



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

Processo Johnny Depp vs Amber Heard:

“sentiment analysis” dei tweet

Relatore
Prof. Gianmaria Silvello

Laureanda
Sofia Radio
Matricola n. 2019229

Anno Accademico 2021-2022

Indice

INTRODUZIONE	4
1. IL PROCESSO JOHNNY DEPP VS AMBER HEARD	6
1.1 LE PRIME ACCUSE E LE CITAZIONI IN GIUDIZIO	6
1.2 IL PROCESSO PER DIFFAMAZIONE IN VIRGINIA	7
1.3 IL PROCESSO COME EVENTO MEDIATICO VIRALE SUI SOCIAL NETWORK	11
2. TWITTER, LA SENTIMENT ANALYSIS E PYTHON	16
2.1 COS'È TWITTER	16
2.2 GLI UTENTI E LE CIFRE RAGGIUNTE DA TWITTER NEL 2021	20
2.3 L'API DI TWITTER	22
2.4 COS'È LA SENTIMENT ANALYSIS	26
2.5 COS'È PYTHON	28
2.4.1 NLTK (<i>Natural Language Processing</i>)	30
2.4.2. NumPy	30
2.4.3 VADER (<i>Valence Aware Dictionary and Sentiment Reasoner</i>)	31
3. METODOLOGIA DELLA TWITTER SENTIMENT ANALYSIS CON PYTHON	32
3.1 COLLEZIONE DEI TWEET	32
3.1.1 <i>Scelta delle variabili d'interesse</i>	33
3.2 PREELABORAZIONE DEI TWEET	35
3.3 SENTIMENT ANALYSIS	37
3.4 ANALISI DEI RISULTATI	38
3.4.1 <i>Analisi dei tweet nel primo arco temporale</i>	38
3.4.2 <i>Analisi dei tweet nel secondo arco temporale</i>	47
3.4.3 <i>Analisi dei teweet nel terzo arco temporale</i>	56
3.4.4 <i>Confronto tra le tre sezioni</i>	63
CONCLUSIONI	73
BIBLIOGRAFIA E SITOGRAFIA	77
RINGRAZIAMENTI	83

Introduzione

Con l'emergere dei nuovi media e delle tecnologie digitali, numerosi servizi di social networking online, ad esempio Twitter, Facebook, TikTok e YouTube, sono sempre più popolari ed apprezzati dagli utenti, perché permettono loro di esprimersi liberamente su qualsiasi argomento costruendo e condividendo un mondo globale¹. Questo è possibile poiché i social network consentono uno scambio di comunicazione da molti a molti, in cui gli interlocutori sono caratterizzati da background differenti. Per tali ragioni, essi sono diventati fonti utilissime per le aziende per monitorare, attraverso la *sentiment analysis* sui commenti pubblicati online, l'opinione dei loro clienti sui prodotti o servizi che offrono, potendoli così migliorare e possibilmente modificare la loro strategia di marketing.

La *sentiment analysis* dei dati di Twitter sta guadagnando ogni giorno sempre più importanza e attenzione nel campo della ricerca. La piattaforma infatti, con l'aiuto dell'elaborazione del linguaggio naturale (NLP) e dell'intelligenza artificiale, permette di ricavare dati attraverso le sue API ottenendo così informazioni sull'opinione pubblica attraverso lo studio dei tweet e classificando la loro polarità in diverse classi di sentiment (positiva, negativa o neutra) automaticamente. Questa opportunità attrae sempre più studiosi grazie al fatto che essi sono contenuti pubblici visibili anche agli utenti non registrati alla piattaforma, sono brevi (massimo 280 caratteri), spesso evidenziati da parole chiave, gli hashtag, ed espressi da un ampio bacino eterogeneo di utenti.

In questo elaborato, utilizzando il linguaggio di Python e il *Valence Aware Dictionary for Sentiment Reasoner* (VADER), si vuole comprendere il modo in cui gli utenti di Twitter si sono espressi in merito ai processi per diffamazione che vedevano Johnny Depp e Amber Heard come protagonisti in tre precisi archi temporali:

- dal 1° dicembre 2018 al 31 dicembre 2019 (dal primo processo per diffamazione avuto luogo in tribunale a Londra)
- dal 1° aprile 2022 al 10 giugno 2022 (durata del processo per diffamazione in Virginia)

¹ Marshall McLuhan, *Understanding Media: The Extensions of Man*, 2001 (1964)

- dall'11 giugno 2022 all'11 settembre 2022 (tre mesi dopo la fine del processo in Virginia).

In aggiunta, oltre a verificare se gli utenti si esprimano più a favore o meno per una o l'altra parte in causa, si intende studiare anche se e come questi nel corso dei mesi e degli avvenimenti abbiano modificato il modo di scrivere ed esprimersi in merito a temi come la violenza domestica e gli abusi ricevuti dai propri partner, quindi anche come differiscono semanticamente nella loro comunicazione online.

Per riuscire nello scopo della ricerca, ci si è avvalsi delle documentazioni di Python e dei suoi vari pacchetti per costruirvi il codice corretto, e in più di DataCamp, piattaforma di apprendimento online che offre videolezioni, corsi e successivi esercizi per mettere in pratica e padroneggiare quanto appreso.

Questo studio è strutturato come segue: il primo capitolo definisce una visione generale su Johnny Depp, Amber Heard e il loro travagliato matrimonio, descrivendo in aggiunta i vari processi giuridici per diffamazione. Un ulteriore approfondimento sarà eseguito in merito al successo e la viralità online ottenuta dalle vicende nei vari social network.

Il secondo capitolo intende osservare e descrivere dettagliatamente il social network usato per compiere le successive analisi, ovvero Twitter, cosa si intende per *sentiment analysis*, Python, i vari pacchetti utilizzati per costruire il codice e VADER per compiere la *sentiment analysis*.

Infine, nel terzo capitolo saranno effettuate le varie fasi per la pulizia e le analisi dei tweet scaricati in modo da ricavarne le polarità e poterli confrontare in base alle query (combinazioni di hashtag che i tweet devono contenere) scelte e ai periodi temporali a cui appartengono.

1. Il processo Johnny Depp vs Amber Heard

1.1 Le prime accuse e le citazioni in giudizio

Johnny Depp (58 anni) e Amber Heard (36 anni), attori di fama internazionale, si sono incontrati per la prima volta nel 2008 sul set del film “*The Rum Diary – Cronache di una passione*” (film uscito nelle sale nel 2011), luogo nel quale ha avuto inizio la loro relazione. Questa è stata tenuta segreta per un lungo periodo, fino a quando i due attori hanno fatto la loro prima apparizione pubblica in coppia al settimo Heaven Gala nel 2014, riferendo alla stampa di essersi fidanzati ufficialmente. Dopo un breve periodo, il 5 febbraio 2015, i due attori si sono sposati e hanno trascorso la loro luna di miele sull’isola privata di Depp alle Bahamas. Dopo solamente quindici mesi dal matrimonio si sono separati e, il 23 maggio 2016, Amber Heard ha chiesto il divorzio. L’attrice ha giustificato l’atto per “differenze inconciliabili” con Depp e ha richiesto al giudice un’ordinanza restrittiva momentanea contro l’attore hollywoodiano con l’accusa che egli avesse abusato di lei. Infatti, la Heard si è recata in Tribunale a Los Angeles con un livido al viso, asserendo che fosse stato perpetrato dall’ex marito dopo averla colpita con il cellulare. Pochi giorni prima che si tenesse l’udienza, a metà agosto del 2016, Depp e la Heard hanno trovato un accordo di divorzio e l’attrice ha ritirato le sue accuse di violenza e la richiesta di ordinanza restrittiva con la clausola che non avrebbe più potuto ripresentarle, ottenendo 7 milioni di dollari dall’ex marito. Insieme, i due attori, hanno sostenuto che la loro relazione è comunque stata “intensamente passionale e a tratti volatile, ma sempre piena di amore” e che nessuna delle due parti ha rivolto all’altra false accuse a scopi economici. La pace però ha avuto durata breve: nel 2018, le due star sono tornate in tribunale, a Londra, siccome Depp ha denunciato per diffamazione la società che possiede il giornale *The Sun*, la *News Group Newspapers*, dopo la pubblicazione di un articolo in cui veniva definito un “wife beater” cioè “picchiatore di mogli”. La legge britannica prevede che chi è accusato di diffamazione, in questo caso il *The Sun*, debba dimostrare la veridicità dei fatti citati, così la Heard è stata chiamata in tribunale a testimoniare, raccontando nuovi episodi di aggressione fisica e verbale. A luglio 2020 il processo a Londra si è concluso con la vittoria del *The Sun*. Amber Heard, per dimostrare che Johnny Depp fosse “un uomo spesso fuori controllo” ha portato come prove in tribunale alcune foto in cui si vedeva l’attore a terra privo di sensi, ubriaco e con della

droga accanto. Egli comunque non ha mai smentito che abusasse di droghe ma allo stesso tempo ha giurato anche che non fosse un violento con le donne. La sentenza del processo aveva sancito il declino di Depp, ma egli non si è arreso e nel 2019 ha fatto causa all'ex coniuge per 50 milioni di dollari asserendo di averlo diffamato in un articolo del *Washington Post*, intitolato infatti *“Ho parlato contro la violenza sessuale e ho affrontato l'ira della nostra cultura. Questo deve cambiare”*. Nonostante la Heard non avesse mai citato direttamente il nome dell'ex marito, per molte persone è stato chiaro che si stesse riferendo proprio a lui. L'attrice, infatti, ha sostenuto che:

«Due anni fa sono diventata una figura pubblica che rappresentava la violenza domestica, e ho sentito tutta la forza della collera della società nei confronti delle donne che non restano in silenzio e parlano.²»

In seguito alla pubblicazione dell'editoriale, Johnny Depp è stato licenziato dalla collaborazione con la Disney per la saga di *“Pirati dei Caraibi”* e rimosso dalla Warner Bros dal cast anche di *“Animali Fantastici”* affrontando, così, vari problemi finanziari. Anche la casa di produzione americana MGM, che aveva comprato i diritti dell'ultimo film *“Minamata”* dell'attore, ha deciso di non distribuirlo negli Stati Uniti. Per queste ragioni, nel marzo 2019, il team legale di Johnny Depp ha accusato l'attrice per diffamazione per un valore di 50 milioni di dollari, sostenendo che quelle affermazioni fossero *“categoricamente false”*. Il team ha dichiarato inoltre di possedere numerosi testimoni oculari, filmati, registrazioni e fotografie che avrebbero dimostrato che l'attrice avesse mentito per quanto riguardava gli abusi e che, contrariamente, la vittima delle violenze fosse proprio Depp.

1.2 Il processo per diffamazione in Virginia

In seguito a migliaia di articoli giornalistici in merito alla relazione dei due attori, l'11 aprile 2022 ha preso luogo, presso il tribunale di Fairfax in Virginia, il processo per le accuse di diffamazione rivolte da Johnny Depp all'ormai ex moglie Amber Heard con una richiesta di risarcimento di 50 milioni di dollari. Da molti questo processo è stato

² https://www.washingtonpost.com/opinions/ive-seen-how-institutions-protect-men-accused-of-abuse-heres-what-we-can-do/2018/12/18/71fd876a-02ed-11e9-b5df-5d3874f1ac36_story.html

definito come “uno dei più cruenti di Hollywood”³, ricco di retroscene di un matrimonio violento.

Durante le prime tre settimane del processo, trasmesso in diretta in vari canali YouTube, il team legale di Depp ha citato come testimoni circa una dozzina di soggetti e l'attore ha reso spontanee dichiarazioni sostenendo che il suo ipotetico capo di imputazione, oltretutto aver perpetrato violenza domestica nei confronti dell'ex coniuge, era errato in quanto non sussistevano i presupposti per la configurazione del reato in questione, come è risultato anche dalla decisione finale della giuria, sottolineando il fatto che tale questione ha inciso molto sulla sua carriera. Fin dall'inizio del processo, Depp ha dichiarato infatti che:

«Mi ha ucciso il fatto che tutte queste persone che avevo incontrato nel corso degli anni hanno pensato che fossi un imbroglione. È strano quando un giorno sei Cenerentola, per così dire, e poi in 0.6 secondi sei Quasimodo. Non me lo meritavo. Sono orgoglioso dell'onestà e della verità. La verità è l'unica cosa che mi interessa, le bugie non ti porteranno da nessuna parte. [...] Indipendentemente dall'esito di questo processo, ho già perso. Ho perso quando sono state fatte queste accuse, perché rimarranno con me per sempre. La mia vita è rovinata per sempre.⁴»

Depp ha anche replicato che l'ex moglie avesse fatto la performance della sua vita al processo, rievocando in aula le violenze che avrebbe subito dall'ex marito durante la loro relazione, però commettendo alcuni errori. Infatti, il suo racconto ha continuato a crescere con nuovi particolari, *«mentre i ricordi di Depp sono rimasti esattamente gli stessi nei sei dolorosi anni da quando sono state fatte le prime accuse. [...] La sua verità, che è la verità, è rimasta la stessa a prescindere dal contesto in cui è stata presentata⁵»*, hanno sostenuto i legali dell'attore.

Depp e la Heard hanno portato in tribunale dichiarazioni capaci di distruggere la reputazione di entrambi, descrivendo la loro relazione come un matrimonio da incubo. Alcune prove che la Heard ha portato sono state:

³ <https://www.vanityfair.it/article/quanti-danni-ha-fatto-davvero-editoriale-di-amber-heard-alla-carriera-di-johnny-depp-sentenza>

⁴ <https://www.stylise.it/in-evidenza/bozza-automatica-dichiarazione-di-johnny-depp-durante-il-processo/#:~:text=Mi%20ha%20ucciso%20il%20fatto,dell'onest%C3%A0%20e%20della%20verit%C3%A0>

⁵ <https://tg24.sky.it/spettacolo/cinema/approfondimenti/johnny-depp-amber-heard-storia#18>

- messaggi da Johnny Depp in cui vi era scritto che voleva “bruciarla come una strega” dopo averla accusata di tradirlo con James Franco;
- un’aggressione avvenuta sul loro jet privato sempre per il medesimo motivo;
- un’ulteriore aggressione da parte di lui perché l’attrice avrebbe riso di un suo nuovo tatuaggio;
- la minaccia di Depp che avrebbe detto di voler bruciare in microonde il cane di lei.

Dal lato suo, Johnny Depp ha dichiarato che:

- le minacce citate precedentemente fossero uno scherzo;
- la Heard avesse defecato nel loro letto. Una sera, infatti, Depp arrivò tardi al trentesimo compleanno di lei (nel 2016) provocando così un litigio. La mattina Depp trovò nella sua parte del letto delle feci e la Heard diede la colpa ai loro cani (in seguito durante il processo, l’attrice confermò che fossero le sue feci);
- *“Non sono mai arrivato al punto di colpire la signora Heard in alcun modo né ho mai colpito alcuna donna in vita mia”*⁶;
- che la persona violenta fosse la Heard, portando come prova un audio del 2015 registrato durante una terapia di coppia in cui l’ex moglie ha ammesso di aver lanciato contro Depp pentole e vasi. L’attore ha portato in tribunale anche delle foto del suo dito mozzato dall’ex moglie con i cocci di una bottiglia di vodka in Australia con tanto di audio in cui Depp urlava «Non trovo la punta del dito, non trovo la punta del dito⁷». L’attore l’ha definito come il momento in cui si è sentito più vicino ad un crollo psicologico⁸.

Una testimonianza cruciale è stata quella di Kate Moss, top model con cui Depp ha avuto una relazione negli anni Novanta, in apertura della ventiduesima giornata del processo. La Heard a tal proposito aveva insinuato che l’ex marito avesse spinto la super modella e l’avesse fatta cadere dalle scale. Le sue accuse però sono state smontate in tre minuti in

⁶ <https://inews.co.uk/inews-lifestyle/people/when-amber-heards-testimony-continue-date-johnny-depp-trial-restart-1615032>

⁷ <https://www.vanityfair.it/people/mondo/2020/04/09/johnny-depp-dito-mozzato-amber-heard-audio-divorzio-gossip>

⁸ <https://www.ilpost.it/2022/04/28/processo-depp-heard/>

un collegamento video in diretta dal Regno Unito dalla stessa Kate Moss che ha negato il fatto sotto giuramento.

Il primo giugno 2022 si è concluso il processo con il verdetto della giuria, dopo sei settimane e 13 ore di deliberazioni in camera di consiglio: Amber Heard ha diffamato Johnny Depp e dovrà risarcirlo con 10 milioni di dollari. Anche l'attore dovrà risarcire la stessa per un importo pari a 2 milioni di dollari poiché inizialmente il suo Avvocato ha definito un "imbroglio" le accuse di lei. I due attori hanno parlato:

Depp ha detto:

«La giuria mi ha ridato la vita»

mentre Amber Heard ha sostenuto che:

«La delusione che provo oggi supera qualsiasi parola. [...] Avevo portato una montagna di prove, ma non sono state abbastanza per resistere allo sproporzionato potere ed influenza del mio ex marito⁹. [...] Il verdetto rappresenta un ritorno all'epoca in cui una donna che avesse osato parlare contro la violenza domestica veniva pubblicamente umiliata.»

Nel 2006 è stato fondato dalla sopravvissuta e attivista Tarana Burke, il movimento "Me Too", con lo scopo di apportare e rendere disponibili risorse, supporto e percorsi di guarigione conseguenti alla violenza sessuale, per interromperla ovunque essa avvenga. Nel 2017 l'hashtag¹⁰ #MeToo è diventato virale nei social network¹¹, sviluppandosi come un movimento globale e rendendo la violenza sessuale evidente e consapevole agli occhi di tutti. Molte persone, infatti, hanno cominciato a pubblicare post con questo hashtag, evidenziandosi davanti a tutti come dei sopravvissuti alla ricerca di aiuto. Le vittime più frequenti che si appoggiano a Me Too sono giovani, queer, trans, disabili, donne e ragazze

⁹ <https://tg24.sky.it/spettacolo/cinema/approfondimenti/johnny-depp-amber-heard-storia#24>

¹⁰ L'hashtag è un'etichetta utilizzata su alcuni servizi web e social network come aggregatore tematico per rendere più semplice agli utenti trovare i contenuti pubblicati su un tema specifico.

¹¹ Un social network è una rete sociale fatta da individui connessi tra loro da dei legami sociali, quindi tendenzialmente da rapporti lavorativi, da legami familiari, da interessi comuni o anche da conoscenza casuale.

nere e tutte le comunità di colore. Un ulteriore scopo del movimento è quello di distruggere i Sistemi che consentono alla violenza sessuale di profilarsi nel mondo, attuando strategie per sostenere il cambiamento sistemico a lungo termine, in modo che un giorno nessuno debba più dire “me too” (“anche io”).

Amber Heard cita proprio il movimento MeToo, di come questo insegni le modalità in cui funziona il potere degli uomini per cui le donne si confrontano in tutte le istituzioni come luoghi di lavoro, di culto e comunità. L’attrice, ambasciatrice dei diritti delle donne presso l’American Civil Liberties Union, aggiunge anche che, grazie a questo processo, ha avuto ancora una volta la prova di come le Istituzioni proteggano gli uomini accusati di abusi, sostenuti dal potere sociale, economico e culturale.

1.3 Il processo come evento mediatico virale sui social network

Il processo Johnny Depp vs Amber Heard si è trasformato in un vero e proprio show televisivo. Da quando il giudice Penney Azcarate ha permesso a Court TV di mandare in streaming live i video delle due telecamere presenti in aula, ha amplificato l’effetto mediatico del caso in esame, attirando grandi attenzioni sui tabloid e sui social network per la notorietà dei protagonisti e la presenza di deposizioni ricche di storie e avvenimenti intimi e violenti. Alcuni video hanno cominciato anche a circolare in rete, in particolare su YouTube, Instagram, Twitter e TikTok, nei quali i commenti pubblicati dagli utenti erano maggiormente pro-Johnny Depp. Per esempio, nel mese di aprile su TikTok, l’hashtag #justiceforjohnnydepp aveva oltre 18 miliardi di visualizzazioni, mentre l’hashtag #justiceforamberheard ne aveva solamente 39 milioni. Nello stesso mese vi erano anche già hashtag che esprimevano odio verso l’attrice come: #amberheardtrash, #amberheardcancelled, #amberheardsucks e #amberheardbitch anch’essi con milioni di visualizzazioni¹². In questi mesi è stato facile imbattersi in video pubblicati dagli utenti che ripropongono momenti precisi delle disposizioni in tribunale, editati in modo da suscitare empatia e simpatia nei confronti dell’attore, ridicolizzando e umiliando la Heard. Come ha scritto Amanda Hess, critica di Internet e della cultura pop per il New York Times:

¹² <https://www.elle.com/it/magazine/women-in-society/a40086948/metoo-processo-depp-amber-heard/>

«Io non ho seguito il processo di diffamazione tra Johnny Depp e Amber Heard, è stato lui a seguire me»¹³.

L'interesse nei confronti della battaglia legale ha guadagnato seguito settimana dopo settimana e infatti, secondo il magazine Usa Today, l'ultima testimonianza di Johnny Depp è stata seguita da oltre un milione e duecento spettatori, superando il doppio dei numeri della prima avuta luogo ad aprile. Permettere al pubblico di assistere alle disposizioni, conoscere ogni particolare e visionare le prove, ha reso il processo a tutti gli effetti uno show televisivo. In aggiunta, tra meme¹⁴ che riprendono le risate, le espressioni sarcastiche, le risposte sagaci di Depp e i post pubblicati e rilanciati centinaia di volte sotto forma di clip online dagli utenti nei vari canali social, il processo ha generato più interazioni online rispetto a notizie più importanti ed urgenti negli Stati Uniti come, ad esempio, alla decisione della Corte Suprema sull'aborto negli USA, all'inflazione e alla guerra della Russia in Ucraina. Secondo la dottoressa Pamela Rutledge, psicologa dei media e scienziata sociale che applica la sua esperienza nel comportamento umano e nelle neuroscienze¹⁵, Depp è un "facile oggetto di memeficazione" grazie al collegamento che le persone vi fanno con i personaggi iconici da esso interpretati durante la sua carriera cinematografica. La dottoressa, infatti, aggiunge che:

«Nonostante si stia discutendo di questioni di abusi domestici molto serie, la persona pubblica di Depp, i suoi ruoli ben noti e la sua reputazione, conferiscono agli atti una qualità surreale che contribuisce ai contenuti fancam¹⁶»¹⁷.

