compagnie aeree. In questo modo, si è riusciti ad includere esclusivamente i tweet provenienti dagli utenti.

Nello specifico, il numero di tweet totale costituente i due corpora è 176133 (33519 per il corpus 1; 142614 per il corpus 2), distribuiti come segue:

NUMERO DI TWEET DEGLI ACCOUNT UFFICIALI DELLE COMPAGNIE AEREE						
COMPAGNIA AEREA	servizio baggage/luggage	servizio booking	servizio check-in	servizio flight	N° totale di tweet pubblicati dalla singola compagnia	
British Airways	2980	8445	568	4333	16326	
EasyJet	1014	4046	240	2869	8169	
KLM	1684	2533	43	994	5254	
Lufthansa	1596	787	89	532	3004	

¹⁷³ Lo user ID delle compagnie aeree qui analizzate è inserito nelle tabelle 16 e 17.

Ryanair	155	72	146	393	766	
					33519	N° totale di tweet pubblicati da tutte le compagnie

Tabella 19: numero di tweet pubblicati dagli account ufficiali delle compagnie aeree diviso per servizio (Corpus 1).

NUMERO DI TWEET PUBBLICATI DAGLI UTENTI						
COMPAGNIA AEREA DI CUI PARLANO GLI UTENTI	servizio baggagge/luggage	servizio booking	servizio check-in	servizio flight	N° totale di tweet relativi alla singola compagnia	
British Airways	10185	7540	4643	16152	38520	
EasyJet	6606	3951	1833	15712	28102	
Klm	7114	1663	860	14892	24529	
Lufthansa	7879	2053	919	16025	26876	
Ryanair	4762	1835	1954	16036	24587	
					142614	N° totale dei tweet pubblicati dagli utenti

Tabella 20: numero di tweet pubblicati dagli utenti relativamente a ogni compagnia aerea e ai vari servizi (Corpus 2).

4.3.2 Visualizzazione dei dati non riguardanti la sentiment analysis

Dopo la fase di raccolta dei dati è già possibile ottenere delle informazioni riguardo il numero di tweet pubblicati. La libreria Seaborn permette di costruire dei grafici a barre e di visualizzare:

- Il numero di tweet totali pubblicati da o su una compagnia. Ad ogni compagnia viene assegnato un colore diverso.

```

Import seaborn as sns

cols={'britishair': 'darkred', 'easyjet': 'darkorange', 'klm':
'deepskyblue', 'lufthansa': 'slategray', 'ryanair': 'mediumblue'}

sns.countplot(x='airline', data=df_tweets, palette=cols, order
= df_tweets['airline'].value_counts().index)
sns.set(rc={'figure.figsize' (5, 5)})

```

```
plt.title("NUMERO DI TWEET RIFERITO ALLE COMPAGNIE AEREE")
plt.xlabel('compagnie aeree', fontsize = 12)
plt.ylabel('n di tweet', fontsize = 12)
```

Tabella 21: codice per realizzare un grafico a barre che mostra i tweet raggruppati per compagnia.

- Il numero di tweet totali pubblicati da o su una compagnia diviso per servizio specifico. Ad ogni compagnia viene assegnato un colore diverso.

```
Import seaborn as sns

cols={'britishair':'darkred', 'easyjet': 'darkorange','klm':
'deepskyblue','lufthansa':'slategray','ryanair': 'mediumblue'}

sns.countplot(x='service', hue = 'airline', data=df_tweets,
palette=cols)

sns.set(rc={'figure.figsize' (10,5)})
plt.title("NUMERO DI TWEET SU OGNI COMPAGNIA AEREA DIVISO PER
SERVIZIO")
plt.xlabel('servizio', fontsize = 12)
plt.ylabel('n di tweet', fontsize = 12)
```

Tabella 22: codice per realizzare un grafico a barre che mostra i tweet raggruppati per servizio e divisi per compagnia.

- L'indice di correlazione tra il numero di tweet delle compagnie aeree e degli utenti, tramite la funzione `corr()` di Pandas: l'indice di correlazione di Pearson¹⁷⁴ (chiamato anche coefficiente di correlazione lineare) ha un valore compreso tra -1 e +1 ed è una misura della relazione di linearità tra due variabili (in questo caso: n° di tweet pubblicati dalle compagnie vs n° di tweet pubblicati dagli utenti). Il valore +1 corrisponde a una perfetta correlazione lineare positiva; -1 a una correlazione lineare negativa e 0 a un'assenza di correlazione.

```
correlazione =
df_correlazione['tweet_x'].corr(df_correlazione['tweet_y'])
```

Tabella 23: codice utilizzato per effettuare una correlazione.

- Le menzioni all'interno dei tweet, trovate tramite la funzione `str.contains()`.

```
import pandas as pd
df_ok = df_RYANAIR
# si filtrano le righe che contengono "@easyJet"
mention_tweets =
df_ok[df_ok['tweet'].str.contains('@askryanair')]
# si stampano i tweet
```

¹⁷⁴ https://it.wikipedia.org/wiki/Indice_di_correlazione_di_Pearson

```
for tweet in mention_tweets['tweet']:  
    print(tweet)
```

Tabella 24: codice utilizzato per trovare le menzioni all'interno dei tweet.

4.3.3 Pre-processing dei tweet (prima parte)

Prima di poter effettuare una sentiment analysis sul corpus 2, è stato necessario pre-processare i tweet, ovvero “trasformare i dati grezzi in un formato comprensibile [per il programma che svolgerà la sentiment analysis]” (Agarwal 2015: 30). L'obiettivo della fase di pre-processing è quello di favorire la velocità di elaborazione da parte del programma, migliorare la qualità dei dati testuali su cui effettuare l'analisi e, di conseguenza, avere dei risultati più accurati.

In questa ricerca, il testo è stato pre-processato in due fasi differenti: nella prima sono stati eliminati tutti quegli elementi superflui ai fini della sentiment analysis. Nella seconda fase, invece, che segue la sentiment analysis, sono state eseguite delle operazioni specifiche per la successiva visualizzazione dei dati testuali.

Nella prima fase di pre-processing dei tweet, grazie all'importazione del pacchetto *NeatText* (import neattext.functions as nfx), si sono potuti rimuovere tutti quegli elementi che non veicolano alcuna informazione riguardo il sentiment degli utenti, ovvero:

- gli hashtag

```
df_BRITISH['clean_tweet']=  
df_BRITISH['tweet'].apply(nfx.remove_hashtags)
```

Tabella 25: esempio di codice utilizzato per rimuovere gli hashtag.

- le menzioni

```
df_BRITISH['clean_tweet']=  
df_BRITISH['clean_tweet'].apply(lambda x:  
nfx.remove_userhandles(x))
```

Tabella 26: esempio di codice utilizzato per rimuovere le menzioni.

- i doppi spazi

```
df_BRITISH['clean_tweet']=  
df_BRITISH['clean_tweet'].apply(nfx.remove_multiple_spaces)
```

Tabella 27: esempio di codice utilizzato per rimuovere i doppi spazi.

- gli URL

```
df_BRITISH['clean_tweet']=  
df_BRITISH['clean_tweet'].apply(nfx.remove_urls)
```

Tabella 28: esempio di codice utilizzato per rimuovere gli URL.

Al contrario, invece, la punteggiatura, le stopwords, i caratteri maiuscoli e le emoji sono stati mantenuti perché sono degli elementi rilevanti per la sentiment analysis svolta con VADER. Si noti che ogni volta che il tweet è stato “pulito” da un certo elemento, il risultato della pulizia è stato riportato in una colonna del dataframe chiamata “clean_tweet” in modo che potessero essere visibili sia il tweet originario che quello pre-processato.

Inoltre, all’interno del corpus sono stati conservati i retweet¹⁷⁵, poiché, come fanno notare anche Thelwall et al. (2010), essi possono essere considerati un indicatore di come una specifica informazione o opinione espressa nel tweet originale abbia ottenuto consenso e diffusione da parte di più utenti.

4.3.4 Sentiment analysis con VADER

Dopo aver pre-processato i tweet, si è pronti per eseguire la sentiment analysis con VADER (Valence Aware Dictionary and sEntiment Reasoner), un modulo della libreria NLTK. È stato scritto un programma che tramite VADER rileva il sentiment dei singoli tweet e ne fornisce una misura tramite il *compound*, una metrica che corrisponde alla somma dei punteggi delle singole parole presenti nel tweet normalizzati in una scala tra -1 e +1 (dove -1 è il massimo della negatività e +1 il massimo della positività). Inoltre, per ogni tweet l’output restituisce un punteggio percentuale della positività, negatività e neutralità della frase, la cui somma è 1.

Le soglie standardizzate del valore del compound per classificare il testo come positivo, negativo e neutro sono:

- Sentiment positivo ↔ $\text{compound} \geq 0.05$
- Sentiment neutro ↔ $-0.05 < \text{compound} < 0.05$
- Sentiment negativo ↔ $\text{compound} \leq -0.05$

¹⁷⁵ Un retweet è la ripubblicazione di un tweet.

Per svolgere la sentiment analysis, è stata importata la libreria NLTK e scaricato il lessico “vader_lexicon” su cui poi si basa l’analisi che cerca una corrispondenza tra le parole presenti nei tweet e quelle presenti nel lessico già etichettate come positive o negative. Successivamente, si è creato un oggetto “sia” corrispondente al modulo *sentiment intensity analyzer* che viene utilizzato per analizzare i tweet. La funzione `lambda`¹⁷⁶ è stata utilizzata per applicare la funzione `sia.polarity_scores()` a ciascun tweet del dataframe. Questa funzione restituisce come output un dizionario¹⁷⁷ le cui chiavi “neg”, “neu”, “pos” e “compound” corrispondono rispettivamente alla percentuale di parole negative, neutre e positive nel tweet e al compound. Sono state poi create delle nuove colonne nel dataframe che presentano i vari valori relativi al sentiment.

Esempio di un testo pre-processato e sottoposto a una sentiment analysis:

```
'Basically nobody was wearing mask on my BA flight LHR-VCE this evening either. They also lost my luggage which their staff made me check in at boarding, even though there was space in the overhead compartment (also printed my name wrong in its receipt)..'
```

Codice utilizzato per effettuare la sentiment analysis:

```
import nltk
#si scarica vader da nltk
nltk.download('vader_lexicon')
from nltk.sentiment.vader import SentimentIntensityAnalyzer
#creo un oggetto relativo alla sentiment analysis
sia = SentimentIntensityAnalyzer()

#si creano delle nuove colonne che presentano il punteggio della sentiment analysis
df_ok['scores']=df_ok['clean_tweet'].apply(lambda tweet:
sia.polarity_scores(str(tweet)))
```

Risultato della sentiment analysis:

clean_tweet	scores	compound	pos	neu	neg	sentiment
Basically nobody was wearing mask on my BA flight LHR-VCE this evening either. They also lost my luggage which their staff made me check in at boarding, even though there was space in the overhead compartment (also printed my name wrong in its receipt)..	{'neg': 0.114, 'neu': 0.886, 'pos': 0.0, 'comp...}	-0.6597	0.000	0.886	0.114	negative

Tabella 29: esempio di tweet sottoposto alla sentiment analysis e relativi risultati del compound e delle singole percentuali di positività, neutralità e negatività.

¹⁷⁶ Una funzione lambda è una funzione molto semplice che può essere definita con una sola riga di codice.

¹⁷⁷ Un dizionario è un insieme non ordinato di coppie chiave-valore (Deitel 2021).

4.3.5 Visualizzazione dei dati provenienti dalla sentiment analysis

I risultati ottenuti dalla sentiment analysis possono essere visualizzati tramite dei grafici realizzati con le librerie *Matplotlib* e *Seaborn* o analizzati ulteriormente svolgendo delle operazioni con i dataframe di *Pandas*. In questo modo si può avere una visione d'insieme del valore del compound rispetto a un determinato periodo di tempo, servizio o compagnia aerea. Le visualizzazioni grafiche e le operazioni eseguite sui dataframe sono:

- Grafico a torta: utilizzato per visualizzare la percentuale di tweet positivi, negativi e neutri relativi a una compagnia aerea nel complesso. Nel codice utilizzato si è creato un dizionario dei colori associati al tipo di sentiment e si sono definiti dei parametri per la funzione `plot()`, tra cui `autopct` che permette di mostrare i valori in percentuale.

```
#sentiment solo di una linea
import matplotlib.pyplot as plt

df_ok = df_tweets_S

colors = {'positive': 'lime',
         'negative': 'lightcoral',
         'neutral': 'thistle'}

df_ok[df_ok['airline']=='ryanair']['sentiment'].value_counts()
.plot(kind='pie', figsize=(5, 5), autopct='%1.0f%%',
      colors=[colors[v] for v in
df_ok['sentiment'].value_counts().keys()])
```

Tabella 30: esempio di codice utilizzato per mostrare i dati sul sentiment con un grafico a torta.

- Grafico a barre, utilizzato per:
 - o Confrontare il numero di tweet positivi, negativi e neutri rispetto ai singoli servizi di una compagnia aerea o di tutte le compagnie aeree nel complesso. Anche in questo caso si è definito un dizionario per i colori relativi ai vari servizi e si sono personalizzate le etichette degli assi e le dimensioni del grafico.

```
Cols={'positive': 'lime',
      'negative': 'lightcoral',
      'neutral': 'thistle'}

sns.countplot(x='service', hue='sentiment',
             data=df_LUFTHANSA_S, palette=cols)
```

```
sns.set(rc={'figure.figsize': (10,5)})
plt.title("SENTIMENT DEI SERVIZI DI LUFTHANSA")
plt.xlabel('servizio', fontsize = 12)
plt.ylabel('\n di tweet', fontsize = 12)
```

Tabella 31: esempio di codice utilizzato per mostrare i dati sul sentiment con un grafico a barre.

- Confrontare il valore medio del compound delle compagnie aeree in tutto il periodo: si è definito il dizionario relativo ai colori che identificano le varie compagnie aeree, si sono raggruppati i valori medi del compound per compagnia aerea, si sono ordinati tali valori in ordine decrescente e si è personalizzato l'aspetto del grafico.

```
Colorss= ('darkred', 'mediumblue', 'deepskyblue',
'slategray', 'darkorange')

df_tweets_S.groupby(['airline']).mean().compound.sort_
lues(ascending=False).plot(kind='bar', title =
'SENTIMENT BASATO SULLA POLARITÀ MEDIA', edgecolor =
'black', color = colorss)

plt.xticks(rotation = 0, weight = 'bold')
plt.xlabel('compagnia aerea', fontsize = 12)
plt.ylabel('compound', fontsize = 12)
```

Tabella 32: esempio di codice utilizzato confrontare i valori del compound delle cinque compagnie aeree.

- Individuazione del valore medio del compound per ogni mese: i tweet nel dataframe sono stati ordinati per data, è stata aggiunta una colonna che identifica il mese di pubblicazione del tweet, si sono raggruppati i valori del dataframe per mese e si è calcolato il valore medio del compound per ogni mese.

```
# si converte la colonna delle date in tipo datetime
df_sort['created_at'] = pd.to_datetime(df_sort['created_at'])
# si aggiunge una nuova colonna con il mese
df_sort['month'] = df_sort['created_at'].dt.month
# si raggruppa il dataframe per mese e si calcola il valore
medio di ogni mese
valore_medio_per_mese =
df_sort.groupby('month')['compound'].mean()
# si stampa il valore medio per ogni mese
print(valore_medio_per_mese)
```

Tabella 33: esempio di codice utilizzato per individuare il valore medio del compound di ogni mese.

- Grafico a linee: dopo aver individuato il mese con il compound medio più negativo, si è osservato come il valore del compound sia variato nel corso di quel mese specifico e quali siano state le date con i picchi più negativi. Per fare ciò, si sono selezionate le righe del dataframe che corrispondono all'intervallo di tempo considerato e la colonna relativa al valore del compound.

```

Import pandas as pd
import matplotlib.pyplot as plt

#plt.axhline permette di evidenziare y=0, in modo da
visualizzare meglio quando il sentiment è positivo o negativo

plt.figure(figsize=(20,10))
plt.plot(df_okk, color='deeppink', marker='o')
plt.title('VARIAZIONE DEL COMPOUND AD AGOSTO - BRITISH
AIRWAYS', fontsize=14)
plt.axhline(y=0, color='k')
plt.xlabel('date', fontsize=14)
plt.ylabel('compound medio', fontsize=14)
plt.xticks(rotation=45)
plt.grid(True)
plt.show()

```

Tabella 34: esempio di codice utilizzato per visualizzare la variazione del compound nel corso di un intervallo di tempo specifico.

- Ricerca specifica dei tweet più negativi: è possibile rintracciare nel dataframe il tweet o i tweet che presentano il valore minimo di compound, tramite il metodo `nsmallest()`.

```

N = 10
smallest = df_sort.nsmallest(n, 'compound')

print(smallest)

```

Tabella 35: esempio di codice utilizzato per individuare i 10 tweet con il compound più basso in un dataframe.

4.3.6 Pre-processing dei tweet (seconda parte)

Nella seconda fase di pre-processing, il testo è stato ulteriormente elaborato per migliorare la successiva visualizzazione e analisi dei dati testuali tramite delle word cloud e il grafico della frequenza dei bigrams. Per questo tipo di visualizzazioni, infatti, servono solo le *parole piene* presenti nei tweet, ovvero tutte quelle parole che sono portatrici di contenuto e significato (Bolasco 2013), come i sostantivi e i verbi. Al contrario, le *parole*

*vuote*¹⁷⁸, le stopword (come le preposizioni, gli articoli determinativi e indeterminativi, i pronomi e aggettivi possessivi, gli aggettivi dimostrativi e interrogativi), la punteggiatura, i caratteri maiuscoli, i caratteri speciali e le emoji vengono trascurate, in quanto non esprimono alcun contenuto utile ai fini dell'analisi testuale. Ad esempio, si consideri il seguente tweet relativo a British Airways, che era già stato sottoposto alla prima fase di pre-processing:

```
' BA you are an embarrassment. Despite our party being underweight on ALL of our luggage and only bringing minimal hand baggage, because YOU managed to make your flight too heavy you INTENTIONALLY left 2 of our cases behind in London. Shocking 😞 '
```

Si sono rimossi gli elementi superflui:

- Punteggiatura: rimossa tramite la funzione `remove_puncts` del pacchetto NeatText.

<pre>df_BRITISH_S['clean_tweet']= df_BRITISH_S['clean_tweet'].apply(nfx.remove_puncts)</pre>
<p>Tweet pulito:</p> <pre>' BA you are an embarrassment Despite our party being underweight on ALL of our luggage and only bringing minimal hand baggage because YOU managed to make your flight too heavy you INTENTIONALLY left 2 of our cases behind in London Shocking 😞 '</pre>

Tabella 36: codice utilizzato per rimuovere la punteggiatura dal testo.

- Caratteri maiuscoli: rimossi tramite il metodo `lower()`, che fa parte di `str`, una classe built-in di Python.

<pre>df_tweets_S['clean_tweet'] = df_tweets_S['clean_tweet'].str.lower()</pre>
<p>Tweet pulito:</p> <pre>' ba you are an embarrassment despite our party being underweight on all of our luggage and only bringing minimal hand baggage because you managed to make your flight too heavy you intentionally left 2 of our cases behind in london shocking 😞 '</pre>

Tabella 37: codice utilizzato per rimuovere i caratteri maiuscoli dal testo.

- Emoji e caratteri speciali: è stata creata una funzione per sostituire due pattern di espressioni regolari con uno spazio vuoto. Questa funzione ha permesso di

¹⁷⁸ Le *parole vuote* non esprimono in sé alcun significato (Bolasco 2013).

rimuovere i numeri, gli URL e tutti i caratteri che non fossero lettere maiuscole o minuscoli, compresi i caratteri speciali e le emoji.

Viene definita e applicata la funzione:

```
def function(string):  
    return re.sub ("\d+\http?\S+|[^A-Za-z0-9]+", " ", string)
```

```
df_BRITISH_S['clean_tweet']=  
df_BRITISH_S['clean_tweet'].apply(function)
```

Tweet pulito:

```
' ba you are an embarrassment despite our party being underw  
eight on all of our luggage and only bringing minimal hand b  
aggage because you managed to make your flight too heavy you  
intentionally left 2 of our cases behind in london shocking  
'
```

Tabella 38: esempio di applicazione della funzione a un tweet in cui sono presenti delle emoji.

- Stopword¹⁷⁹: rimosse tramite la funzione `remove_stopwords` del pacchetto `NeatText`.

```
df_tweets_S['clean_tweet'] =  
df_tweets_S['clean_tweet'].apply(nfx.remove_stopwords)
```

```
'ba embarrassment despite party underweight luggage bringin  
g minimal hand baggage managed flight heavy intentionally l  
eft 2 cases london shocking'
```

Tabella 39: codice utilizzato per rimuovere le stopwords dal testo.

Successivamente, sono state effettuate altre due operazioni:

- Tokenizzazione: il testo dei tweet presenti in un dataframe è stato suddiviso in singole parole, chiamate *token*, tramite il metodo `split()` e si è formata una lista.

```
for line in neg_tweetB:  
    print(line)  
    for token in line.split():  
        print(token)
```

```
neg_tokens_B = [token for line in neg_tweetB for token in  
line.split()]
```

Tabella 40: codice utilizzato per dividere il testo dei tweet in token.

- Lemmatizzazione: dopo aver eseguito la tokenizzazione, si è eseguita la lemmatizzazione, che svolge un'analisi morfologica delle parole presenti nella lista e le riduce in lemmi, ovvero nella forma base del paradigma verbale o

¹⁷⁹ L'elenco completo delle stopwords considerate da `NeatText` è presente nella documentazione ufficiale su GitHub: <https://github.com/Jcharis/neattext/blob/master/neattext/neattext.py> [ultima consultazione: 15/04/23]

nominale con cui i lessemi vengono inseriti nel vocabolario¹⁸⁰ (Enciclopedia Treccani). Si è eseguita la lemmatizzazione solo dei sostantivi e dei verbi, in quanto nella lingua inglese le altre parti del discorso presenti nei tweet preprocessati (aggettivi e avverbi) possono presentare delle flessioni che si è deciso di mantenere (forma comparativa o superlativa). Per eseguire una lemmatizzazione, è stato usato WordNetLemmatizer, una classe del modulo `nltk.stem` della libreria NLTK. WordNetLemmatizer si basa su WordNet, un database lessicale di lingua inglese, per associare le parole dei tweet ai lemmi corrispondenti. È stata definita una funzione che usa il metodo `lemmatize()`, i cui parametri permettono di definire la parte del discorso (`pos = part of speech`) sui cui effettuare la lemmatizzazione (`v = verb; n = noun`).

```
wn = nltk.WordNetLemmatizer()

def lemmatizationV(token_text):
    text = wn.lemmatize(token_text, pos = 'v')
    return text
```

Tabella 41: codice utilizzato per la lemmatizzazione dei verbi.

