UNIVERSITÀ
DEGLI STUDI
DI PADOVA

DIPARTIMENTO
DI INGEGNERIA
DELL'INFORMAZIONE

DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

CORSO DI LAUREA IN INGEGNERIA INFORMATICA

STUDIO DELLA POLARIZZAZIONE DELLE OPINIONI NEI SOCIAL NETWORK ATTRAVERSO TECNICHE DI INTELLIGENZA ARTIFICIALE

Relatore: Prof. ANDREA LOREGGIA

Laureando: JACOPO ZECCHIN

Matricola: 1187374

ANNO ACCADEMICO 2021 – 2022

Data di laurea 7/03/2022

Abstract

Con l'avvento della tecnologia moderna, basata principalmente sull'Intelligenza Artificiale, la quantità di informazioni disponibili è aumentata in modo esponenziale rispetto al passato. Tra i vari inconvenienti, possiamo anche identificare la non affidabilità di molte fonti di informazione. I Social Network hanno dato un forte impulso alla condivisione: la libertà di pensiero e l'eterogeneità dei contenuti espressi hanno portato ad avere risorse diverse sullo stesso argomento, senza che queste vengano approfonditamente verificate.

In questo lavoro, analizziamo alcune metriche proposte in letteratura per la misurazione del fenomeno della polarizzazione dell'opinione e, come caso di studio, applichiamo queste metriche a un set di dati contenente i risultati di un sondaggio americano che ha raccolto le opinioni degli utenti su varie questioni sociali.

L'applicazione delle varie metriche ha permesso di identificare un certo grado di polarizzazione nel dataset che può essere mappato sui due principali allineamenti politici americani. Infine, data la correlazione tra le varie variabili coinvolte nell'analisi, abbiamo sviluppato una rete neurale in grado di predire l'allineamento politico dell'utente con una precisione molto elevata.

Sommario

Introduzione

Con l'avvento dei moderni mezzi di comunicazione la mole di informazioni reperibili è aumentata in maniera esponenziale rispetto ad un tempo, e fra le varie conseguenze negative emerge la non attendibilità delle fonti utilizzate.

In particolare, la nascita dei Social Network ha dato un forte slancio alla condivisione di dati: la libertà di espressione e l'eterogeneità dei contenuti hanno ampliato il ventaglio di risorse disponibili su un particolare argomento, senza che queste vengano necessariamente verificate (Barrett-Maitland e Lynch 2019). Il numero giornaliero di utenti delle piattaforme digitali è in crescita, solo nell'ultimo anno l'incremento di utenza a livello mondiale è stata di circa il 13.2% e più della metà della popolazione degli utenti di queste piattaforme risulta attiva[1].

Gli utenti entrano dunque in contatto con un quantitativo di risorse un tempo impensabile. A loro volta, producono essi stessi informazione, lasciando traccia delle proprie preferenze, opinioni, necessità, spesso inconsapevoli di tale lascito. Questi dati vengono usati per la profilazione dell'utenza, processo che permette di classificare gli utenti in ambiti specifici al fine di mostrare informazioni, pubblicità, prodotti o attività più appetibili, senza che questo entri in contatto con soluzioni diverse o discordanti rispetto alle proprie preferenze (Tam e Ho 2006).

Non avendo la possibilità di interfacciarsi con pensieri diversi dal proprio, l'utente perde la possibilità di un confronto critico che porta ad uno scambio di idee e spesso ad una crescita costruttiva (Sîrbu, et al. 2019). Il risultato è che si rafforzano le convinzioni preesistenti e la posizione iniziale si cementifica (Levy 2021) (Nguyen, et al. 2011). Per polarizzazione si intende la divisione in due o più gruppi sulla base di un insieme di opinioni e/o credenze contrastanti[2].

Diverse metriche sono state proposte in letteratura per la misura della polarizzazione nelle comunità digitali (Guerra, Cardie e Kleinberg 2021). In questo lavoro analizziamo alcune metriche proposte per la rivelazione ed il calcolo del fenomeno e applichiamo tali metriche come case-study ad un dataset contenente i risultati di un sondaggio americano che ha collezionato le opinioni degli utenti su diverse tematiche inerenti questioni sociali.

[1] https://datareportal.com/reports/digital-2021-global-overview-report

[2] https://en.oxforddictionaries.com/definition/polarization

L’applicazione delle diverse metriche ha permesso di identificare un certo grado di polarizzazione mappabile nei due principali schieramenti politici americani. Infine, vista la correlazione esistente tra le diverse variabili implicate, abbiamo sviluppato una rete neurale in grado di predire con un’accuratezza molto elevata lo schieramento politico dell’utente.