Secondo Google Trends, il numero di ricerche nel mese di maggio su Google per Amber Heard è stato il doppio rispetto a quelle per Elon Musk e quattro volte di più rispetto a quelle sull'aborto. Anche il canale YouTube di Law&Crime ha visto numeri da record: 50 volte di più rispetto a prima del processo, quindi circa un milione di spettatori in più

¹³ <https://www.nytimes.com/2022/05/26/arts/amber-heard-tiktok-johnny-depp.html>

¹⁴ Un meme è un singolo elemento di una cultura o di un sistema di comportamento, replicabile e trasmissibile per imitazione da un individuo a un altro o da uno strumento di comunicazione ed espressione a un altro (giornale, libro, pellicola cinematografica, sito internet, ecc...).

¹⁵ <https://www.pamelarutledge.com/dr-pamela-rutledge/>

¹⁶ Una *fancam* è una ripresa/filmato di una celebrità da parte di un suo fan

¹⁷ <https://www.vice.com/it/article/wxdmey/processo-depp-heard-meme>

ogni giorno. Infine, secondo i dati di SmilarWeb¹⁸, anche i siti come *People*, *Us* e *New York Post* hanno registrato nel mese di aprile un aumento del traffico rispettivamente del 9%, del 16% e del 22%¹⁹.

I commenti sotto i video del processo in diretta su YouTube sono stati molto critici e svilenti nei confronti di Amber Heard e della sua credibilità, definendola la “regina del dramma con lacrime di cocodrillo” ed evidenziando le sue ottime doti di attrice narcisista in tribunale.

I video in cui piange durante l’interrogatorio sono stati utilizzati su TikTok per ridicolizzarla e per ritenere che fosse tutta una finzione con lo scopo di guadagnare soldi e visibilità. Sul social network Snapchat²⁰ è stato creato inoltre un filtro che deforma il volto degli utenti in una smorfia di pianto ridicola, soprannominato come il “filtro di Amber Heard”. Ma è su TikTok che questo filtro sta riscuotendo più successo: ad oggi esso conta già più di 117 milioni di visualizzazioni²¹.



¹⁸ SimilarWeb è una società informatica che compie analisi di dati e ricerche per business intelligence e aziende internazionali.

¹⁹ <https://www.rainews.it/articoli/2022/05/processo-depp-heard-in-attesa-del-verdetto-a429d3d5-6186-47c0-8d34-e60c2c3d1103.html>

²⁰ Snapchat, lanciata nel 2011, è un’app di messaggistica per cellulari e tablet. Permette di condividere foto e brevi filmati che scompaiono, distruggendosi automaticamente, pochi secondi dopo essere stati ricevuti dal destinatario.

²¹ <https://vm.tiktok.com/ZMN81aUYC/>

Secondo un'esperta di movimenti d'odio online, Hilde Van den Bulck, la comunità di utenti che produce contenuti contro la Heard si è formata negli anni in cui gli attori si stavano separando, fondandosi su sentimenti negativi. È come se i follower della vicenda avessero compiuto un vero e proprio lavoro di investigazione e analisi di tutti i contenuti che hanno trovato online, per partecipare attivamente nelle varie comunità aumentando il loro senso di appartenenza e di identità comune. Un altro motivo per cui meno persone sono a favore dell'attrice è perché molto frequentemente alcune donne denunciano i propri ex partner famosi per ottenere soldi e fama mediatica. Infine, altre persone provano questa antipatia verso di lei, screditandola come una donna che mente e finge, perché sostengono sinceramente il movimento MeToo e tutte le vere vittime donne di violenza, ma anche estendendosi a uomini vittime come in questo caso Depp. Perciò i sostenitori di Depp ritengono che la sua vittoria non rappresenti la fine di MeToo, ma una sua espansione²², una demolizione dello stigma contro gli uomini che si identificano come vittime di abusi.

Per altri commentatori, la vittoria di Johnny Depp al processo è ritenuta come un fatto pericoloso per le vittime di abusi e un passo indietro per i movimenti femministi. Essi, infatti, ritengono che sia un'ulteriore vittoria del Sistema che protegge gli uomini violenti e declassa le donne che li denunciano (movimento MeToo), spingendole conseguentemente a non prenderne più parola e a non testimoniare le loro esperienze, come se avesse un effetto paralizzante. Una sopravvissuta alla violenza domestica, infatti, ha dichiarato che:

«Questo caso è la mia peggiore paura che si manifesta su un palcoscenico pubblico. [...] Mi dice che il mio ex aveva ragione. Se avesse deciso di farlo, avrebbe potuto distruggermi e umiliarmi in modo irreparabile. [...] Mi sento davvero stupida a pensare che avrei potuto vincere il mio caso. Gli uomini vincono sempre²³».

²² <https://www.vox.com/culture/23150632/johnny-depp-amber-heard-trial-verdict-me-too-backlash>

²³ <https://www.rollingstone.com/culture/culture-news/amber-heard-johnny-depp-verdict-metoo-trial-1361356/>

La dottoressa Jessica Taylor, psicologa e autrice di libri sulla misoginia e sugli abusi, dice su *Rolling Stone*²⁴ che «*Questa è fondamentalmente la fine di MeToo. [...] È la morte dell'intero movimento*», perché dimostra quanto una vittima possa essere ancora incolpata ed isolata». «*Un vero e proprio contraccolpo al movimento MeToo, in quanto Depp è sempre stato più potente, più ricco e più apprezzato rispetto all'ex coniuge*», sostiene Constance Grady su *Vox*²⁵.

Il movimento MeToo ha rilasciato una dichiarazione a sostegno dei sopravvissuti in generali, senza allinearsi né con Depp né con la Heard esplicitamente. Sul suo profilo Twitter scrive infatti:

«Quello che abbiamo vissuto nel processo Depp-Heard è un caso studio su come i movimenti sociali e politici vengono usati in modo improprio e strumentalizzati contro le stesse persone a cui dovrebbero servire»²⁶.

²⁴ <https://www.rollingstone.com/culture/culture-news/amber-heard-johnny-depp-verdict-metoo-trial-1361356/>

²⁵ <https://www.vox.com/culture/23150632/johnny-depp-amber-heard-trial-verdict-me-too-backlash>

²⁶ <https://twitter.com/MeTooMVMt/status/1530567171998367745>

2. Twitter, la Sentiment Analysis e Python

Twitter è diventato molto popolare perché permette alle persone di esprimere online e condividere in tempo reale le proprie opinioni riguardo qualsiasi argomento. Questo consente a ricercatori e aziende di ricavare e classificare informazioni da tutto il mondo sull'opinione pubblica attraverso l'analisi dei tweet e classificando la polarità del loro "sentiment".

Oggi giorno, molti strumenti offrono dei mezzi per ricavare funzionalità avanzate dai testi. In questo studio, infatti, grazie all'uso di Python, dei suoi pacchetti e librerie, e del Valence Aware Dictionary and Sentiment Reasoner (VADER), si è riuscito ad estrarre i tweet necessari e studiarli in base agli scopi della ricerca.

La domanda di ricerca di questa tesi sussiste nel comprendere come gli utenti abbiano comunicato e interagito all'interno della piattaforma relativamente ai fatti accaduti durante i processi Johnny Depp vs Amber Heard. Oltre a studiare la polarità dei tweet nei tre periodi in esame (prime accuse nel 2019, processo in Virginia nel 2022 e periodo subito dopo la fine del processo del 2022), interessante sarà anche capire se i termini utilizzati dagli utenti nei tweet sono cambiati e se anche la preferenza per uno o per l'altro attore hollywoodiano si è invertita dopo l'emergere nei tribunali di vari episodi. Inoltre, due ulteriori analisi con i medesimi scopi sono state effettuate in base a degli avvenimenti precisi che caratterizzano le discordie dei due attori, la violenza domestica e il ritrovamento di feci appartenenti a Heard nel letto di Depp.

Prima di analizzare le varie procedure effettuate per ottenere delle risposte, è fondamentale osservare cosa siano Twitter, Python ed infine cosa si intenda per *Sentiment Analysis*.

2.1 Cos'è Twitter

Twitter è una piattaforma, un servizio di notizie e microblogging gratuito che permette di comunicare aggiornamenti e notizie, esprimere opinioni e mostrare i propri sentimenti in diverse occasioni attraverso messaggi brevi, foto, GIF²⁷ animate, sondaggi, link e video,

²⁷ GIF, *Graphics Interchange Format*, è un formato di file per immagini digitali utilizzato nella grafica digitale e inventato il 15 giugno 1987.

pubblicabili da pc, smartphone e tablet. Il servizio è fornito dalla società Twitter, Inc. con sede a San Francisco negli Stati Uniti ed è stato creato da Jack Dorsey nel marzo 2006, guadagnando subito popolarità in tutto il mondo: ad oggi si stima abbia più di 330 milioni di utenti e che questi generino circa 500 milioni di tweet al giorno²⁸. Attraverso un pubblico così ampio, Twitter attira quotidianamente gli utenti a trasmettere le proprie opinioni e punti di vista su qualsiasi tematica, creando di conseguenza materiali informativi per molte aziende, organizzazioni ed istituzioni.

Twitter è anche definito come “l’SMS di Internet”. Grazie alla rete Internet, infatti, gli utenti possono pubblicare messaggi di breve lunghezza (massimo 280 caratteri), chiamati “tweet” che vengono visualizzati nel loro profilo personale e nel feed²⁹ principale. Ogni utente può iscriversi ai tweet di altri divenendo dei veri e propri follower. Traducendo in italiano “tweet” significa “cinguettii” da qui il logo del servizio che rappresenta un uccellino.

Il primo tweet pubblicato è stato proprio da parte del creatore della piattaforma alle 12:50 UTC-8 del 21 marzo 2006 e recitava: “Just setting up my Twttr” ovvero “sto semplicemente impostando il mio Twitter”.



Dall’11 dicembre 2009 le lingue attive in Twitter sono il francese, l’italiano, il tedesco e lo spagnolo insieme all’inglese e al giapponese che erano presenti fin dalla nascita della piattaforma.

²⁸ <https://www.factslices.com/s-Twitter>

²⁹ Un *feed* è una sequenza di contenuti personalizzata per l’utente che gli vengono proposti secondo i suoi interessi e gusti grazie ad un algoritmo all’interno della home page dell’applicazione.

Il 22 gennaio 2010 è stato il giorno del primo tweet pubblicato dallo spazio dall'astronauta della NASA Timothy Creamer dalla Stazione Spaziale Internazionale (ISS).

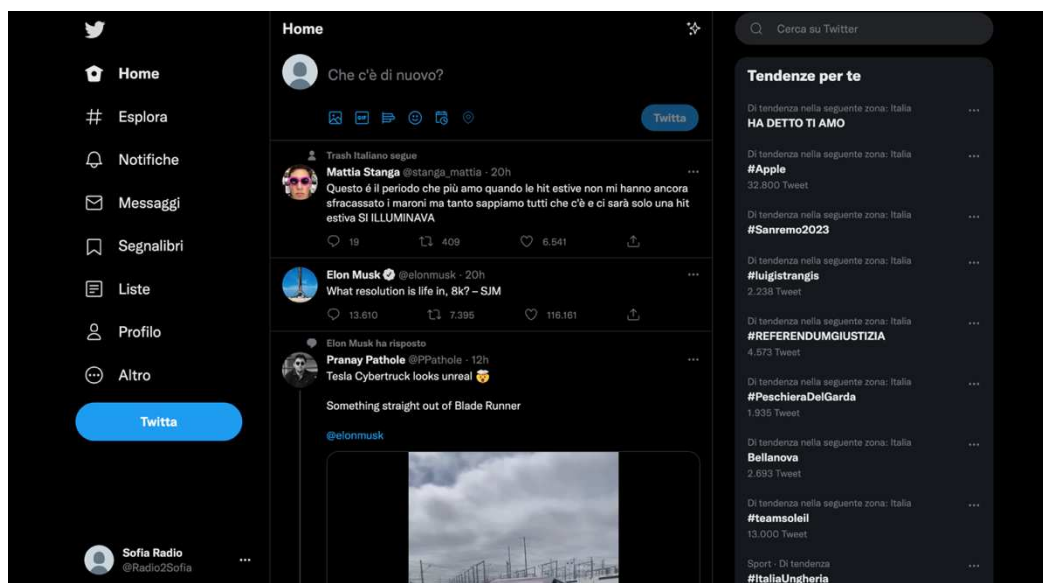
Il 25 aprile 2022 è stato definito l'accordo per l'acquisizione di Twitter da parte del miliardario Elon Musk per 44 miliardi di dollari.

La piattaforma, grazie alla sua semplicità ed immediatezza (proprio come il cinguettio di un uccellino) d'utilizzo, è diventato un vero rivale di Facebook, in cui non esistono gli "amici" ma i "follower" e i "following", ovvero persone che seguono gli aggiornamenti di altre e persone di cui si seguono gli aggiornamenti. In entrambi i casi non si è costretti a ricambiare il "follow". Inoltre, data l'enorme quantità di tweet, e quindi di materiale presente sulla piattaforma, questa viene usata spesso dalle aziende per comunicare con i propri clienti e per monitorare la reputazione dei propri marchi, ma anche dalle università e dalle pubbliche amministrazioni a scopi didattici.

Lo slogan di Twitter è "*What's happening?*" ossia "Cosa sta accadendo?" perché esso stesso si definisce come "ciò che sta accadendo e ciò di cui le persone stanno parlando in questo momento". Per Twitter è importante che le persone abbiano uno spazio libero e sicuro per parlare ed esprimersi, promuovendo una conversazione pubblica. Uno dei valori principali del servizio è, infatti, credere che il cambiamento inizi con la conversazione, con uno scopo in cui ogni voce conta in una conversazione pubblica³⁰. La sua mission è di promuovere conversazioni sane, trovando un equilibrio tra libertà di espressione e la creazione di un servizio sicuro per una conversazione difatti pubblica e partecipata³¹.

³⁰ <https://about.twitter.com/en>

³¹ <https://about.twitter.com/content/dam/about-twitter/en/company/global-impact-2020.pdf>



I tweet sono visibili pubblicamente a chiunque per default, ma gli utenti possono limitare la visualizzazione anche solo ai loro seguaci. I mittenti possono twittare tramite il sito Web di Twitter, applicazioni esterne compatibili (come quelle per gli smartphone) o tramite SMS (Short Message Service), funzionalità disponibile solamente in alcuni paesi. Ad ogni tweet si può mettere il famoso “mi piace” detto anche “like” e possono contenere al massimo 280 caratteri. Ciò consegue che gli utenti debbano massimizzare le loro opinioni in testi brevi avvalendosi anche di:

- Testo
- Immagini
- Video
- GIF animate
- Emoji
- Menzioni
- Hashtag³²

Nel 2009, Steven Johnson³³ descrive nel Time la meccanica di Twitter come “molto semplice”:

³² L’hashtag è un’etichetta utilizzata su alcuni servizi web e social network come aggregatore tematico per rendere più semplice agli utenti trovare i contenuti pubblicati su un tema specifico.

³³ Steven Johnson è un giornalista e scrittore statunitense per Wired, Slate, Discover Magazine e speaker alla conferenza TED.

«Twitter ha una profondità insospettata. In parte perché sapere cosa hanno mangiato i vostri amici a colazione è più interessante di quanto sembri. [...] Seguendo questi rapidi e abbreviati resoconti di stato dei membri della vostra rete sociale estesa, ottenete uno sguardo stranamente soddisfacente sulla loro routine quotidiana. Twitter vi dà le stesse informazioni senza che dobbiate chiederlo direttamente alla persona interessata.»³⁴

I tweet possono essere etichettati con uno o più hashtag, ossia combinazioni di parole concatenate precedute dal simbolo cancelletto (#) per far comprendere agli altri utenti che ci si sta riferendo ad un determinato argomento nel tweet pubblicato. In questo modo, si crea un collegamento ipertestuale a tutti i messaggi che citano lo stesso hashtag. Nel 2010 Twitter ha introdotto nella home la sezione “tendenze”, ovvero l’elenco degli hashtag e dei temi più utilizzati e più popolari in quel momento, e nel 2012 vi ha aggiunto un filtro che potesse selezionarli in base alla posizione di paesi o città.

Un altro simbolo utilizzato all’interno dei tweet è la chiocciola (@) seguito da un nome utente per menzionarlo o rispondere nei commenti ad altre persone.

I retweet, infine, sono ripubblicazioni di messaggi di altri utenti per condividerli con i propri seguaci nel feed personale e tali ripubblicazioni sono contrassegnate con la sigla “RT” nel messaggio.

2.2 Gli utenti e le cifre raggiunte da Twitter nel 2021

Da quando è stato inviato il primo tweet, Twitter ha accumulato numeri considerevoli. Il terzo trimestre del 2021³⁵, il servizio ha contato circa 211 milioni di utenti attivi al giorno monetizzabili (mDAU), il 12,8% in più rispetto all’anno precedente. A differenza di altre piattaforme, Twitter considera solamente i membri che vedono gli annunci pubblicitari e non quindi tutti i visitatori giornalieri. Differentemente, il numero totale di account presenti nella piattaforma è stato circa 1.3 miliardi ma attivi solamente 330 milioni (quasi la stessa quantità dell’intera popolazione degli Stati Uniti). Al giorno sono stati pubblicati 500 milioni di tweet e 350 mila ogni minuto, mentre il 44% degli utenti registrati non ha mai tweetato³⁶.

³⁴ <https://content.time.com/time/subscriber/article/0,33009,1902818,00.html>

³⁵ https://s22.g4cdn.com/826641620/files/doc_financials/2021/q3/Final-Q3'21-S-Shareholder-letter.pdf

³⁶ <https://www.factslikes.com/s-Twitter>

Numeri importanti li hanno ottenuti gli Stati Uniti, paese leader per numero di utenti (77.75 milioni), seguiti dall'India con 24.45 milioni, dal Brasile con 19.05 milioni e il Regno Unito con 19.05 milioni.

Gli utenti del servizio sono chiamati “tweeter” cioè “coloro che twittano” e, nell'ottobre 2021, erano composti per il 70.4% da uomini e per il 29.6% da donne. In America, il 42% del pubblico della piattaforma apparteneva alla fascia d'età compresa tra i 18 e i 29 anni, seguiti dal 27% di adulti tra i 30 e i 49 anni. Altri dati interessanti sono quelli derivanti dal titolo di studio degli utenti americani: secondo delle statistiche del 2021 di Twitter, il 33% degli americani con una laurea ha usufruito del servizio, tra i quali il 45% per ricevere notizie ed informarsi. Differentemente, il 26% frequentava l'università e il 14% possedeva massimo un diploma superiore. Per quanto riguarda la demografia, il 27% viveva nelle città, il 23% nei paesi suburbani e il 18% nei paesi rurali. Infine, tra tutti gli utenti USA, il 52% è entrato nella piattaforma giornalmente, l'84% settimanalmente e il 96% mensilmente.

Come anticipato nel paragrafo 2.1, il potere di Twitter viene sfruttato³⁷ da molte aziende e pubbliche amministrazioni per *data mining*³⁸, *opinion mining*³⁹, fare *branding*⁴⁰ e per strategie lavorative come aumentare la propria credibilità ed estendere il messaggio del proprio marchio. Solo negli Stati Uniti nel 2021 il 65,8% delle aziende con più di 100 dipendenti ha utilizzato la piattaforma per il settore marketing digitale⁴¹ e, secondo eMarketer⁴², circa il 66% delle aziende mondiali con più di 100 dipendenti era dotato di un account Twitter.

Possedere un account all'interno del servizio è utile anche per fidelizzare i propri clienti e creare un rapporto con essi. Il 77% degli utenti, infatti, ha riferito che quando ai propri tweet riceve risposta dal brand, aumenta di conseguenza la loro positività nei confronti del marchio.

³⁷ <https://zephoria.com/twitter-statistics-top-ten/>

³⁸ Il *Data Mining* è l'individuazione di informazioni di diversa natura tramite tecniche ed operazioni automatizzate di estrapolazione di grandi moli di dati.

³⁹ L'*Opinion Mining* è una funzionalità della sentiment analysis e si basa su aspetti in NLP (Elaborazione del Linguaggio Naturale), fornendo informazioni sulle opinioni in base alle parole nel testo.

⁴⁰ Fare Branding significa utilizzare delle tecniche di marketing e comunicazione per aumentare il valore percepito dal mercato di un prodotto, di una marca, di un'azienda.

⁴¹ <https://www.websiterating.com/it/research/twitter-statistics/#references>

⁴² <https://blog.hootsuite.com/twitter-statistics/>

Secondo il rapporto sulle tendenze digitali di Hootsuite⁴³, a gennaio 2022, tra gli utenti Internet compresi dai 16 ai 64 anni, il 3,3% ha dichiarato che Twitter fosse il loro social media preferito, classificandosi così settimo e superando Snapchat, Pinterest, Telegram e Messenger (Figura 1).

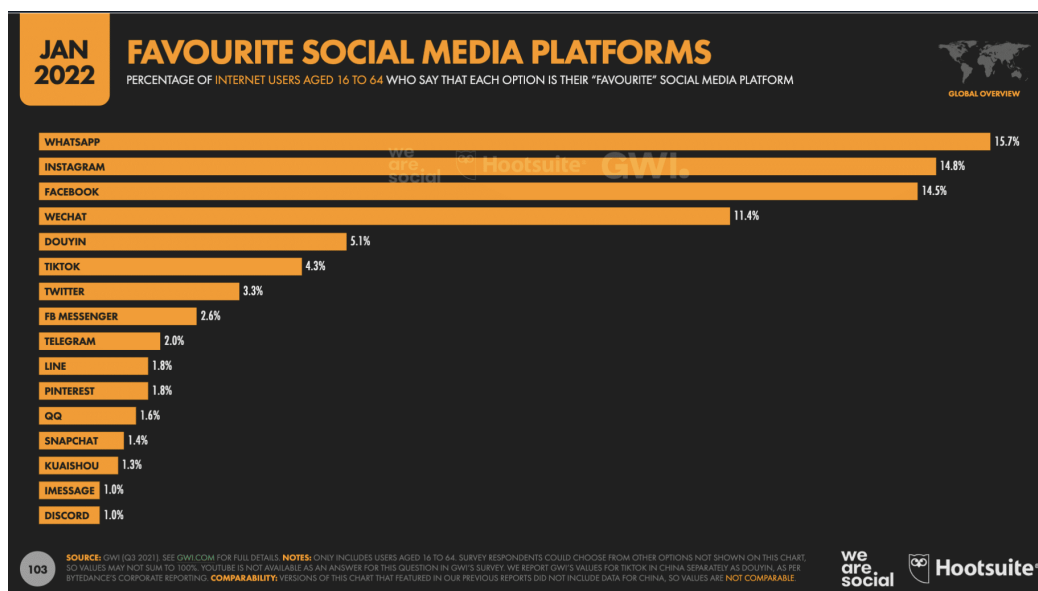


Figura 1: Fonte <https://www.hootsuite.com/resources/digital-trends>

2.3 L'API di Twitter

Per condividere informazioni nel modo più ampio possibile, Twitter offre l'accesso programmatico ai suoi dati attraverso le sue Interfacce di Programmazione delle Applicazioni (API, *Application Programming Interfaces*)⁴⁴.