4.3.7 Visualizzazione e analisi dei dati testuali

Le singole parole (i *token*), ricondotte ai propri lemmi, sono state utilizzate per una breve analisi testuale, che ha permesso di individuare e visualizzare le parole più frequenti all'interno del dataset di riferimento. Nello specifico, sono state create due tipi visualizzazioni grafiche:

- Nuvola di parole (chiamata anche *word cloud*): essa è un'immagine che mostra un insieme di parole, la cui dimensione varia in base alla frequenza delle parole all'interno del corpus considerato. Più una parola è frequente e più aumenterà la sua dimensione rispetto alle altre. In questo caso, è stata creata una funzione per visualizzare la nuvola di parole grazie alla libreria WordCloud. Essa permette di impostare diversi parametri per personalizzare l'aspetto dell'immagine, come ad esempio il numero massimo di parole da visualizzare, il colore dello sfondo, la larghezza e l'altezza della figura e la mappa di colori da utilizzare.

```
def plot_wordcloud(docx):
```

¹⁸⁰ Ad esempio, i verbi vengono ridotti al modo infinito; i sostantivi vengono ridotti alla forma singolare.

```
plt.figure(figsize=(20,10))
mywordcloud = WordCloud(max_words = 100, background_color
= "white", width = 1600, height = 800,
colormap='tab10').generate(docx)
plt.imshow(mywordcloud, interpolation = 'bilinear')
plt.axis('off')
plt.show()
```

Tabella 42: codice utilizzato per la creazione di una word cloud.

- Grafico a barre orizzontali che mostra i bi-gram più frequenti: i bi-gram¹⁸¹ sono unità di due parole che compaiono una accanto all'altra nel testo di riferimento (Collins Dictionary). In questo caso, è stato realizzato un grafico a barre orizzontali che mostra i 15 bi-gram più frequenti all'interno del dataset considerato, ordinati per frequenza decrescente.

```
bigrams_series = (pd.Series(nltk.ngrams(words,
2)).value_counts())[:15]

bigrams_series.sort_values().plot.barh(color='rebeccapurple',
width=.9, figsize=(12, 8))
plt.title('Bigram più frequenti')
plt.ylabel('Bigram')
plt.xlabel('frequenza')
```

Tabella 41: codice utilizzato per la creazione di un grafico a barre orizzontali che mostra i bigram più frequenti.

- Fusione (*Merging*) di più dataframe per individuare le parole in comune: viene importata la funzione `reduce` dal modulo `functools`. Essa considera due dataframe alla volta e li fonde (*merge*) in base alla colonna “word” utilizzando la funzione `pd.merge` di Pandas.

```
data_frames_BAG = [neg_df_B, neg_df_K, neg_df_L]

from functools import reduce

df_merged_BAG = reduce(lambda left, right:
pd.merge(left, right, on=['word']), data_frames_BAG)
```

Tabella 42: codice utilizzato per la fusione di più dataframe.

¹⁸¹ I bigram sono un tipo di n-gram, ovvero un'unità di n parole, lettere o simboli che ricorrono in sequenza di testo (Collins Dictionary).

CAPITOLO 5 – RISULTATI DELL’ANALISI

Attraverso l’ausilio di grafici e tabelle, in questo capitolo vengono presentati i risultati dell’analisi dei due corpora.

Innanzitutto, viene fornita una panoramica della situazione delle compagnie aeree sia dal punto di vista degli account Twitter ufficiali delle stesse, sia dal punto di vista degli utenti. Viene fatta un’analisi comparativa tra il numero di tweet pubblicati dagli account ufficiali delle compagnie aeree e il numero di tweet pubblicati dagli utenti che si esprimono su tali compagnie aeree. In questo caso, i risultati riguardano l’intero arco temporale a cui appartengono i tweet scaricati e permettono di indagare i servizi su cui vengono pubblicati più tweet da parte di entrambi gli attori. Inoltre, viene calcolato il coefficiente di correlazione tra il numero di tweet pubblicati dalle compagnie aeree e il numero di tweet pubblicati dagli utenti.

Successivamente, si presentano i risultati relativi alla sentiment analysis dei tweet pubblicati dagli utenti, in modo da indagare la reputazione e il contenuto delle opinioni negative sulle cinque compagnie aeree. I risultati mostrati fanno riferimento all’intero arco temporale a cui appartengono i tweet, ma anche a dei periodi temporali specifici per ogni compagnia aerea che sono risultati essere più critici rispetto ad altri.

I risultati relativi al sentiment degli utenti sono quindi ottenuti considerando sia il contesto generale, ovvero la situazione di tutte le compagnie aeree nell’intero arco temporale a cui appartengono i tweet scaricati, sia il contesto caratteristico di ogni singola compagnia aerea in un arco temporale più ristretto. In tal modo, è possibile individuare delle tendenze all’interno dell’arena competitiva delle compagnie aeree ma anche rilevare informazioni più specifiche riguardanti una singola azienda.

5.1 Numero di tweet

5.1.1 Numero di tweet totale: compagnie vs utenti

Sono stati prodotti due grafici a barre che mostrano il numero di tweet pubblicati dagli account ufficiali di ogni compagnia aerea (figura 22) e il numero di tweet pubblicati dagli utenti riguardo le compagnie aeree (figura 23) durante tutto il periodo considerato (23 maggio – 2 ottobre). In entrambi i grafici, l’asse x riporta le cinque compagnie aeree considerate, mentre l’asse y indica il numero di tweet pubblicati. Le barre sono colorate

in modo diverso a seconda della compagnia aerea e sono rappresentate in ordine decrescente.

La figura 22 permette di confrontare il numero di tweet pubblicati dalle singole compagnie aeree con i propri account ufficiali, mentre la figura 23 permette di confrontare il numero di tweet pubblicati dagli utenti rispetto una determinata compagnia aerea. Affiancare questi due grafici¹⁸² permette, per ogni compagnia aerea, di confrontare l'attività degli account ufficiali delle compagnie con quella degli utenti.

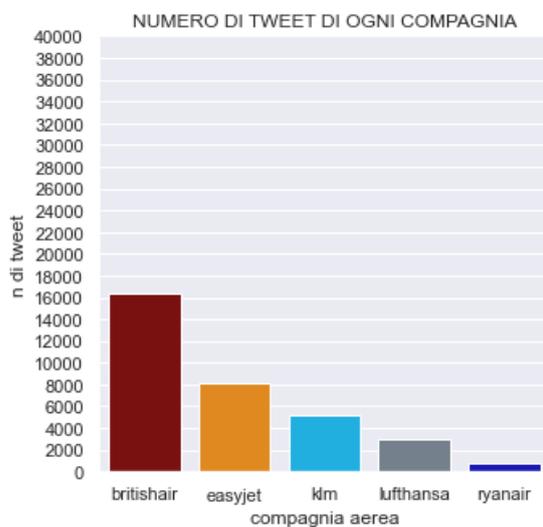


Figura 22: numero di tweet pubblicati dagli account ufficiali di ogni compagnia aerea.

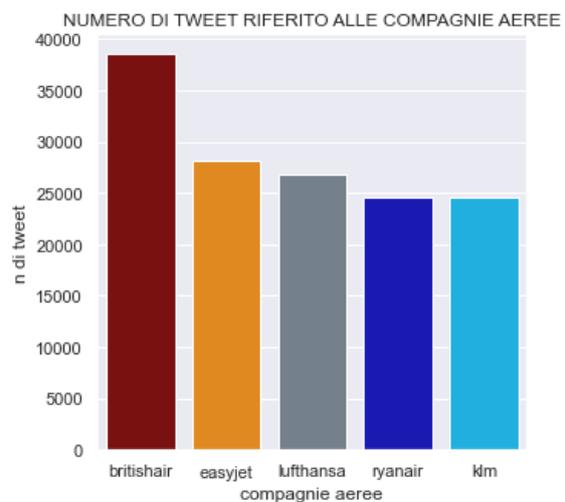


Figura 23: numero di tweet pubblicati dagli utenti relativamente a ogni compagnia aerea.

Dai singoli grafici si può osservare che:

- L'azienda che ha pubblicato più tweet nel periodo considerato è British Airways, seguita, in ordine decrescente, da EasyJet, KLM, Lufthansa e Ryanair (figura 22).
- L'ordine delle compagnie aeree su cui viene pubblicato il maggior numero di tweet corrisponde quasi a quello della figura 22, a eccezione di KLM, che in questo caso si posiziona in ultima posizione (figura 23).

Confrontando i due grafici, si nota che:

¹⁸² Per confrontare facilmente i risultati presentati nei due grafici, in entrambi è stata impostata la stessa scala di valori sull'asse y.

- In tutti i casi, il numero di tweet pubblicati dagli utenti è maggiore rispetto a quello pubblicato dagli account ufficiali.
- La differenza più netta tra l'attività della compagnia aerea e quella degli utenti si ha nel caso di Ryanair che, nell'intero periodo considerato, ha pubblicato meno di 2000 tweet, a fronte dei quasi 25000 tweet pubblicati dagli utenti.

Un basso numero di tweet pubblicati dall'account ufficiale di una compagnia aerea potrebbe indicare che tale compagnia non utilizza attivamente il proprio servizio clienti su Twitter (pur dichiarando di essere disponibile a rispondere alle richieste degli utenti), oppure che la compagnia aerea risponda a molte richieste simili con un unico tweet o, ancora, che molte richieste vengono processate tramite messaggi privati: alcune compagnie, infatti, invitano i clienti a scrivere dei *direct message* (DM) per le richieste più specifiche che implicano la comunicazione di dati privati (come nome e cognome e numero di prenotazione).



Official global account of KLM. FAQ about flying from AMS: klmf.ly/Schiphol. Please share personal details only in DMs.

Figura 24: esempio di descrizione in cui la compagnia aerea invita a scrivere messaggi diretti (DM).

Inoltre, bisogna considerare che i tweet pubblicati dagli utenti possono essere più numerosi di quelli della compagnia aerea, in quanto il corpus potrebbe includere sia i tweet in cui la compagnia aerea viene taggata (in cui quindi c'è una specifica richiesta a cui deve rispondere) sia quelli che contengono semplicemente il nome della compagnia e che potrebbero non rivolgere alcuna richiesta diretta al servizio clienti.

Compagnia aerea	n° totale dei tweet degli utenti	n° dei tweet che contiene una menzione (tag) della compagnia aerea
British Airways	38520	27659
EasyJet	28102	17140
KLM	24529	819
Lufthansa	26876	15038
Ryanair	24587	1343

Tabella 43: dettaglio del corpus 2 che mostra il numero di tweet totali e il numero di tweet contenenti una menzione della compagnia aerea.

In ogni caso, sia i tweet degli utenti in cui la compagnia aerea è taggata, sia i tweet in cui la compagnia aerea è semplicemente nominata possono contribuire a diffondere informazioni, opinioni e, infine, a plasmare la reputazione aziendale.

5.1.2 Numero di tweet diviso per servizio: compagnie aeree vs utenti

Il confronto tra il numero di tweet dei due corpora è stato fatto anche in un'ottica più specifica, andando a visualizzare, tramite un grafico a barre raggruppate, quanti tweet sono stati pubblicati rispetto ai quattro servizi considerati (baggage, booking, flight, check-in). Le figure 25 e 26 mostrano rispettivamente il numero di tweet pubblicati dagli account ufficiali delle compagnie aeree e dagli utenti che si esprimono sulle compagnie aeree, divisi per servizio specifico. In entrambi i grafici, l'asse x raggruppa per ogni servizio le barre che rappresentano le cinque compagnie aeree, mentre l'asse y indica il numero di tweet pubblicati¹⁸³. Le barre sono colorate in modo diverso a seconda della compagnia aerea e sono rappresentate sempre nello stesso ordine.

Comparare questi due grafici permette di individuare i servizi per cui le compagnie aeree pubblicano più o meno tweet rispetto agli utenti.

¹⁸³ Per confrontare facilmente i risultati presentati nei due grafici, in entrambi è stata impostata la stessa scala di valori sull'asse y.

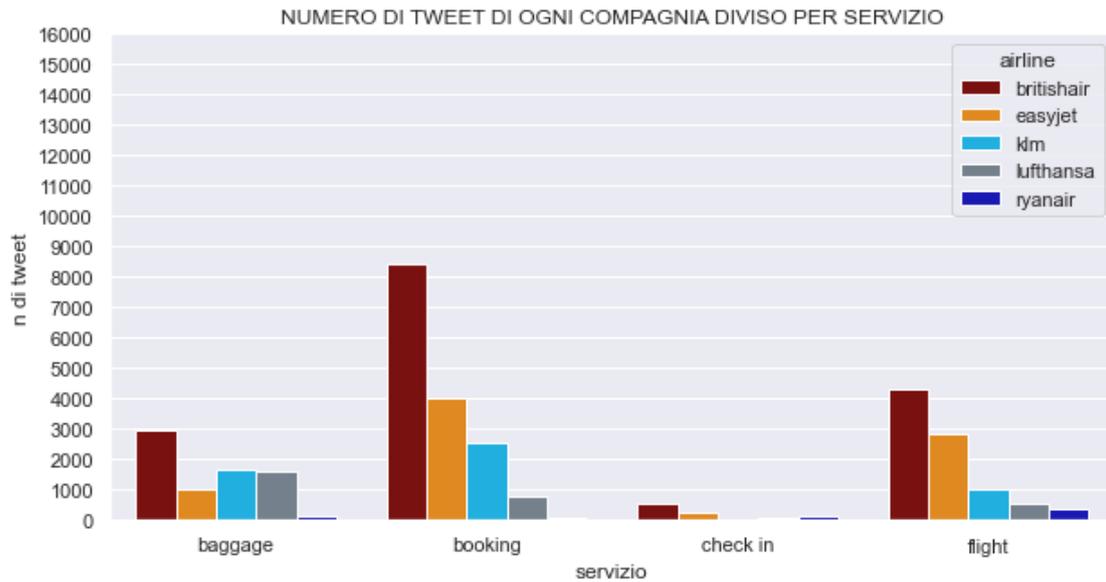


Figura 25: numero di tweet pubblicati dagli account ufficiali delle compagnie aeree relativamente ai propri servizi.

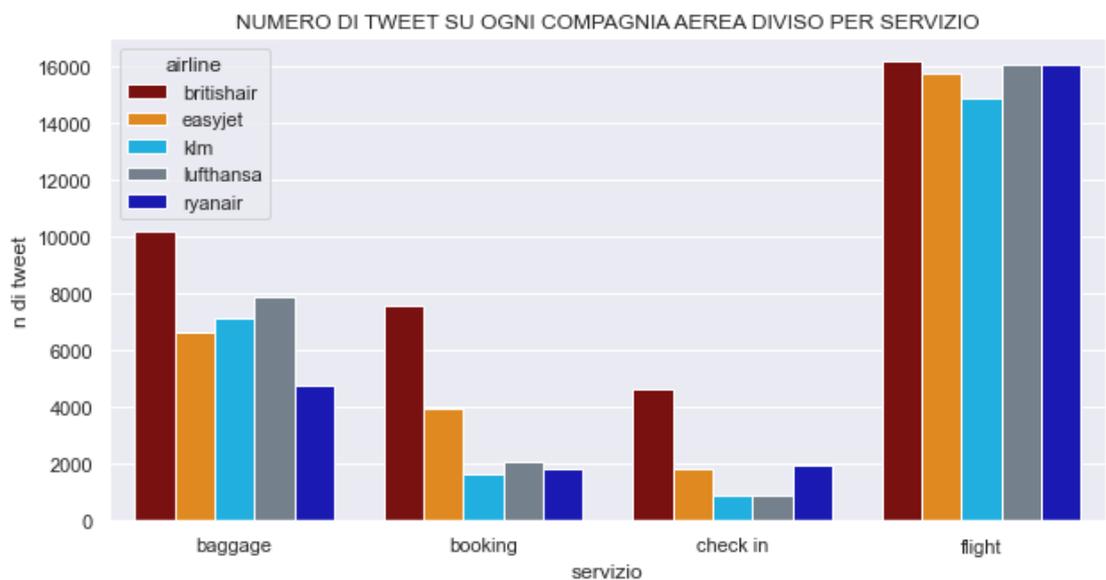


Figura 26: numero di tweet pubblicati dagli utenti relativamente ai servizi di ogni compagnia aerea.

Si può osservare che:

- Nella figura 25, il servizio che riceve maggiore attenzione nei tweet di tutte le compagnie aeree, tranne Ryanair e Lufthansa, è quello di prenotazione (“booking”). Lufthansa, invece, pone la propria attenzione primariamente al servizio bagagli (“baggage”) e Ryanair al servizio di volo (“flight”).

- Nella figura 26, i tweet pubblicati dagli utenti invece seguono quasi la stessa tendenza: nel caso di tutte le compagnie, tranne Ryanair, viene posta maggiore attenzione sul servizio “flight”, seguito dai servizi “baggage”, “booking” e “check in”. Diversamente, i tweet su Ryanair riguardano primariamente il servizio “flight”, seguito da “baggage”, e da “check in” e “booking” che presentano quasi lo stesso numero di tweet. Per tutte e cinque le compagnie, quindi, i due servizi su cui vengono pubblicati più tweet sono il volo e il bagaglio.

Si nota quindi che, mentre l’attenzione delle compagnie aeree rispetto ai propri servizi varia caso per caso, nei tweet degli utenti il livello di attenzione rispetto ai servizi delle compagnie aeree ha quasi lo stesso andamento in tutti i sub-corpus e, in tutti i casi, i primi due servizi su cui vengono pubblicati più tweet sono il volo e il bagaglio.

Nell’intero periodo considerato le uniche compagnie aeree che twittano primariamente su volo e bagaglio sono rispettivamente Ryanair e Lufthansa. Nel caso delle altre compagnie, quindi, si nota un disallineamento tra gli argomenti trattati dagli utenti e tra quelli trattati dagli account ufficiali.

Anche in questo caso, un basso numero di tweet della compagnia aerea riguardo un determinato servizio può indicare che la compagnia non utilizzi il proprio servizio clienti su Twitter per affrontare temi o problematiche legati a quei servizi o che aiuti i clienti tramite altri canali e forme di comunicazione (messaggio privato, call center, sito web, aeroporti). Inoltre, se più tweet pubblicati dagli utenti segnalano uno stesso problema rispetto a un servizio, la compagnia aerea potrebbe riuscire, in un dialogo da uno a molti, a rispondere agli utenti con un unico tweet e questo giustificherebbe il numero minore di contenuti pubblicati.

5.1.3 Correlazione

Per ogni compagnia aerea si è calcolato il coefficiente di correlazione lineare di Pearson (r), che può fornire informazioni sulla relazione tra il numero di tweet pubblicati dagli account ufficiali delle compagnie e il numero di tweet pubblicati dagli utenti relativamente ai quattro servizi (baggage, flight, booking e check-in).

Si è cercato quindi di osservare se le due variabili (numero di tweet pubblicati dagli account ufficiali e numero di tweet pubblicati dagli utenti) siano correlate positivamente

(valore del coefficiente vicino a +1), negativamente (valore del coefficiente vicino a -1) o non siano correlate (valore del coefficiente pari a 0).

In questa ricerca è interessante osservare se il coefficiente di correlazione è positivo e vicino a +1, ovvero se il numero di tweet pubblicati dall'account ufficiale di una compagnia aerea aumenta in risposta all'aumento del numero di tweet pubblicati dagli utenti e viceversa. Ciò, infatti, potrebbe suggerire che le compagnie aeree stiano attivamente monitorando e rispondendo alle esigenze degli utenti su Twitter. Se il coefficiente di correlazione invece è pari a 0, significa che non vi è alcuna correlazione tra il numero di tweet pubblicati dalle compagnie aeree e dagli utenti relativi ai servizi specifici, il che potrebbe indicare la presenza di fattori esterni che influenzano l'attività su Twitter di entrambi gli attori.

Bisogna comunque ricordare che la correlazione tra due variabili non è sinonimo di causalità, in quanto potrebbero esistere altre variabili non osservate¹⁸⁴ che influenzano l'associazione tra le variabili di interesse di uno studio (Agresti & Franklin 2016). La correlazione fornisce informazioni sull'andamento di una variabile in relazione a un'altra ma non dà informazioni su un eventuale rapporto di causa-effetto tra esse.

Compagnia aerea	Coefficiente di correlazione (r) tra numero di tweet degli account ufficiali delle compagnie aeree e numero di tweet degli utenti
British Airways	0,19436925
EasyJet	0,340356843
KLM	-0,057971412
Lufthansa	0,199156856
Ryanair	0,976899954
Correlazione media	0,330562298

Tabella 44: valore della correlazione tra il numero di tweet degli utenti e il numero di tweet degli account ufficiali per ogni compagnia aerea.

L'unica compagnia aerea per cui si ha un dato significativo è Ryanair, il cui coefficiente di correlazione è pari a 0,98, quindi prossimo alla perfetta correlazione positiva. C'è quindi una forte correlazione positiva tra il numero di tweet pubblicati dalla

¹⁸⁴ Le variabili non osservate vengono chiamate "variabili confondenti" (Agresti & Franklin 2016).

compagnia aerea e il numero di tweet pubblicati dagli utenti, il che potrebbe indicare che lo staff di Ryanair monitora i tweet degli utenti e, di conseguenza, pubblica dei contenuti fornendo delle risposte o delle informazioni per colmare determinate esigenze.

5.2 Sentiment rispetto a tutto il periodo e a tutte le compagnie aeree

Il sentiment degli utenti, rappresentato dal valore del *compound*, viene inizialmente indagato da un punto di vista macroscopico considerando:

- la percentuale di tweet positivi, negativi e neutri pubblicati dagli utenti nell'intero arco temporale.
- il numero dei tweet positivi, negativi, neutri pubblicati dagli utenti nell'intero arco temporale, divisi per servizio.
- il compound medio di tutti i tweet pubblicati dagli utenti relativi a una determinata compagnia aerea.

5.2.1 Percentuale di tweet negativi, positivi e neutri

I seguenti diagrammi a torta mostrano la percentuale di tweet con sentiment positivo, negativo e neutro presenti nel corpus totale comprendente tutti i tweet su tutte le compagnie aeree e nei sub-corpus specifici relativi a una determinata compagnia aerea. In ogni grafico, il tipo di sentiment è associato ad un colore specifico della sezione della torta e questo schema cromatico è ripetuto in tutti i grafici.

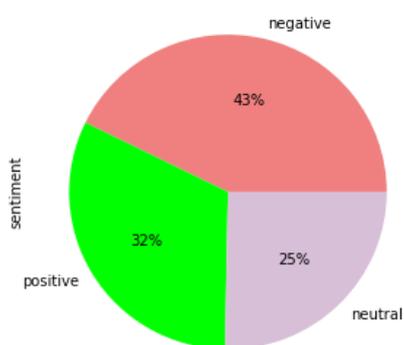


Figura 27: percentuale di tweet positivi, negativi e neutri presenti nel corpus 2.

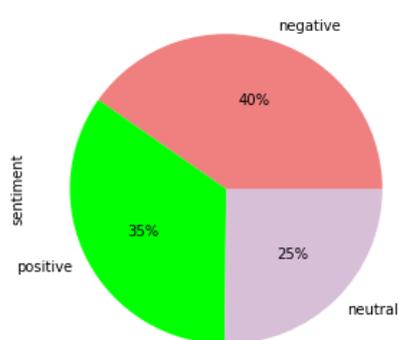


Figura 28: percentuale di tweet positivi, negativi e neutri presenti nel sub-corpus di tweet relativi a British Airways.