Le API sono il modo in cui i programmi informatici “comunicano” tra di loro per richiedere e fornire informazioni e questo avviene attraverso gli “endpoint” cioè indirizzi, generalmente univoci, che corrispondono ad una tipologia specifica di informazioni da fornire o da ricevere. Si tratta perciò di applicazioni che utilizzano le funzionalità di altre applicazioni per agevolare la programmazione di servizi resi disponibili. La nuova versione 2 della piattaforma API (API v2) di Twitter, offre ampio accesso ai suoi dati

⁴³ <https://www.hootsuite.com/resources/digital-trends>

⁴⁴ <https://help.twitter.com/it/rules-and-policies/twitter-api#:~:text=Per%20condividere%20informazioni%20nel%20modo,interfacce%20di%20programmazione%20delle%20applicazioni>

pubblici, ovvero a quelli che i suoi utenti hanno deciso di condividere con tutti. Essa include diversi endpoint, raggruppabili in cinque gruppi principali:

- account ed utenti: il servizio permette agli sviluppatori di gestire programmaticamente il profilo e le impostazioni di un account, come i suoi follower e le informazioni sulle sue attività. Ad esempio, questi endpoint possono supportare servizi per i cittadini;
- tweet e risposte: la piattaforma fornisce agli sviluppatori tweet pubblici e le risposte associate, consentendo loro di pubblicarli. Ad esempio, questi endpoint vengo utilizzati da organizzazioni non governative per individuare e contrastare la disinformazione sulla salute pubblica;
- messaggi diretti: questi endpoint garantiscono l'accesso alle conversazioni dirette di utenti che hanno concesso esplicitamente l'autorizzazione. Ad esempio, chatbot⁴⁵ per comunicare con i clienti durante le assistenze online;
- annunci: questi endpoint consentono agli sviluppatori di aiutare le aziende a creare e gestire automaticamente campagne pubblicitarie su Twitter, identificando argomenti ed interessi targettizzati;
- strumenti e SDK (*Software Development Kit*) per i publisher: questi endpoint consentono di incorporare, nelle pagine web di publisher e sviluppatori di software, strumenti semplici ed immediati come le cronologie, i pulsanti di condivisione e altri.

Twitter, quindi, si può definire un'ottima piattaforma con cui fare mining di dati essendo pubblico, immediato, gratuito e dotato di una buona documentazione per le API. La raccolta dei dati è utile per comprendere il fattore umano nei contenuti dei messaggi e, perciò, come utenti che non si conoscono interagiscono tra di loro creando connessioni aperte, comunità e quali sono le loro opinioni su vari argomenti.

Ad oggi, vi sono circa 14mila API pubbliche utilizzate da vari servizi e software di uso comune e quelle più conosciute sono quelle fornite per l'appunto da Twitter, ma anche Facebook, Google, eBay e Amazon⁴⁶. Si può quindi sostenere che condividere le proprie

⁴⁵ Un *Chatbot* è un software che simula le conversazioni umane scritte o parlate, permettendo agli utenti di interagire con i dispositivi digitali come se stessero comunicando con persone reali.

⁴⁶ <https://www.zerounoweb.it/software/erp-crm-scm/cosa-sono-le-api-e-quale-impatto-hanno-sul-business/>

API con altri sviluppatori consegue per le aziende fornitrici un ampliamento della loro portata, dei ricavi ed infine della fidelizzazione degli utenti.

Le sei macrocategorie di casi d'uso⁴⁷ dell'API di Twitter sono le seguenti:

- a. Costruire per la propria azienda, aiutandola ad ascoltare, agire e scoprire il mercato ed il pubblico. Con questa opzione le aziende possono⁴⁸:
 - comprendere, monitorare e confrontare le percezioni che ruotano attorno al loro marchio, individuare tempestivamente eventi (come, ad esempio, le crisi) che potrebbero influire sulla loro reputazione;
 - gestire ed osservare la loro presenza sui social con strumenti e ROI⁴⁹ in tempo reale;
 - interagire con i propri clienti fornendo rapidamente servizi di assistenza;
 - scoprire nuove tendenze, preferenze, feedback ed opportunità per le proprie strategie aziendali;
 - migliorare le proprie soluzioni pubblicitarie con Twitter Ads⁵⁰.
- b. Costruire per il pubblico nuovi strumenti e prodotti per migliorare ed integrare la loro esperienza sulla piattaforma nelle seguenti aree⁵¹:
 - conversazioni pubbliche sulla salute e sicurezza su Twitter;
 - ampliare i modi in cui le persone possono interagire sulla piattaforma come, ad esempio, la *gamification*⁵² e i bot⁵³;
 - accompagnare le aziende a comprendere e analizzare le informazioni ricavate dai loro contenuti di conversazione pubblica per migliorare le proprie prestazioni e decisioni da adottare;
 - garantire agli utenti un'esperienza soddisfacente all'interno di tutte le comunità di Twitter;

⁴⁷ <https://developer.twitter.com/en>

⁴⁸ <https://developer.twitter.com/en/use-cases/build-for-businesses>

⁴⁹ Il *ROI* (Ritorno sull'Investimento o Indice di Redditività del capitale investito) è un indice di bilancio che determina la redditività e l'efficienza economica di un'azienda, ovvero quanto rende il suo capitale investito.

⁵⁰ *Twitter ADS* è un sistema di annunci che consente di scegliere il pubblico, il mezzo per raggiungerlo, l'oggetto della campagna promozionale e il suo obiettivo.

⁵¹ <https://developer.twitter.com/en/use-cases/build-for-consumers>

⁵² La *Gamification* è la pratica di utilizzare componenti di game design o elementi provenienti dai giochi in contesti diversi da quelli ludici.

⁵³ Un *Bot* è un programma autonomo che accede alla rete nello stesso modo in cui accedono gli utenti compiendo compiti automatizzati.

- creare esperienze personalizzate e mirate per far sì che gli utenti scoprano nuovi contenuti a loro interessanti.
- c. Fare ricerca per recuperare dati storici in tempo reale per i propri progetti di ricerca scientifici, per il giornalismo investigativo e la ricerca indipendente, per le ricerche di mercato e le ricerche applicate per le ONG⁵⁴.
- d. Insegnare e apprendere abilità utili come la programmazione e l'elaborazione di dati⁵⁵.
- e. Costruire, sperimentare, esplorare per divertimento⁵⁶ come:
 - creare robot per la condivisione di informazioni utili;
 - costruire esperienze interattive e creative da condividere;
 - ampliare le proprie prospettive;
 - pianificare e automatizzare i propri tweet.
- f. Costruire e rendere il mondo e Twitter dei posti migliori⁵⁷ condividendo automaticamente informazioni utili e sane per gli utenti (come, ad esempio, avvisi per la salute pubblica).

Twitter afferma che:

«Con una base moderna, nuove funzionalità ed endpoint e un accesso ancora più gratuito, l'API di Twitter v2 rende più facile che mai aiutare a migliorare la salute della conversazione pubblica, scoprire informazioni dettagliate, aiutare le persone a esprimersi e costruire il prossimo futuro»⁵⁸.

L'API v2 offre tre tipologie di accesso gratuite in base all'utilizzo che si prevede di necessitare:

- a. Essenziale: offre la possibilità di studiare fino a 500 mila tweet al mese attraverso un'app per progetto ed è considerata la tipologia migliore per iniziare rapidamente a testare e creare su tutti gli endpoint;
- b. elevato: offre la possibilità di analizzare due milioni di tweet al mese attraverso tre app per progetto;

⁵⁴ <https://developer.twitter.com/en/use-cases/do-research>

⁵⁵ <https://developer.twitter.com/en/use-cases/teach-and-learn>

⁵⁶ <https://developer.twitter.com/en/use-cases/build-for-fun>

⁵⁷ <https://developer.twitter.com/en/use-cases/build-for-good>

⁵⁸ <https://developer.twitter.com/en/products/twitter-api>

- c. ricerca Accademica: permette l'accesso ai dati pubblici riguardanti qualsiasi argomento per riuscire a soddisfare agli obiettivi di ricerca di studenti, istituzioni accademiche e universitarie, recuperando fino a dieci milioni di tweet al mese attraverso un'app per progetto.

Questa introduzione alle funzionalità di Twitter vuole rendere più comprensibile le procedure e i risultati ottenuti in questo elaborato presenti in seguito.

2.4 Cos'è la Sentiment Analysis

I social media oramai appartengono alla nostra quotidianità e rappresentano un luogo in continua crescita esponenziale in cui sempre più utenti esprimono e condividono nel Web le loro opinioni sui più disparati argomenti, servizi, prodotti, aziende... oggi giorno queste opinioni rappresentano una mole considerevole di dati utili e necessari per chi offre questi servizi o prodotti per compiere monitoraggi dei cambiamenti di sentimento rispetto ad un determinato evento, comprendere la loro reputazione online, il grado di soddisfazione dei propri clienti riguardo il loro marchio o il rilascio di un prodotto o servizio, i propri punti di forza e di debolezza. Per queste ragioni, è utile e necessario per le aziende compiere periodicamente la *Sentiment Analysis* o *Opinion Mining*⁵⁹.

Sentiment Analysis in italiano significa “analisi del sentimento/delle opinioni” ed è un insieme di tecniche di analisi procedurale di calcolo di informazioni testuali al fine di recuperare i sentimenti e le opinioni espresse nei testi online dagli utenti su un servizio, prodotto, azienda, evento o brand. Essa consente di cogliere e classificare l'autentica natura dei tipi di interazioni che avvengono tra gli utenti, definiti per l'appunto “*opinion holder*” o “possessore dell'opinione” in un determinato contesto e periodo temporale, e queste interazioni possono avere polarità positiva, neutra o negativa. Si tratta di una tecnica vera e propria di *data mining* e *Natural Language Processing* in cui, attraverso quindi un'analisi computazionale, vengono estratti dai testi online parole chiave alla base di un sentimento. Negli ultimi decenni la ricerca e la classificazione dei sentimenti veicolati su Twitter è cresciuta costantemente ed ha considerevoli applicazioni in campo economico, politico e sociale.

⁵⁹ <https://lapennadelweb.com/sentiment-analysis-cosa-quali-tool-usare/>

Esistono differenti tipologie di approcci in base alla classificazione dei dati:

- Rilevamento delle keyword (parole chiave): vi sono alcune parole a cui vengono associate delle categorie di emozioni come la felicità, la tristezza, la rabbia, la preoccupazione...
- metodi statistici: l'analisi precisa del testo in esame, valuta la correlazione grammaticale tra le parole riuscendo così ad associarle ad una polarità;
- affinità lessicale: in aggiunta alle keyword, vengono analizzate parole che lessicalmente sono associate a determinati sentimenti.

Quando si compie la *sentiment analysis*, il software utilizzato deve essere in grado di leggere e tradurre in polarità le seguenti caratteristiche:

- La rilevanza: i testi devono essere valutati all'interno dei determinati contesti e quindi devono essere coerenti al tema oggetto di analisi altrimenti risultano senza rilevanza;
- il tono di voce: per comprendere l'opinione degli opinion holder è importante che il software utilizzato riesca ad intendere il registro (amichevole, aggressivo, comico...) di quest'ultimi per attribuirvi una polarità;
- l'intensità: è rilevante riuscire a considerare le parole espresse dagli utenti anche dal punto di vista della loro intensità che può rilevarsi intensa e passionata sia in positivo che in negativo, ma anche disinteressata;
- l'emotività: è importante che il software riesca anche a rilevare le modalità con cui sono scritti i testi quindi, ad esempio, la qualità e la quantità della punteggiatura, delle emoticon e degli aggettivi presenti.

I passaggi di base per eseguire la *sentiment analysis* includono⁶⁰:

- la raccolta dei dati;
- la preelaborazione dei dati;
- l'estrazione delle funzionalità;
- la selezione delle caratteristiche di base;

⁶⁰ <https://www.intelligenzaartificialeitalia.net/post/sentiment-analysis-con-python-la-guida-completa>

- il rilevamento del sentimento;
- la classificazione.

Compiere la *sentiment analysis* è impegnativo su dati di Twitter perché:

1. I tweet possono contenere solo 280 caratteri, quindi vi sono opinioni espresse in modo compatto e con diverse sfumature (in base alla soggettività e al lessico di ogni utente);
2. l'uso di abbreviazioni e slang (linguaggi differenti), simboli ripetuti per esprimere un'opinione o un'emozione (gli emoticon);
3. presenza nei tweet di hashtag, link, nomi utenti (user ID) che richiedono una diversa elaborazione rispetto ad un testo.

Queste complicanze devono essere risolte e filtrate nella sezione di preelaborazione.

2.5 Cos'è Python

Python è un linguaggio di programmazione adatto specialmente a sviluppare applicazioni, scripting⁶¹, computazione numerica⁶² e system testing⁶³. È stato creato dall'informatico olandese Guido van Rossum all'inizio degli anni Novanta, il quale scelse questo nome per la sua passione verso i Monty Python, un gruppo comico britannico degli anni Settanta.

Il linguaggio è molto popolare per la leggibilità del codice e viene utilizzato spesso anche per la creazione di software grazie alla sua flessibilità, semplicità e dinamicità di sperimentazione che consente. Python è fornito con licenza Open-Source⁶⁴, quindi il suo download ed utilizzo è gratuito e libero anche per prodotti commerciali, può essere modificato e ridistribuito secondo le regole di licenza. Nonostante ciò, riceve miglioramenti costanti dalla sua community che lo mantiene sempre aggiornato. È anche multiplatforma, ovvero funziona sui principali sistemi operativi: Linux, MacOS,

⁶¹ Lo *Scripting* è un linguaggio di programmazione che compie lavori automatizzati evitando che essi vengano svolti da un operatore umano.

⁶² La *Computazione Numerica* (o Calcolo Numerico) è un'analisi matematica approssimativa che si avvale degli algoritmi con una velocità di calcolo molto elevata.

⁶³ Il *System Testing* è un test del sistema nel suo complesso con lo scopo di verificare che tutti i moduli del software funzionino correttamente.

⁶⁴ <https://www.python.org/about/>

Windows e Unix. Esso è, infine, fornito di una libreria di funzioni *built-in*⁶⁵, di una sintassi chiara e leggibile⁶⁶.

Ogni installazione di Python include la “standard library” ovvero una collezione di moduli per svolgere varie applicazioni come l’elaborazione del linguaggio naturale, l’apprendimento automatico, l’analisi di dati, e utilizzando il Python Package Index è possibile scaricare ed installare migliaia di moduli aggiuntivi creati e mantenuti dalla sua comunità. I programmi all’interno di questo linguaggio sono codificati in formato *bytecode*⁶⁷ prima di essere eseguiti in modo da renderli compatti, efficienti e performanti. Python è un linguaggio ad alto livello che gestisce automaticamente la memoria tramite un meccanismo chiamato *garbage collection*, consentendo al suo utilizzatore di non dover dichiarare e rilasciare spazi di memoria manualmente.

Alcune caratteristiche di Python sono:

- Presenta una varietà dei tipi di dati disponibili: numeri (a virgola mobile, complessi, interi a lunghezza illimitata), stringhe (sia ASCII che Unicode), liste, tuple⁶⁸ e dizionari;
- la sintassi e la struttura del linguaggio sono semplici da memorizzare;
- è versatile, infatti ricopre vari ambiti di utilizzo come il machine learning, sviluppo del web, applicazioni desktop, giochi e grafica 3D...;
- supporta la programmazione ad oggetti con classi ed ereditarietà multipla;
- il codice può essere raggruppato in moduli e pacchetti;
- supporta un vario sistema delle eccezioni, risultando chiaro nella gestione degli errori;
- i tipi di dati sono fortemente e dinamicamente tipicizzati. Mischiare tipi incompatibili (es. cercare di sommare una stringa ed un numero), scatena un'eccezione;

⁶⁵ Le funzioni *built-in* sono funzioni già presenti all’interno di Python che consentono di compiere operazioni di base.

⁶⁶ <https://www.python.it/about/>

⁶⁷ Il formato *bytecode* è un codice binario leggibile dal processore hardware di un computer.

⁶⁸ La *Tupla* è una lista immutabile di valori separati da una virgola.

- Python contiene interessanti caratteristiche avanzate come i generatori e le list comprehensions⁶⁹.

2.4.1 NLTK (Natural Language Processing)

L'elaborazione del linguaggio naturale (NLP) è una parte della computer science⁷⁰ che si occupa di sviluppare sistemi che permettono ai programmi informatici di riconoscere, comprendere, interpretare e riprodurre il linguaggio umano. La Natural Language Toolkit (NLTK) è un pacchetto open source gratuito in Python⁷¹ (sviluppato nel 2001 da Steven Bird e il suo assistente all'insegnamento Edward Loper al Department of Computer and Information Science dell'Università della Pennsylvania⁷²) che, utilizzando l'NLP, fornisce la base e varie librerie per l'elaborazione programmatica e la classificazione del testo, ad esempio la tokenizzazione, lo stemming, la classificazione, il filtraggio, la manipolazione del testo e il ragionamento semantico. NLTK viene utilizzato per creare un modello di bag-words, ovvero un modello unigram per il testo che viene convertito in un vettore, semplificando il conteggio della frequenza totale delle parole più comunemente utilizzate nel corpus⁷³. NLTK è adatto a linguisti, ricercatori, studenti, ingegneri, utenti industriali ed è disponibile per Mac OS, Windows e Linux.

2.4.2. NumPy

NumPy è un pacchetto open source, creato nel 2005 da Travis Oliphant, fondamentale per rendere possibile il calcolo scientifico in Python⁷⁴. Esso fornisce un oggetto array⁷⁵ multidimensionale ad alte prestazioni e strumenti per lavorare velocemente su tali array come compiere operazioni matematiche e logiche, algebra lineare di base, manipolazione di forme, operazioni statistiche di base....

⁶⁹ Le *list comprehension* sono liste con sintassi più breve utilizzabili quando si desidera creare una nuova lista basta sui valori di una lista esistente.

⁷⁰ La *Computer Science* è lo studio dell'informazione, del calcolo e dell'automazione attraverso discipline teoriche (come gli algoritmi) e pratiche (come la programmazione di software).

⁷¹ <https://www.nltk.org/>

⁷² <https://github.com/nltk/nltk/wiki/FAQ>

⁷³ Un *Corpus* è una raccolta di file di testo.

⁷⁴ <https://numpy.org/doc/stable/>

⁷⁵ Un *Array*, detto anche vettore (quando monodimensionale) o matrice (quando bidimensionale), indica una struttura dati complessa, omogenea e statica. Consente di definire nuovi tipi di dati a partire da dati preesistenti.

2.4.3 VADER (Valence Aware Dictionary and Sentiment Reasoner)

VADER è uno strumento di analisi del sentimento, completamente open source, basato su lessico e regole adatto ai sentimenti espressi nei social media ma che funziona in modo corretto anche su testi di altra tipologia. VADER incorpora un lessico di sentiment “gold-standard” particolarmente adatto a contesti di tipo microblog. Il suo lessico è sensibile sia alla polarità che all'intensità dei sentimenti espressi⁷⁶.

Creare manualmente un lessico del sentiment è un processo che richiede molto tempo e lavoro, per questa ragione molti ricercatori si affidano a strumenti di sentiment preesistenti come VADER. Esso è stato creato raggruppando in una lista le parole di sentiment già studiate come il vocabolario LIWC⁷⁷, aggiungendovi il lessico comune nel sentiment presente nei microblog come l'elenco completo delle emoticon più utilizzate nelle regioni occidentali (per esempio :(indica un sentimento negativo), gli acronimi e sigle legate al sentimento (ad esempio LOL indica un sentimento preponderante) ed il gergo comune (ad esempio “nah” indica un sentimento di disapprovazione).

⁷⁶ <https://github.com/cjhutto/vaderSentiment>

⁷⁷ *The Linguistic Inquiry and Word Count (LIWC)* è un dizionario psicolinguistico che raggruppa le parole in base ad un'emozione specifica a cui sono legate. Esso legge un file di testo e conta la percentuale di parole che riflettono differenti emozioni, stili di pensiero, parti del discorso.

3. Metodologia della Twitter sentiment analysis con Python

Per eseguire l'analisi del sentiment dei tweet, occorre raccogliere i dati dalla fonte in considerazione, in questo studio Twitter. Il primo step, quindi, consiste nell'estrarre i testi e sottoporli a varie fasi di preelaborazione per renderli più sensibili alla macchina d'apprendimento rispetto alla loro forma iniziale e riuscire così, infine, a recuperare i punteggi positivi, neutri o negativi associati ad una determinata opinione.

3.1 Collezione dei tweet

La raccolta dei tweet pertinenti all'area di interesse, avviene utilizzando l'API di Twitter. Questa interfaccia può essere installata su Python usando il comando pip: `pip install tweepy`, che installerà la libreria Tweepy e acconsentirà l'accesso all'API di Twitter.

Innanzitutto, però, per recuperare i tweet dall'API di Twitter è necessario registrare un'app tramite il proprio account Twitter eseguendo i seguenti passaggi:

- a. Aprire <https://apps.twitter.com/> e cliccando “Crea nuova app”
- b. Compilare i dati richiesti nel modulo fornito
- c. La pagina si caricherà automaticamente e apparirà la scheda “chiavi e token di accesso” che contiene quattro informazioni segrete e strettamente personali che consentono l'autorizzazione per l'accesso al database dei tweet: “consumer key”, “consumer secret”, “access token” e “access token secret”
- d. Copiare le seguenti chiavi ed i token generati

Queste chiavi vengono successivamente inserite nel codice di Python, rendendo così possibile l'estrazione dei tweet attraverso la libreria Pandas, che deve essere importata su Python con il comando: `import pandas as pd`.

Pandas è uno strumento open source di analisi e manipolazione di set di dati usufruito maggiormente in settori accademici e commerciali come la finanza, la statistica, la pubblicità, economia e altri. Esso viene utilizzato per gestire operazioni⁷⁸ come:

- l'estrazione, la scrittura e la lettura di una data set da Excel, CSV e file di testo;
- selezionare di righe e colonne particolari per la mutabilità delle dimensioni;
- convertire il set di dati in un data frame⁷⁹;
- rimuovere i dati indesiderati;

⁷⁸ <https://pandas.pydata.org/>

⁷⁹ Il *Data Frame* è una lista utilizzata per rappresentare degli insiemi di dati.

- indicizzazione e suddivisione di grandi insiemi di dati;
- eseguire operazioni di divisione, applicazione e combinazione su un insieme di dati;
- unire insiemi di dati;
- generare intervalli temporali senza la perdita di dati;
- esaminare e gestire valori nulli/ dati mancanti.

Successivamente, è necessario importare altri pacchetti su Python utili per applicare determinati algoritmi di apprendimento automatico per le analisi da compiere successivamente. Questi pacchetti sono:

- a. Numpy per compiere calcoli numerici, integrazioni e operazioni array (matrice);
- b. Matplotlib, una libreria completa per creare rappresentazioni grafiche, statistiche, interattive e schematiche dei dati⁸⁰;
- c. WordCloud per estrarre visivamente solo le parole chiave del set di dati utilizzate dalle persone per esprimere le proprie opinioni⁸¹;
- d. NLTK per elaborare ed estrarre le caratteristiche delle opinioni nei tweet;
- e. VADER (Valence Aware Dictionary and Sentiment Reasoner) è uno strumento di analisi del sentimento basato sul lessico delle parole e su regole che sono attinenti con i sentimenti espressi sui social media. Esso restituisce un punteggio di polarità per ogni tweet che può possedere qualsiasi valore in decimale, ma più esso si avvicina all'1, allora più positivo sarà considerato il tweet e viceversa.

3.1.1 Scelta delle variabili d'interesse

Per estrarre solamente i tweet d'interesse, bisogna specificare diversi parametri, come i periodi di tempo della ricerca, la lingua e gli hashtag che i tweet devono contenere per essere scaricati.

Come anticipato in precedenza, i periodi di tempo scelti sono:

- a. dal 1° dicembre 2018 al 31 dicembre 2019 (dal primo processo per diffamazione avuto luogo in tribunale a Londra)

⁸⁰ <https://matplotlib.org/>

⁸¹ <https://pypi.org/project/wordcloud/>

- b. dal 1° aprile 2022 al 10 giugno 2022 (durata del processo per diffamazione in Virginia)
- c. dall'11 giugno 2022 all'11 settembre 2022 (tre mesi dopo la fine del processo in Virginia)

Sono stati scelti questi tre periodi differenti per riuscire ad analizzare e confrontare come, agli albori delle complicità e delle prime accuse tra la relazione di Depp e della Heard, le persone comunicassero su Twitter a riguardo, per poi verificare se durante e dopo il processo definitivo di quest'anno gli utenti abbiano cambiato modalità di espressione sul social e la preferenza verso uno dei due attori.

Come lingua dei tweet da recuperare è stato scelto solamente l'inglese questo perché, dopo diversi tentativi di recupero dati in lingua italiana, si è osservato che questi non erano di un numero sufficiente per compiere una analisi soddisfacente e non contenevano testi con significato utile alla domanda di ricerca, ciò probabilmente dovuto dal fatto che le vicende hanno interessato maggiormente il popolo americano.

Infine, lo step più significativo per recuperare in modo corretto i tweet è la definizione degli hashtag che questi devono includere.

Ognuna delle tre query (espressioni interrogative per il data frame con combinazioni di hashtag) presenti nel codice, rappresenta una tematica di studio, per cui la scelta degli hashtag stessi è molto importante perché influenzerà l'intera analisi del progetto. La prima query è volta a rappresentare generalmente i tweet contenenti gli hashtag #AmberHeard, #JohnnyDepp, #JohnnyDeppVsAmberHeard, #JohnnyDeppVsAmberHeardTrial e #DeppVsHeard. Quindi tweet che esprimessero vicende inerenti ai processi dei due attori. La scelta di questi hashtag è avvenuta attraverso uno studio all'interno di Twitter in base alla ricerca avanzata aggiungendo dei filtri per data e parole chiave che dovessero contenere. Infine, sono stati scelti quelli più utilizzati nei tre periodi temporali. La seconda query, invece, è volta al recupero di tweet inerenti alla violenza domestica, uno degli argomenti preponderanti nei loro processi. Per il recupero di questi tweet sono stati utilizzati gli hashtag: #DomesticAbuse, #DomesticViolence, #JohnnyDepp, #AmberHeard, #JohnnyDeppVsAmberHeardTrial, #JohnnyDeppVsAmberHeard e #DeppVsHeard. Per concludere, la terza query contiene gli hashtag #MePoo, #amberpoopedthebed, #Poop, #AmberHeard e #JohnnyDepp inerenti al ritrovamento di feci appartenenti alla Heard nel letto di Depp.

Da specificare è che, utilizzando la versione accademica dell'API di Twitter, è possibile recuperare a richiesta nella propria query un massimo di 500 tweet. Per tale ragione si è deciso di dividere gli intervalli di tempo in diversi sottoperiodi ai quali corrispondeva un determinato numero di giorni (e anche per recuperare un numero di tweet approssimativamente simile per tutti i periodi).

Per il primo intervallo (dal 1° dicembre 2018 al 31 dicembre 2019) il sottoperiodo corrisponde a due giorni; quindi, sono state mandate 500 richieste di recupero dei tweet ogni due giorni (48 ore), riuscendone così a recuperare un totale di **4156** tweet per la prima query, per la seconda **539** e per la **406** terza.

Per il secondo periodo (dal 1° aprile 2022 al 10 giugno 2022) il sottoperiodo corrisponde a 12 giorni; quindi, ogni 12 giorni sono stati richiesti 500 tweet riuscendo così a recuperare un totale di **10730** tweet per la prima query, **3993** per la seconda e **7718** per la terza. Infine, per il terzo periodo (dall'11 giugno 2022 all'11 settembre 2022) il sottoperiodo corrisponde a 15 giorni, recuperando un totale di **12216** tweet per la prima query, **979** per la seconda e **1688** per la terza.

Di seguito è riportata una tabella in cui vengono illustrati per ogni periodo e le corrispettive query, il numero totali di tweet ricavati senza alcun filtraggio.

Periodo	Query	N° tweet ricavati
dal 1° dicembre 2018 al 31 dicembre 2019	((#AmberHeard #JohnnyDepp) OR #JohnnyDeppVsAmberHeard OR #JohnnyDeppVsAmberHeardTrial)	4156
	((#DomesticAbuse OR #DomesticViolence) (#JohnnyDepp OR #AmberHeard OR #JohnnyDeppVsAmberHeardTrial OR #JohnnyDeppVsAmberHeard OR #DeppVsHeard))	539
	(#MePoo OR #amberpoopedthebed OR (#Poop #AmberHeard) OR (#Poop #JohnnyDepp))	406
dal 1° aprile 2022 al 10 giugno 2022	((#AmberHeard #JohnnyDepp) OR #JohnnyDeppVsAmberHeard OR #JohnnyDeppVsAmberHeardTrial)	10730
	((#DomesticAbuse OR #DomesticViolence) (#JohnnyDepp OR #AmberHeard OR #JohnnyDeppVsAmberHeardTrial OR #JohnnyDeppVsAmberHeard OR #DeppVsHeard))	3993
	(#MePoo OR #amberpoopedthebed OR (#Poop #AmberHeard) OR (#Poop #JohnnyDepp))	7718
dall'11 giugno 2022 all'11 settembre 2022	((#AmberHeard #JohnnyDepp) OR #JohnnyDeppVsAmberHeard OR #JohnnyDeppVsAmberHeardTrial)	12216
	((#DomesticAbuse OR #DomesticViolence) (#JohnnyDepp OR #AmberHeard OR #JohnnyDeppVsAmberHeardTrial OR #JohnnyDeppVsAmberHeard OR #DeppVsHeard))	979
	(#MePoo OR #amberpoopedthebed OR (#Poop #AmberHeard) OR (#Poop #JohnnyDepp))	1688

Tabella 1: Numero di tweet ricavati

3.2 Prelaborazione dei tweet

La prelaborazione dei tweet è uno step fondamentale affinché la macchina apprenda la struttura dei dati stessi determinando, in questo modo, l'efficienza dei passaggi successivi. Questa comporta la correzione sintattica dei tweet rendendoli più leggibili alla macchina d'apprendimento in modo da ridurre l'ambiguità nell'estrazione dei risultati e,

quindi, è strettamente correlata allo scopo dell'analisi. Di seguito sono riportati i passaggi compiuti sul dataset per la preelaborazione dei tweet, per renderli più "puliti" e che contengano solo parole significative e per ridurre il volume dei testi in modo da diminuire il tempo impiegato per le esecuzioni successive:

- Importare il pacchetto re con il comando: `import re`
- Rimozione dei re-tweet (eseguito avvalendosi dell'espressione regolare " `is:retweet` ")
- Rimozione degli hashtag (eseguito avvalendosi dell'espressione regolare " `r'#[a-zA-Z0-9(_){1,}]'` ")
- Rimozione dei link (eseguito avvalendosi dell'espressione regolare " `r'http\S+'` ")
- Rimozione delle menzioni (eseguito avvalendosi dell'espressione regolare " `r'@[a-zA-Z0-9(_){1,}]'` ")
- Rimozione dei tweet doppi (`filt_tw=list(dict.fromkeys(filt_tw))`) poiché considerare due o più volte gli stessi tweet influirebbe sugli attributi dell'analisi semantica e sui parametri all'interno dei grafici, non fornendo quindi alcun miglioramento
- Rimozione di caratteri speciali (entità HTML) come `&` e `>` (eseguito avvalendosi dell'espressione regolare " `r'(\&\;)'` " e " `r'(\>\;)'` ")
- Scaricare solo tweet con testo (non immagini/video) (`if len(t.strip()) > 0: print(t.strip())`)
- Convertire consecutivi e multipli spazi vuoti in un unico spazio vuoto (`filt_tw.append(t.strip())`)

Infine, fondamentale è salvare i tweet filtrati a questo livello in un file di testo .txt .

Di seguito è riportata una tabella in cui vengono illustrati per ogni periodo e le corrispondenti query, il numero totale di tweet ricavati con il filtraggio iniziale appena spiegato.

Periodo	Query	N° tweet ricavati	N° tweet ricavati filtrati
dal 1° dicembre 2018 al 31 dicembre 2019	((#AmberHeard #JohnnyDepp) OR #JohnnyDeppVsAmberHeard OR #JohnnyDeppVsAmberHeardTrial)	4156	3647
	((#DomesticAbuse OR #DomesticViolence) (#JohnnyDepp OR #AmberHeard OR #JohnnyDeppVsAmberHeardTrial OR #JohnnyDeppVsAmberHeard OR #DeppVsHeard))	539	474
	(#MePoo OR #amberpooedthebed OR (#Pooop #AmberHeard) OR (#Pooop #JohnnyDepp))	406	380
dal 1° aprile 2022 al 10 giugno 2022	((#AmberHeard #JohnnyDepp) OR #JohnnyDeppVsAmberHeard OR #JohnnyDeppVsAmberHeardTrial)	10730	10042
	((#DomesticAbuse OR #DomesticViolence) (#JohnnyDepp OR #AmberHeard OR #JohnnyDeppVsAmberHeardTrial OR #JohnnyDeppVsAmberHeard OR #DeppVsHeard))	3993	3819
	(#MePoo OR #amberpooedthebed OR (#Pooop #AmberHeard) OR (#Pooop #JohnnyDepp))	7718	6361
dall'11 giugno 2022 all'11 settembre 2022	((#AmberHeard #JohnnyDepp) OR #JohnnyDeppVsAmberHeard OR #JohnnyDeppVsAmberHeardTrial)	12216	10508
	((#DomesticAbuse OR #DomesticViolence) (#JohnnyDepp OR #AmberHeard OR #JohnnyDeppVsAmberHeardTrial OR #JohnnyDeppVsAmberHeard OR #DeppVsHeard))	979	827
	(#MePoo OR #amberpooedthebed OR (#Pooop #AmberHeard) OR (#Pooop #JohnnyDepp))	1688	1505

Tabella 2: Numero di tweet ricavati post filtraggio

Si può osservare come nei tre periodi differenti, considerando la stessa query, vi siano circa lo stesso numero di tweet scaricati. Interessante è notare come, per il primo periodo, essi siano di un numero nettamente inferiore rispetto agli altri due periodi, questo perché le vicende di violenza domestica (seconda query) e l'accaduto delle feci nel letto (terza query) sono avvenute ed emerse dal secondo periodo temporale e quindi hanno avuto maggior impatto e tendenza su Twitter.

3.3 Sentiment analysis

Principalmente la *sentiment analysis* dipende dalla polarità del testo, la quale contiene sentimenti ed emozioni e per questa ragione deve essere calcolata. VADER per ogni tweet vi calcola la positività, la neutralità, la negatività ed il valore di compound.

I punteggi positivi, neutri e negativi sono le metriche più utili se lo scopo di ricerca è quello di analizzare il contesto e la prestazione del modo in cui il sentimento è trasmesso nella retorica di una determinata frase. Essi possono avere valori che vanno da 0 a 1.

Il valore di compound (o punteggio composito normalizzato/ponderato) è calcolato sommando i punteggi di valenza di ogni parola del lessico e normalizzando questi in modo da essere compresi tra punteggi da -1 (massimo negativo) e +1 (massimo positivo). Il compound è la metrica più utile e più utilizzata dai ricercatori perché permette di calcolare una misura unica del sentiment di una frase standardizzandola in positiva, negativa o neutra. I valori di soglia sono:

- Sentimento positivo di compound per valori $\geq 0,05$
- Sentimento neutro di compound per valori $> -0,05$ e $< 0,05$
- Sentimento negativo di compound per valori $\leq -0,05$

Di seguito sono riportati due esempi di tweet di cui VADER ha studiato le varie polarità:

I do not like that Amber Heard.
I do not like her bed of turds.
I do not like her in the house.
I do not like how she abused her spouse.
I do not like her playing in the sand.
I will not watch her in Aquaman!

`{ 'neg': 0.277, 'neu': 0.723, 'pos': 0.0, 'compound': -0.9142 }`

Il tweet sopra illustrato è stato classificato da VADER con un valore di compound quasi al massimo negativo (-0.9142), infatti al suo interno vi sono molte espressioni di dissenso e sdegno per la Heard come “do not like” e “turds”.

All of 's Supporters want to receive All of the Justice that she truly
deserves

`{ 'neg': 0.184, 'neu': 0.451, 'pos': 0.365, 'compound': 0.8225 }`

Questo secondo tweet, invece, è un esempio di compound quasi al massimo positivo. Infatti, il lessico utilizzato rimanda ad un senso di appartenenza e sostegno per la Heard.

3.4 Analisi dei risultati

Questo paragrafo verrà suddiviso in sezioni in base agli studi fatti per ogni periodo temporale e le rispettive query. Infine, verranno compiuti dei confronti tra esse.

3.4.1 Analisi dei tweet nel primo arco temporale

Questa sezione conterrà le analisi svolte per i tweet nel periodo temporale dal 1° dicembre 2018 al 31 dicembre 2019.

A. La prima parte vedrà lo studio dei tweet contenenti gli hashtag #AmberHeard #JohnnyDepp ed in più almeno uno tra #JohnnyDeppVsAmberHeard, #JohnnyDeppVsAmberHeardTrial o #DeppVsHeard (prima query).

Dopo il recupero e il filtraggio iniziale dei tweet citato in precedenza, è stata svolta la *sentiment analysis* per i 3647 tweet. Per gli scopi della ricerca si è deciso di tenere in considerazione solamente i valori di compound, più precisi e soddisfacenti.

In questo set di dati sono stati calcolati:

- Il valore medio di compound: -0.15410967918837393
- Il valore massimo di compound: 0.9943
- Il valore minimo di compound: -0.9867

Successivamente, la visualizzazione dei dati processati è stata compiuta avvalendosi di alcuni grafici. Il primo grafico sottostante (grafico 1) rappresenta la distribuzione di tutti i valori di compound. Si può notare come i tweet abbiano ottenuto in media circa lo stesso numero di valori positivi e negativi, anche se c'è maggior affluenza nella parte inferiore del grafico, ovvero quella negativa. Infatti, già la media negativa poteva far prevedere questa caratteristica.

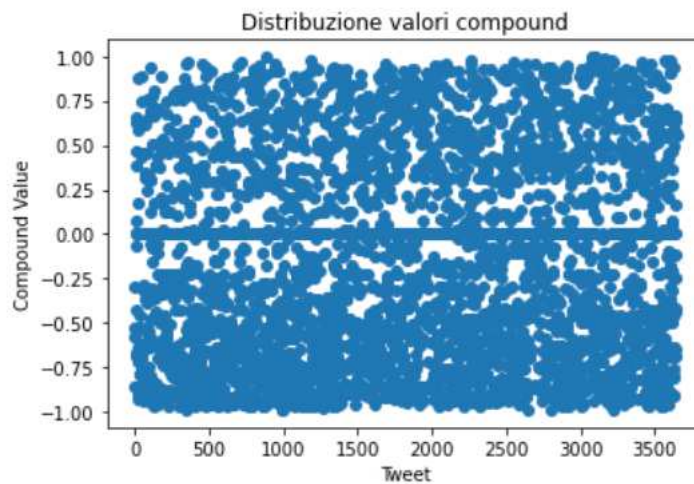


Grafico 1: Distribuzione dei valori di compound

Il grafico 2 invece rappresenta un istogramma e sempre i tweet e i loro valori di compound, ma questi sono raggruppati in barre in cui l'altezza (asse delle y) rappresenta il numero di tweet mentre l'asse delle x i valori di compound da -1 a 1. Anche attraverso questo grafico si può osservare che più tweet siano negativi.

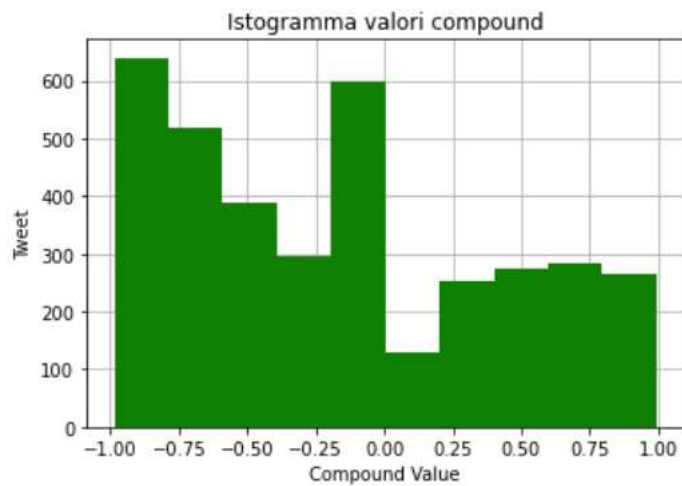


Grafico 2: Istogramma valori di compound

Un'ulteriore analisi è stata compiuta per classificare tutti i valori di compound in tre classi generali: estremamente positivi, neutri, estremamente negativi. Ciò che ne è emerso è che 1188 tweet sono estremamente positivi (il 32.6%), 468 sono neutri (il 12.8%) e 1991 sono estremamente negativi (il 54.6%) (grafico 3).

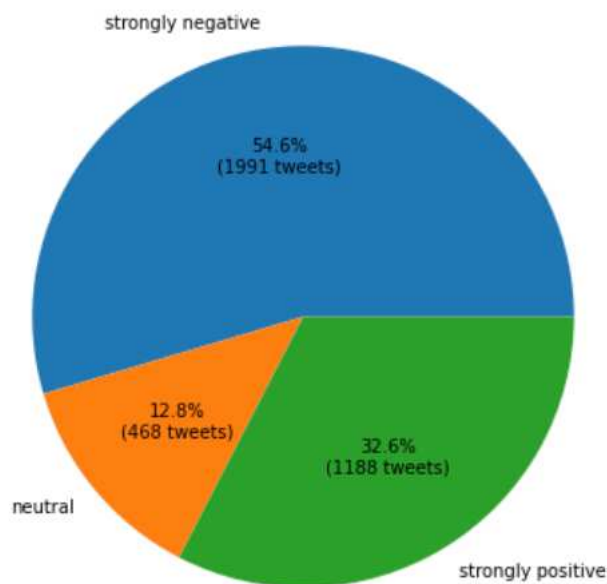


Grafico 3: Grafico a torta della suddivisione per categoria dei valori di compound

È stata creata anche una word-cloud che rappresenta le parole più utilizzate in tutto il corpus.