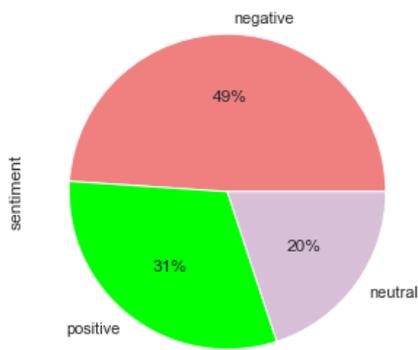


Figura 29: percentuale di tweet positivi, negativi e neutri presenti nel sub-corpus di tweet relativi a EasyJet.

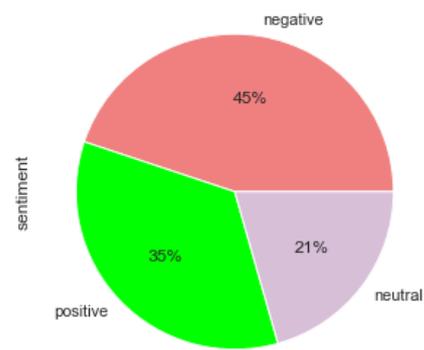


Figura 30: percentuale di tweet positivi, negativi e neutri presenti nel sub-corpus di tweet relativi a KLM.

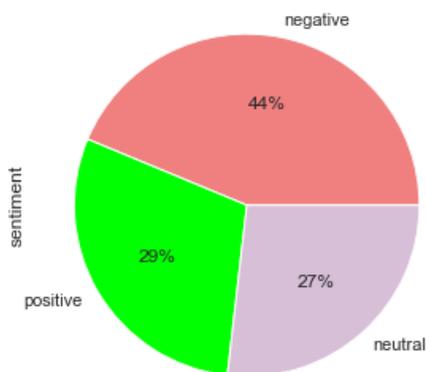


Figura 31: percentuale di tweet positivi, negativi e neutri presenti nel sub-corpus di tweet relativi a Lufthansa.

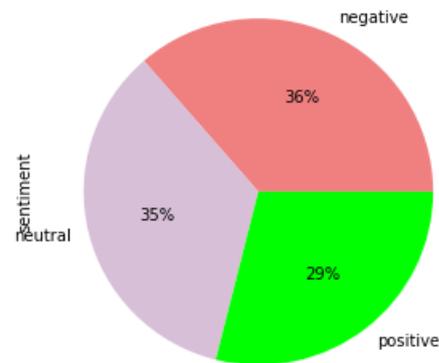


Figura 32: percentuale di tweet positivi, negativi e neutri presenti nel sub-corpus di tweet relativi a Ryanair.

Da questi diagrammi a torta si può notare che:

- In tutti i casi, la percentuale maggiore è quella di tweet negativi.
- La percentuale di tweet negativi più alta è di EasyJet e corrisponde al 49%.
- La percentuale di tweet negativi più bassa è di Ryanair e corrisponde al 36%.

Si può quindi affermare che, complessivamente, nell'intero periodo considerato e rispetto a tutte le compagnie aeree, il sentiment degli utenti è negativo. Il quadro teorico proposto da Hennig-Thurau et al. (2004) può essere applicato a questo caso specifico e aiutare a identificare diverse ragioni che portano un utente a pubblicare un tweet negativo:

- si diffonde una notizia, un'opinione o un'informazione per prevenire l'esperienza negativa di altri clienti delle compagnie aeree;
 - si vuole utilizzare Twitter come strumento di potere per influenzare la reputazione aziendale;
 - si vogliono chiedere pareri o opinioni ad altri utenti;
 - si vuole chiedere aiuto all'azienda o ad altri utenti;
 - si vuole ridurre il proprio senso di frustrazione esprimendo un'emozione negativa.
- Il sentiment negativo potrebbe rappresentare il malcontento degli utenti che, dopo aver avuto esperienza di uno o di più servizi offerti dalle compagnie aeree, vogliono ridurre il proprio senso di frustrazione esprimendo un'emozione negativa.

5.2.2 Numero di tweet positivi, negativi e neutri diviso per servizio

Si è realizzato un grafico (figura 33) che presenta il numero di tweet con sentiment positivo, negativo o neutro, raggruppato per servizio specifico. Anche in questo caso, il corpus di riferimento è quello contenente tutti i tweet pubblicati dagli utenti relativamente a tutte le compagnie aeree (corpus 2). L'asse x raggruppa per ogni servizio le barre che rappresentano il tipo di sentiment, mentre l'asse y indica il numero di tweet pubblicati. Il colore della barra corrispondente a un determinato tipo di sentiment segue lo stesso schema cromatico utilizzato per i diagrammi a torta mostrati precedentemente.

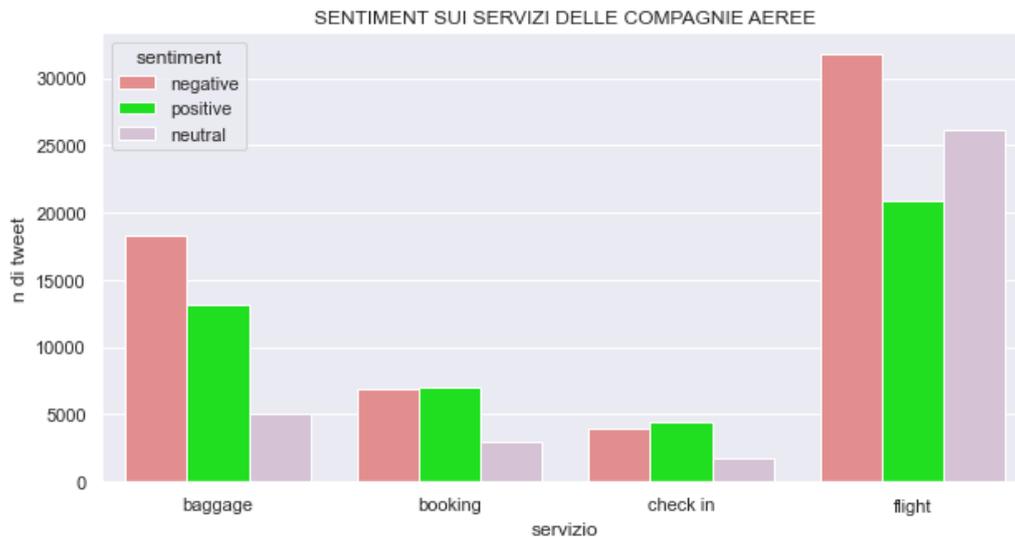


Figura 33: numero di tweet con sentiment positivo, negativo e neutro presenti nel corpus 2 e raggruppati per servizio.

Il numero maggiore di tweet negativi è quello corrispondente ai servizi di “flight” e “baggage” che, come si è osservato nella figura 26, sono anche quelli per cui gli utenti pubblicano più tweet. Ciò conferma quanto evidenziato da Rodriguez-Diaz et al. (2018), ovvero che i clienti insoddisfatti tenderebbero a esprimere maggiormente la propria opinione rispetto a quelli soddisfatti. Per avere una percezione reale della reputazione aziendale, quindi, la sola analisi del sentiment espresso nei tweet non è sufficiente, in quanto i dati presenti su Twitter non sono rappresentativi di tutti i clienti e potrebbero non includere eventuali opinioni positive. La reputazione online, quindi, potrebbe non corrispondere sempre alla reputazione reale di un’azienda. Secondo gli autori (ibid.), è perciò necessario utilizzare i dati provenienti da Twitter come integrazione per quelli provenienti da altre fonti che usano campioni statisticamente rappresentativi.

5.2.3 Compound medio

Nel seguente grafico a barre (figura 34) viene mostrato il valore di compound medio dei tweet pubblicati dagli utenti relativamente ad ogni compagnia aerea nell’intero periodo (23 maggio – 2 ottobre 2022). Anche in questo caso, viene considerato l’intero arco temporale a cui appartengono i tweet e ogni compagnia aerea è identificata da un colore univoco.

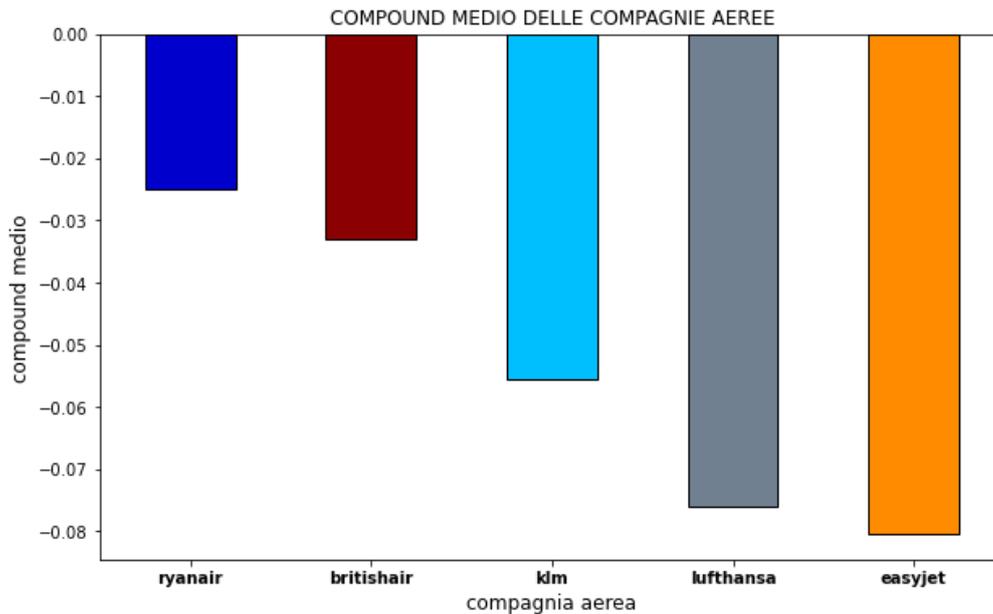


Figura 34: compound medio per i tweet di ogni compagnia aerea nell'intero arco temporale considerato.

I valori di compound medio riferiti a tutto il periodo temporale considerato sono, in ordine dal più negativo al meno negativo:

- Easyjet: -0,080525
- Lufthansa: -0,075973
- KLM: -0,055689
- British Airways: -0,033094
- Ryanair: -0,025049

Easyjet è quindi la compagnia aerea che, secondo il compound rilevato nei tweet, performa peggio, mentre Ryanair è quella che performa meglio. Si sottolinea che in questa ricerca non si cercano di comparare la performance dei diversi modelli di business delle compagnie aeree (low cost vs tradizionale): tuttavia, da questi dati si può osservare che non esiste una chiara tendenza a favore di un modello di business piuttosto che un altro, poiché sia la compagnia aerea che performa meglio che quella che performa peggio sono low-cost.

Nell'analisi di Mabrian Technologies¹⁸⁵ (2022) viene messa in luce la migliore performance di Ryanair rispetto alle altre compagnie in termini di numero di cancellazioni di voli. L'analisi mostra quali sono le compagnie aeree europee che sono state maggiormente colpite dalle cancellazioni dell'ultimo minuto tra il 17 luglio e il 30 settembre 2022. Le compagnie aeree qui considerate hanno visto, in ordine decrescente, le seguenti percentuali¹⁸⁶ di voli cancellati:

- Lufthansa: 3,69%
- KLM: 2,64%
- British Airways: 1,79%
- EasyJet: 1,34%
- Ryanair: 0,5%

Lo studio di Mabrian Technologies conferma quindi come Ryanair abbia performato meglio a livello di cancellazioni. Come riportato da Euronews (Frost 2022), mentre le compagnie aeree che hanno la propria base operativa a Heathrow hanno continuato a cancellare voli, Ryanair, che opera a Stansted, li ha addirittura aumentati. Michael O'Leary, CEO di Ryanair, ha dichiarato che la compagnia aveva previsto la ripresa dei viaggi e l'aumento della domanda di voli e ha agito con sufficiente anticipo per poter disporre di un numero di dipendenti adeguato.

Inoltre, O'Leary ha specificato che uno dei motivi per cui Ryanair è riuscita a prepararsi in tempo all'estate 2022 è stato il fatto di essere un'azienda irlandese che non ha subito le conseguenze negative della Brexit. Quest'ultima, infatti, ha reso più difficoltoso reclutare lo staff per le linee aeree e gli aeroporti inglesi (Bickerton 2022).

Il minor numero di cancellazioni da parte di Ryanair potrebbe aver avuto un effetto positivo sull'esperienza dei clienti e sul sentiment espresso nei tweet. Da ciò potrebbe essere risultato il numero minore di tweet pubblicati e, di conseguenza, il numero minore di tweet con sentiment negativo (data la possibile tendenza degli utenti a pubblicare più contenuti quando si sentono insoddisfatti).

¹⁸⁵ <https://mabrian.com/blog/top-10-european-airlines-cancellations-summer-2022/> [ultima consultazione: 20/04/23]

¹⁸⁶ I valori percentuali fanno riferimento al totale dei voli programmati dalle compagnie aeree.

5.3 Mese più critico

In questa sezione si individua il mese più critico di ogni compagnia aerea, basandosi sul valore medio del compound dei tweet pubblicati dagli utenti e si effettua un'analisi approfondita relativa al mese individuato.

5.3.1 Compound medio per ogni mese

Si è svolta un'analisi temporale del sentiment relativo ai tweet degli utenti di ogni compagnia aerea. A partire dall'intero periodo in cui i tweet sono stati pubblicati, che va dal 23 maggio al 2 ottobre, si è calcolato il compound medio dei tweet pubblicati dagli utenti delle cinque compagnie aeree per ogni mese. Poiché il periodo completo non copre interamente i mesi di maggio e ottobre, si è deciso di non considerare il compound medio di questi due mesi. Infine, per ogni compagnia aerea, si è identificato il mese in cui il compound medio è stato più basso. Nella tabella 45 si mostrano i valori medi del compound relativi ad ogni mese e ogni compagnia e si evidenziano i valori minimi.

VALORE MEDIO DEL COMPOUND PER OGNI MESE				
	Giugno	Luglio	Agosto	Settembre
British Airways	-0,04495	-0,03371	-0,03033	-0,0092
EasyJet	-0,08983	-0,07015	-0,08504	-0,07793
KLM	-0,06746	-0,08123	-0,03083	-0,04918
Lufthansa	-0,03804	-0,10405	-0,07819	-0,0839
Ryanair	-0,03108	-0,01516	-0,02254	-0,03288

Tabella 45: valore medio del compound di ogni compagnia relativo ad ogni mese.

In base al valore medio del compound, i mesi che possono essere considerati più critici sono:

- Giugno, per British Airways ed EasyJet.
- Luglio, per KLM e Lufthansa.
- Settembre, per Ryanair.

Si nota che in tutti i mesi e per tutte le compagnie aeree, i valori medi sono negativi. Il valore più negativo di compound medio è quello di Lufthansa (-0,104), mentre il valore meno negativo è quello di Ryanair (-0,033).

5.3.2 Risultati specifici per ogni compagnia

Per ogni compagnia, relativamente al mese con il valore di compound medio più basso, si indagano:

- A. La variazione del valore del compound medio giornaliero.
- B. Il servizio con compound più negativo in quel mese: per tale servizio si mostrano i bi-gram e le parole più frequenti.
- C. La data con il valore di compound più negativo: si visualizza la word cloud con le parole più frequenti, si leggono i 20 tweet più negativi pubblicati in quella data e si ricercano su Google eventuali episodi particolari connessi a quella giornata, riportati da testate giornalistiche online o da blog di settore.

- **British Airways**

Mese critico: giugno

A. Variazione del compound medio giornaliero

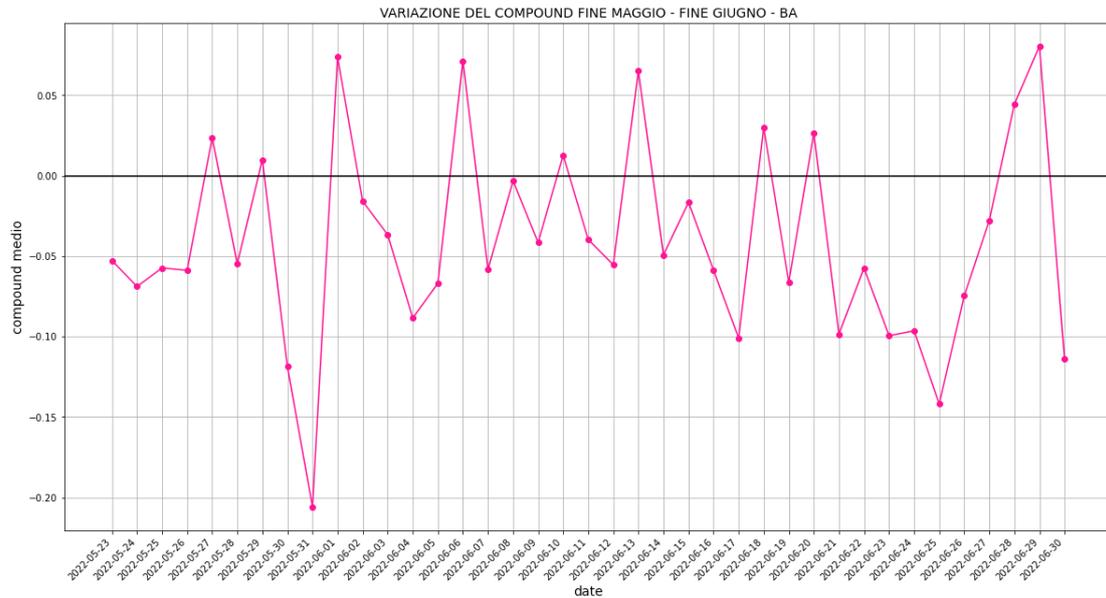


Figura 35: variazione del compound medio dei tweet degli utenti relativi a British Airways pubblicati dal 23 maggio 2022 al 30 giugno 2022.

Il grafico a linee (figura 35) mostra la variazione del compound giorno per giorno: nell'asse x sono riportate le date, mentre nell'asse y il valore medio del compound. Nell'analisi si includono anche le date delle ultime settimane di maggio, nonostante queste non siano state considerate per andare a calcolare il compound medio del mese di giugno.

B. Servizio con il valore di compound più negativo

Il grafico a barre (figura 36) mostra il valore del compound medio per ogni servizio nel periodo tra maggio e giugno (23 maggio – 30 giugno). I tweet che hanno avuto il compound medio più negativo sono quelli relativi al servizio “baggage”. Il servizio “booking”, invece, è l'unico che ha un compound positivo. Ciò potrebbe significare che gli utenti che pubblicano dei tweet riguardanti questo servizio lo fanno per chiedere o diffondere informazioni e non per esprimere opinioni negative.

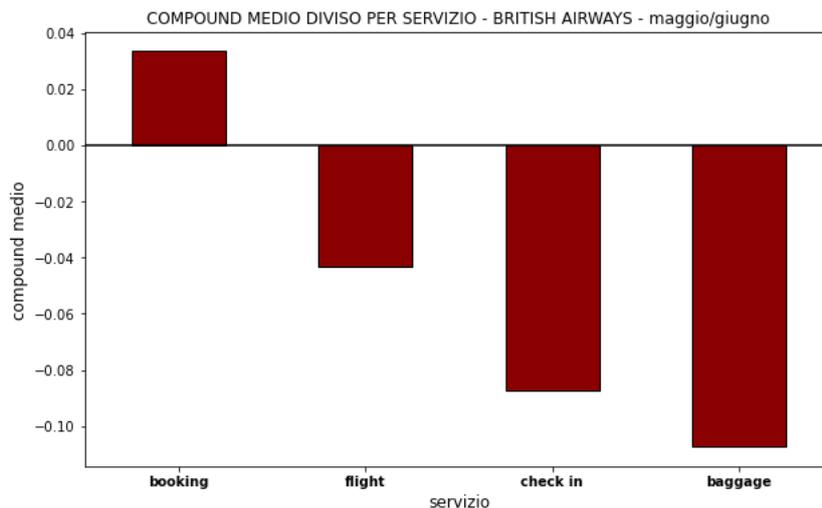


Figura 36: valore medio del compound diviso per servizio, relativo ai tweet pubblicati dal 23 maggio al 30 giugno 2022 riferiti a British Airways.

Per capire quali possono essere state le problematiche connesse al servizio “baggage”, si mostra un grafico a barre orizzontali (figura 37) che mostra i bi-gram più frequenti nei tweet negativi di British Airways dal 23 maggio al 30 giugno. Si nota come il sentiment negativo sia associato alla perdita del bagaglio (“lose, luggage”, “lose, baggage”, “miss, luggage”), ai tempi d’attesa (“2, hour”) e alla difficoltà nel contattare il servizio clienti (“customer,service”), che probabilmente non risponde tempestivamente alle chiamate dei clienti (“answer, phone”).

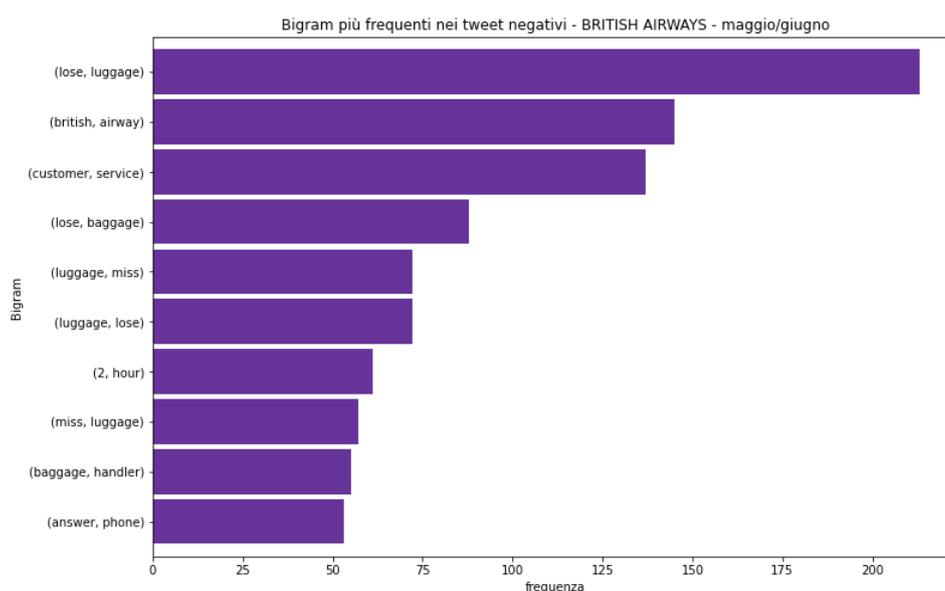


Figura 37: grafico a barre orizzontali che mostra i bi-gram più frequenti presenti nei tweet negativi riferiti al servizio “baggage” di British Airways e pubblicati tra il 23 maggio e il 30 giugno 2022.

Anche la word cloud (figura 38), che si riferisce sempre al servizio “baggage” nello stesso periodo, mette in mostra come il malcontento sia legato alla perdita del bagaglio (“lose”) e all’impossibilità di ricevere aggiornamenti e aiuto dallo staff (“service”, “customer”, “response”, “staff”, “claim”, “update”, “phone”, “help”, “tell”, “wait”, “need”, “call”, “contact”). Inoltre, rispetto al grafico dei bi-gram, risalta anche un sentiment negativo verso il servizio generico “flight” e l’aeroporto di Londra Heathrow (“london”, “heathrow”). Si precisa che per creare questa word cloud sono state rimosse alcune parole non utili ai fini dell’analisi perché l’occorrenza era ovvia o superflua: “baggage”, “luggage”, “bag”, “british”, “airway”, “ba¹⁸⁷”, “amp¹⁸⁸”.