In seguito ad un tentativo, si è notato che all’interno della word-cloud apparivano come parole più utilizzate le “stopwords” ovvero parole prive di significato e molto frequenti nei testi. È stato bisognoso quindi installare su Python il pacchetto STOPWORDS che include le seguenti parole inglesi:

```
['i', 'me', 'my', 'myself', 'we', 'our', 'ours', 'ourselves', 'you',  
, "you're", "you've", "you'll", "you'd", 'your', 'yours', 'yourself',  
, 'yourselves', 'he', 'him', 'his', 'himself', 'she', "she's", 'he',  
r', 'hers', 'herself', 'it', "it's", 'its', 'itself', 'they', 'them',  
, 'their', 'theirs', 'themselves', 'what', 'which', 'who', 'whom',  
, 'this', 'that', "that'll", 'these', 'those', 'am', 'is', 'are', 'wa',  
s', 'were', 'be', 'been', 'being', 'have', 'has', 'had', 'having',  
, 'do', 'does', 'did', 'doing', 'a', 'an', 'the', 'and', 'but', 'if',  
, 'or', 'because', 'as', 'until', 'while', 'of', 'at', 'by', 'for', 'with',  
, 'about', 'against', 'between', 'into', 'through', 'during',  
, 'before', 'after', 'above', 'below', 'to', 'from', 'up', 'down', 'in',  
, 'out', 'on', 'off', 'over', 'under', 'again', 'further', 'then',  
, 'once', 'here', 'there', 'when', 'where', 'why', 'how', 'all', 'any',  
, 'both', 'each', 'few', 'more', 'most', 'other', 'some', 'such',  
, 'no', 'nor', 'not', 'only', 'own', 'same', 'so', 'than', 'too', 'very',  
, 's', 't', 'can', 'will', 'just', 'don', "don't", 'should', "should've",  
, 'now', 'd', 'll', 'm', 'o', 're', 've', 'y', 'ain', 'aren', "aren't",  
, 'couldn', "couldn't", 'didn', "didn't", 'doesn', "doesn't", 'hadn',  
, "hadn't", 'hasn', "hasn't", 'haven', "haven't", 'isn', "isn't", 'ma',  
, 'mightn', "mightn't", 'mustn', "mustn't", 'needn', "needn't", 'shan',  
, "shan't", 'shouldn', "shouldn't", 'wasn', "wasn't", 'weren',  
, "weren't", 'won', "won't", 'wouldn', "wouldn't"], e poi durante l’analisi toglierle dal corpus di riferimento.
```

Le parole che emergono maggiormente sono: “people”, “Amber Heard”, “Johnny Depp”, “victim”, “abuse”, “evidence” (figura 2).



Figura 2: Word-cloud parole più utilizzate

Per osservare in modo migliore il conteggio delle parole più utilizzate, è stato creato un istogramma che rappresentasse le prime dieci in maniera più schematica (grafico 4):

```
[('people', 299), ('abuse', 260), ('She', 229), ('think', 152), ('evidence', 148), ('domestic', 129), ('vs', 129), ('case', 126), ('court', 126), ('He', 124)].
```

Successivamente, dopo aver notato che nel grafico apparivano parole poco utili alla ricerca, esse sono state selezionate ed eliminate dal corpus. Queste sono:

```
("", "I", "And", "😓", "To", "via", "A", "In", "going", "New", "new", "@", "💩", "4", "movement.", "\u2060", "pants.", "really", "?", "don't", "also", "like", "said", "So", "so", "You", "even", "say", "abused", "get", "know", "see", "would", "\n\n", "It", "one", "-", ",", "Amber", "Depp", "Johnny", "Heard", "\nI", ".", "Vs", "VA", "The", "For", "Of", "\n", "This").
```

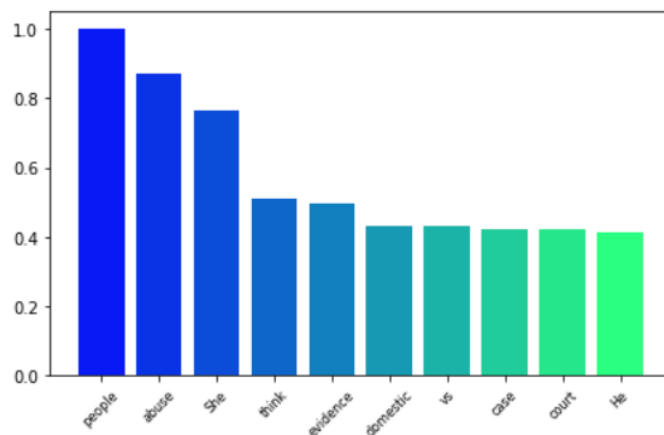


Grafico 4: istogramma dieci parole più utilizzate

Come analisi finale sono stati selezionati i dieci tweet con compound più positivo e dieci con compound più negativo e ne è emerso che i più negativi si riferiscono ad Amber Heard come “una abusatrice e una psicopatica”, mentre i più positivi ripongono fiducia e sostegno a Johnny Depp. Una difficoltà riscontrata in questo studio finale è stato che i tweet, grazie al filtraggio iniziale, non contenessero altri hashtag al loro interno, quindi, è risultato difficile comprendere i tweet a chi due attori si riferissero. Perciò è stata attuata una formula che facesse ricomparire tutti gli hashtag originali presenti.

B. La seconda parte dello studio in questo periodo temporale riguarda i tweet della seconda query, ovvero quelli contenenti gli hashtag #DomesticAbuse o #DomesticViolence e #JohnnyDepp o #AmberHeard o #JohnnyDeppVsAmberHeardTrial o #JohnnyDeppVsAmberHeard o #DeppVsHeard. Dopo il recupero e il primo filtraggio, è stata effettuata la *sentiment analysis* per i 474 tweet ottenuti.

In questo set di dati sono stati calcolati:

- Il valore medio di compound: -0.1260755274261604
- Il valore massimo di compound: 0.9781
- Il valore minimo di compound: -0.9719

Di seguito sono elencati i grafici ottenuti dalle analisi effettuate come per il caso precedente.

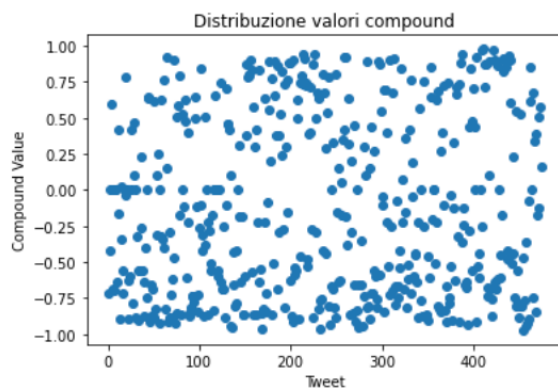


Grafico 5: Distribuzione dei valori di compound

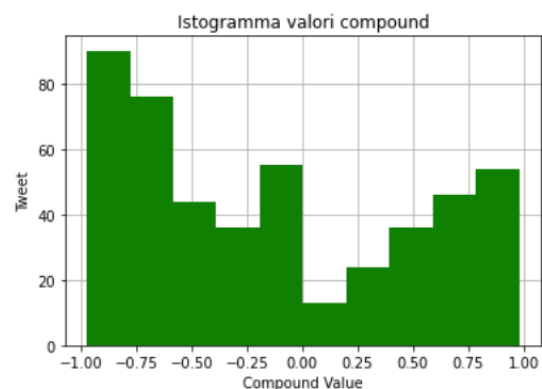


Grafico 6: Istogramma valori di compound

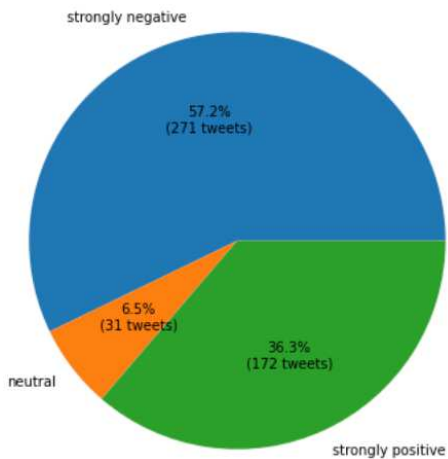


Grafico 7: Grafico a torta della suddivisione per categoria dei valori di compound



Figura 3: Word-cloud parole più utilizzate

Da questi grafici si nota che la maggior parte (il 57.2%) abbiano valore di compound negativo e che tra questi pochi hanno lo stesso valori. Infatti, come si può notare dal grafico 5, la distribuzione è circa omogenea all'interno del grafico, pochi punti si sovrappongono. Nella word-cloud, figura 3, risaltano le parole "victim", "support", "people", "women", "Amber Heard", "survivor" e "abuse".

Le prime dieci parole più utilizzate sono state: ('people', 52), ('She', 40), ('victims', 39), ('don't', 29), ('think', 27), ('support', 27), ('victim', 25), ('abuse', 24), ('women', 23), ('domestic', 21), rappresentate nel grafico seguente (grafico 8).

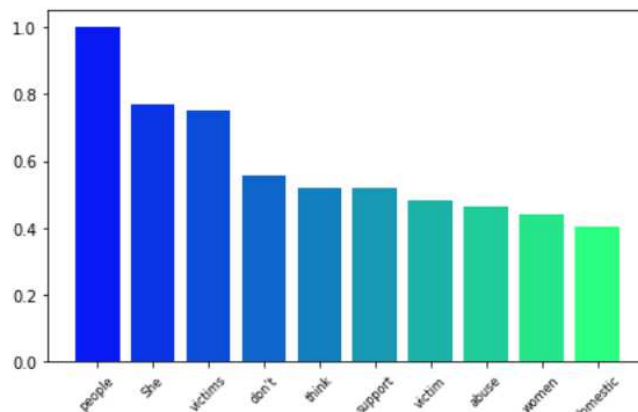


Grafico 8: istogramma dieci parole più utilizzate

Infine, sono stati selezionati i dieci tweet con compound più positivo e dieci con compound più negativo. È stato possibile notare che i più negativi si riferiscono ad Amber Heard come “una abusatrice, una bugiarda, una ladra e una criminale”, mentre i più positivi si riferiscono sempre ad Amber Heard però in una prospettiva di solidarietà femminile, aiuto da attivisti a attivista, presa di coscienza nei suoi confronti per i psicofarmaci di cui l’attrice abusa.

C. La terza parte dell’analisi riguarda la vicenda del ritrovamento delle feci nel letto di Johnny Depp e la query contiene i seguenti hashtag: #MePoo o #amberpoopedthebed o #Poop e #AmberHeard o #Poop e #JohnnyDepp. Siccome è una vicenda delineata e apparsa nel secondo periodo, quindi nel processo in Virginia del 2022, ci si aspetta che il numero di tweet in questo intero anno 2018-2019 siano pochi. Infatti, dopo aver effettuato il filtraggio, il numero totale di tweet ricavati è stato di 380.

In questo set di dati sono stati calcolati:

- Il valore medio di compound: 0.01632236842105263
- Il valore massimo di compound: 0.9776
- Il valore minimo di compound: -0.9503

Di seguito sono elencati i grafici ottenuti dalle analisi effettuate come per i casi precedenti.

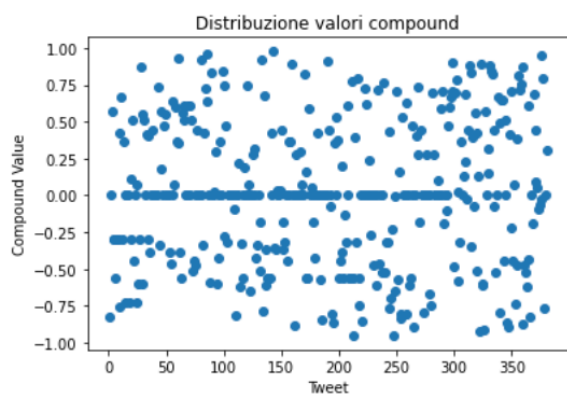


Grafico 9: Distribuzione dei valori di compound

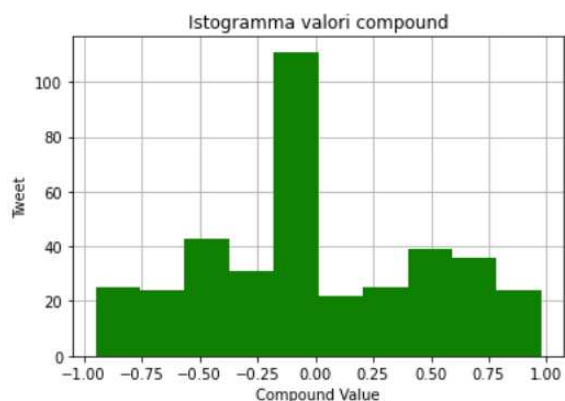


Grafico 10: Istogramma valori di compound

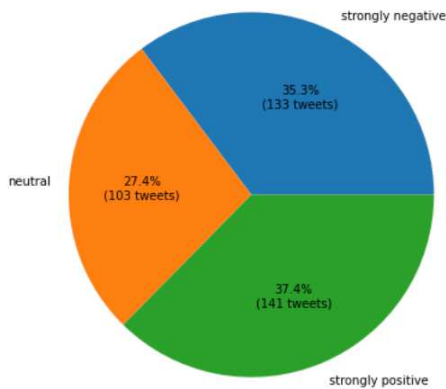


Grafico 11: Grafico a torta della suddivisione per categoria dei valori di compound



Figura 4: Word-cloud parole più utilizzate

Come si può evincere, dato il limitato numero di tweet scaricati, il grafico 9 rappresenta una distribuzione molto espansa con delle sovrapposizioni dei tweet nel valore medio di compound. Esso infatti è neutro. In aggiunta, sono presenti più valori di compound positivi, sebbene la percentuale di valori neutri e negativi non si allontani di molto da quella dei tweet positivi.

Dalla word-cloud (figura 4) si nota come le parole più utilizzate siano “pant”, “shit”, “movement”, “pooped”. Più precisamente, dal grafico seguente si osservano le dieci parole più utilizzate.

('shit', 48), ('pant', 46), ('movement', 14), ('poop', 14), ('Stop', 13), ('people', 13), ('pooped', 13), ('think', 13), ('grind', 10), ('much', 10).

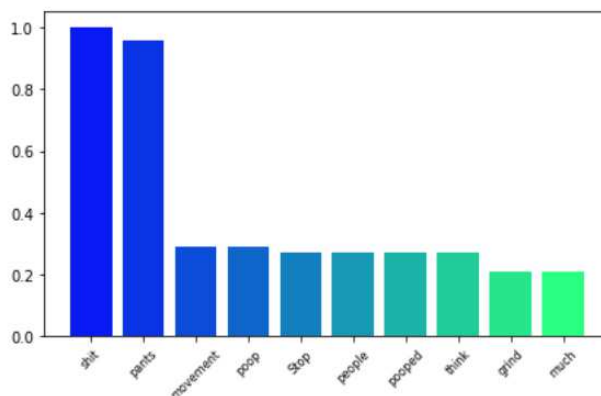


Grafico 12: istogramma dieci parole più utilizzate

Infine, sono stati selezionati i dieci tweet con compound più positivo e dieci con compound più negativo. È stato possibile notare che i più negativi si riferiscono a favore di Johnny Depp ma usando i termini con cui veniva definito inizialmente prima di avere delle prove che lo scagionassero. Il tweet più negativo infatti cita:

```
'Shame on you all who thought this wonderful man would be capable of something as horrible as sexually assaulting someone. Poor man lost his job and got his reputation destroyed. Ay only 50% of me thought you was a creepy ass sex offender!'
```

Ovvero:

```
'Vergogna a tutti voi che avete pensato che quest'uomo meraviglioso fosse capace di una cosa così orribile come l'aggressione sessuale. Il pover'uomo ha perso il lavoro e la sua reputazione è stata distrutta. Ay solo il 50% di me pensava che tu fossi un viscido maniaco sessuale!'
```

I tweet con valori di compound più positivi, invece, contengono un testo molto ironico e provocatorio sulla vicenda, quasi una presa in giro a riguardo.

3.4.2 Analisi dei tweet nel secondo arco temporale

Questa sezione contiene le analisi svolte per i tweet nel periodo temporale dal 1° aprile 2022 al 10 giugno 2022.

A. La prima parte vede lo studio dei tweet contenenti gli hashtag #AmberHeard #JohnnyDepp ed in più almeno uno tra #JohnnyDeppVsAmberHeard, #JohnnyDeppVsAmberHeardTrial o #DeppVsHeard (prima query).

Dopo il recupero e il filtraggio iniziale dei tweet è stata svolta la *sentiment analysis* per i 10.042 tweet ottenuti.

In questo set di dati sono stati calcolati:

- Il valore medio di compound: -0.007694264090818501
- Il valore massimo di compound: 0.9985
- Il valore minimo di compound: -0.9897

('trial', 608), ('She', 525), ('think', 366), ('people', 334), ('case', 296), ('court', 282), ('abuse', 274), ('believe', 265), ('He', 265), ('make', 249).

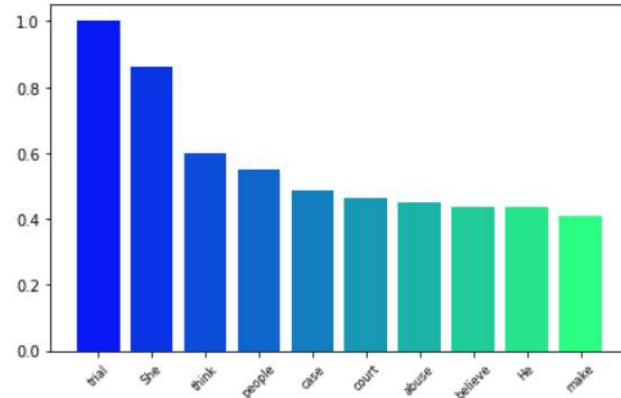


Grafico 16: istogramma dieci parole più utilizzate

Infine, sono stati selezionati i dieci tweet con compound più positivo e dieci con compound più negativo. È stato possibile notare che i più negativi si riferiscono contro Amber Heard, accusandola di essere una abusatrice narcisista, una finta vittima di violenza domestica e, quindi, una bugiarda.

I tweet con valori di compound più positivi, invece, contengono un testo ironico e a favore di Johnny Depp. Tra gli esempi più significativi vi sono questi tweet:

'We all know what a great actor Johnny is but after this trial we all know what a great, humble, kind and courteous person he is \nWE LOVE YOU JOHNNY DEPP'

Ovvero:

"Sappiamo tutti che Johnny è un grande attore, ma dopo questo processo sappiamo tutti che è una persona fantastica, umile, gentile e cortese".

E anche:

(Johnny can play himself but they'll need to get someone who can actually act to play Amber! 🤔🤔🤔🤔🤔🤔🤔🤔🤔🤔🤔🤔🤔🤔🤔🤔🤔', 0.9968)

Ovvero:

'Johnny può interpretare sé stesso, ma devono trovare qualcuno che sappia davvero recitare per interpretare Amber! 🤔🤔🤔🤔🤔🤔🤔🤔🤔🤔🤔🤔🤔🤔🤔🤔🤔🤔'.
 😏'.

B. La seconda parte dello studio in questo periodo temporale riguarda i tweet della seconda query, ovvero quelli contenenti gli hashtag #DomesticAbuse o #DomesticViolence e #JohnnyDepp o #AmberHeard o #JohnnyDeppVsAmberHeardTrial o #JohnnyDeppVsAmberHeard o #DeppVsHeard. Dopo il recupero e il primo filtraggio, è stata effettuata la *sentiment analysis* per i 3.819 tweet scaricati.

In questo set di dati sono stati calcolati:

- Il valore medio di compound: -0.1293809112333074
- Il valore massimo di compound: 0.9879
- Il valore minimo di compound: -0.9831

Di seguito sono elencati i grafici ottenuti dalle analisi effettuate come per i casi precedenti.

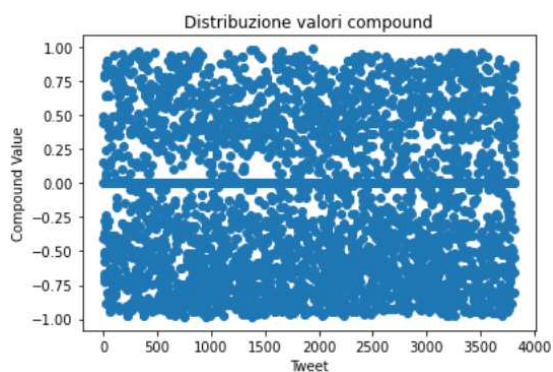


Grafico 17: Distribuzione dei valori di compound

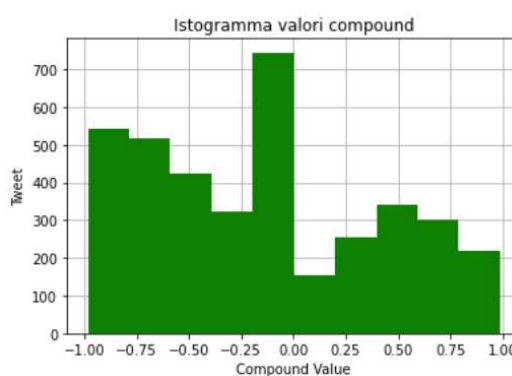


Grafico 18: Istogramma valori di compound

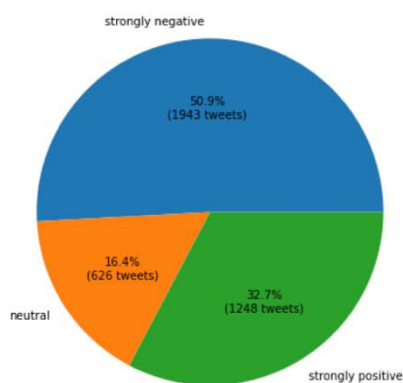


Grafico 19: Grafico a torta della suddivisione per categoria dei valori di compound

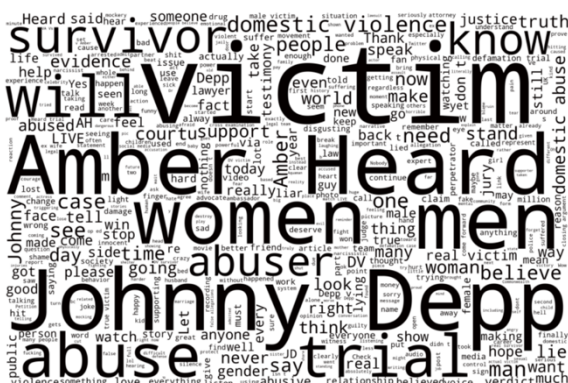


Figura 6: Word-cloud parole più utilizzate

Come si può osservare dal grafico 18, il valore medio di compound è negativo e vi è una maggior distribuzione di tweet negativi. Infatti, con il grafico 19 si è calcolato che i negativi siano il 50,9% mentre i positivi il 32,7%. Dalla word-cloud si evidenziano le parole “victim”, “Johnny Depp”, “Amber Heard”, “women”, “men”, “abuse” e “survivor”.

Piu specificatamente. Il conteggio delle parole più utilizzato è il seguente (grafico 20):

```
('victims', 366), ('domestic', 256), ('abuse', 245), ('men', 218), ('women', 214), ('She', 209), ('trial', 190), ('people', 159), ('believe', 138), ('violence', 137)
```

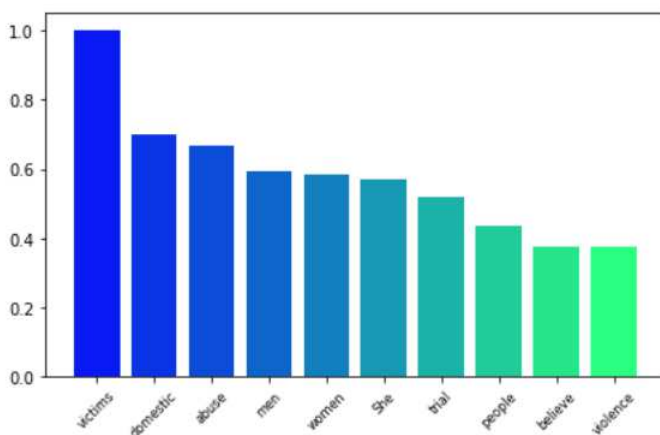


Grafico 20: istogramma dieci parole più utilizzate

In conclusione, i dieci tweet con compound più negativo definiscono la Heard come un aggressore manipolatore e violento. Molti utenti inoltre portano dei loro personali esempi di violenza domestica vissuta per dimostrare gli errori della attrice. Di seguito un esempio:

My dad used to beat my stepmom almost to death my brother I grew up having to pull him off her! He tried to kill her, kill himself threatened to kill all of us if we said anything! lying about has triggered me in a way I can't explain!!'

Ovvero:

Mio padre picchiava la mia matrigna quasi fino alla morte, mio fratello è cresciuto e io dovevo staccarlo da lei! Ha cercato di ucciderla, di uccidersi e ha minacciato di uccidere tutti noi se avessimo detto qualcosa! mentire mi ha scatenato in un modo che non riesco a spiegare!!!

Oppure alcuni utenti giudicano il modo di comportarsi e di esprimersi sempre della stessa in tribunale, argomentando che:

'TEXTBOOK NARCISSISTIC TENDENCY. ABUSE, THEN ACT INNOCENT THEN THREATEN WITH SUICIDE IF THE ABUSED VICTIM WANTS TO LEAVE. THIS IS A TEXTBOOK NARCISSISTIC ABUSER. EVERY WOMAN LIKE HER SHOULD BE IN PRISON.'

Ovvero:

"TENDENZA NARCISSISTICA DA TESTO. ABUSARE, POI FINGERSI INNOCENTE, POI MINACCIARE IL SUICIDIO SE LA VITTIMA DELL'ABUSO VUOLE ANDARSENE. QUESTO È UN ABUSATORE NARCISISTA DA MANUALE. TUTTE LE DONNE COME LEI DOVREBBERO ESSERE IN PRIGIONE"

Mentre, i dieci tweet con compound più positivo si soffermano sul fatto che anche gli uomini possono essere vittime di abusi e che non debba esserci la presunzione a priori di classificarli come abusatori e aggressori solo perché uomini. Uno di questi tweet recita:

'I'm ECSTATIC. DM goes both ways, and I am so thankful the conversation took the stage it did. Guys deserve to be heard be believed to o. They deserve justice too. HUGE win!'

Ovvero:

'Sono ECSTATICA. Il DM va in entrambe le direzioni e sono così grata che la conversazione abbia preso la piega che ha preso. Anche i ragazzi meritano di essere ascoltati e creduti. Anche loro meritano giustizia. ENORME vittoria!''.

C. La terza parte dell'analisi riguarda la vicenda del ritrovamento delle feci nel letto di Johnny Depp e la query contiene i seguenti hashtag: #MePoo o #amberpoopedthebed o #Poop e #AmberHeard o #Poop e #JohnnyDepp. Dopo aver effettuato il filtraggio, il numero totale di tweet ricavati è stato di 6361. Differentemente dal primo arco temporale, in questo il numero dei tweet scaricati è notevolmente maggiore dato che l'episodio è sorto durante questo processo.

In questo set di dati sono stati calcolati:

- Il valore medio di compound: 0.0016778022323533906
- Il valore massimo di compound: 0.9986
- Il valore minimo di compound: -0.9828

Di seguito sono elencati i grafici ottenuti dalle analisi effettuate come per i casi precedenti.

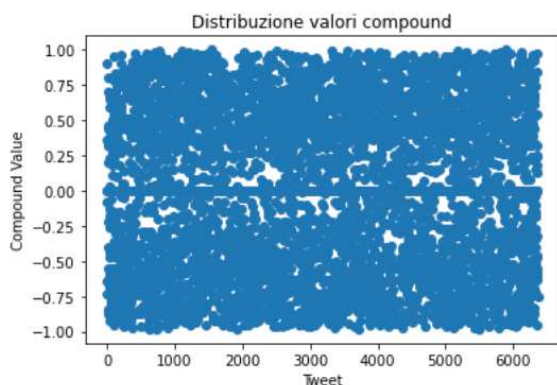


Grafico 21: Distribuzione dei valori di compound

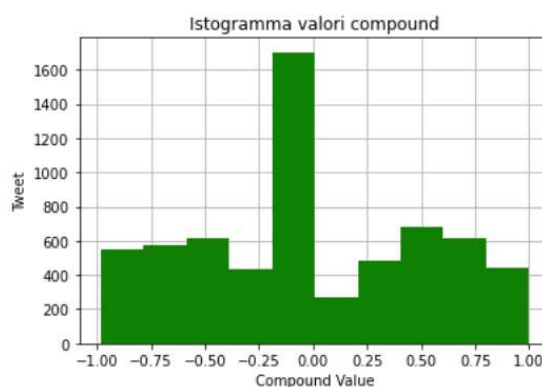


Grafico 22: Istogramma valori di compound

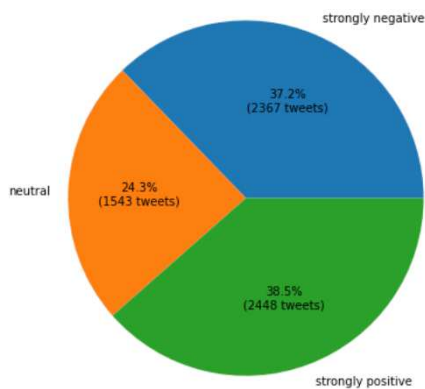


Grafico 23: Grafico a torta della suddivisione per categoria dei valori di compound

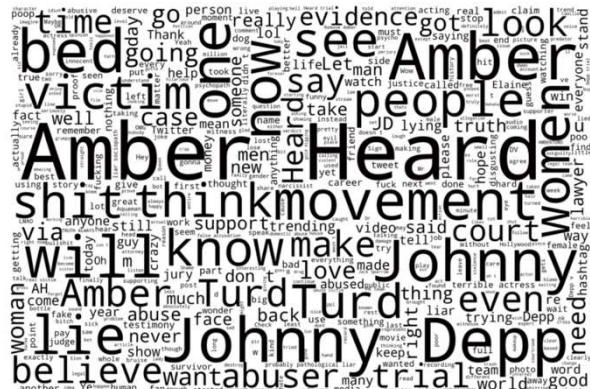


Figura 7: Word-cloud parole più utilizzate

Attraverso questi grafici, si può notare come il numero dei tweet con compound positivo sia lievemente maggiore di quelli negativi, infatti i primi sono il 38,5% mentre i secondi il 37,2% (grafico 23). Questo risultato si può dedurre anche dal grafico 21 in cui la distribuzione dei tweet per valori di compound è circa omogenea. Differentemente dai casi precedenti, è interessante osservare che nella word-cloud (figura 7) risaltino come parole più utilizzate molti termini e non solamente due o tre tra tutti. Quelli ad occhio, infatti, che hanno carattere maggiore sono: “Amber Heard”, “Johnny Depp”, “turd”, “victim”, “bed”, “shit”, “women”, “abuser” e “trial”.

Ma osserviamo in seguito nel dettaglio le parole più utilizzate ed il loro conteggio, ricordando del filtraggio per termini come effettuato in ognuno dei casi. Riepilogando, sono stati sottratti dal conteggio i seguenti termini:

("", "I", "And", "😂", "To", "via", "A", "In", "going", "New", "new", "@", "💩", "4", "movement.", "\u2060", "pants.", "really", "?", "don't", "also", "like", "said", "So", "so", "You", "even", "say", "abused", "get", "know", "see", "would", "\n\n", "It", "one", "-", " ", "Amber", "Depp", "Johnny", "Heard", "\nI", ".", "Vs", "VA", "The", "For", "Of", "\n", "This").

Dunque, i termini ricavati sono:

('She', 436), ('Turd', 210), ('people', 185), ('think', 169), ('bed', 167), ('women', 165), ('movement', 160), ('shit', 156), ('trial', 142), ('abuse', 135)

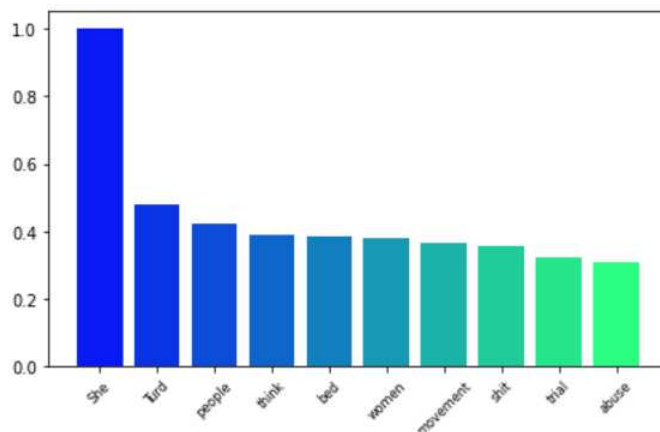


Grafico 24: istogramma dieci parole più utilizzate

Infine, i dieci tweet con compound più negativo si riferiscono alla Heard come una bugiarda che merita il carcere, una signora sociopatica che annega nelle sue lacrime di cocodrillo. In molti inoltre ritengono che il fatto della defecazione nel letto di Depp sia uno di quegli atti a cui conseguiranno altri abusi.

I dieci tweet con compound più positivo, invece, sono ricchi di emoji come 😂, 🤔, 💕 e sono carichi di ironia. In essi, infatti, si ritrovano frasi che prendono in giro il consiglio, i legali di Amber Heard e la stessa. Eccone un esempio:

Objection... making objection on your own question? Ambers council is going mental from this trial.
 😂😂😂😂😂😂😂😂😂😂😂😂😂😂😂😂😂😂😂😂😂😂😂😂😂😂

Ovvero:

Obiezione... fare obiezione alla tua stessa domanda? Il consiglio di Amber sta diventando matto per questo processo.
 😂😂😂😂😂😂😂😂😂😂😂😂😂😂😂😂😂😂😂😂😂😂😂😂😂😂

3.4.3 Analisi dei teweet nel terzo arco temporale

Questa sezione contiene le analisi svolte per i tweet pubblicati nel terzo periodo temporale, ovvero sia dall'11 giugno 2022 all'11 settembre 2022. Questo per osservare e comprendere se, e come, gli utenti di Twitter che hanno seguito le vicende del processo abbiano mantenuto la loro opinione o il contrario dopo l'uscita del verdetto finale e, capire come questi abbiano comunicato ed espresso i loro pensieri a distanza di giorni.

A. La prima parte vede lo studio dei tweet contenenti gli hashtag #AmberHeard #JohnnyDepp ed in più almeno uno tra #JohnnyDeppVsAmberHeard, #JohnnyDeppVsAmberHeardTrial o #DeppVsHeard (prima query).

Dopo il recupero e il filtraggio iniziale dei tweet è stata svolta la *sentiment analysis* per i 10.508 tweet ottenuti.

In questo set di dati sono stati calcolati:

- Il valore medio di compound: -0.04928676722497213
- Il valore massimo di compound: 0.9926
- Il valore minimo di compound: -0.9813

Di seguito sono elencati i grafici ottenuti dalle analisi effettuate come per i casi precedenti.

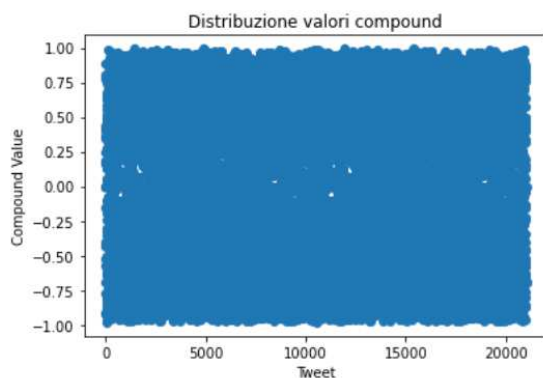


Grafico 25: Distribuzione dei valori di compound

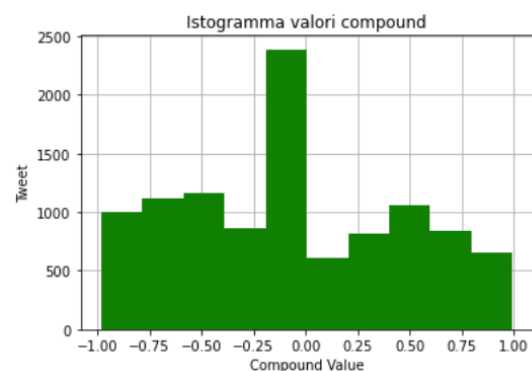


Grafico 26: Istogramma valori di compound

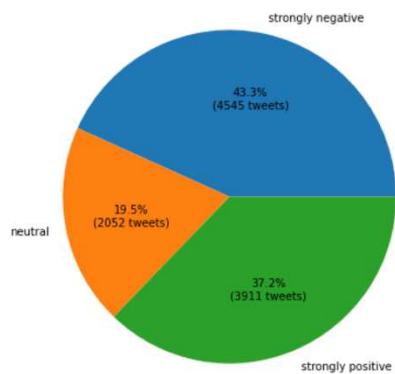


Grafico 27: Grafico a torta della suddivisione per categoria dei valori di compound

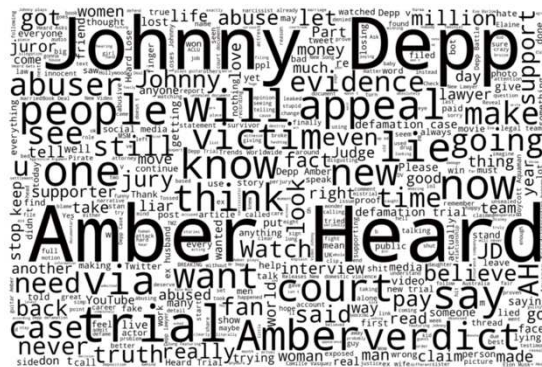


Figura 8: Word-cloud parole più utilizzate

In questo studio, la distribuzione dei valori di compound (grafico 25) non è molto utile perché, data la grande mole di tweet recuperati, essi sono distribuiti omogeneamente all'interno del grafico. Dal grafico 27, però, è più facile evincere quali siano le percentuali di ogni categoria di polarità: il 43,3% dei tweet ha compound negativo, il 37,2% positivo e il 19,5% neutro. Le parole che spiccano maggiormente nella word-cloud sono ridotte rispetto ai casi precedenti e sono: “Amber Heard”, “Johnny Depp”, “trial”, “verdict”, “court”.

Più precisamente, dopo aver compiuto il successivo filtraggio per rimuovere le parole più generali, le dieci più utilizzate sono state (grafico 28):

('She', 658), ('trial', 647), ('people', 373), ('defamation', 366), ('court', 314), ('case', 314), ('think', 309), ('still', 281), ('AH', 273), ('never', 269)

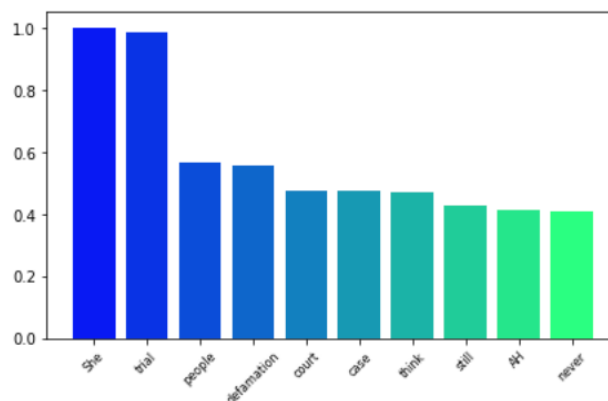


Grafico 28: istogramma dieci parole più utilizzate

Mentre, per quanto riguarda i dieci tweet con compound più negativo e più positivo, si può evincere che: nei primi si nota come gli utenti siano accaniti contro l'attrice, considerandola una abusatrice bugiarda e pericolosa che merita la galera. Molti sostengono che abbia rovinato la vita di Johnny Depp e altri sono indignati del fatto che la Heard abbia ottenuto una intervista dai media nonostante sia stata classificata come colpevole e abusatrice. Il post con compound più negativo, infatti, rimarca proprio questo fatto:

"HOW DOES A WOMAN, PROVEN TO BE AN ABUSER IN A COURT OF LAW, GET A F*ING MAINSTREAM MEDIA INTERVIEW??? SHE IS GUILTY OF TRYING TO DESTROY A MAN'S LIFE. is the victim! NOT. YOU ARE GIVING A GUILTY ABUSER AN INTERVIEW. YOURE HELPING AN ABUSER SPREAD MORE LIES"

Ovvero:

"COME FA UNA DONNA, CHE HA DIMOSTRATO DI ESSERE UN'ABUSATRICE IN UN TRIBUNALE, A OTTENERE UNA FOTTUTA INTERVISTA DAI MEDIA MAINSTREAM?). È COLPEVOLE DI AVER TENTATO DI DISTRUGGERE LA VITA DI UN UOMO. è la vittima! NON... STATE CONCEDENDO UN'INTERVISTA A UN COLPEVOLE DI ABUSI. STATE AIUTANDO UNA ABUSATRICE A DIFFONDERE SEMPRE PIU' BUGIE"

Per quanto concerne, invece, i dieci tweet con compound più positivo, molti di questi fanno riferimento ai meme e alle clip diventate virali nel web che ritraggono Amber mentre fa finta di piangere. Altri, invece, sono sempre ironici ma si riferiscono al fatto che questo processo abbia reso ancora più famoso e ricco Johnny Depp.

B. La seconda parte dello studio in questo periodo temporale riguarda i tweet della seconda query, ovvero quelli contenenti gli hashtag #DomesticAbuse o #DomesticViolence e #JohnnyDepp o #AmberHeard o #JohnnyDeppVsAmberHeardTrial o #JohnnyDeppVsAmberHeard o #DeppVsHeard.

Dopo il recupero e il primo filtraggio, è stata effettuata la *sentiment analysis* per gli 827 tweet scaricati. È interessante notare già in questo step, come gli utenti non si siano più focalizzati su questo tema ma, come si è notato con la query precedente, sul fatto principale (il processo di per sé) e il suo verdetto.

negativa dei grafici. Infatti, i tweet con compound negativo sono il 47,5% mentre quelli positivi solamente il 28,5%.

È doveroso evidenziare che nella word-cloud (figura 9) non spicchino più i nomi dei due attori come parole maggiormente utilizzati, ma termini come “victim”, “men”, “abuser”, “women”, “part”, “violence” e “domestic”.

Più specificatamente e successivamente alla rimozione dei termini che possiamo definire “neutri”, le parole più utilizzate sono state infatti (grafico 32):

('Part', 66), ('victim', 63), ('men', 58), ('She', 54), ('domestic', 40), ("I'm", 38), ('Deposition', 36), ('AH', 33), ('male', 33), ('women', 31)

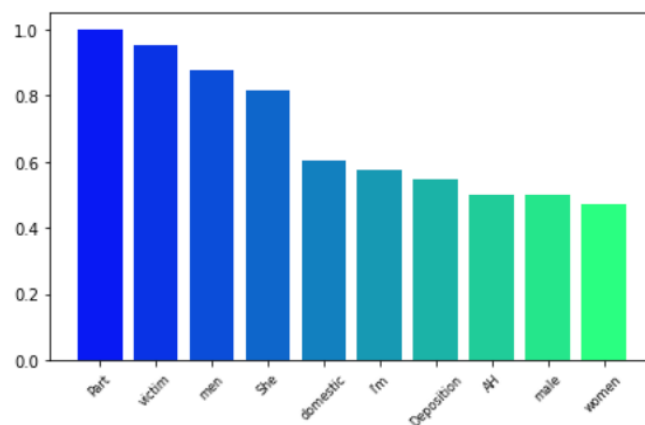


Grafico 32: istogramma dieci parole più utilizzate

Per quanto concerne i dieci tweet con compound più negativo e più positivo, si può affermare che: nei primi viene rimarcato il fatto che sia stata data la possibilità ad Amber Heard di raccontare le sue bugie in tribunale e pubblicamente, cosa che per i veri sopravvissuti agli abusi domestici è stata vista come una minimizzazione dell'accaduto; nei secondi gli utenti richiamano le persone al coraggio di dire la verità, di denunciare se ricevono abusi parlando ad una comunità di vittime e sostenendo che l'onestà e la giustizia trionfano anche se la vittima è un uomo.

C. La terza parte dell'analisi riguarda la vicenda del ritrovamento delle feci nel letto di Johnny Depp e la query contiene i seguenti hashtag: #MePoo o #amberpoopedthebed o #Poop e #AmberHeard o #Poop e #JohnnyDepp. Dopo aver effettuato il filtraggio, il numero totale di tweet scaricati è stato di 1505.

In questo set di dati sono stati calcolati:

- Il valore medio di compound: -0.009252093023255806
- Il valore massimo di compound: 0.9922
- Il valore minimo di compound: -0.9856

Di seguito sono elencati i grafici ottenuti dalle analisi effettuate come per i casi precedenti.