Figura 38: word cloud relativa ai tweet negativi del servizio “baggage” di British Airways nel periodo che va dal 23 maggio al 30 giugno 2022.

C. Data con il valore di compound più negativo

Notando la presenza di un picco negativo a cavallo tra i due mesi (figura 35), si è deciso di indagare cosa sia accaduto nella data del 31 maggio (nonostante il valore del compound di questa giornata non abbia influenzato il valore medio del compound di giugno). Si sono quindi analizzati tutti i tweet negativi pubblicati in questa data (in totale sono 194). Per interpretare questo dato vengono in aiuto le parole più frequenti presenti nei tweet negativi di quella giornata, rappresentati nella seguente word cloud (figura 39), in cui sono state rimosse alcune parole non utili ai fini dell’analisi perché la loro occorrenza era ovvia: “british”, “airway”, “ba¹⁸⁹”.

¹⁸⁷ Acronimo di British Airways.

¹⁸⁸ Parola che compare nei tweet processati e che corrisponde alla E commerciale “&”, in inglese chiamata “ampersand”.

¹⁸⁹ Acronimo di British Airways.



Figura 39: word cloud relativa ai soli tweet negativi riferiti a British Airways pubblicati nella giornata del 31 maggio 2022.

- “pay”, “cut”, “staff”, “member”, “strike”, “battle”: queste parole sembrano rimandare alla situazione difficoltosa in cui si è trovato il personale del trasporto aereo. Gli utenti di Twitter sembrano rilevare come, a causa dell'emergenza sanitaria, dal 2020 le compagnie aeree abbiano dovuto ridurre le proprie operazioni, arrivando anche a tagliare gli stipendi dei dipendenti (“pay”, “cut”, “staff”, “member”). Questa situazione è stata il motivo scatenante dei numerosi scioperi che i dipendenti hanno organizzato in più momenti dell'estate 2022 (“strike”, “battle”). Tra gli utenti di Twitter, come fanno notare Hasson et al. (2019) potrebbero essere inclusi gli stessi dipendenti delle aziende, i quali commentano la situazione in cui si sono ritrovati. Non si può però averne la certezza, dato che Twitter non fornisce alcun dato sociodemografico.
- “covid”: nei tweet negativi sono presenti dei rimandi alla situazione sanitaria che ha ostacolato il settore del trasporto aereo dal 2020.
- “chaos”, “situation”, “luggage”, “baggage”: queste parole rimandano alle situazioni di disagio che si sono manifestate per tutti gli stakeholder di British Airways, dai dipendenti che hanno dovuto lavorare in condizioni svantaggiose, ai clienti finali che hanno avuto esperienza di vari disagi, come quelli legati alla gestione dei bagagli.

I 20 tweet più negativi pubblicati il 31 maggio 2022 contengono opinioni sui tagli salariali dello staff. Uno di questi, ad esempio, allega un articolo di giornale e contiene la seguente frase (qui tradotta):

I tagli salariali introdotti durante la pandemia sono stati revocati per il management della compagnia aerea, ma continuano ad applicarsi al resto del personale. Vergogna! @British_Airways deve sistemare questa cosa per evitare gli scioperi, ma non può comunque mantenere la propria reputazione.

Un altro tweet, invece, menziona un problema relativo alla gestione dei bagagli e alla disponibilità del servizio clienti: una famiglia con un bambino di due anni, arrivata a Gatwick da Palma di Mallorca, ha dovuto aspettare ben due ore per ottenere i propri bagagli e non è riuscita a contattare il servizio clienti. Questo episodio è stato menzionato anche in un articolo del 31 maggio pubblicato da Sussex Live (Carey 2022), che ha riportato anche altre esperienze negative dei clienti con il servizio bagagli di British Airways. Si nota quindi come uno o più tweet possano diffondersi velocemente ed essere riportati e commentati da testate giornalistiche che aumenteranno ulteriormente la visibilità e la diffusione dell'evento. Un singolo tweet può quindi influire sulla reputazione della compagnia aerea coinvolta, evidenziando eventuali carenze nei servizi offerti e portando alla diffusione di un sentimento negativo nei confronti dell'azienda. È quindi importante che l'azienda sia in grado di monitorare e gestire in modo efficace i riscontri negativi da parte dei clienti su Twitter prima che questi si amplifichino.

- **EasyJet**

Mese critico: giugno

A. Variazione del compound medio giornaliero

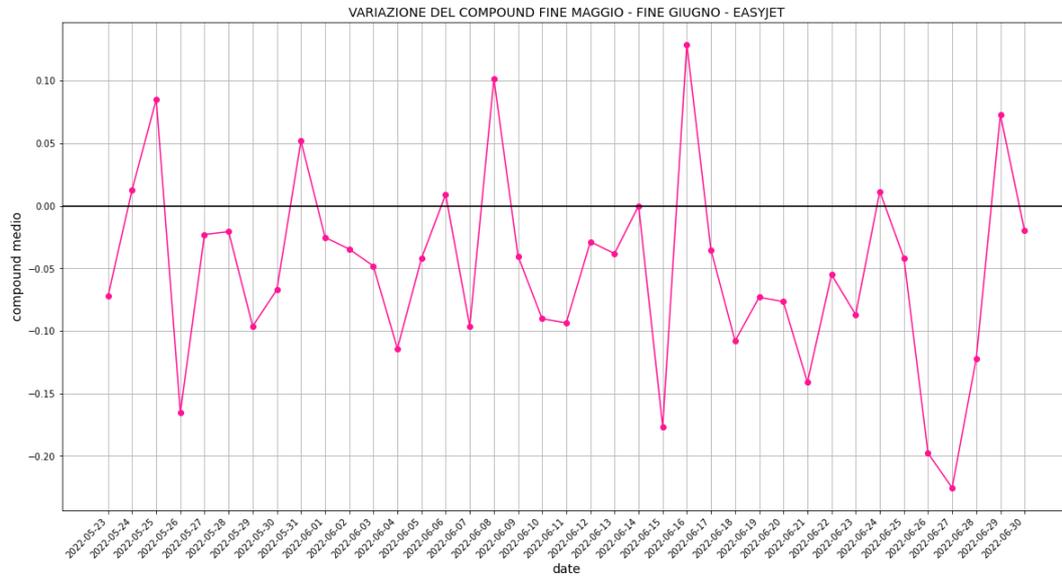


Figura 40: variazione del compound medio dei tweet degli utenti relativi a EasyJet pubblicati dal 23 maggio 2022 al 30 giugno 2022.

Nel caso di EasyJet il mese con il compound medio giornaliero più negativo è stato giugno e la figura 40 ne mostra la variazione giorno per giorno. Come nel caso di British Airways, il grafico mostra sull'asse x le date a partire dal 23 maggio 2022 fino al 30 giugno 2022, mentre sull'asse y il valore medio del compound.

B. Servizio con il valore di compound più negativo

Il grafico a barre (figura 41) mostra il valore del compound medio per ogni servizio nel periodo di maggio e giugno (23 maggio – 30 giugno). I tweet che hanno avuto il compound medio più negativo sono quelli relativi al servizio “flight”. Il servizio “check in”, invece, è l'unico che ha un compound medio positivo. Ciò potrebbe significare che gli utenti che pubblicano tweet riguardanti il check-in lo fanno per chiedere o diffondere informazioni e non per esprimere opinioni negative.

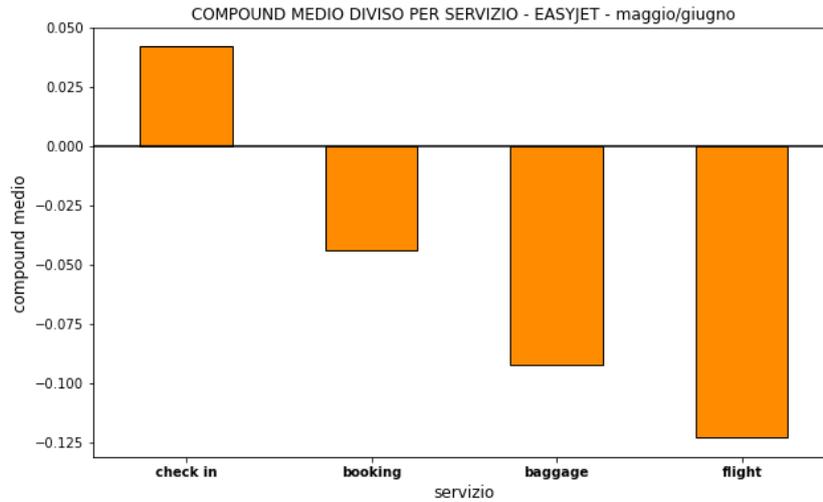


Figura 41: valore medio del compound diviso per servizio, relativo ai tweet pubblicati dal 23 maggio al 30 giugno 2022 riferiti a EasyJet.

Per capire quali possono essere state le problematiche connesse al servizio “flight”, si mostra un grafico a barre orizzontali (figura 42) che mostra i bi-gram più frequenti nei tweet negativi su EasyJet pubblicati tra il 23 maggio e il 30 giugno 2022. Il sentiment negativo è associato alla cancellazione (“cancel,flight”) e al ritardo (“flight,delay”) dei voli, al mese in cui sono stati pubblicati i tweet (“june, 2022”), all’aeroporto di Londra Gatwick (“gatwick, airport”) e alla natura improvvisa degli eventi che hanno cambiato i piani dei viaggiatori (“unexpectedly, collapse”, “suddenly, unexpectedly”).

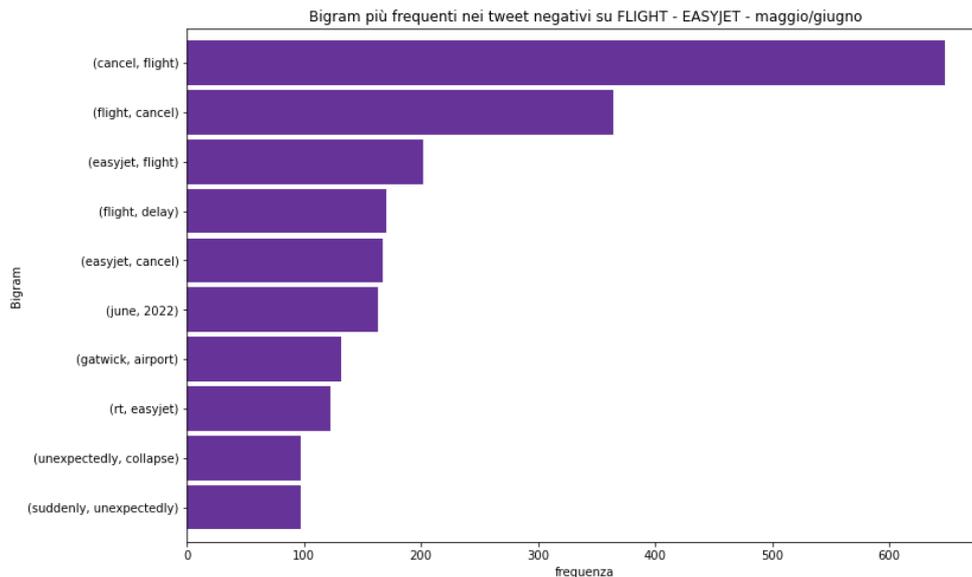


Figura 42: grafico a barre orizzontali che mostra i bi-gram più frequenti presenti nei tweet negativi riferiti al servizio “flight” di EasyJet, pubblicati tra il 23 maggio e il 30 giugno 2022.

La word cloud (figura 43), che si riferisce sempre al servizio “flight” nello stesso periodo, oltre a confermare le informazioni presentate nel grafico dei bi-gram, indica che il sentiment negativo potrebbe essere legato anche al soggiorno in hotel (“hotel”) e al rimborso (“compensation”, “refund”) che le compagnie aeree offrono in caso di cancellazione, allo stress causato dai vari cambi di programma e dalle attese (“stress”, “sit”, “wait”) e dall’affollamento di qualche area in cui i passeggeri si ritrovano a sostare (“fully”). Si precisa che per creare questa word cloud sono state rimosse le parole “easyjet” e “flight”, non utili ai fini dell’analisi perché la loro occorrenza era ovvia.

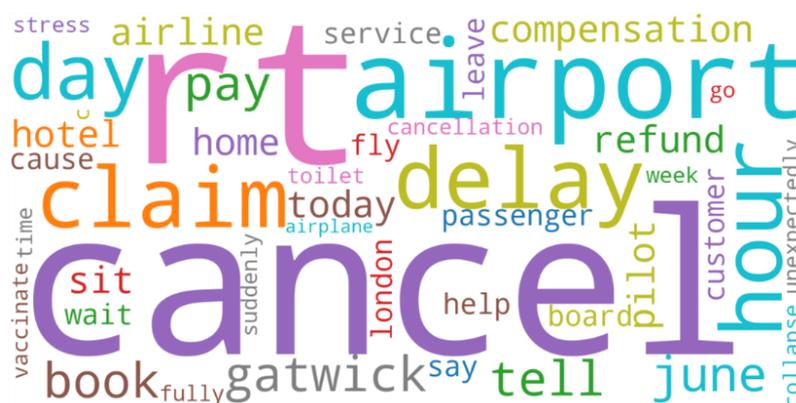


Figura 43: word cloud relativa ai tweet negativi sul servizio “flight” di EasyJet nel periodo che va dal 23 maggio al 30 giugno 2022.

C. Data con il valore di compound più negativo

Dalla figura 40, si nota che tra maggio e giugno la data con il valore di compound più negativo relativo a tutti i servizi di British Airways è il 27 giugno. Si analizzano quindi tutti i tweet negativi pubblicati in questa giornata (in totale sono 536). Per interpretare questo dato vengono in aiuto le parole più frequenti presenti nei tweet negativi di quella giornata, visibili nella figura 44.



Figura 44: word cloud relativa ai soli tweet negativi riferiti a EasyJet pubblicati nella giornata del 27 giugno 2022.

In questo caso, le parole superflue che sono state rimosse dalla word cloud sono “easyjet” e “amp¹⁹⁰”. Si nota come anche in questo caso il sentiment negativo sia legato alle parole:

- “airport”, “flight”, “delay”, “cancel”, “amsterdam”, “london”, “gatwick”, “bristol”: esse si riferiscono ai problemi legati ai voli (ritardi e cancellazioni) e agli aeroporti in cui si potrebbero essere verificati (Londra Gatwick, Bristol, Amsterdam).
- “hotel”, “refund”, “compensation”, “bag”, “luggage”, “lose”, “hour: tramite i tweet gli utenti potrebbero aver chiesto informazioni o criticato le azioni delle compagnie relativamente ai rimborsi dei voli, all'alloggio offerto, all'attesa o alla perdita dei bagagli.
- “email”, “help”, “service”: probabilmente le compagnie aeree non hanno gestito in modo adeguato le richieste di aiuto arrivate via e-mail o altri tipi di canali del servizio

¹⁹⁰ Parola che compare nei tweet processati e che corrisponde alla E commerciale “&”, in inglese chiamata “ampersand”.

clienti e per questo gli utenti hanno cercato di contattarle o criticarle tramite la pubblicazione dei tweet.

I 20 tweet più negativi pubblicati il 27 giugno mettono in luce le seguenti problematiche:

- Non è stato erogato il rimborso ai passeggeri in diversi casi, tra cui uno in cui il ritardo del volo non superava la soglia minima di ore di ritardo per ottenere un rimborso e un altro in cui, dopo la cancellazione del volo, i passeggeri non avevano ancora ricevuto nessuna notizia da parte della compagnia aerea.
- È stato difficile contattare la compagnia aerea attraverso alcuni canali, come la chat online o il telefono.
- Sono stati smarriti i bagagli e non è stata fornita una prova del loro smarrimento da parte della compagnia aerea: ciò ha impedito ai passeggeri di usufruire della propria assicurazione bagaglio.

A differenza dei tweet negativi che riguardavano la compagnia Ryanair che avevano attirato l'attenzione dei media, tramite la ricerca su Google non è stato trovato alcun articolo che ponesse una particolare attenzione sui disagi che stavano caratterizzando EasyJet il 27 giugno. Probabilmente, sebbene ci siano stati alcuni problemi per i clienti, non si sono verificati eventi così straordinari da meritare una particolare attenzione da parte dei media. Gli unici articoli trovati in questa data presentano la questione delle cancellazioni che ha interessato diverse compagnie aeree, tra cui EasyJet (Seveno 2022).

- **KLM**

Mese critico: luglio

A. Variazione del compound medio giornaliero

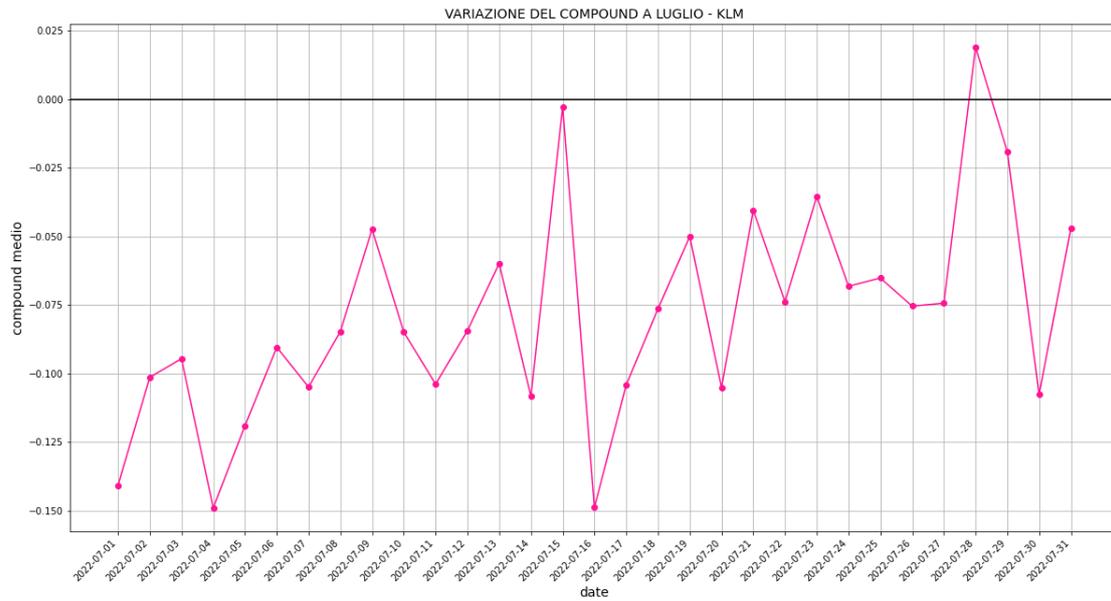


Figura 45: variazione del compound medio dei tweet degli utenti relativi a KLM pubblicati dal primo luglio al 31 luglio 2022.

Il grafico a linee (figura 45) mostra la variazione del compound medio giornaliero nel mese di luglio: sull'asse x vengono mostrate le date a partire dal primo luglio fino al 31 luglio, mentre sull'asse y vengono mostrati i valori del compound. Si nota come il compound medio giornaliero sia sempre stato negativo, tranne il 28 luglio.

B. Servizio con il valore di compound più negativo

Il grafico a barre (figura 46) mostra il valore del compound medio per ogni servizio nel mese di luglio (1-31 luglio 2022). I tweet che hanno avuto il compound medio più negativo sono quelli relativi al servizio “baggage”. Il servizio “booking” è l'unico che ha un compound medio positivo, segnalando così che probabilmente gli utenti che pubblicano tweet riguardo a quel servizio lo fanno per chiedere o diffondere informazioni e non per esprimere opinioni negative.

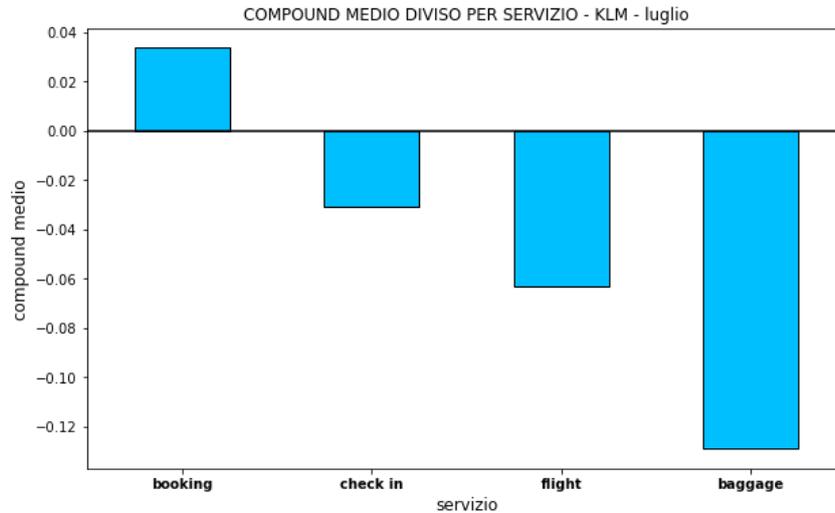


Figura 46: valore medio del compound diviso per servizio, relativo ai tweet pubblicati dal primo luglio al 31 luglio 2022.

Il grafico a barre orizzontali (figura 47) mostra i bi-gram più frequenti nei tweet negativi sul servizio “baggage” di KLM pubblicati tra il primo luglio e il 31 luglio e aiuta a individuare quali possono essere stati i problemi incontrati dai viaggiatori. Il sentiment negativo è associato alla perdita del bagaglio (“lose, luggage”, “lose, baggage”, “miss, baggage”, “miss, luggage”, “luggage, lose”, “lose, bag”) e al servizio clienti (“customer, service”). Si può presumere quindi che il servizio clienti non sia stato in grado di fornire l'assistenza necessaria ai passeggeri riguardo ai bagagli smarriti.

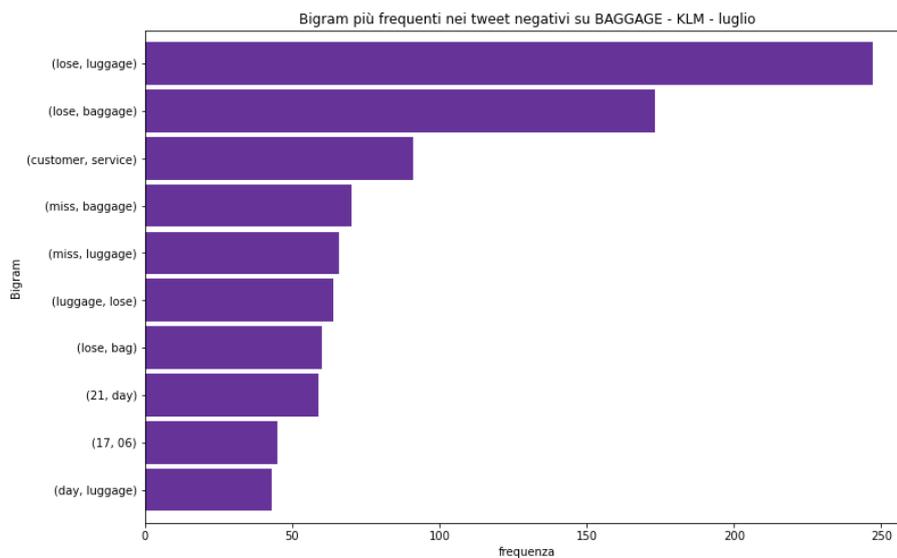


Figura 47: grafico a barre orizzontali che mostra i bi-gram più frequenti presenti nei tweet negativi riferiti al servizio “baggage” di KLM pubblicati tra il primo luglio e il 31 luglio 2022.