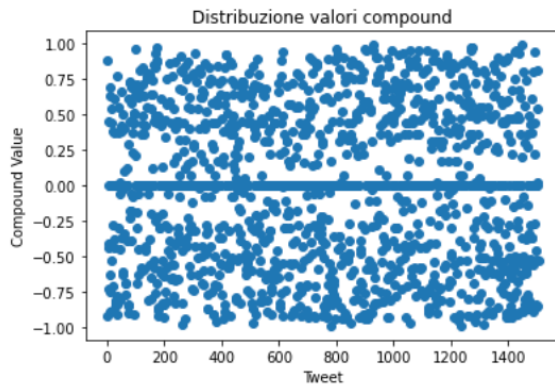


Grafico 33: Distribuzione dei valori di compound

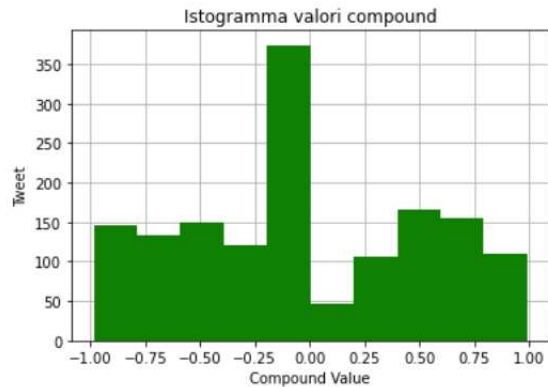


Grafico 34: Istogramma valori di compound

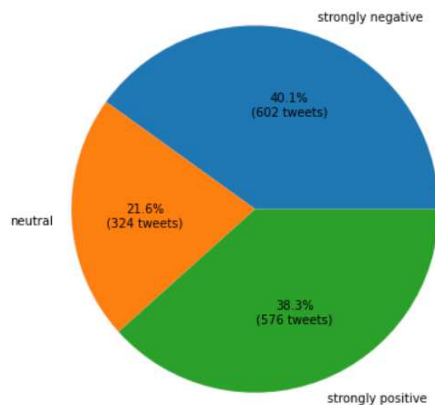


Grafico 35: Grafico a torta della suddivisione per categoria dei valori di compound

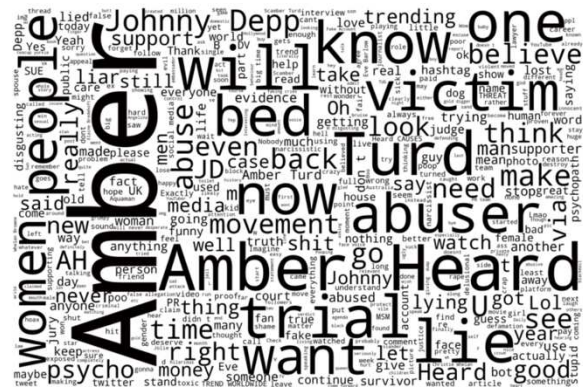


Figura 10: Word-cloud parole più utilizzate

Dai grafici 33 e 34 si può dedurre che il numero di tweet con compound negativo e positivo sia all'incirca simile; infatti, si nota una distribuzione di essi equilibrata. La conferma viene data dal grafico 35 dove è stato calcolato che la percentuale dei tweet negativi è del 40,1% mentre quella dei positivi poco inferiore, il 38,3%. Anche in questo caso, non vi sono parole che spicchino più di altre nella word-cloud (figura 10) ma tra queste vi sono: “Amber”, “turd”, “bed”, “lie”, “victim”, “abuser” e “trial”. Precisamente, dal grafico 36 si possono osservare le dieci parole più frequenti ed il loro conteggio:

('She', 105), ('Turd', 69), ('people', 49), ('want', 46), ('trial', 42), ('lies', 41), ('make', 40), ('bed', 39), ('women', 37), ('media', 36)

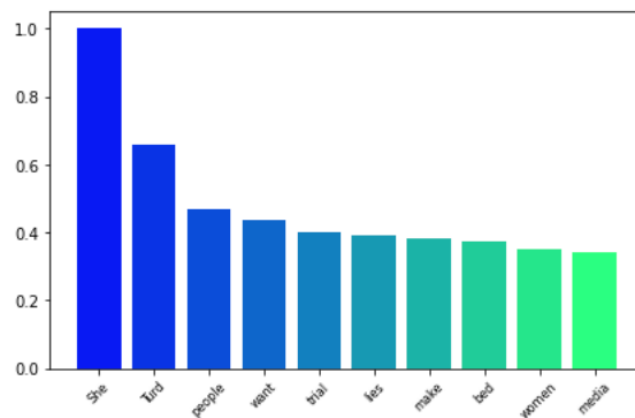


Grafico 36: istogramma dieci parole più utilizzate

Successivamente, come analisi finale sono stati studiati i dieci tweet con compound più negativo e positivo. I primi si riferiscono sempre alla Heard con toni e termini negativi. Di seguito alcuni esempi:

'We Heard lies for YEARS!!!! She lost because she LIED! Shes a liar , abuser, thief, criminal and a FRAUD!!!!\n😞\U0001f978🤢🤮🤢🤮'

Ovvero:

"Abbiamo sentito bugie per ANNI!!!! Ha perso perché ha mentito! È una bugiarda, un'abusatrice, una ladra, una criminale e una FRODE!!!!\n😞\U0001f978🤢🤮🤢🤮"

"Someone with control issues has no control herself. Weird"

Ossia:

"Una persona con problemi di controllo non ha il controllo di sé stessa. Strano"

'Never trust Amber Heard supporters! This moronic flying monkey admits to having 8 fake accounts to spread heard lies and misinformation. Sadly pathetic!'

Tradotto:

"Mai fidarsi dei sostenitori di Amber Heard! Questa stupida scimmia volante ammette di avere 8 account falsi per diffondere bugie e disinformazione. Tristemente patetico"

Per quanto riguarda, invece, i tweet con compound positivo, gli utenti esprimono molto divertimento riguardo alla faccenda e che dopo questo episodio amano ancora di più Johnny Depp. Inoltre, i tweet sono carichi di emoji divertite come 😂, 😊, 😄 e 🍌.

3.4.4 Confronto tra le tre sezioni

Questa ultima analisi vede il confronto tra i tweet della stessa query ma nei tre archi temporali differenti, dal 1° dicembre 2018 al 31 dicembre 2019, dal 1° aprile 2022 al 10 giugno 2022 e dall'11 giugno 2022 all'11 settembre 2022.

A. Questa parte riguarda i tweet della query #AmberHeard #JohnnyDepp ed in più almeno uno tra #JohnnyDeppVsAmberHeard, #JohnnyDeppVsAmberHeardTrial o #DeppVsHeard.

Tweet tot filtrati	Valore medio di compound	Valore max di compound	Valore min di compound
3647	-0.15410967918837393	0.9943	-0.9867
10042	-0.007694264090818501	0.9985	-0.9897
10508	-0.04928676722497213	0.9926	-0.9813

Tabella 3: tweet prima query

Dalla tabella 3 si può notare come nel primo periodo, dal 1° dicembre 2018 al 31 dicembre 2019, le prime accuse e vicende che vede i due attori come protagonisti, non fossero

diventate molto virali e discusse su Twitter. Infatti, il numero di tweet ricavati in un anno intero è solo di 3647, mentre nei seguenti due periodi di durata di circa due mesi ne sono stati raccolti più di 10 mila.

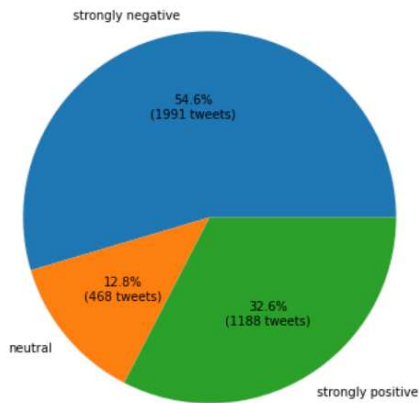


Grafico 3

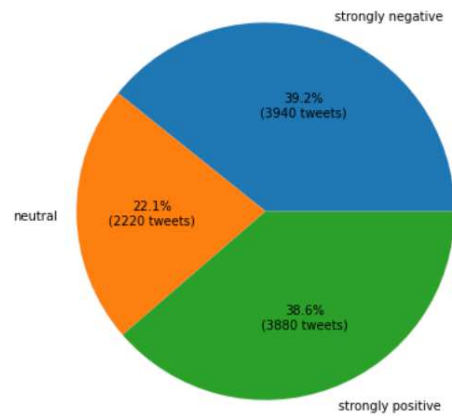


Grafico 15

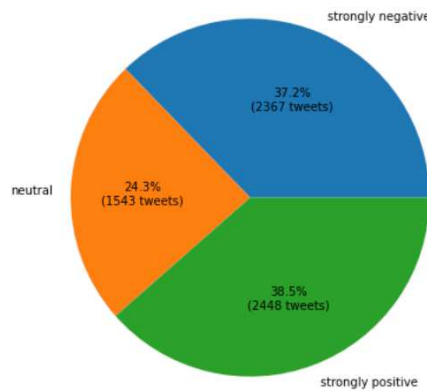


Grafico 23

Un altro aspetto che sottolinea il fatto che inizialmente vi fossero poche prove a favore di Johnny Depp o Amber Heard e che le vicende non fossero quindi molto approfondite, è il fatto che la percentuale di tweet con compound negativo nel primo periodo è più della metà rispetto agli altri due. Quindi si può presumere che gli utenti esprimessero le loro opinioni impulsivamente in base ai fatti emersi.

Per quanto concerne le parole più utilizzate all'interno dei tweet (figura 2, figura 5 e figura 8), si può notare che in tutti e tre i periodi ricorrono gli stessi termini come i nomi dei due attori e “victim”, “people”, “abuse”.

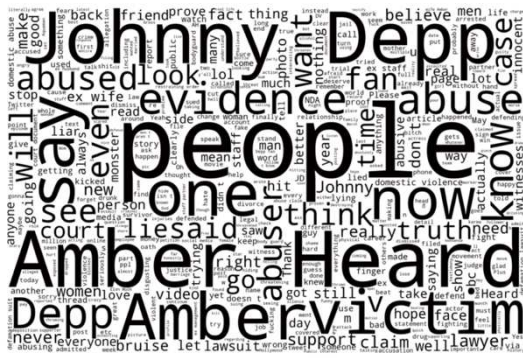


Figura 2

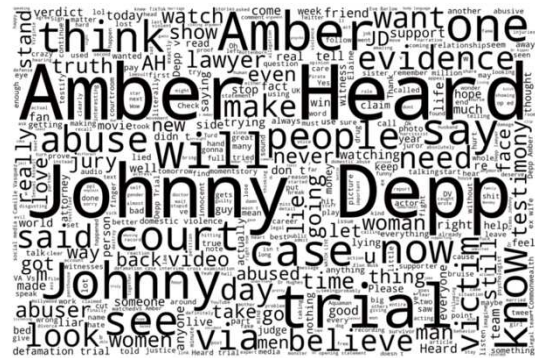


Figura 5

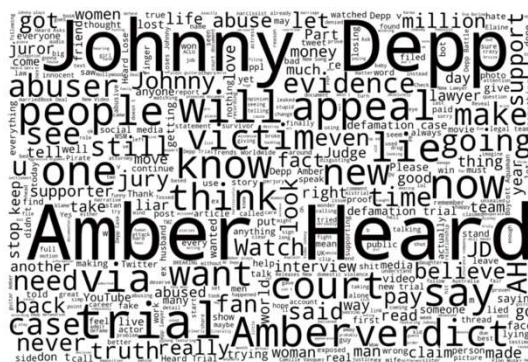


Figura 8

Più precisamente, però, nel secondo e nel terzo arco temporale emergono parole inerenti al processo come “trial”, “defamation”, “court” e “case” (grafico 4, grafico 16, grafico 28).

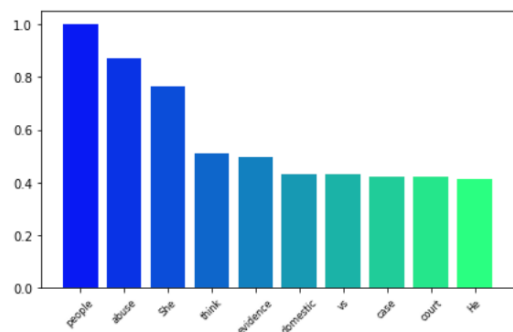


Grafico 4

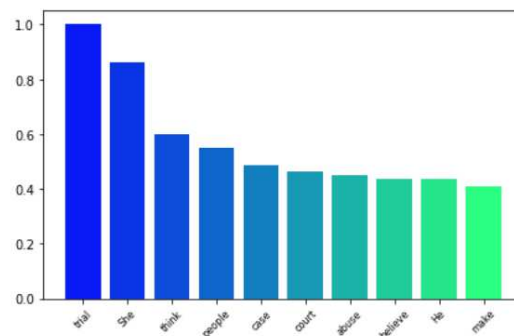


Grafico 16

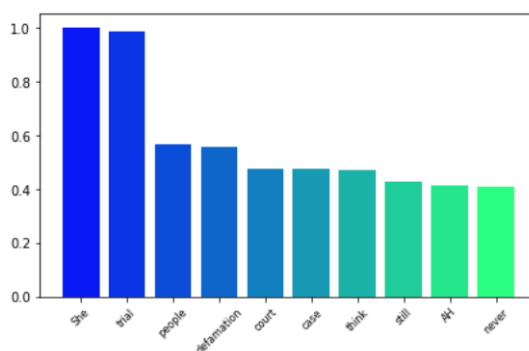


Grafico 28

Si può perciò concludere i tweet del secondo e terzo periodo con tale query siano più precisi in termini di terminologia, dovuto dalla presa di coscienza, dalla riflessione e dalle prove confutanti l'attrice.

B. La seconda parte di confronto riguarda i tweet con query #DomesticAbuse o #DomesticViolence e #JohnnyDepp o #AmberHeard o #JohnnyDeppVsAmberHeardTrial o #JohnnyDeppVsAmberHeard o #DeppVsHeard.

Anche in questo caso, dalla tabella 4 si può osservare come nel primo periodo il numero totali di tweet sia nettamente inferiore rispetto agli altri due. In aggiunta però si noti come anche nel terzo periodo si siano recuperati pochi tweet. Senza confrontare i tre periodi, si può affermare che anche nel secondo periodo, quello in pieno processo, il numero di tweet sia molto basso. Questo potrebbe essere dovuto da una scorretta scelta di combinazione degli hashtag o per il fatto che gli utenti si siano concentrati su altri topic.

Tweet tot filtrati	Valore medio di compound	Valore max di compound	Valore min di compound
474	-0.1260755274261604	0.9781	-0.9719
3819	-0.1293809112333074	0.9879	-0.9831
827	-0.13107424425634842	0.9725	-0.9759

Tabella 4: tweet seconda query

Per quanto riguarda il valore di compound dei tweet nei tre casi, le percentuali non si allontanano molto uno dall'altro. In tutti e tre, infatti, la maggior parte dei tweet ha compound negativo e si differenziano soltanto per un massimo del 10%.

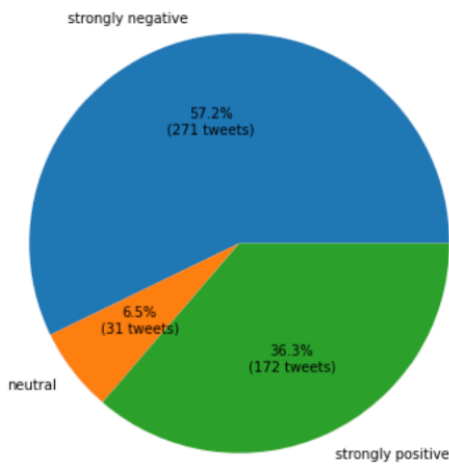


Grafico 7

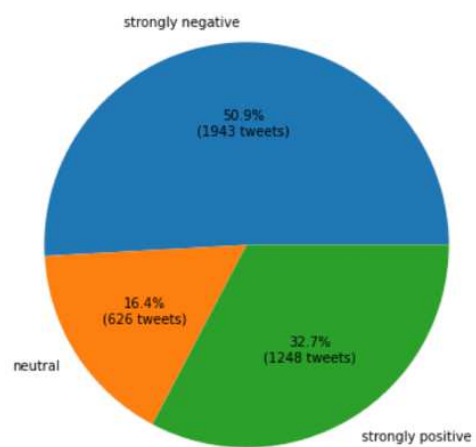


Grafico 19

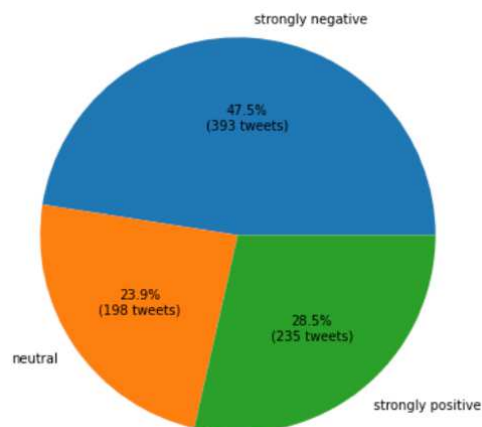


Grafico 31

Per quanto concerne le parole più utilizzate (figura 3, figura 6, figura 9) spiccano termini inerenti ad atti di violenza come “survivor”, “victim”, “men”, “women”, “support” ed “abuser”.



Figura 3

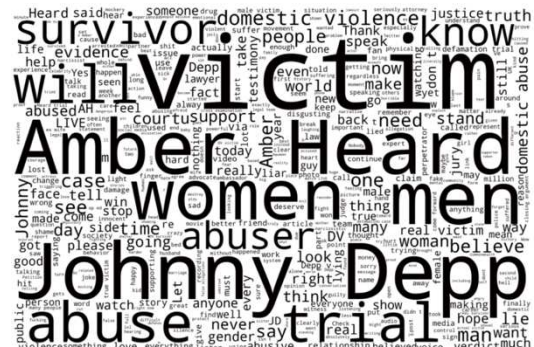


Figura 6

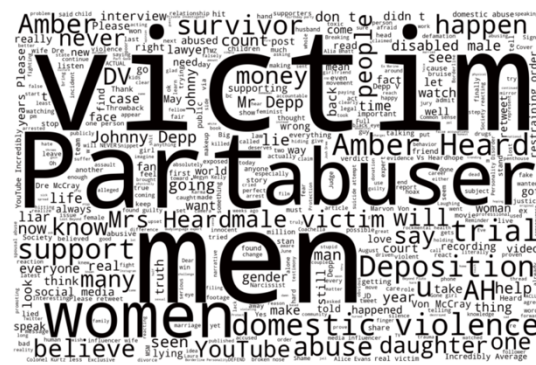


Figura 9

Precisamente, visibile nei grafici 9, 20 e 32, i termini maggiormente utilizzati nei tre periodi sono sempre ricorrenti.

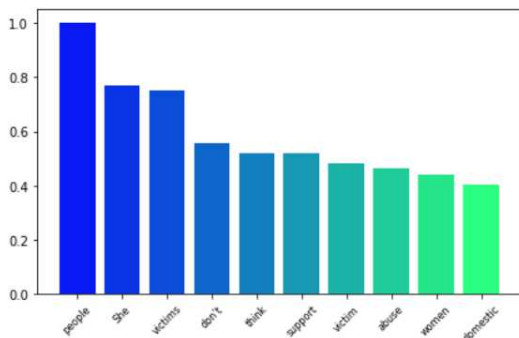


Grafico 9

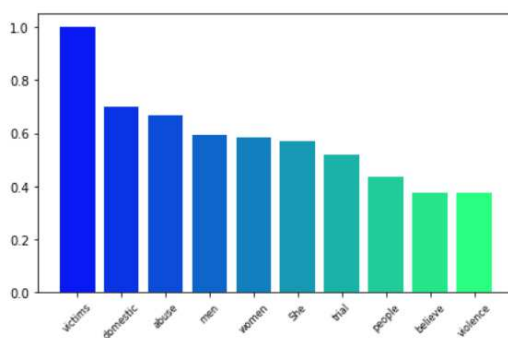


Grafico 20

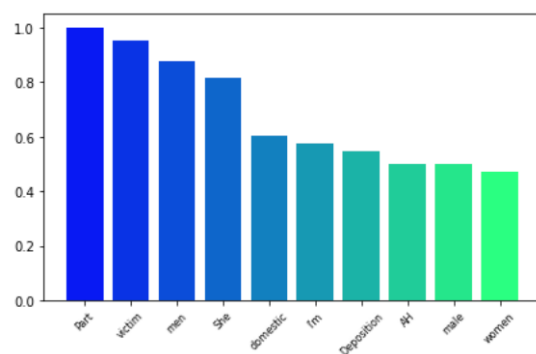


Grafico 32

C. La terza parte di confronto vede lo studio dei tweet contenenti gli hashtag riguardanti l'episodio delle feci di Amber Heard nel letto Johnny Depp, ovvero #MePoo o #amberpoopedthebed o #Poop e #AmberHeard o #Poop e #JohnnyDepp.

La tabella 5 mostra che i tweet scaricati nel primo periodo, da sottolineare nuovamente di un anno intero, sono solamente 380, diversamente dal secondo, periodo in cui è uscita pubblicamente e diventata virale anche nei social network la vicenda, e 1505 nel terzo.

Tweet tot filtrati	Valore medio di compound	Valore max di compound	Valore min di compound
380	0.01632236842105263	0.9776	-0.9503
6361	0.0016778022323533906	0.9986	-0.9828
1505	-0.009252093023255806	0.9922	-0.9856

Tabella 5: tweet terza query

Per quanto riguarda il valore di compound dei tweet nei tre casi, le percentuali sono molto simili. L'unica differenza risiede nel fatto che nel primo e nel secondo periodo i valori di compound positivi sono di percentuale maggiore, mentre nel terzo periodo quelli negativi.

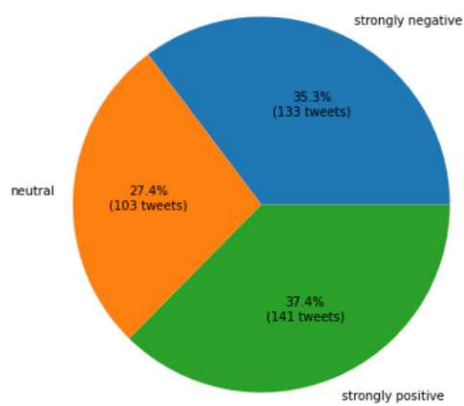


Grafico 11

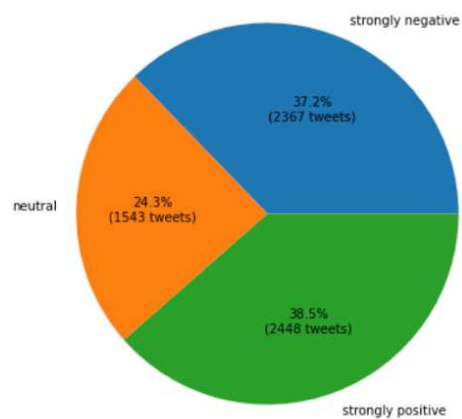


Grafico 23

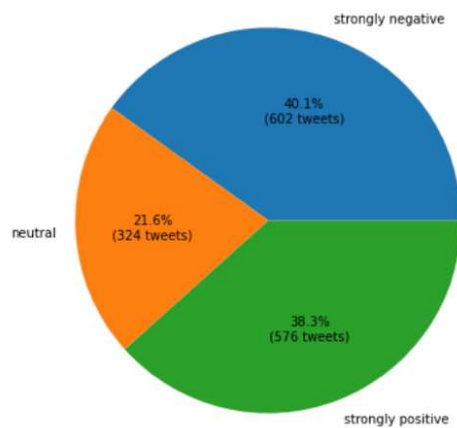


Grafico 35

Le parole più utilizzate, non considerando i nomi dei due attori, sono diverse in ognuno dei tre archi temporali. Nella figura 4, quindi nel primo periodo, emergono “shit”, “pants”, “shit” e “movement”. Nella figura 7, secondo periodo, si evidenziano poche parole significando che molte sono state scritte con frequenza simile. Si notato: “movement”, “turd”, “bed” e “victim”. Nella figura 10, terzo periodo, si sottolineano “turd”, “victim”, “abuser”, “bed” e “lie”.