- **Lufthansa**

Mese critico: luglio

A. Variazione del compound medio giornaliero

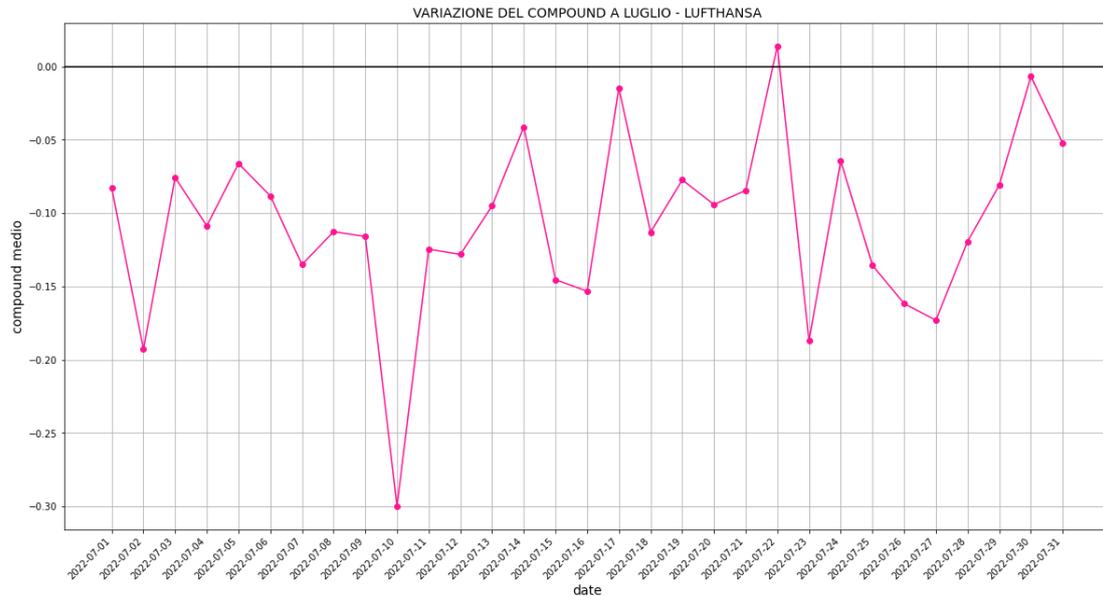


Figura 50: variazione del compound medio dei tweet degli utenti relativi a Lufthansa pubblicati dal primo luglio al 31 luglio 2022.

La figura 50 mostra la variazione del compound medio giornaliero di Lufthansa nel mese di luglio: sull'asse x vengono mostrate le date a partire dal primo luglio fino al 31 luglio, mentre sull'asse y vengono mostrati i valori del compound. Si nota come il valore del compound sia sempre stato negativo, tranne il 22 luglio 2022 in cui supera leggermente il valore 0.

B. Servizio con il valore di compound più negativo

Il grafico a barre (figura 51) mostra il valore del compound medio per ogni servizio nel mese di luglio (1-31 luglio 2022). Tutti i servizi hanno un valore medio di compound negativo, ma il servizio “baggage” è quello che ha il valore più negativo.

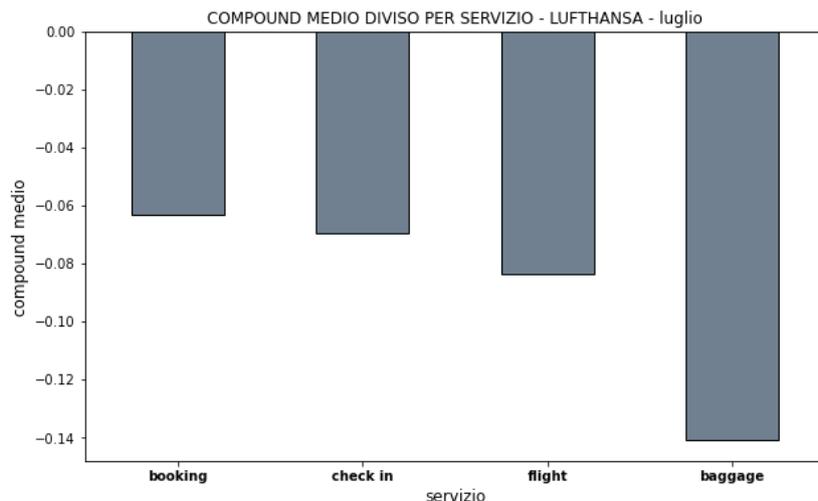


Figura 51: valore medio del compound diviso per servizio, relativo ai tweet pubblicati su Lufthansa dal primo luglio al 31 luglio 2022.

Nella figura 52 a barre orizzontali si possono leggere i bi-gram più frequenti nei tweet negativi sul servizio “baggage” pubblicati tra il primo luglio e il 31 luglio. Il sentiment negativo è associato alla perdita del bagaglio (“lose, luggage”, “miss, baggage”), al ritardo nella ricezione del bagaglio (“delay, baggage”), al servizio clienti (“customer, service”) e alla modalità di contatto (“phone, number”), alla cancellazione del volo (“cancel, flight”). Inoltre, sembra esserci stato un problema con uno strumento musicale (“musical, instrument”). Cercando online (Gallagher 2022), si scopre che gli strumenti musicali persi sono quelli di Andy Irvine, musicista irlandese che per un viaggio da Dublino a Skaagen, in Danimarca, è salito su tre voli (due di questi di Lufthansa). Inizialmente gli strumenti non sono arrivati alla destinazione finale insieme al musicista. Solo in un secondo momento, la linea aerea Scandinavian Airlines è riuscita a riconsegnarli al proprietario, sollevando così Lufthansa da qualsiasi responsabilità.

del trasporto aereo, non sono disposti ad accettare questi problemi. Il tweet con il valore di compound più negativo è il seguente:

```
@lufthansa sette giorni e quattro voli da Francoforte a Nairobi e ancora nessun bagaglio?! Veramente?! Nessuna chiamata, nessuna informazione... questo è ridicolo! Con tutta la tecnologia disponibile, non si tratta solo di carenza di manodopera. Quanti altri sono nella stessa situazione?!
```

Ricercando ulteriori informazioni su Google riguardo agli eventi del 10 luglio, si trova un'intervista del 11 luglio 2022 che Jens Ritter, CEO di Lufthansa, ha rilasciato a Pilot Career News. Egli ha sottolineato che l'intero settore del trasporto aereo stava vivendo una crisi operativa senza precedenti e a livello globale. Secondo Ritter, riuscire a rispondere in tempi brevi alla ripresa dei viaggi aerei era impossibile, perché a livello operativo erano coinvolti troppi attori diversi che necessitavano di coordinazione. Per questi motivi, Lufthansa ha deciso di ridurre in modo significativo i voli operati da Francoforte e Monaco tra l'8 e il 14 luglio. Tra i problemi descritti da Ritter ci sono stati la mancanza di personale di terra, gli alti tassi di malattia a causa del covid, le carenze di materie prime e lo spazio aereo ridotto a causa della guerra in Ucraina. Si nota come Ritter, con le proprie dichiarazioni, abbia messo in atto una strategia di risposta alla crisi che può essere interpretata tramite il modello della *Image Restoration Theory* (Benoit 1997 citato in Splendiani 2022): egli cerca di eludere la responsabilità di Lufthansa, affermando che rispondere in tempi brevi alla ripresa dei viaggi fosse impossibile e cerca di ridurre l'offensività affermando che la compagnia aeree diminuirà in modo significativo il numero di voli operati in modo da non provocare più disagi ai viaggiatori.

- **Ryanair**

Mese critico: settembre

A. Variazione del compound medio giornaliero

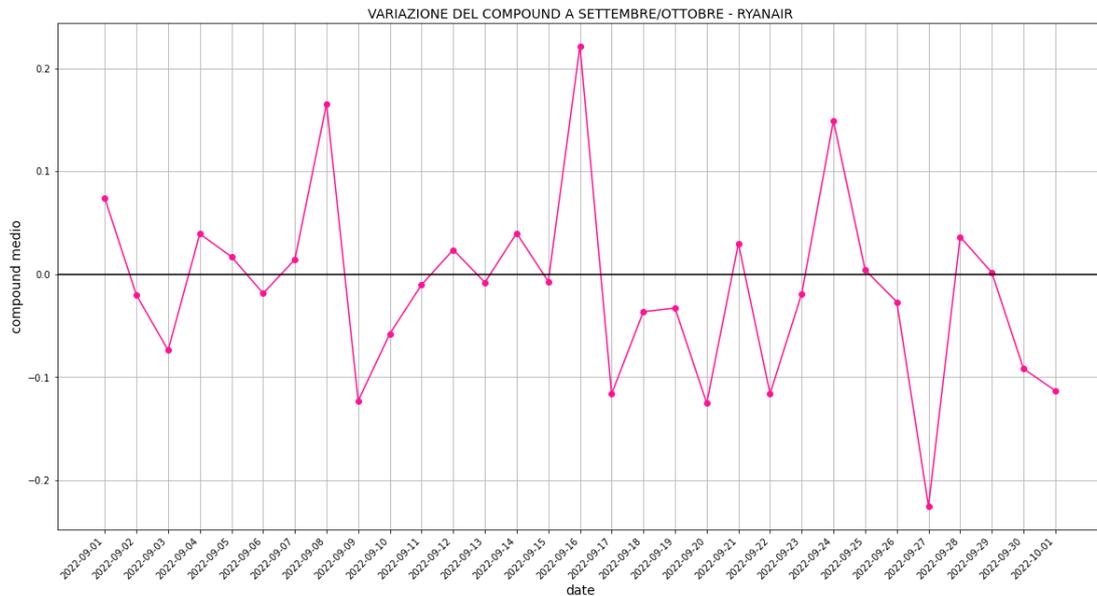


Figura 55: variazione del compound medio giornaliero dei tweet degli utenti relativi a Ryanair pubblicati dal 1° settembre al 1° ottobre 2022.

Per Ryanair il mese con il valore di compound medio più basso è stato settembre e la figura 55 ne mostra la variazione giorno per giorno (si è considerato anche il primo giorno di ottobre, che è l'ultima data in cui sono stati scaricati i tweet). A differenza delle altre compagnie aeree, quindi, il mese più critico per Ryanair è stato verso la fine dell'estate. Si nota come nel corso di settembre il valore del compound abbia visto diverse fluttuazioni, con picchi sia positivi che negativi.

B. Servizio con il valore di compound più negativo

Il grafico a barre (figura 56) mostra il valore del compound medio per ogni servizio nel mese di settembre. Ryanair si differenzia rispetto alle altre compagnie aeree viste finora poiché registra dei valori di compound medio positivi per tutti i servizi, tranne che per il servizio "flight". Si potrebbe ipotizzare quindi che i tweet più negativi di questo mese, che hanno contribuito alla media negativa del compound mensile, riguardino questo servizio.

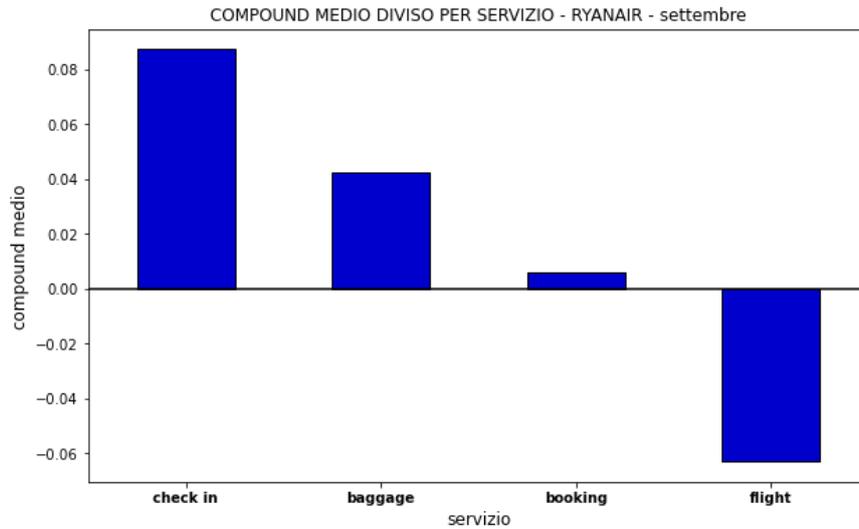


Figura 56: valore medio del compound diviso per servizio, relativo ai tweet pubblicati su Ryanair dal primo settembre al 30 settembre 2022.

Il grafico a barre orizzontali (figura 57) mostra i bi-gram più frequenti nei tweet negativi sul servizio “flight” di Ryanair pubblicati nel mese di settembre e consente di indagare più nel dettaglio quale può essere stata la causa del sentiment negativo degli utenti. Innanzitutto, si notano i problemi che riguardano il ritardo (“flight, delay”) e la cancellazione (“cancel, flight”) dei voli e gli aeroporti (“airport, claim”). Ciò che però risalta ed è insolito, rispetto alle altre compagnie aeree viste finora, sono i bi-gram relativi a un possibile violenza sessuale (“man, arrest”, “sexually, assault”) che ha attirato l’attenzione degli utenti di Twitter. Si potrebbe quindi trattare di un evento specifico avvenuto a settembre che ha coinvolto direttamente la compagnia aerea.

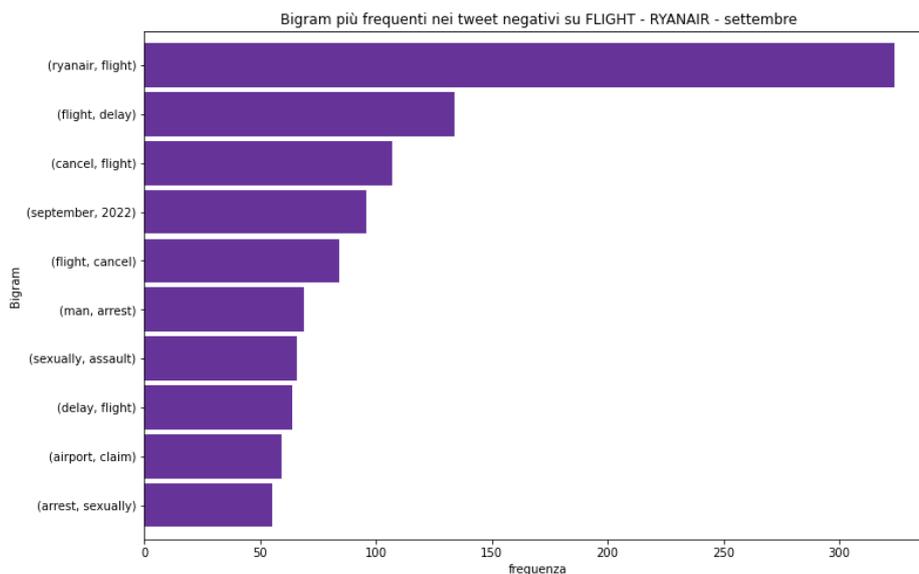


Figura 57: grafico a barre orizzontali che mostra i bi-gram più frequenti presenti nei tweet negativi riferiti al servizio “flight” di Ryanair pubblicati tra il 1° settembre e il 30 settembre 2022.

La word cloud (figura 58) permette di visualizzare altre parole e di approfondire le informazioni mostrate dai bi-gram. Anche in questo caso, ci sembrano essere tre argomenti-chiave diversi:

- Cancellazioni e i ritardi dei voli: i viaggiatori potrebbero aver visto i propri voli cancellati o ritardati (“cancel”, “hour”, “time”, “delay”, “fly”) e potrebbero aver richiesto dei rimborsi (“voucher”, “refund”, “compensation”).
- Problemi con determinati aeroporti, tra cui Manchester e Dublino (“manchester”, “dublin” “airport”, “claim”).
- Violenza sessuale: questa violenza (“sexually”, “assault”) potrebbe riguardare uno o più passeggeri (“passenger”, “man”) che l’hanno subita o compiuta e potrebbe esserci stato un arresto (“arrest”). Si verificherà se questo evento corrisponda alla data con più tweet negativi del mese di settembre.

All’interno della word cloud sono state rimosse le parole “ryanair” e “flight” perché ritenute superflue ai fini dell’analisi.



Figura 58: word cloud relativi ai tweet negativi sul servizio “flight” di Ryanair nel mese di settembre.

C. Data con il valore di compound più negativo

Nel mese di settembre si può notare che la data con il valore di compound medio più negativo è il 27 settembre (figura 55), in cui sono stati pubblicati 135 tweet negativi. Dalla seguente word cloud (figura 59) emerge come il sentiment negativo sia legato all’episodio di molestia sessuale già rilevato nell’analisi dell’intero mese di settembre. Si possono ricavare delle informazioni aggiuntive: molto probabilmente l’episodio di violenza ha comportato un arresto (“arrest”) e tra i luoghi interessati ci sono la Spagna, in particolare Palma di Maiorca (“spain”, “mallorca”, “majorca”), e il Regno Unito, nello specifico le città di Liverpool e l’aeroporto di Londra Stansted (“brit”, “british”, “london”, “stansted”, “liverpool”). L’accaduto sembra essere connesso al servizio di volo (“flight”) e ora compare anche una seconda persona coinvolta, di genere femminile (“woman”, “female”).



Figura 59: word cloud relativa ai soli tweet negativi riferiti a Ryanair, pubblicati il 27 settembre 2022.

Analizzando i 20 tweet più negativi del 27 settembre e cercando informazioni su Internet, si scopre che il 25 settembre un uomo britannico è stato arrestato per aver molestato una donna seduta accanto a lui su un volo Ryanair da Liverpool a Palma di Maiorca. Secondo quanto riportato dal The Independent (Sly & Penza 2022) e dal Liverpool Echo e dichiarato da un portavoce di Ryanair, il personale di volo ha agito contattando immediatamente la polizia, che ha arrestato l'uomo non appena l'aereo è atterrato a Palma di Maiorca. I tweet analizzati mostrano come diversi utenti abbiano diffuso la notizia due giorni dopo l'evento, il 27 settembre. Si nota come Ryanair, che non era direttamente responsabile dell'accaduto, abbia gestito la situazione in modo efficace, informando tutti i possibili stakeholder su cosa fosse stato fatto per aiutare la vittima della molestia. In questo caso, il portavoce di Ryanair ha preservato la reputazione della compagnia aerea sottolineando come questa non fosse responsabile dell'accaduto. Tuttavia, si osserva come la compagnia aerea abbia rilasciato le proprie dichiarazioni alle testate giornalistiche (The Independent, Liverpool Echo) ma abbia scelto di non menzionare l'accaduto sul proprio profilo Twitter.

5.4 Elementi comuni tra le varie compagnie

Per offrire una panoramica della sentiment analysis effettuata per le cinque compagnie aeree, la seguente tabella riporta in modo sintetico alcuni dati:

Compagnia aerea	Mese critico	Servizio peggiore nel mese critico	Data critica	N° di tweet negativi pubblicati nella data critica	Eventi nella data critica e narrazione giornalistica
British Airways	Giugno	Baggage	31 maggio ¹⁹³	194	<ul style="list-style-type: none"> • Bagagli persi o in ritardo. • Compagnia aerea non reperibile. • Articolo che narra l'esperienza

¹⁹³ Notando la presenza di un picco negativo a cavallo tra i due mesi (figura 35), si è deciso di indagare cosa sia accaduto nella data del 31 maggio (nonostante il valore del compound di questa giornata non abbia influenzato il valore medio del compound di giugno).

					negativa di alcuni clienti.
EasyJet	Giugno	Flight	27 giugno	536	<ul style="list-style-type: none"> • Ritardo dei voli e mancato rimborso. • Compagnia aerea non reperibile. • <u>Nessun articolo di giornale.</u>
KLM	Luglio	Baggage	4 luglio	219	<ul style="list-style-type: none"> • Bagagli persi o in ritardo. • Compagnia aerea non reperibile. • Articolo su un uomo che non disponeva di medicine salvavita.
Lufthansa	Luglio	Baggage	10 luglio	46	<ul style="list-style-type: none"> • Problemi con servizio bagagli. • Compagnia aerea non reperibile. • Articolo con dichiarazione di Lufthansa.
Ryanair	Settembre	Flight	27 settembre	135	<ul style="list-style-type: none"> • Violenza sessuale a bordo di un aereo Ryanair. • Articoli che narrano l'evento e riportano dichiarazioni di Ryanair.

Tabella 46: sintesi della sentiment analysis effettuata sulle cinque compagnie aeree.

Considerando i risultati ottenuti dall'analisi delle cinque compagnie aeree, si possono fare le seguenti osservazioni:

- Le date in cui sono stati rilevati dei picchi negativi nei valori del compound possono corrispondere sia ad eventi eccezionali (come il caso della molestia sessuale a bordo del volo Ryanair) sia a un insieme di eventi diffusi (ritardo o cancellazione voli, perdita bagagli, scarsa reperibilità della compagnia aerea), da considerarsi quasi eventi ordinari nella situazione in cui si è trovato il settore del trasporto aereo e che hanno colpito più passeggeri in uno o più aeroporti.
- In molte word cloud compare la parola “rt”, che segnala che molti tweet sono dei retweet, ovvero ripubblicazioni di tweet da parte di altri utenti. Ciò fa comprendere che l’opinione espressa da un utente tramite un tweet è spesso condivisa (nel senso letterale della parola) anche da molti altri account di Twitter. Insieme, i tweet e i retweet enfatizzano e diffondono una determinata opinione o informazione sulla compagnia aerea e contribuiscono a plasmare la sua reputazione online.
- Come si è notato dall’analisi della maggior parte delle compagnie aeree, la diffusione di notizie e opinioni su un evento può avvenire contemporaneamente sia attraverso i tweet degli utenti che tramite gli articoli di giornale. Questi due canali, insieme, contribuiscono a plasmare la reputazione aziendale. Pertanto, un’azienda dovrebbe monitorare più canali contemporaneamente, in modo da capire con che modalità e con che influenza vengono diffuse le notizie e le opinioni online.
- In tutti i sub-corpus analizzati, molti tweet negativi si riferiscono alla scarsa reperibilità delle compagnie aeree tramite i propri siti web e call center. Come riportato anche dal The Telegraph (Beaumont 2022), dopo i numerosi ritardi, le cancellazioni dei voli e i disagi con i bagagli, un numero elevato di utenti ha cercato di rivolgersi ai call center o all’assistenza clienti online delle compagnie. Anche in questo caso però, i sistemi informatici e telefonici e il numero di operatori non erano adeguati a fronteggiare un alto numero di richieste. I clienti hanno così avuto esperienza di linee telefoniche costantemente occupate, tempi di attesa eccessivi e siti web fuori servizio. In questa situazione, EasyJet, a luglio 2022 ha dichiarato di aver prolungato gli orari di disponibilità dei call center e di aver assunto altri 350 operatori (Calder 2022). British Airways si è scusata su Twitter dei disagi e ha cercato di dirottare i clienti verso altri canali:



Figura 60: tweet di British Airways¹⁹⁴.

Si nota come, nei momenti in cui si registra un alto volume di richieste di assistenza e i canali di comunicazione principali non si dimostrano efficaci, Twitter possa essere utilizzato dalle compagnie aeree e dai clienti come canale di comunicazione alternativo e secondario. Come ha fatto notare Matthews (2022), in situazioni di crisi come questa, in cui è a rischio la reputazione collettiva dell'intero settore, una compagnia aerea che vuole riacquistare la fiducia dei propri clienti e risaltare rispetto ai concorrenti dovrebbe concentrare i propri sforzi sull'implementazione dei servizi di comunicazione e assistenza. Nel caso delle compagnie aeree EasyJet e British Airways, come strategia di risposta alla crisi¹⁹⁵ la prima ha adottato un'azione correttiva, mentre la seconda si è dichiarata mortificata.

- Ryanair si differenzia rispetto alle altre compagnie aeree:
 - Come mostrato in precedenza, questa compagnia aerea ha il compound medio più alto (-0,025) relativo a tutto il periodo di pubblicazione dei tweet (23 maggio – 2 ottobre) e, come rilevato dalla ricerca di Mabrian Technologies (2022), ha avuto un tasso di cancellazione dei voli inferiore rispetto alle altre compagnie aeree considerate.
 - Dall'analisi della compagnia aerea nel mese di settembre (che è quello con la performance peggiore in termini di sentiment) emerge che il compound

194

https://twitter.com/British_Airways/status/1544710142771478540?ref_src=twsrc%5Etfw%7Ctwcamp%5Etweetembed%7Ctwterm%5E1544710142771478540%7Ctwgr%5E5ff77631fca82e9f6f3bafb14aab4bf33aafc3a7%7Ctwcon%5Es1_&ref_url=https%3A%2F%2Fwww.dailymail.co.uk%2Fnews%2Farticle-10988229%2FBritish-Airways-axes-10-300-short-haul-flights-end-October.html [ultima consultazione: 15/04/23]

¹⁹⁵ Si fa riferimento alla *Image Restoration Theory* (Benoit 1997 citato in Splendiani 2022).

medio relativo a tutti i servizi di Ryanair, tranne il volo, è positivo, mentre nel caso delle altre compagnie aeree il valore del compound era negativo per la maggior parte dei servizi.

- Nel caso del servizio di volo di Ryanair, bisogna ricordare che molti dei tweet che hanno contribuito ad abbassare il valore del compound si riferiscono all'episodio di molestia sessuale commessa da un passeggero, che non è quindi imputabile alla compagnia aerea. Inoltre, la compagnia aerea ha dimostrato di aver agito tempestivamente, collaborando con le autorità per far arrestare l'autore del reato e informando tutti i possibili stakeholder tramite delle dichiarazioni rilasciate a diverse testate giornalistiche. In aggiunta, però, la compagnia aerea avrebbe potuto rilasciare delle dichiarazioni anche tramite il proprio profilo Twitter.
- I servizi che registrano il compound medio più negativo sono il servizio “baggage” (per British Airways, KLM e Lufthansa) e il servizio “flight” (per EasyJet e Ryanair).
 - Le parole più frequenti in comune ai tweet negativi di British Airways, KLM e Lufthansa relativi al servizio “baggage” sono rappresentate sinteticamente tramite una word cloud (figura 61). In sintesi, si può osservare che il sentiment negativo sembra essere legato alla perdita o al ritardo nella ricezione del bagaglio (“miss”, “lose”, “delay”, “wait”) che può portare i viaggiatori ad aspettare ore o addirittura giorni (“hour”, “delay”) prima di riaverlo. Oltre alla perdita del bagaglio, per i clienti di queste compagnie aeree uno dei problemi principali è la difficoltà nel mettersi in contatto con il servizio clienti (“customer”, “service”, “answer”, “contact”, “update”), soprattutto tramite la linea telefonica (“number”, “phone). Di conseguenza, molti utenti utilizzano Twitter come canale di comunicazione alternativo e inseriscono il numero del volo (“flight”) per cui stanno cercando informazioni.

service: in tutti questi casi non è necessaria la presenza fissa del personale e perciò la sua mancanza potrebbe aver avuto un effetto meno negativo sulla qualità del servizio.

CAPITOLO 6 – CONCLUSIONI

L'obiettivo di questo elaborato sperimentale è stato quello di monitorare la reputazione aziendale su Twitter attraverso la tecnica della sentiment analysis. Dopo aver fornito una base teorica e metodologica per la definizione e la misurazione della reputazione, si è applicata la tecnica della sentiment analysis a un caso studio specifico: si è individuata l'opinione dei clienti di British Airways, EasyJet, KLM, Lufthansa, Ryanair rispetto a quattro servizi specifici (baggage, booking, check-in, flight), nell'estate del 2022. In totale sono stati scaricati e analizzati 176133 tweet (33519 per il corpus 1; 142614 per il corpus 2), ognuno dei quali è stato categorizzato in base alla compagnia aerea e al servizio di riferimento.

Il concetto astratto di reputazione è stato quindi scomposto in una variabile empirica, ovvero il tipo di opinione dei clienti che, a seconda dei casi, può essere positivo, negativo o neutro e può essere rilevato tramite una sentiment analysis. L'analisi dei dati si è focalizzata sull'individuazione di eventuali opinioni negative, con l'obiettivo di offrire una riflessione su come queste potessero influire sulla reputazione aziendale online.

Si è precisato che i corpora su cui è stata effettuata l'analisi non contengono un campione di dati statisticamente rappresentativo, in quanto non sono conosciuti i dati sociodemografici della popolazione indagata (gli utenti di Twitter), né la sua ampiezza (numero totale di utenti che hanno espresso la propria opinione su Twitter). I risultati ottenuti, quindi, non sono generalizzabili ma, in un'ottica esplorativa, possono aiutare a individuare delle aree problematiche o di possibile miglioramento che necessitano di ulteriori ricerche di tipo statistico, oppure, in un'ottica confermativa, possono rendere più solidi i risultati già ottenuti in altre ricerche.

La ricerca ha confermato l'ipotesi di partenza, ovvero che su Twitter sia possibile rilevare l'opinione degli utenti che si esprimono sulle compagnie aeree e quindi poter ottenere dei dati al fine di monitorare la reputazione aziendale. Alla luce dei risultati dell'analisi, le risposte R alle domande di ricerca D1, D2, D3, sono le seguenti:

R1: scaricando i tweet pubblicati dagli account ufficiali delle compagnie aeree è stato possibile rilevare il numero di tweet pubblicato da ogni compagnia aerea e individuare i servizi su cui sono stati pubblicati più tweet. I risultati hanno permesso di confrontare l'attività su Twitter degli account ufficiali delle cinque compagnie aeree.

R2: grazie alla *sentiment analysis*, è stato possibile rilevare il sentiment degli utenti (corrispondente al valore del *compound*) nei confronti delle cinque compagnie aeree e dei loro servizi. Si è prima considerata l'intera arena competitiva, ottenendo dei risultati riguardanti tutte le compagnie aeree nell'intero arco temporale a cui appartengono i tweet scaricati. Successivamente, si sono potute individuare delle tendenze caratteristiche di ogni singola compagnia aerea (*compound* medio, parole più frequenti) in un arco temporale più ristretto (un mese o una data specifica). A partire da alcune date specifiche, in cui si sono registrati dei picchi negativi nei valori del *compound*, si sono potuti rintracciare degli eventi critici che hanno ricevuto attenzione anche da parte delle testate giornalistiche.

R3: l'unica relazione tra i due corpora che è stata indagata è quella tra il numero di tweet pubblicati dagli account ufficiali delle compagnie aeree e il numero di tweet pubblicati dagli utenti, misurata tramite il coefficiente di correlazione di Pearson. Si è potuto quindi osservare se le due variabili (numero di tweet pubblicati dagli account ufficiali e numero di tweet pubblicati dagli utenti) fossero correlate positivamente, negativamente o non correlate affatto. Ciò ha permesso di osservare se il numero di tweet pubblicati dall'account ufficiale di una compagnia aerea aumentasse in risposta all'aumento del numero di tweet pubblicati dagli utenti e viceversa. Una correlazione positiva potrebbe infatti suggerire che la compagnia aerea stia monitorando e rispondendo alle esigenze degli utenti su Twitter.

Nello specifico, i risultati ottenuti dall'analisi portano alle seguenti conclusioni:

1. Il numero di tweet pubblicato dagli utenti è sempre maggiore rispetto a quello pubblicato dagli account ufficiali.

Alla base di questa constatazione potrebbero esserci diverse ragioni¹⁹⁶, ma soprattutto bisogna considerare che i tweet pubblicati dagli utenti sono più numerosi perché il corpus 2 include sia i tweet in cui la compagnia aerea viene taggata (in cui c'è quindi una specifica richiesta a cui deve rispondere) sia i tweet che contengono semplicemente il nome della compagnia. Un *reputation manager*, quindi, dovrebbe prestare attenzione a

¹⁹⁶ La compagnia aerea potrebbe non utilizzare attivamente il proprio servizio clienti su Twitter, oppure potrebbe rispondere a più richieste simili pubblicando un unico tweet o, ancora, molte richieste potrebbero essere processate tramite messaggi privati in cui gli utenti possono comunicare i propri dati personali.

tutto ciò che viene detto in rete riguardo l'azienda, ovvero sia alle richieste dirette che i clienti fanno alle compagnie aeree, sia ai tweet che non si rivolgono direttamente alle compagnie ma che contengono un'opinione su di esse. Tutti i contenuti, diretti o indiretti, su media *owned* e *not owned*, possono contribuire a diffondere informazioni e opinioni e, infine, plasmare la reputazione aziendale.

2. In generale, non vi è alcuna correlazione tra il numero di tweet pubblicati dagli account ufficiali delle compagnie aeree e il numero di tweet pubblicati dagli utenti.

Solo nel caso di Ryanair è stata rilevata una correlazione positiva significativa ($r = 0,98$), suggerendo che questa compagnia aerea potrebbe essere l'unica a monitorare attivamente e rispondere alle esigenze degli utenti su Twitter. Tuttavia, le ragioni che portano gli account ufficiali a pubblicare dei tweet possono essere molteplici e non dipendere esclusivamente dal volume dei contenuti pubblicati dagli utenti.

3. L'esperienza negativa vissuta dai viaggiatori nell'estate del 2022 si è riflessa anche nel contesto online e nei tweet pubblicati dagli utenti.

Nei casi di tutte le compagnie aeree, la percentuale maggiore di tweet è quella di tweet negativi, con un valore massimo del 49% nel caso di EasyJet e un valore minimo del 36% nel caso di Ryanair. Anche il valore di compound medio dei tweet pubblicati dagli utenti in tutto il periodo considerato è sempre negativo con, anche in questo caso, EasyJet che performa peggio e Ryanair che performa meglio. Si è mostrato quindi come il passaparola online (e-wom) sia un riflesso e un ampliamento di quello offline (wom) e possa fornire importanti indicazioni circa il grado di soddisfazione dei viaggiatori e la reputazione delle compagnie aeree.

4. Le date in cui si registrano picchi negativi nei valori del compound possono corrispondere sia ad eventi eccezionali che ad eventi ordinari.

A seconda dei casi, il picco negativo di compound può corrispondere ad eventi eccezionali in cui la compagnia aerea si è trovata coinvolta o a un insieme di eventi diffusi che, sebbene negativi e numerosi, possono essere considerati ordinari nella situazione di crisi strisciante in cui si è trovato il settore del trasporto aereo. In entrambi i casi, però, si possono individuare i fattori che hanno causato la pubblicazione di tweet estremamente

negativi, valutare l'eventuale coinvolgimento della compagnia aerea e la relativa ricaduta reputazionale.

5. Oltre ai tweet, i retweet sono fondamentali nella diffusione delle informazioni e delle opinioni.

In molte word cloud è comparsa la parola “rt”, che segnala che molti tweet sono dei retweet. Ciò fa comprendere che l'opinione espressa da un utente è spesso condivisa (nel senso letterale della parola) anche da molti altri account di Twitter. Insieme, i tweet e i retweet enfatizzano e diffondono una determinata opinione o informazione sulla compagnia aerea e contribuiscono a plasmare la reputazione online. Ciò conferma quanto affermato da Thelwall et al. (2010) nel proprio studio, in cui i retweet vengono mantenuti perché considerati un indicatore di come una specifica opinione abbia ottenuto consenso tra gli utenti.

6. La diffusione di opinioni e notizie può avvenire contemporaneamente sia attraverso i tweet che attraverso i media tradizionali.

Gli eventi a cui fanno riferimento i tweet negativi degli utenti trovano spesso riscontro anche in degli articoli di giornale. Come fa notare Norsa (2002a citato in Poma e Vecchiato 2012: 180), i contenuti condivisi su Internet sono incontrollabili e veloci, possono raggiungere un pubblico molto ampio ed essere una fonte di informazione anche per i media tradizionali. Si nota quindi come tutto ciò che viene pubblicato nel Web entra a far parte di un “archivio permanente” (Sernovitz 2011: 74), in cui sarà possibile rintracciare le informazioni anche a distanza di molto tempo.

7. In tutti i sub-corpus analizzati, gli utenti si sono lamentati della scarsa reperibilità delle compagnie aeree, in quanto queste non riuscivano ad essere contattate tramite l'assistenza nel sito web o il call center.

Innanzitutto, si nota come, nei momenti in cui si registra un alto volume di richieste di assistenza e i canali di comunicazione principale non si dimostrano efficaci, Twitter possa essere utilizzato dai clienti come canale di comunicazione alternativo e secondario per provare a contattare le compagnie aeree. Inoltre, le lamentele riguardanti la reperibilità delle compagnie aeree mettono in luce le caratteristiche del Web 2.0 che hanno richiesto

delle modifiche nell'approccio alle situazioni di crisi: come fa notare Splendiani (2022), nel contesto online, l'avvento di una crisi e la propagazione di notizie e opinioni a riguardo sono simultanei. Allo stesso modo, la risposta da parte dell'azienda dovrebbe essere immediata, in quanto gli utenti hanno "un grande appetito di notizie" (Poma & Vecchiato 2012: 185) e si aspettano che le aziende presenti online siano in grado di soddisfarlo. Inoltre, come fa notare Matthews (2022), in situazioni di crisi come questa, in cui è a rischio la reputazione collettiva dell'intero settore, una compagnia aerea che vuole riacquistare la fiducia dei propri clienti e risaltare rispetto ai concorrenti dovrebbe concentrare i propri sforzi sull'implementazione dei servizi di comunicazione e assistenza.

Dall'analisi si è potuto rilevare il bisogno dei viaggiatori di ricevere una maggiore assistenza prima, durante e dopo il viaggio. Le compagnie aeree possono utilizzare le informazioni ottenute dall'analisi per adattare le proprie strategie di comunicazione, creando, ad esempio, campagne pubblicitarie che mettano in risalto il loro impegno nel migliorare la reperibilità e l'assistenza tempestiva. In tal modo, le compagnie aeree potranno migliorare la propria reputazione, emergere rispetto alle aziende concorrenti e costruire una relazione di fiducia con i clienti, dimostrando di averli ascoltati e aver compreso le loro esigenze.

8. Nell'intero periodo considerato, i due servizi su cui gli utenti pubblicano più tweet sono quelli del volo e del bagaglio. Questi servizi sono anche quelli che registrano il maggior numero di tweet negativi e un valore medio di compound più negativo.

Il numero maggiore di tweet negativi è quello corrispondente ai servizi di "flight" e "baggage" che sono anche i due servizi per cui gli utenti pubblicano più tweet. Ciò conferma la tesi di Rodriguez-Diaz et al. (2018) per cui gli utenti insoddisfatti tenderebbero a esprimere maggiormente la propria opinione rispetto ai clienti soddisfatti. È importante tener presente che i dati provenienti da Twitter potrebbero non includere eventuali opinioni positive e che essi non sono rappresentativi di tutti i clienti. La reputazione online potrebbe quindi non corrispondere sempre alla reputazione reale di un'azienda. È necessario utilizzare questi dati come integrazione per quelli rilevabili tramite altre ricerche che usano campioni di dati statisticamente rappresentativi.

Le uniche compagnie aeree che hanno pubblicato dei tweet primariamente su volo e bagaglio sono rispettivamente Ryanair e Lufthansa. Bisognerebbe quindi verificare la motivazione di questo disallineamento di interessi da parte delle altre compagnie aeree e se esse forniscono assistenza e informazioni per questi due servizi tramite altri canali (call center, sito web, messaggi privati, aeroporti).

9. Nei mesi più critici delle singole compagnie aeree, i servizi che hanno registrato il compound medio più negativo sono il servizio “baggage” (per British Airways, KLM e Lufthansa) e il servizio “flight” (per EasyJet e Ryanair).

L’analisi testuale dei tweet e la visualizzazione del contenuto tramite una word cloud hanno permesso di rilevare alcuni dettagli sulle cause del malcontento degli utenti. Per quanto riguarda il bagaglio, gli utenti hanno espresso opinioni negative riguardo la perdita del bagaglio, i lunghi tempi di attesa per ritrovarlo o la mancanza di informazioni da parte del servizio clienti che, come anticipato, secondo gli utenti di Twitter non è stato efficiente. Nel caso dei voli, invece, il malcontento degli utenti è stato causato primariamente dalla cancellazione e secondariamente dal ritardo dei voli. Non sorprende quindi che Ryanair, che ha registrato un tasso di cancellazione minore rispetto ad altre compagnie, sia la compagnia aerea che ha registrato il valore di compound meno negativo.

10. Ryanair è la compagnia aerea che ha performato meglio. EasyJet è la compagnia aerea che ha performato peggio (relativamente alla percentuale di tweet negativi e al valore di compound medio rispetto a tutto il periodo considerato).

Ryanair è la compagnia aerea che, nell’intero periodo considerato, ha registrato la percentuale minore di tweet negativi e il compound medio più alto. Al contrario, EasyJet è la compagnia aerea che, nell’intero periodo considerato, ha registrato la percentuale maggiore di tweet negativi e il compound medio più basso.

Nel caso di Ryanair, si nota una performance migliore rispetto alle altre compagnie considerate anche in termini di numero di cancellazioni di voli nell’estate del 2022 (come mostrato nella ricerca di Mabrian Technologies). Nel caso di EasyJet, invece, nonostante il valore del compound sia il più basso tra le compagnie aeree considerate, la percentuale di voli cancellati non è stata la più alta. Pertanto, le cancellazioni possono essere solo uno

dei fattori che ha contribuito alla generazione di un sentiment negativo da parte degli utenti.

Inoltre, nonostante questa ricerca non avesse come obiettivo quello di comparare la performance dei diversi modelli di business delle compagnie aeree (low cost vs tradizionale), si può osservare che non esiste una chiara tendenza a favore di un modello di business piuttosto che un altro, poiché sia la compagnia aerea che performa meglio che quella che performa peggio sono low-cost.

Infine, si è sottolineato come tutte le compagnie aeree si trovano in difficoltà di fronte ai *trend breakers*, ovvero tutti quegli eventi che interrompono i modelli di servizio consolidati (De Neufville 2017: 62), e solo chi è in grado di adattarsi velocemente al cambiamento può superare le crisi: l'aumento della domanda di voli nell'estate 2022 ha messo in difficoltà tutte le compagnie, ma solo alcune (come nel caso di Ryanair) sono state in grado di adattarsi velocemente e garantire la fornitura dei servizi.

Rispetto ad alcuni studi presenti in letteratura (come quelli di Misopoulous et al. 2014; Kwon et al. 2021; Monika et al. 2019) che hanno applicato la sentiment analysis limitandosi all'individuazione delle opinioni negative e alla valutazione dell'esperienza dei clienti delle compagnie aeree, questa ricerca ha voluto anche indagare come le opinioni espresse su Twitter potessero influire sulla reputazione, offrendo una panoramica sia del contesto competitivo che del contesto specifico in cui opera ogni compagnia aerea.

In particolare, monitorando la reputazione online, si sono provate a colmare alcune delle carenze descritte da Rodriguez-Diaz et al. (2018) nel *gap model*¹⁹⁷: l'analisi ha permesso di rilevare l'intensità della comunicazione aziendale negli account ufficiali delle compagnie aeree (quella che Cuomo et al. 2012 chiamano *brand presence*), di indagare l'opinione dei clienti sulla qualità dei servizi e di fornire una valutazione generale del contesto competitivo.

Ulteriori sviluppi dell'analisi potrebbero consistere nella valutazione del grado di convergenza dell'opinione dei clienti tra le varie piattaforme e nell'integrazione dei risultati ottenuti con altre ricerche che considerano un campione di dati statisticamente

¹⁹⁷ Il *gap model* aiuta a individuare i problemi che un'azienda potrebbe avere nella gestione della reputazione online

rappresentativo di tutti i clienti (ibid.). Inoltre, secondo le indicazioni fornite da Rappaport (2012), se si monitorasse quotidianamente l'opinione online si potrebbero anticipare i problemi e le preoccupazioni future degli stakeholder, coinvolgere i settori aziendali che possono influire sulla relazione con il cliente (in questo caso, ad esempio, si può coinvolgere chi gestisce i call center e i siti web), capire chi sono gli stakeholder che esprimono una determinata opinione e coinvolgerli in progetti creati ad hoc per ristabilire la reputazione. A partire dai dati ottenuti, ad esempio, si potrebbero creare dei progetti di comunicazione riguardanti i temi e le preoccupazioni che interessano di più gli utenti (come la reperibilità delle compagnie aeree, la cancellazione dei voli e le politiche di rimborso).

In conclusione, questa ricerca è stata esemplificativa di come la tecnica *data-driven* della sentiment analysis possa essere impiegata per rilevare le opinioni che gli utenti esprimono online, per poi riflettere sull'impatto che esse hanno sulla reputazione e utilizzarle per indirizzare opportuni interventi correttivi e di miglioramento dei servizi. Tale metodologia può essere estesa a casi studio riguardanti altri settori merceologici, canali di comunicazione o periodi temporali. In ogni caso, la sentiment analysis pone le persone al centro della comunicazione, aiuta a sviluppare una relazione bidirezionale simmetrica con gli utenti, orientata all'ascolto, e permette di individuare eventuali problematiche prima che si amplifichino e provochino una crisi.