Figura 4

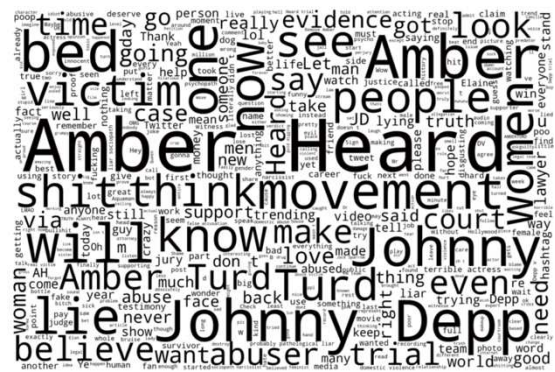


Figura 7

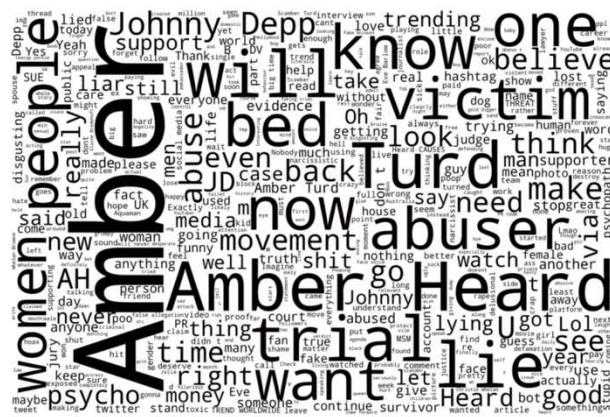


Figura 10

Grazie al conteggio delle dieci parole più frequenti, è interessante notare come nei primi due periodi queste si riferiscano all'atto in sé, alla vicenda, mentre nel terzo si fa più riferimento alle opinioni degli utenti nei confronti della Heard dopo il verdetto finale.

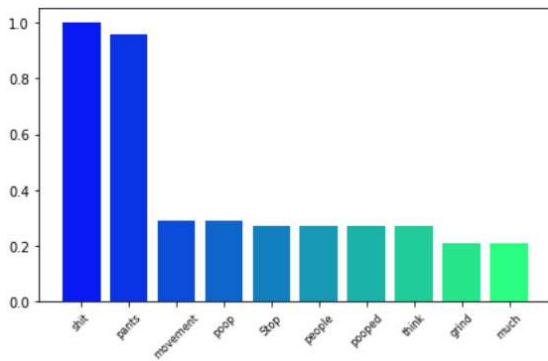


Grafico 12

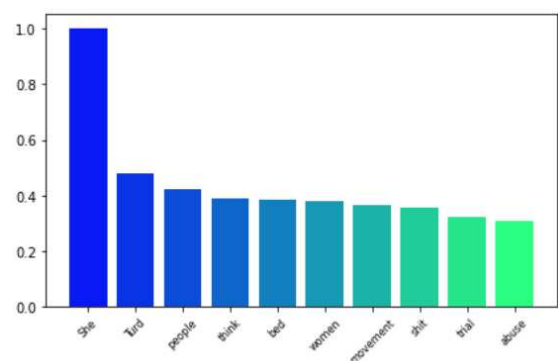


Grafico 24

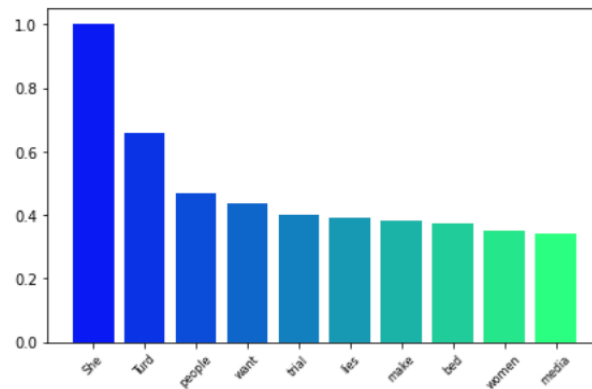


Grafico 36

Per entrambe le tre sezioni si può concludere che, avendo osservato precedentemente per ogni periodo e ogni query i dieci tweet con compound più negativo e più positivo, la maggior parte degli utenti sin dal 2018 abbiano sostenuto Johnny Depp come innocente e che le azioni che compiva fossero state delle conseguenze dovute dagli abusi della ex moglie Amber Heard, definita sempre in questi tweet come una abusatrice bugiarda e sociopatica che merita il carcere. I tweet con compound più positivo sono inoltre stati scritti con molta ironia e ricchi di emoji in cui l'attrice veniva derisa.

Conclusioni

Attraverso Twitter, le sue API e Python, è stato possibile comprendere come, in particolare con precisi filtri e scelte di raccolta dati, gli utenti di Twitter si siano espressi in merito ai processi per diffamazione Johnny Depp vs Amber Heard diventati virali nel mondo del Web e dei social network, toccando anche temi importanti come la violenza domestica, la violenza sulle donne e sugli uomini, e gli abusi perpetrati dai propri partner. Con lo scopo di comprendere come fossero caratterizzati lessicalmente e semanticamente i tweet degli utenti, nel corso di questo elaborato si è voluto analizzare le loro caratteristiche più rilevanti.

Inizialmente, il primo capitolo mirava a descrivere le battaglie legali a cui hanno preso parte i due attori Hollywoodiani fin dagli esordi del loro matrimonio iniziato nel 2015 e concluso solamente quindici mesi dopo. La loro è stata una relazione molto travagliata, caratterizzata da violenze fisiche e verbali, abusi di alcol e droghe e da accuse di diffamazione attraverso interviste uscite in giornali molto famosi in America. Data la popolarità dei due attori, il fatto che i processi fossero in streaming live e che questi contenessero testimonianze ed accuse a tratti divertenti nonostante i temi molto delicati, questi sono diventati, nel Web e nel mondo dei social network, molto virali raggiungendo persone da tutto il mondo. Basti pensare ad esempio che solo nel mese di aprile su TikTok, ovvero pochi giorni dopo l'inizio del processo in Virginia, l'hashtag #justiceforjohnnydepp aveva oltre 18 miliardi di visualizzazioni, mentre l'hashtag #justiceforamberheard ne aveva 39 milioni. Addirittura, i video in cui Amber Heard piange durante l'interrogatorio sono stati utilizzati su TikTok per ridicolizzarla ritenendo che fosse tutta una finzione con lo scopo di guadagnare soldi e visibilità. Su Snapchat è stato pure creato un filtro che deforma il volto degli utenti in una smorfia di pianto ridicola, soprannominato come il "filtro di Amber Heard". Ma è su TikTok che questo filtro ha riscosso più successo: ad oggi esso conta già più di 117 milioni di visualizzazioni.

Successivamente, nel secondo capitolo dell'elaborato sono state riportate dettagliatamente le caratteristiche di Twitter e i motivi per cui è sempre più utilizzato per compiere *data mining*. Un'ulteriore descrizione utile alla comprensione dei passaggi effettuati per recuperare i dati dalla piattaforma, è quella delle sue tecnologie API, ovvero le Interfacce di Programmazione delle Applicazioni. In seguito, per comprendere i

passaggi e le tecniche per la creazione del codice e l'elaborazione dei dati scaricati è stata descritta la *Sentiment Analysis* e le fasi che la contraddistinguono, e il linguaggio di programmazione Python con annesse tutte le librerie e i pacchetti utili impiegati in seguito.

Nel terzo capitolo sono state illustrate dettagliatamente tutte le fasi della metodologia per eseguire la *sentiment analysis* dei tweet, ovvero:

1. la collezione o recupero dei tweet in forma originale e grezza grazie all'API di Twitter; la scelta delle variabili di interesse come il periodo temporale in cui si desidera che i tweet siano stati pubblicati, gli hashtag che tali tweet devono contenere e la lingua in cui essi devono essere scritti;

2. la preelaborazione dei tweet raccolti per renderli più "puliti" e semplici per le analisi successive come la rimozione dei re-tweet, di hashtag non utili nel testo, di tweet doppi, di caratteri speciali di HTML inutili alla ricerca...;

Si è potuto notare come i tweet grezzi iniziali fossero 42.425 mentre una volta puliti 37.563;

3. l'analisi dei risultati. Inizialmente sono stati studiati i tweet ricavati nel medesimo periodo temporale per ognuna delle tre query, ovvero la prima conteneva gli hashtag #AmberHeard #JohnnyDepp ed in più almeno uno tra #JohnnyDeppVsAmberHeard, #JohnnyDeppVsAmberHeardTrial o #DeppVsHeard, la seconda #DomesticAbuse o #DomesticViolence e #JohnnyDepp o #AmberHeard o #JohnnyDeppVsAmberHeardTrial o #JohnnyDeppVsAmberHeard o #DeppVsHeard ed infine la terza #MePoo o #amberpoopedthebed o #Poop e #AmberHeard o #Poop e #JohnnyDepp.

Per il primo periodo, dal 1° dicembre 2018 al 31 dicembre 2019, si è potuto notare come le prime due query contenessero tweet più negativi, a differenza della terza in cui la percentuale maggiore risiedeva in quelli positivi. Questo perché la terza query corrisponde ad una accusa e una vicenda che agli occhi degli utenti era ritenuta sconcertante ma in modo bizzarro ed ironico, mentre le prime due riguardavano temi generali come la violenza domestica e le accuse di per sé. In ognuno dei tre casi, però, gli utenti si sono espressi positivamente sempre in favore di Johnny Depp e negativamente

per la Heard, con una piccola percentuale di tweet che richiamavano la solidarietà femminile da altre vittime di abusi.

Per quanto concerne il secondo periodo, quello che va dal 1° aprile 2022 al 10 giugno 2022 in pieno processo in Virginia, si è potuto evincere che anche in questo caso le prime due query contenessero tweet per lo più negativi, mentre la terza appena più positivi. Infatti, quelli negativi si esprimono maggiormente contro l'attrice accusandola di essere una "bugiarda violenta narcisista e una finta vittima" mentre i più positivi si esprimono a favore di Johnny Depp, in modo sarcastico elogiandolo come persona e attore, ma anche in modo serio classificandolo come vera vittima delle vicende.

Infine, nel terzo periodo dall'11 giugno 2022 all'11 settembre 2022 e quindi a processo concluso, si può osservare che in tutte e tre le query la maggior parte dei tweet sia negativa. Questi, infatti, recitano di come la Heard abbia rovinato la vita dell'attore e di come meriti la galera. Interessante invece sono i tweet più positivi che sempre in modo sarcastico fanno notare come questi processi abbiano reso Johnny Depp ancora più ricco e famoso, e che la giustizia trionfa sempre, che tu sia uomo o donna.

Come ultima analisi è stato eseguito il confronto tra i tweet con stessa query ma nei tre archi temporali differenti. Ne è emerso che:

1. i tweet con query #AmberHeard #JohnnyDepp ed in più almeno uno tra #JohnnyDeppVsAmberHeard, #JohnnyDeppVsAmberHeardTrial o #DeppVsHeard, nei primi due periodi sono maggiormente negativi e più specificatamente nel primo, sono più della metà (54,6%) e generalmente il numero di dati ricavati è nettamente inferiore rispetto a quelli del secondo e terzo periodo, probabilmente dato dal fatto che nel primo periodo le vicende non fossero molto emerse ed approfondite. Per quanto riguarda il modo lessicale di esprimersi, si è notato che nel primo periodo fossero più frequenti parole generiche come "abuse", "victim", "lie" mentre nei seguenti due appaiono termini più tecnici ed inerenti al processo come "trial", "court", "case" ...

2. i tweet recuperati con query #DomesticAbuse o #DomesticViolence e #JohnnyDepp o #AmberHeard o #JohnnyDeppVsAmberHeardTrial o #JohnnyDeppVsAmberHeard o #DeppVsHeard, sono di un numero molto limitato in ogni periodo. Questo potrebbe

essere dovuto da una combinazione errata di hashtag nella query o perché gli utenti hanno utilizzato altri hashtag più popolari e virali su Twitter riguardo le vicende. In ogni caso, nei tre periodi i tweet sono maggiormente negativi e i termini più utilizzati riguardano atti di violenza tipici degli abusi domestici e di genere come “victim”, “survivor”, “support”, “abuser” ...

3. anche i tweet scaricati con query #MePoo o #amberpoopedthebed o #Poop e #AmberHeard o #Poop e #JohnnyDepp sono di un numero molto limitato. Nel primo e nel secondo arco temporale la percentuale di tweet positivi è maggiore e si riferiscono all'atto in sé della defecazione nel letto di Johnny Depp da parte dell'ex moglie, mentre nel terzo prevalgono i tweet negativi in cui prevalgono le opinioni personali degli utenti nei confronti della Heard in seguito al verdetto finale.

Per concludere, si può ritenere che fin dagli inizi delle prime accuse e dei primi processi gli utenti si siano sempre espressi maggiormente a favore di Johnny Depp, ritenendolo innocente e che le azioni violente (non causanti di danni fisici a persone) e l'abuso di droga ed alcol siano state una conseguenza del nervoso, dello stress, delle accuse e delle bugie di Amber Heard. Maggiormente invece, i tweet più positivi sono sempre stati quelli carichi di ironia, emotività e sarcasmo, ricchi di emoji divertenti nei confronti delle parole e smorfie di Amber.

Bibliografia e Sitografia

Marshall McLuhan, *Understanding Media: The Extensions of Man*, 2001 (1964)

Amber Heard, 13 dicembre 2018, *Amber Heard: I spoke up against sexual violence — and faced our culture's wrath. That has to change.*, The Washington Post

https://www.washingtonpost.com/opinions/ive-seen-how-institutions-protect-men-accused-of-abuse-heres-what-we-can-do/2018/12/18/71fd876a-02ed-11e9-b5df-5d3874f1ac36_story.html

Roberta Mercuri, 1 giugno 2022, *Quanti danni ha fatto (davvero) l'editoriale di Amber Heard alla carriera di Johnny Depp?*, Vanity Fair

<https://www.vanityfair.it/article/quant-danni-ha-fatto-davvero-editoriale-di-amber-heard-alla-carriera-di-johnny-depp-sentenza>

1° giugno 2022, *Johnny Depp e Amber Heard, dalla storia d'amore al processo: cosa è successo.*, Sky Tg24

<https://tg24.sky.it/spettacolo/cinema/approfondimenti/johnny-depp-amber-heard-storia#18>

Alex Finnis, 6 maggio 2022, *When will Amber Heard's testimony continue? Date Johnny Depp trial will restart and when it's expected to end*, I News

<https://inews.co.uk/inews-lifestyle/people/when-amber-heards-testimony-continue-date-johnny-depp-trial-restart-1615032>

Claudia Casiraghi, 9 aprile 2022, *Johnny Depp, Amber Heard, il «dito mozzato» e l'ultimo atto di un matrimonio da incubo*, Vanity Fair

<https://www.vanityfair.it/people/mondo/2020/04/09/johnny-depp-dito-mozzato-amber-heard-audio-divorzio-gossip>

28 aprile 2022, *Il seguitissimo processo tra Johnny Depp e Amber Heard*, Il Post

<https://www.ilpost.it/2022/04/28/processo-depp-heard/>

1° giugno 2022, *Johnny Depp e Amber Heard, dalla storia d'amore al processo: cosa è successo.*, Sky Tg24

<https://tg24.sky.it/spettacolo/cinema/approfondimenti/johnny-depp-amber-heard-storia#24>

Carlotta Sisti, 26 maggio 2022, *A che cosa è servito il #MeToo se non crediamo a "vittime imperfette" come Amber Heard?*, Elle

<https://www.elle.com/it/magazine/women-in-society/a40086948/metoo-processo-depp-amber-heard/>

<https://www.pamelarutledge.com/dr-pamela-rutledge/>

Charlotte Colombo, 6 maggio 2022, *Il processo Depp contro Heard sta producendo tantissimi meme. Ecco perché è un problema*, Vice

<https://www.vice.com/it/article/wxdmey/processo-depp-heard-meme>

Vanessa Quinto, 1° giugno 2022, *Processo Depp-Heard: in attesa del verdetto, la giuria torna a riunirsi oggi*, Rai News

<https://www.rainews.it/articoli/2022/05/processo-depp-heard-in-attesa-del-verdetto-a429d3d5-6186-47c0-8d34-e60c2c3d1103.html>

2 giugno 2022, Costance Grady, *The Me Too backlash is here*, Vox

<https://www.vox.com/culture/23150632/johnny-depp-amber-heard-trial-verdict-me-too-backlash>

1 giugno 2022, Ej Dickson, *'Men Always Win': Survivors 'Sickened' by the Amber Heard Verdict*, Rolling Stone

<https://www.rollingstone.com/culture/culture-news/amber-heard-johnny-depp-verdict-metoo-trial-1361356/>

TikTok's Amber Heard Hate Machine

<https://www.nytimes.com/2022/05/26/arts/amber-heard-tiktok-johnny-depp.html>

TikTok

<https://vm.tiktok.com/ZMN81aUYC/>

Twitter

<https://twitter.com/MeTooMVMt/status/1530567171998367745>

Twitter Facts

<https://www.factslides.com/s-Twitter>

Steven Johnson, 5 giugno 2009, *How Twitter Will Change the Way We Live*, Time

<https://content.time.com/time/subscriber/article/0,33009,1902818,00.html>

Twitter

<https://about.twitter.com/en>

Jack Dorsey, 2020, *2020 Global Impact Report*, Twitter

<https://about.twitter.com/content/dam/about-twitter/en/company/global-impact-2020.pdf>

Q3 2021 Letter to Shareholders, Twitter

https://s22.q4cdn.com/826641620/files/doc_financials/2021/q3/Final-Q3'21-Shareholder-letter.pdf

Agosto 2021, *Top Twitter Statistics- Updated August 2021*, Zephoria digital marketing

<https://zephoria.com/twitter-statistics-top-ten/>

Matt Ahlgren, 31 ottobre 2022, *50+ Statistiche e fatti su Twitter per 2022*,

Websiterating

<https://www.websiterating.com/it/research/twitter-statistics/#references>

Claire Beveridge, 16 marzo 2022, *33 Twitter Stats That Matter to Marketers in 2022*,

Hootsuite

<https://blog.hootsuite.com/twitter-statistics/>

Hootsuite

<https://www.hootsuite.com/resources/digital-trends>

Twitter

<https://help.twitter.com/it/rules-and-policies/twitter-api#:~:text=Per%20condividere%20informazioni%20nel%20modo.interfacce%20di%20programmazione%20delle%20applicazioni>

Laura Zanotti, 21 giugno 2022, *Cosa sono le API in informatica e quale impatto hanno sul business*, Zero Uno

<https://www.zerounoweb.it/software/erp-crm-scm/cosa-sono-le-api-e-quale-impatto-hanno-sul-business/>

Miriam De Falco, 24 aprile 2022, *La dichiarazione di Johnny Depp durante il processo*, StyLise

<https://www.stylise.it/in-evidenza/bozza-automatica-la-dichiarazione-di-johnny-depp-durante-il-processo/#:~:text=Mi%20ha%20ucciso%20il%20fatto.dell'onest%C3%A0%20e%20della%20verit%C3%A0>

Twitter

<https://developer.twitter.com/en>

Twitter

<https://developer.twitter.com/en/use-cases/build-for-businesses>

Twitter

<https://developer.twitter.com/en/use-cases/build-for-consumers>

Twitter

<https://developer.twitter.com/en/use-cases/do-research>

Twitter

<https://developer.twitter.com/en/use-cases/teach-and-learn>

Twitter

<https://developer.twitter.com/en/use-cases/build-for-fun>

Twitter

<https://developer.twitter.com/en/use-cases/build-for-good>

Twitter

<https://developer.twitter.com/en/products/twitter-api>

Python

<https://www.python.org/about/>

Python

<https://www.python.it/about/>

Nltk

<https://www.nltk.org/>

Tom Aarsen, 7 aprile 2022, *FAQ*, GitHub

<https://github.com/nltk/nltk/wiki/FAQ>

NumPy

<https://numpy.org/doc/stable/>

GitHub

<https://github.com/cjhutto/vaderSentiment>

Valentina D'Antonio, 27 aprile 2021, *Sentiment Analysis: come studiare e migliorare la percezione del tuo brand*, La Penna del Web

<https://lapennadelweb.com/sentiment-analysis-cosa-quali-tool-usare/>

Team I.a. Italia, Intelligenza Artificiale Italia

<https://www.intelligenzaartificialeitalia.net/post/sentiment-analysis-con-python-la-guida-completa>

Pandas

<https://pandas.pydata.org/>

Matplotlib

<https://matplotlib.org/>

PyPi

<https://pypi.org/project/wordcloud/>

Ringraziamenti

Ringrazio sinceramente il mio relatore, il professore Gianmaria Silvello, per essere sempre stato tempestivamente disponibile, per avermi consigliato e supportato nella stesura di questo elaborato.

Ringrazio i miei genitori e mio fratello per avermi permesso ed essermi stati accanto in questo percorso di studi e per avermi spronata a non arrendermi mai.

Ringrazio i miei amici più cari, Giorgia, Beatrice, Nicole, Maria Elena, Maria Vittoria e Davide per avermi ricordato ogni giorno di non perdere la grinta e la tenacia che mi caratterizzano e per essermi sempre stati vicino nei momenti in cui ho creduto di non potercela fare.