BIBLIOGRAFIA

- Aced Toledano, C. (2013). Web 2.0: the origin of the word that has changed the way we understand public relations. In: *Barcelona international PR conference, Barcelona, Spain*.
https://www.researchgate.net/publication/266672416_Web_20_the_origin_of_the_word_that_has_changed_the_way_we_understand_public_relations
- Agarwal, V. (2015). Research on data preprocessing and categorization technique for smartphone review analysis. In: *International Journal of Computer Applications*, 131(4), pp. 30-36. DOI: 10.5120/ijca2015907309
- Agresti, A., Franklin, C. (2016). *Statistica: l'arte e la scienza d'imparare dai dati*. Milano: Pearson.
- Albers, S., & Rundshagen, V. (2020). European airlines' strategic responses to the COVID-19 pandemic (January-May, 2020). In: *Journal of air transport management*, 87, 101863. <https://doi.org/10.1016/j.jairtraman.2020.101863>
- Alton, L. (n.d.). *Creative roundup: 4 brands that take Twitter customer service to the next level*. *Twitter Business*. <https://business.twitter.com/en/blog/creative-roundup-customer-service.html> [ultima consultazione: 08/04/23]
- Armstrong, P. (2010, 20 marzo). Greenpeace, Nestlé in battle over Kit Kat viral. *CNN*. <http://edition.cnn.com/2010/WORLD/asiapcf/03/19/indonesia.rainforests.orangutan.nestle/index.html> [ultima consultazione: 25/04/23]
- Arnoldi, B. (2023, 2 febbraio). Twitter, ora anche le API sono a pagamento: gli sviluppatori storcono il naso. *Everyeye.it*. <https://tech.everyeye.it/notizie/twitter-api-pagamento-sviluppatori-storcono-naso-632832.html> [ultima consultazione: 26/04/23]
- Arokiasamy, L. & Kwaider, S., & Balaraman, R. A. (2019). Best Practices for Crisis Communication: A Qualitative Study. In: *Global Business & Management Research*, 11(2). <https://www.proquest.com/scholarly-journals/best-practices-crisis-communication-qualitative/docview/2236675646/se-2>
- Austin, M & Wittes Schlack, J. (2010). *Leaving our comfort zone: 21st Century Market Research*. C space. <https://www.slideshare.net/wazek/21st-century-market-research-50678392> [ultima consultazione: 25/04/23]

- Aziona Digitale (2021, 5 ottobre). Come funzionano le API, gli agenti alla base del funzionamento delle app e servizi informatici? *Aziona digitale*.
<https://www.azionadigitale.com/api-cosa-sono-e-come-funzionano/> [ultima consultazione: 09/05/23]
- Bagadiya, J. (2023, 7 febbraio). 500+ Social Media Statistics You Must Know in 2023. *Socialpilot*. <https://www.socialpilot.co/blog/social-media-statistics> [ultima consultazione: 25/02/23]
- Barnett, M. L., Jermier, J. M., & Lafferty, B. A. (2006). Corporate reputation: The definitional landscape. In: *Corporate reputation review*, 9, pp. 26-38.
<https://doi.org/10.1057/palgrave.crr.1550012>
- Beaumont, E. (2022, 2 settembre). The best and worst airlines when it comes to speaking to a human (and getting a refund). *The Telegraph*.
<https://www.telegraph.co.uk/travel/advice/best-worst-airlines-customer-service-calls-getting-refunds/> [ultima consultazione: 26/04/23]
- Beers, B. (2022, 28 luglio). How does load factor impact airline profitability? *Investopedia*. <https://www.investopedia.com/ask/answers/041515/how-can-i-use-load-factor-indicator-profitability-airline-industry.asp#:~:text=The%20load%20factor%20is%20a,over%20a%20low%20load%20factor> [ultima consultazione: 20/03/23]
- Bennett, R., & Kottasz, R. (2000). Practitioner perceptions of corporate reputation: An empirical investigation. In: *Corporate Communications: An International Journal*, 5(4), pp. 224-235. <https://doi.org/10.1108/13563280010357349>
- Berberi, L. (2022, 29 maggio). Estate 2022, tanti passeggeri, poco personale: In Europa rischiano di essere cancellati 114 mila voli. *Corriere Della Sera*.
https://www.corriere.it/economia/aziende/22_maggio_29/estate-2022-tanti-passeggeri-poco-personale-europa-rischiano-essere-cancellati-114-mila-voli-ac3e1f12-de99-11ec-aa4f-a6eb11d53611.shtml [ultima consultazione: 02/04/23]
- Beri, M. (2010). *Python. Una cosa accomuna Google, Star Wars e la NASA: Python!* Milano: Apogeo.
- Bertolin, A., Beria, P., Ferrara, E., Debernardi, A., Grimaldi, R., & Laurino, A. (2018). *Atlante dei trasporti italiani. Infrastrutture, offerta, domanda*. Novara: Libreria Geografica.

- Bhatia, S., Li, J., Peng, W., & Sun, T. (2013). Monitoring and analyzing customer feedback through social media platforms for identifying and remedying customer problems. In *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pp. 1147-1154. DOI: 10.1145/2492517.2500287
- Bickerton, J. (2022, 11 luglio). Brexit staff shortages are causing travel chaos, explains aviation expert. *LBC*. <https://www.lbc.co.uk/opinion/brexit-staff-shortages-are-causing-travel-chaos-explains-aviation-expert/> [ultima consultazione: 24/04/23]
- Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with Python: analyzing text with the natural language toolkit*. Sebastopol (CA): O'Reilly Media.
- Bird, S., Klein, E., Loper, E., & Baldridge, J. (2008). Multidisciplinary instruction with the natural language toolkit. In: *Proceedings of the Third Workshop on Issues in Teaching Computational Linguistics*, pp. 62-70. <https://aclanthology.org/W08-0208>
- Bisignani, G. (2009). *L'hub, un modello di organizzazione del trasporto aereo*. Enciclopedia Treccani. https://www.treccani.it/enciclopedia/un-modello-di-organizzazione-del-trasporto-aereo-l-hub_%28XXI-Secolo%29/ [ultima consultazione: 22/03/23]
- Blinov P. D. ,Klekovkina M. V. , Kotelnikov E. V. , & Pestov O. A. (2013). *Research of lexical approach and machine learning methods for sentiment analysis*. Kirov: Vyatka State Humanities University. <https://www.semanticscholar.org/paper/Research-of-lexical-approach-and-machine-learning-Pestov/8b3f7fe44673a82003d7b536ff6430f2ddda633b>
- Blunt, C. & Hill-Wilson, M. (2013). *Delivering effective social customer service: How to redefine the way you manage customer experience and your corporate reputation*. West Sussex (UK): John Wiley & Sons.
- Bodell, L. (2022, 15 luglio). KLM reaches pay agreement with Dutch unions. *Simple Flying*. <https://simpleflying.com/klm-reaches-pay-agreements/> [ultima consultazione: 03/04/23]
- Bolasco, S (2013). *L'analisi automatica dei testi: fare ricerca con il text mining*. Roma: Carocci editore.

- Bonomi, I., Masini, A., Morgana, S., & Piotti, M. (2010). *Elementi di linguistica italiana* (Vol. 103). Roma: Carocci editore.
- Bonta, V., & Janardhan, N. K. N. (2019). A comprehensive study on lexicon-based approaches for sentiment analysis. In: *Asian Journal of Computer Science and Technology*, 8(S2), pp. 1-6. <http://dx.doi.org/10.51983/ajcst-2019.8.S2.2037>
- Boyd, D. M., & Ellison, N. B. (2007). Social network sites: Definition, history, and scholarship. In: *Journal of computer-mediated Communication*, 13(1), pp. 210-230. <https://doi.org/10.1111/j.1083-6101.2007.00393.x>
- Brancale, Federica (2018). *Data-driven marketing. Lasciati guidare dai dati e crea strategie vincenti per il tuo business*. Palermo: Flaccovio Editore.
- Budd, L. (2017). Airlines, information communication technology and social media. In: Budd, L., & Ison, S. (a cura di). *Air transport management: an international perspective*. Londra: Routledge, pp. 265-281.
- Budd, L., & Ison, S. (a cura di.). (2017). *Air transport management: an international perspective*. Londra: Routledge.
- Burchfiel, A. (2022, 16 maggio). What is NLP (Natural Language Processing) Tokenization? *Tokenex*. <https://www.tokenex.com/blog/ab-what-is-nlp-natural-language-processing-tokenization/> [ultima consultazione: 10/03/23]
- Calder, S. (2022, 18 luglio). EasyJet promises 'new initiatives to help customers' as flight cancellations continue. *The Independent*. <https://www.independent.co.uk/travel/news-and-advice/easyjet-call-centre-customers-flights-cancelled-b2125471.html> [ultima consultazione: 26/04/23]
- Carey, J. (2022, 31 maggio). Gatwick Airport: British Airways passengers complain about 'poor customer service' after waiting hours for bags. *Sussex Live*. <https://www.sussexlive.co.uk/news/gatwick-airport-british-airways-passengers-7150753> [ultima consultazione: 20/04/23]
- Cennamo, D., & Fornaro, C. (2017). *Professione Brand Reporter: Brand journalism e nuovo storytelling nell'era digitale*. Milano: Hoepli.
- Chaudhari, A. (2012). *Greenpeace, Nestlé, and the palm oil controversy: Social media driving change?* IBS Center for Management Research. https://search.alexanderstreet.com/view/work/bibliographic_entity%7Cbibliographic_details%7C2778258

- Chen, J. (2022, 7 dicembre). Turbulence: Europe's Summer of Airline Strikes. *Harvard International Review*. <https://hir.harvard.edu/turbulence-europes-summer-of-airline-strikes/> [ultima consultazione: 02/04/23]
- Chiluka, V. (2022, 22 settembre). What is a Python module? How is it different from libraries? *Tutorialspoint*. <https://www.tutorialspoint.com/what-is-a-python-module-how-is-it-different-from-libraries#:~:text=A%20module%20is%20a%20set,community%20members%2C%20developers%20and%20researchers> [ultima consultazione: 10/04/23]
- Chowdhary, K. R. (2020). *Fundamentals of artificial intelligence*. New Delhi: Springer India.
- Chun, R. (2005). Corporate reputation: Meaning and measurement. In: *International journal of management reviews*, 7(2), pp. 91-109. <https://doi.org/10.1111/j.1468-2370.2005.00109.x>
- Citarella, P., Martello, S., Vecchiato, G., & Zicari, S. (2010). *Come comunicare il Terzo Settore. Che la mano sinistra sappia quel che fa la destra: Che la mano sinistra sappia quel che fa la destra*. Milano: FrancoAngeli.
- Colleoni, E., Arvidsson, A., Hansen, L. K., & Marchesini, A. (2011). Measuring corporate reputation using sentiment analysis. In: *Proceedings of the 15th international conference on corporate reputation: navigating the reputation economy*, New Orleans, USA. https://research-api.cbs.dk/ws/files/58852755/Colleoni_Arvidsson_Hansen_Marchesin.pdf
- Collins Dictionary (n.d.). What is the difference between luggage and baggage? *Collins Dictionary*. <https://grammar.collinsdictionary.com/english-usage/what-is-the-difference-between-luggage-and-baggage> [ultima consultazione: 20/04/23]
- Collins Dictionary (s.d). Collins dictionary online. <https://www.collinsdictionary.com/> [ultima consultazione: 05/05/23].
- Conant, S. (2017, 4 gennaio). 5 things to watch in Python in 2017. *O'Reilly*. <https://www.oreilly.com/radar/5-things-to-watch-in-python-in-2017/> [ultima consultazione: 03/03/23]
- Corbetta, P. (2014). *Metodologia e tecniche della ricerca sociale*. Bologna: il Mulino.
- Cortelazzo, M. A. (2013). Metodi qualitativi e quantitativi di analisi dei testi. In: *Contemporanea*, 16(2), pp. 299-310. <http://www.jstor.org/stable/24653899>

- Cover, L. (2021, 17 giugno). 7 statistics that prove the importance of social media marketing in business. *Sproutsocial*. <https://sproutsocial.com/insights/importance-of-social-media-marketing-in-business/> [ultima consultazione: 28/03/23]
- Cullen, R. (2022, 4 agosto). Andy Irvine reunited with missing musical instruments worth over €16,000. *Dublin Live*. <https://www.dublinlive.ie/news/dublin-news/andy-irvine-reunited-missing-musical-24674021> [ultima consultazione: 26/04/23]
- Cuomo, M. T., Metallo, G., & Tortora, D. (2012). *Corporate reputation management: analisi e modelli di misurazione*. Torino: Giappichelli Editore.
- Dardi, F. (2011). *Twitter: I edizione*. Milano: Apogeo Editore.
- Davies, G., Chun, R., da Silva, R. V., & Roper, S. (2004). A corporate character scale to assess employee and customer views of organization reputation. In: *Corporate reputation review*, 7, pp. 125-146. <https://doi.org/10.1057/palgrave.crr.1540216>
- De Neufville R. (2017). Airport systems planning and designs. In: Budd, L., & Ison, S. (a cura di). *Air transport management: an international perspective*. Londra: Routledge, pp. 61-78.
- Deitel, H., & Deitel, P., (2021). *Introduzione a Python. Per l'informatica e la data science*. Milano: Pearson Italia.
- Devika, M. D., Sunitha, C., & Ganesh, A. (2016). Sentiment analysis: a comparative study on different approaches. In: *Procedia Computer Science*, 87, pp. 44-49. <https://doi.org/10.1016/j.procs.2016.05.124>
- Dincer, N. (2020, 21 dicembre). Evolution of Web from 1.0 to 5.0. *Myeltcafe*. <http://myeltcafe.com/articles/evolution-of-web-from-1-0-to-5-0/> [ultima consultazione: 28/01/23]
- Driscoll, M. (n.d.). Jupyter Notebook: An Introduction. *Real Python*. <https://realpython.com/jupyter-notebook-introduction/> [ultima consultazione: 07/03/23]
- Edwards, A. (2021, 16 giugno). An Extensive Guide to collecting tweets from Twitter API v2 for academic research using Python 3. *Towards Data Science*. <https://towardsdatascience.com/an-extensive-guide-to-collecting-tweets-from-twitter-api-v2-for-academic-research-using-python-3-518fcb71df2a> [ultima consultazione: 04/04/23]

- Eurocontrol (2022, 1 gennaio). EUROCONTROL Think Paper #15 – Charting the European Aviation recovery: 2021 COVID-19 impacts and 2022 outlook. *Eurocontrol*. <https://www.eurocontrol.int/press-release/eurocontrol-think-paper-15-charting-european-aviation-recovery-2021-covid-19-impacts> [ultima consultazione: 02/04/23]
- Eurocontrol (2022, 31 dicembre). Analysis Paper: 2022 – The year European aviation bounced back. *Eurocontrol*. <https://www.eurocontrol.int/publication/eurocontrol-analysis-paper-2022-year-european-aviation-bounced-back> [ultima consultazione: 02/04/23]
- Eurostat (2022, 10 agosto). Young air transport workers impacted by COVID crisis. *Eurostat*. <https://ec.europa.eu/eurostat/web/products-eurostat-news/-/ddn-20200525-1> [ultima consultazione: 02/04/23]
- Foiaia, L. G. (2007). *Web 2.0: Guida al nuovo fenomeno della rete*. Milano: Hoepli.
- Fombrun, C. J., & Gardberg, N. A., Ponzi, L. J., (2011). RepTrak™ pulse: Conceptualizing and validating a short-form measure of corporate reputation. In: *Corporate reputation review*, 14, pp. 15-35. <https://doi.org/10.1057/crr.2011.5>
- Fombrun, C. J., Gardberg, N. A., & Sever, J. M. (2000). The Reputation Quotient SM: A multi-stakeholder measure of corporate reputation. In: *Journal of brand management*, 7, pp. 241-255. <https://doi.org/10.1057/bm.2000.10>
- Fombrun, C., & Van Riel, C. (1997). The reputational landscape: a convergence of research and practice. In: *Corporate reputation review*, 1(1 and 2), pp. 1-16. <https://doi.org/10.1057/palgrave.crr.1540008>
- FreeCodeCamp (2022, 28 marzo). Linguaggi di programmazione interpretati e compilati: Qual è la differenza? *Free Code Camp*. <https://www.freecodecamp.org/italian/news/linguaggi-di-programmazione-interpretati-e-compilati-qual-e-la-differenza/#:~:text=Nel%20linguaggio%20compilato%2C%20la%20macchina,legge%20ed%20esegue%20il%20codice> [ultima consultazione: 13/03/2022]
- Frost, R. (2022, 24 agosto). Ryanair and Jet2 revealed as most reliable airlines. How have they avoided delays and cancellations? *Euronews*. <https://www.euronews.com/travel/2022/08/23/how-have-some-of-this-summer->

[most-reliable-airlines-avoided-cancellations-and-delays](#) [ultima consultazione: 18/04/23]

Gaggi, M. (2022, 28 ottobre). Musk perfeziona l'acquisto di Twitter e licenzia subito 4 super-manager. *Corriere Della Sera*.

https://www.corriere.it/esteri/22_ottobre_28/musk-perfezione-l-acquisto-twitter-licenzia-subito-4-super-manager-6c7b2fda-565f-11ed-be15-822086495e48.shtml
[ultima consultazione: 26/02/23]

Gallagher, B. (2022, 12 luglio). I've lost prize possessions worth €16k travelling through 'chaotic' Dublin Airport, musician Andy Irvine says. *The Irish Sun*.

<https://www.thesun.ie/travel/9082992/lost-prize-possessions-thousands-chaotic-dublin-airport/#:~:text=AN%20Irish%20musician%20has%20pleaded,prize%20guitar%20Dbouzouki%20and%20mandola> [ultima consultazione: 26/04/23]

Garzanti Linguistica. (s.d.). Dizionario online Garzanti Linguistica.

<https://www.garzantilinguistica.it/> [ultima consultazione: 05/05/23]

Georgiadis, P., & Chavez, S. (2022, 24 aprile). Back in business: Are airlines ready for a summer travel boom? *Financial Times*. <https://www.ft.com/content/d418f1e5-fbd0-4766-9389-cbfd3d338275> [ultima consultazione: 02/04/23]

Graham A. (2017). Airport management and performance. In: Budd, L., & Ison, S. (a cura di). *Air transport management: an international perspective*. Londra: Routledge, pp. 79-93.

Grunig, J. E., & Hunt, T. T. (1984). *Managing public relations*. New York: Holt, Rinehart and Winston.

Harvey G. & Turnbull P. (2017). Human resource management and industrial relations. In: Budd, L., & Ison, S. (a cura di). *Air transport management: an international perspective*. Londra: Routledge, pp. 307-318.

Hasson, S. G., Piorkowski, J., & McCulloh, I. (2019). Social media as a main source of customer feedback: Alternative to customer satisfaction surveys. In: *Proceedings of the 2019 IEEE/ACM International conference on advances in social networks analysis and mining*, pp. 829-832. <https://doi.org/10.1145/3341161.3345642>

Haynes, T. (2022, 6 luglio). 'I spent £300 replacing lost clothes – then my luggage went missing again on the way home'. *The Telegraph*.

- <https://www.telegraph.co.uk/money/consumer-affairs/husbands-life-saving-medication-lost-weeks-heathrows-baggage/> [ultima consultazione: 21/04/23]
- Hennig-Thurau, T., Gwinner, K. P., Walsh, G., & Gremler, D. D. (2004). Electronic word-of-mouth via consumer-opinion platforms: what motivates consumers to articulate themselves on the internet? In: *Journal of interactive marketing*, 18(1), pp. 38-52. <https://doi.org/10.1002/dir.10073>
- Huete-Alcocer, N. (2017). A literature review of word of mouth and electronic word of mouth: Implications for consumer behavior. In: *Frontiers in psychology*, 8, 1256. DOI: 10.3389/fpsyg.2017.01256
- Hutto, C., & Gilbert, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In: *Proceedings of the international AAAI conference on web and social media*, Vol. 8, No. 1, pp. 216-225. <https://doi.org/10.1609/icwsm.v8i1.14550>
- IATA (2022, 1 marzo). Air Passenger Numbers to Recover in 2024. *IATA*. <https://www.iata.org/en/pressroom/2022-releases/2022-03-01-01/> [ultima consultazione: 02/04/23]
- ICAO (2023). Effects of Novel Coronavirus (COVID-19) on Civil Aviation: Economic Impact Analysis. *ICAO*. https://www.icao.int/sustainability/Documents/COVID-19/ICAO_Coronavirus_Econ_Impact.pdf [ultima consultazione: 03/04/23]
- IFC (2020, 13 maggio). The Impact of COVID-19 on Airports: An Analysis. *International Finance Corporation*. https://www.ifc.org/wps/wcm/connect/industry_ext_content/ifc_external_corporate_site/infrastructure/resources/the+impact+of+covid-19+on+airports [ultima consultazione: 21/03/23]
- Iglesias Moreno, A. (2020, 27 maggio). The Python Standard Library — Modules you should know as a data scientist. *Towards Data Science*. <https://towardsdatascience.com/the-python-standard-library-modules-you-should-know-as-a-data-scientist-47e1117ca6c8> [ultima consultazione: 09/03/23]
- Ingargiola, A. (2015). 1. What is the Jupyter Notebook? *Read The Docs*. <https://jupyter-notebook-beginner-guide.readthedocs.io/en/latest/index.html> [ultima consultazione: 27/03/23]

- Inside Marketing (s.d) Glossario online Inside Marketing.
<https://www.insidemarketing.it/glossario/> [ultima consultazione: 05/05/23]
- Invernizzi, E., & Romenti, S. (2020). *Corporate communication e relazioni pubbliche*. Milano: McGraw-Hill.
- Jackson J. (2017). Airline finance. In: Budd, L., & Ison, S. (a cura di). *Air transport management: an international perspective*. Londra: Routledge, pp. 169-190.
- Jansen, B. J., Zhang, M., Sobel, K., & Chowdury, A. (2009). Twitter power: Tweets as electronic word of mouth. In: *Journal of the American society for information science and technology*, 60(11), pp. 2169-2188. <https://doi.org/10.1002/asi.21149>
- Jones, B., Temperley, J., & Lima, A. (2009). Corporate reputation in the era of Web 2.0: the case of Primark. In: *Journal of marketing management*, 25(9-10), pp. 927-939. <https://doi.org/10.1362/026725709X479309>
- Kane, G. C. (2014, 30 aprile). Reimagining Customer Service at KLM Using Facebook and Twitter. *MIT Sloan Management Review*.
<https://sloanreview.mit.edu/article/reimagining-customer-service-at-klm-using-facebook-and-twitter/> [ultima consultazione: 10/04/23]
- Kaschesky, M., Sobkowicz, P., Hernández-Lobato, J., Bouchard, G., Archambeau, C., Scharioth, N., Manchin, R., Gschwend, A., & Riedl, R. (2013). Bringing Representativeness into Social Media Monitoring and Analysis. In: *46th Hawaii International Conference on System Sciences*, pp. 2003-2012, DOI: 10.1109/HICSS.2013.120
- Kirwin, E. (2022, 27 settembre). Woman sexually assaulted on Ryanair flight from Liverpool to Spain. *Liverpool Echo*.
<https://www.liverpoolecho.co.uk/news/liverpool-news/woman-sexually-assaulted-ryanair-flight-25114311> [ultima consultazione: 23/04/23]
- Kontopoulos, E., Berberidis, C., Dergiades, T., & Bassiliades, N. (2013). Ontology-based sentiment analysis of twitter posts. In: *Expert systems with applications*, 40(10), 4065-4074, <https://doi.org/10.1016/j.eswa.2013.01.001>
- Kotler, P., & Armstrong, G. (2011). *Principles of Marketing*. (14th ed.). New Jersey: Prentice Hall.

- Kumar, R. (2021, 7 maggio). What is Bearer token and how it works? *DevOps School*.
<https://www.devopsschool.com/blog/what-is-bearer-token-and-how-it-works/>
[ultima consultazione: 27/02/23]
- Kwon, H.-J., Ban, H.-J., Jun, J.-K., & Kim, H.-S. (2021). Topic Modeling and Sentiment Analysis of Online Review for Airlines. In: *Information*, 12(2), 78. MDPI AG. <https://dx.doi.org/10.3390/info12020078>
- Lamberti, L. (2013). Customer centricity: the construct and the operational antecedents, In: *Journal of Strategic Marketing*, 21:7, 588-612, DOI: 10.1080/0965254X.2013.817476
- Lampignano, S. P. (2016). *Digital reputation management. Come gestire, promuovere e difendere la propria reputazione online*. Santarcangelo di Romagna (RN): Maggioli Editore.
- Lévy, P. (1996). *L'intelligenza collettiva, Per un'antropologia del cyberspazio*. Milano: Feltrinelli.
- Liddy, E. D. (2001). Natural language processing. In: M. A. Drake (Ed.), *Encyclopedia of library and information science* (2nd ed.). New York: Marcel Decker, Inc.
- Lim, J. C. (2022, 15 luglio). Amid labor shortages, KLM reaches deal with ground staff unions for pay increase. *Aero Time Hub*.
<https://www.aerotime.aero/articles/31597-klm-dutch-airline-reach-agreement-union-ground-staff> [ultima consultazione: 03/04/23]
- Liu, B. (2010). Sentiment analysis and subjectivity. Handbook of natural language processing, In: Indurkha, N., & Damerau, F.J. (Eds.). *Handbook of Natural Language Processing* (2nd ed.). Chapman and Hall/CRC, pp. 627-666.
<https://doi.org/10.1201/9781420085938>
- Liu, B. (2012). Sentiment analysis and opinion mining. In: *Synthesis lectures on human language technologies*, 5(1), 1-167.
<https://doi.org/10.2200/S00416ED1V01Y201204HLT016>
- Live Agent (n.d.). Cos'è il servizio clienti? *Live Agent*. <https://www.live-agent.it/accademia/che-cosa-e-servizio-clienti/> [ultima consultazione: 08/04/23]
- Mabrian Technologies (2022, 22 dicembre). Top 10 European airlines by cancellations for summer 2022. *Mabrian Technologies*. <https://mabrian.com/blog/top-10-european-airlines-cancellations-summer-2022/> [ultima consultazione: 20/04/23]

- Martin, M. (2023, March 13). 29 Twitter Stats That Matter to Marketers in 2023. *Hootsuite*. <https://blog.hootsuite.com/twitter-statistics/> [ultima consultazione: 02/04/23]
- Masini, M., Pasquini, J., & Segreto, G. (2017). *Marketing e comunicazione: Strategie, strumenti, casi pratici*. Milano: Hoepli.
- Matson, M. (2022, 18 dicembre). Word-of-Mouth Marketing: A Complete Guide (2023). *ReferralCandy*. <https://www.referralcandy.com/blog/word-of-mouth-marketing> [ultima consultazione: 20/02/23]
- Matthews, J. (2022, 2 dicembre). Comment: How should airlines rebuild reputation and trust with passengers? *Travel Weekly*. <https://travelweekly.co.uk/in-depth/comment/comment-how-should-airlines-rebuild-reputation-and-trust-with-passengers> [ultima consultazione: 27/04/23]
- McEnery, T & Wilson, A (2001). *Corpus Linguistics. An introduction*. 2nd edition. Edinburgh University Press.
- McIntire, G., Washington, L., & Martin, B. (n.d.). Python Pandas Tutorial: A Complete Introduction for Beginners. *Learn Data Science*. <https://www.learndatasci.com/tutorials/python-pandas-tutorial-complete-introduction-for-beginners/> [ultima consultazione: 09/03/23]
- Melotti, E. (2017, 4 settembre). Funzioni in Python. *HTML.it*. <https://www.html.it/pag/15618/funzioni/> [ultima consultazione: 08/03/23]
- Messina, C. (2007, 23 agosto). *Twitter*. <https://twitter.com/chrismessina/status/223115412> [ultima consultazione: 30/03/23]
- Misopoulos, F., Mitic, M., Kapoulas, A., & Karapiperis, C. (2014). Uncovering customer service experiences with Twitter: the case of airline industry. In: *Management Decision*, 52(4), pp. 705-723. <https://doi.org/10.1108/MD-03-2012-0235>
- Monika, R., Deivalakshmi, S., & Janet, B. (2019). Sentiment analysis of US airlines tweets using LSTM/RNN. In: *2019 IEEE 9th International Conference on Advanced Computing (IACC)* (pp. 92-95). IEEE, Tiruchirappalli, India, DOI: 10.1109/IACC48062.2019.8971592

- Nausheen, F., & Begum, S.H. (2018). Sentiment analysis to predict election results using Python. In: 2nd International Conference on Inventive Systems and Control (ICISC), 1259-1262, DOI:10.1109/ICISC.2018.8399007
- Newton, G. (2014, 22 maggio). The impact of reputation. *Airlines*.
<https://airlines.iata.org/analysis/the-impact-of-reputation> [ultima consultazione: 27/04/23]
- Nissim, M. & Pannitto, L. (2022). *Che cos'è la linguistica computazionale*. Roma: Carocci editore.
- Oldini, R. (2002). L'azione volontaria e il comportamento prosociale: Elementi per una definizione. In: *Studi di sociologia*, pp. 283-302.
<http://www.jstor.org/stable/23004986>
- Olegario, R., & McKenna, C. (2013). Introduction: Corporate reputation in historical perspective. In: *Business History Review*, 87(4), pp. 643-654.
 DOI:10.1017/S0007680513001074
- Oliveira, D. J. S., Bermejo, P. H. D. S., & Dos Santos, P. A. (2017). Can social media reveal the preferences of voters? A comparison between sentiment analysis and traditional opinion polls. In: *Journal of Information Technology & Politics*, 14(1), pp. 34-45, <https://doi.org/10.1080/19331681.2016.1214094>
- Open Education Polito (n.d.) *Impariamo a programmare Python. Libro Interattivo*.
<https://openededucation.polito.it/libro/index.html> [ultima consultazione: 10/04/23]
- Oracle (n.d.). Cos'è il machine learning? Oracle <https://www.oracle.com/it/artificial-intelligence/machine-learning/what-is-machine-learning/> [ultima consultazione: 04/03/23]
- O'Reilly, Tim (2007). What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software. In: *International Journal of Digital Economics*, No. 65, pp. 17-37. <https://mpira.ub.uni-muenchen.de/id/eprint/4580>
- Pak, A., & Paroubek, P. (2010). Twitter as a corpus for sentiment analysis and opinion mining. In: *LREc*, Vol. 10, No. 2010, pp. 1320-1326. http://www.lrec-conf.org/proceedings/lrec2010/pdf/385_Paper.pdf
- Pande, P., & Hardiman, J. (2022, 27 ottobre). What is an airline's load factor and why does it matter? *Simple Flying*. <https://simpleflying.com/load-factor/> [ultima consultazione: 20/03/23]

- Pastore, A., & Vernuccio, M. (2008). *Impresa e comunicazione. Principi e strumenti per il management*. Milano: Apogeo Editore.
- Philip, S. V., Jasper, C., & Wilkes, W. (2022, 1 luglio). How Europe became the epicenter for this summer's travel chaos. *Bloomberg*.
<https://www.bloomberg.com/news/features/2022-07-01/record-flight-delays-cancellations-make-europe-2022-s-worst-place-to-travel> [ultima consultazione: 02/04/23]
- Pianigiani, O. (1907). Vocabolario etimologico della lingua italiana.
<https://www.etimo.it/> [ultima consultazione: 03/05/23]
- Pilot Career News (2022, 11 luglio). Lufthansa boss explains why airlines are having such a tough time. *Pilot Career News*.
<https://www.pilotcareernews.com/lufthansa-boss-explains-why-airlines-are-having-such-a-tough-time/> [ultima consultazione: 23/04/23]
- Poma, L., & Grandoni, G. (2021). *Il Reputation management spiegato semplice*. Torino: Celid.
- Poma, L., & Vecchiato, G. (2012). *La guida del Sole 24 Ore al crisis management*. Milano: Gruppo 24 Ore.
- Powers, B. (2023, 2 febbraio). Charging for Twitter API access risks killing the things that made Twitter great. *Grid News*.
<https://www.grid.news/story/technology/2023/02/02/charging-for-twitter-api-access-risks-killing-the-things-that-made-twitter-great/> [ultima consultazione: 26/02/23]
- Pulcini, A. (2020, January 22). Le 50 'World's Most Admired Companies': Apple davanti a tutte. *Fortune Italia*. <https://www.fortuneita.com/2020/01/22/le-50-worlds-most-admired-companies-apple-davanti-a-tutte/> [ultima consultazione: 25/03/23]
- Rajput, A. (2020). Natural language processing, sentiment analysis, and clinical analytics. In: *Innovation in health informatics*, Academic Press, pp. 79-97.
<https://doi.org/10.1016/B978-0-12-819043-2.00003-4>
- Rappaport, S. (2012). *Listen First: Ascoltate la Rete per migliorare il vostro business*. Milano: Hoepli.

- Rathee, N., Joshi, N., & Kaur, J. (2018). Sentiment analysis using machine learning techniques on Python. In: *2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS)*, Madurai, India, pp. 779-785. IEEE. DOI: 10.1109/ICCONS.2018.8663224
- Reggio, R. (2022, 18 maggio). Intervista a Oliviero Baccelli. Il settore aereo tra crisi e nuovi modelli. *Il Sole 24 ore*. <https://fb.watch/jbnQhh5jMa/> [ultima consultazione: 05/03/23]
- Reuters (2006, 22 settembre). Hasbro's Playskool recalls toy after deaths. *NBC News* <https://www.nbcnews.com/id/wbna14952221> [ultima consultazione: 12/02/23]
- Reuters (2022, 14 luglio). Dutch airline KLM and unions reach agreement. *Reuters*. <https://www.reuters.com/business/aerospace-defense/dutch-airline-klm-unions-reach-agreement-2022-07-14/> [ultima consultazione: 03/04/23]
- Rey-Martí, A., Ribeiro-Soriano, D., & Palacios-Marqués, D. (2016). A bibliometric analysis of social entrepreneurship. In: *Journal of business research*, 69(5), pp. 1651-1655. <https://doi.org/10.1016/j.jbusres.2015.10.033>
- Rodríguez-Díaz, M., Rodríguez-Voltes, C. I., & Rodríguez-Voltes, A. C. (2018). Gap analysis of the online reputation. In: *Sustainability*, 10(5), 1603. <https://doi.org/10.3390/su10051603>
- Ruchko, E. (n.d.). Why Learn Python? Here Are 8 Data-Driven Reasons. *Dan Bader. Improve Your Python Skills*. <https://dbader.org/blog/why-learn-python> [ultima consultazione: 08/03/23]
- Rudy, L. J. (2021, 31 dicembre). Tangible vs. Intangible Assets: What's the Difference? *The Balance*. <https://www.thebalancemoney.com/tangible-vs-intangible-assets-5199067#:~:text=Tangible%20assets%20are%20generally%20anything,up%20around%20your%20company's%20reputation> [ultima consultazione: 16/02/23]
- Russell, S. J., & Norvig, P. (2010). *Artificial intelligence: A modern approach* (3rd ed.). Prentice Hall.
- Sakaki, T., Okazaki, M., & Matsuo, Y. (2010). Earthquake shakes Twitter users: real-time event detection by social sensors. In: *Proceedings of the 19th international conference on World wide web (WWW '10)*. Association for Computing Machinery, New York, NY, USA, pp. 851–860. <https://doi.org/10.1145/1772690.1772777>

- Salesforce (2022, 5 novembre). Che cos'è una Application Programming Interface (API). *Salesforce*. <https://www.salesforce.com/it/blog/2022/09/definizione-application-programming-interface-API.html> [ultima consultazione: 22/02/23]
- SAS (n.d.). Data Science. Che cos'è e perché è importante. *SAS*. https://www.sas.com/it_it/insights/analytics/data-science.html [ultima consultazione: 06/04/23]
- Sbalchiero, S. (2021). *Dal metodo all'esperienza. Fare ricerca con la sociologia comprendente*. Padova University Press.
- Sensini S. (2019). *Analisi del linguaggio con Python*. Milano: Apogeo Editore.
- Sernovitz, A. (2011). *L'arte del Passaparola. Buzz marketing: regole pratiche per far parlare del vostro business*. Milano: Corbaccio.
- Serrano, F., & Kazda, A. (2020). The future of airports post COVID-19. In: *Journal of Air Transport Management*, 89, 101900. <https://doi.org/10.1016/j.jairtraman.2020.101900>
- Seveno, V. (2022, 27 giugno). From KLM to easyJet: How cancelled flights at Schiphol might affect your holiday. *I AM EXPAT*. <https://www.iamexpat.nl/expat-info/dutch-expat-news/klm-easyjet-how-cancelled-flights-schiphol-might-affect-your-holiday> [ultima consultazione: 21/04/23]
- Shafi, A. (2023, marzo). How to Use Jupyter Notebooks: The Ultimate Guide. *DataCamp*. <https://www.datacamp.com/tutorial/tutorial-jupyter-notebook#gs.Z8Pcnv4> [ultima consultazione: 15/04/23]
- Sharma, S.K., Daga, M., & Gemini, B. (2020). Twitter Sentiment Analysis for Brand Reputation of Smart Phone Companies in India. In: Singh, P., Panigrahi, B., Suryadevara, N., Sharma, S., Singh, A. (eds) *Proceedings of ICETIT 2019. Lecture Notes in Electrical Engineering*, vol 605. Springer, Cham. https://doi.org/10.1007/978-3-030-30577-2_75
- Shin, Anys (2006, 23 settembre). 2 Deaths Prompt Toy Recall. *The Washington Post* <https://www.washingtonpost.com/archive/business/2006/09/23/2-deaths-prompt-toy-recall/ab3a6ee8-b6e2-4b60-bf25-19ea49ebaa6d/> [ultima consultazione: 12/02/23]
- Sly, E., & Penza, N. (2022, 27 settembre). Man arrested after being accused of sexual assault on Ryanair flight. *Independent*.

- <https://www.independent.co.uk/travel/news-and-advice/ryanair-flight-sexual-assault-mallorca-liverpool-b2176204.html> [ultima consultazione: 23/04/23]
- Sorbo, A. (2015, 4 novembre). Hashtag: storia di un mito. Chi lo ha inventato e perché? *Ninja Marketing*. <https://www.ninjamarketing.it/2015/11/04/hashtag-storia-di-un-mito-chi-lo-ha-inventato-e-perche/> [ultima consultazione: 16/03/23]
- Speranza, G. (n.d.). Cos'è Python e perché è così popolare? *AuLab*. <https://aulab.it/notizia/300/cose-python-e-perche-e-cosi-popolare> [ultima consultazione: 08/03/23]
- Splendiani, S. (2022). *La comunicazione di crisi ei social media: Principi, modelli teorici e applicazioni pratiche*. Milano: FrancoAngeli.
- Statista Research Department (2022). Leading airlines in Europe in 2021, based on passenger traffic. *Statista*. <https://www.statista.com/statistics/1094759/largest-airlines-in-europe-based-on-passengers/?locale=en> [ultima consultazione: 03/04/23]
- Strano, A. (2021, 24 ottobre). EasyJet apre la sua prima lounge a Londra, è la prima volta che una lowcost offre questo servizio. *The Flight Club*. <https://www.theflightclub.it/2021/10/easyjet-apre-la-sua-prima-lounge-a-londra-e-la-prima-volta-che-una-lowcost-offre-questo-servizio/> [ultima consultazione: 21/03/23]
- Street, F. (2022, 3 agosto). 'Unsustainable and shambolic': Flight attendants speak out on summer travel chaos. *CNN*. <https://edition.cnn.com/travel/article/flight-attendants-summer-travel-chaos-2022/index.html> [ultima consultazione: 02/04/23]
- The Economist (2022, January 22). Where next for air travel? *The Economist* <https://www.economist.com/business/2022/01/22/where-next-for-air-travel> [ultima consultazione: 16/03/23]
- Thelwall, M., Buckley, K., & Paltoglou, G. (2011). Sentiment in Twitter events. In: *Journal of the American Society for Information Science and Technology*, 62(2), pp. 406-418. <https://doi.org/10.1002/asi.21462>
- Thepchalerm, T., & Ho, P. (2021). Impacts of COVID-19 on airline business: An overview. In: *Journal of Business and Economics Review*, 6(1), pp. 81–91. [https://doi.org/10.35609/jber.2021.6.1\(1\)](https://doi.org/10.35609/jber.2021.6.1(1))

- Tibone, F. (2020). *Progettare e programmare con Python. Programmazione orientata agli oggetti, Linguaggi per il web, Database relazionali*. Bologna: Zanichelli
- Tissoni, F., & Rossi, G. (2014). *Social network. Comunicazione e marketing*. Santarcangelo di Romagna (RN): Maggioli Editore.
- Toal, R. (n.d.). 7 Popular Software Programs Written in Python. *Code Institute*. <https://codeinstitute.net/global/blog/7-popular-software-programs-written-in-python/> [ultima consultazione: 08/03/23]
- Treccani (s.d.). Enciclopedia Treccani online. <https://www.treccani.it/enciclopedia/> [ultima consultazione: 05/05/23]
- Treccani. (s.d.). Vocabolario Treccani online. <https://www.treccani.it/vocabolario/> [ultima consultazione: 05/05/23]
- Twitter (2016, 19 settembre). Customer Service on Twitter and the impact on brands. *Twitter Blog*. https://blog.twitter.com/en_gb/a/en-gb/2016/customer-service-on-twitter-and-the-impact-on-brands [ultima consultazione: 23/03/23]
- Udacity Team (2021, 4 gennaio). What is a Python package? *Udacity*. <https://www.udacity.com/blog/2021/01/what-is-a-python-package.html> [ultima consultazione: 09/03/23]
- University of Washington (2023, 2 febbraio). Twitter's API access changes could mark 'end of an era' in academic research on the platform. *Center for an Informed Public*. <https://www.cip.uw.edu/2023/02/02/twitters-api-access-changes-academic-research/> [ultima consultazione: 13/02/23]
- Van den Broek-Altenburg, E. M., & Atherly, A. J. (2019). Using Social Media to Identify Consumers' Sentiments towards Attributes of Health Insurance during Enrollment Season. In: *Applied Sciences*, 9(10), 2035. MDPI AG. <http://dx.doi.org/10.3390/app9102035>
- Vecchiato, G. (2004). *Relazioni pubbliche e comunicazione. Strumenti concettuali. Metodologia. Case history* (Vol. 560). Milano: FrancoAngeli.
- Vecchiato, G. (2015). *Manuale operativo di relazioni pubbliche. Metodologia e case history: Metodologia e case history*. Milano: FrancoAngeli.
- Veh, A., Göbel, M., & Vogel, R. (2019). Corporate reputation in management research: a review of the literature and assessment of the concept. In: *Business research*, 12, pp. 315-353. <https://doi.org/10.1007/s40685-018-0080-4>

- Webster, T. (2010, 28 ottobre). Six steps to a successful social media survey. *Social Media Today*. <https://www.socialmediatoday.com/content/six-steps-successful-social-media-survey> [ultima consultazione: 28/02/23]
- Whyte, R. & Lohmann, G. (2017). Airline business models. In: Budd, L., & Ison, S. (a cura di). *Air transport management: an international perspective*. Londra: Routledge, pp. 107-122.
- Wikipedia (s.d). L'enciclopedia libera. https://it.wikipedia.org/wiki/Pagina_principale [ultima consultazione: 05/05/23]
- Wittmer, A., Bieger, T., & Müller, R. (a cura di.). (2011). *Aviation systems: management of the integrated aviation value chain*. Heidelberg: Springer.
- WTTC (2022, 21 gennaio). WTTC predicts a strong recovery of the European Travel & Tourism sector in 2022. In: *World Travel and Tourism Council*. <https://wttc.org/news-article/wttc-predicts-a-strong-recovery-of-the-european-travel-and-tourism-sector-in-2022> [ultima consultazione: 02/04/23]
- Yakub, M. (2023, 13 aprile). How Many Tweets per Day 2022 (New Data). *Business DIT*. <https://www.businessdit.com/number-of-tweets-per-day/> [ultima consultazione: 20/04/23]
- Zandt, F. (2021, 8 ottobre). The Government Tailwind for Air Travel Recovery. *Statista*. <https://www.statista.com/chart/25930/covid-19-relief-for-airlines-by-region/> [ultima consultazione: 02/04/23]
- Zhang, H., Gan, W., & Jiang, B. (2014). Machine learning and lexicon-based methods for sentiment classification: A survey. In: *2014 11th web information system and application conference*, pp. 262-265. IEEE. DOI: 10.1109/WISA.2014.55
- Zhang, H., Gan, W., & Jiang, B. (2014). Machine learning and lexicon-based methods for sentiment classification: A survey. In: *2014 11th web information system and application conference*, pp. 262-265. IEEE, doi: 10.1109/WISA.2014.55

RINGRAZIAMENTI

*The future belongs to those who believe
in the beauty of their dreams.*

Eleanor Roosevelt

Se sto scrivendo questi ringraziamenti significa che sono giunta alla conclusione del mio percorso di studi magistrali, che mi ha spesso messo a dura prova ma che è stato fondamentale per la mia crescita personale e accademica.

Innanzitutto, desidero ringraziare il mio relatore, il Professor Silvello, che con grande competenza, disponibilità e gentilezza mi ha guidato durante la stesura di questo elaborato. Un ringraziamento va anche al professor Vecchiato, che con il suo corso di Relazioni Pubbliche e i suoi consigli è stato una preziosa fonte di conoscenza.

Ringrazio mia mamma per credere sempre in me, sostenermi in ogni momento e spingermi ad essere propositiva nei confronti della vita. E ringrazio anche mio papà: questo è un altro di quei momenti in cui sento profondamente la tua mancanza, ma se sono arrivata fin qui è stato anche grazie a tutti gli insegnamenti che mi hai donato e che metto in pratica ogni giorno della mia vita.

Ringrazio la nonna Luisa che con una sola telefonata riesce a raddrizzare una giornata storta, Elisa per capirmi senza che ci sia bisogno di troppe parole e tutti i miei zii e cugini per essere una grande fonte di affetto e sicurezza.

Ringrazio Chiara, la mia quasi coinquilina padovana, che ha condiviso con me gran parte dei momenti di questi ultimi anni, sia di gioia che di tristezza. Mi hai insegnato tantissime cose e con le nostre lunghe conversazioni mi hai sempre spronata a riflettere e migliorarmi.

Ringrazio il super team di Uniferpi, in particolare Klair, Alessia, Andrea, Jessica e Davide. Grazie a voi mi sono trovata in un contesto in cui poter crescere libera e senza paura, esprimendo sempre la mia opinione e confrontandomi con dei grandi amici, pronti a farmi sorridere di fronte alle difficoltà.

Ringrazio Asia e Alessandro per aver condiviso con me tantissime risate, avermi insegnato un po' di cultura veneta ed essere stati dei compagni di banco sempre pronti ad aiutarmi e a gustare le prelibatezze della Pio X.

Ringrazio il gruppo di Musical che da diversi anni crea un'atmosfera magica che alimenta la mia passione per la danza e il teatro.

Infine, ringrazio gli amici che, da tanti anni a questa parte, mi sostengono e donano la carica necessaria per affrontare tutte le sfide che mi si presentano. Ringrazio Tala, che sa bene quanto è stato fondamentale in questo periodo. Grazie a Ylenia, Susanna, Federica, Giada, Marta e Sara che dal periodo delle scuole medie e superiori mi sostenete in qualsiasi scelta io possa fare. Grazie Stefani e Debb che da compagni universitari siete ormai diventati amici preziosi che continuano ad accompagnarmi e sostenermi man mano che la vita va avanti.

Avere tutti voi nella mia vita è una grande fortuna e fonte d'ispirazione.

Giulia