In conclusione, i risultati evidenziano la presenza di polarizzazione nel dataset e permettono di identificare quali metriche forniscano una migliore evidenziazione del fenomeno. Inoltre, i risultati ottenuti ci permettono di riportare all’attenzione l’aspetto etico di questo fenomeno, in particolare di come sia possibile estrarre informazioni sensibili attraverso l’inferenza che può essere fatta su dati apparentemente slegati tra loro, ponendo serie domande sulla responsabilità che le piattaforme digitali e i detentori di tali informazioni dovrebbero avere nella governance di tale conoscenza.

Background Tecnico

La seguente sezione si focalizza sul problema analizzato e sugli strumenti utilizzati per la stesura di questo documento. Viene fornita una definizione formale di Social Network (SNS), con una breve analisi riguardo gli algoritmi utilizzati per creare consenso all'interno di quest'ultimi. Successivamente si esamina il problema della polarizzazione e delle echo-chambers (casse di risonanza) all'interno delle piattaforme SNS. Verrà descritto il dataset e gli strumenti usati per l'analisi dei dati. Descriveremo le metriche adottate, con la spiegazione dei risultati che queste forniscono. Infine, descriviamo la rete neurale implementata per la predizione dello schieramento politico in base ai dati forniti in input.

Social Network

I Social Network, o piattaforme SNS, hanno visto i loro primi sviluppi verso la fine del secolo scorso con Six Degrees, fondato nel 1996[3] da Andrew Weinreich.

Da allora sia le piattaforme che gli utilizzatori giornalieri sono aumentati a dismisura. Nonostante le varie differenze, i social media hanno in comune diverse proprietà. Sono servizi basati sul web che consentono agli utenti di:

- creare un profilo che può essere pubblico o semi-pubblico all'interno di un ambiente circoscritto;
- interagire con un certo numero di utenti con i quali condividono delle caratteristiche;
- visualizzare e interagire con la loro lista di connessioni e con quelle sviluppate da altri profili all'interno del sistema.

La natura e la nomenclatura di queste connessioni dipendono dalla piattaforma utilizzata (Boyd e Ellison 2008). Per favorire l'interazione tra gli utenti sono stati sviluppati diversi algoritmi il cui scopo è quello di identificare contenuti simili sulla base delle preferenze espresse dagli utenti, dai contenuti condivisi in precedenza e quelli a cui si è prestato maggiore interesse. Per esempio, in base agli interessi mostrati in precedenza e alla rete di connessioni create dall'account. Gli algoritmi sviluppati creano dei profili di similarità proponendo all'utente i contenuti con una valutazione più elevata [4].

[3] https://www.cbsnews.com/pictures/then-and-now-a-history-of-social-networking-sites/2/

[4] https://engineering.fb.com/2021/01/26/ml-applications/news-feed-ranking/

In questo modo si effettua una selezione dei contenuti: non è l'utilizzatore a decidere cosa vedere ma è la piattaforma stessa che decide. Il fruitore viene guidato nella fruizione delle informazioni, in quanto quelle più simili alle sue preferenze hanno una posizione prioritaria. In questo modo, si evita o si marginalizza l'effetto dei contenuti che contengono opinioni discordanti. Tutto è filtrato al fine di massimizzare i risultati di selezione dei post effettuati dall'algoritmo, il fruitore non ha voce in capitolo sulla scelta, ma è spettatore passivo nella scelta dei contenuti.

Polarizzazione ed Echo-Chambers

Teoricamente le piattaforme SNS consentono di ottenere una connessione con un numero illimitato di utenti; per questo motivo è necessario analizzare gli effetti che queste causano a livello collettivo, ossia la polarizzazione e le echo-chambers.

Polarizzazione

Come affermato dagli studi di Sunstein, la polarizzazione è un processo in cui un gruppo sociale o politico è diviso in due sottogruppi che presentano posizioni, obiettivi e punti di vista conflittuali e contrastanti, con pochi individui che restano neutrali o detengono una posizione intermedia (Sunstein 2002).

Questo fenomeno viene accentuato dalle moderne tecnologie. Mentre una volta le relazioni sociali erano limitate da diversi fattori, tra cui la distanza geografica ed il tempo necessario per effettuare una comunicazione, con l'avvento delle moderne tecnologie queste restrizioni sono venute meno. Avendo a disposizione un bacino di utenza potenzialmente illimitato, si rafforza la teoria secondo la quale la tendenza iniziale dei singoli membri verso una data direzione viene rafforzata e resa più estrema dal confronto con gli altri membri (Myers e Lamm 1975). La polarizzazione di gruppo è un fenomeno presente in diversi ambiti, tra cui quello sociale, religioso e politico. Questo lavoro si focalizza prevalentemente sull'ambito politico.

Per polarizzazione politica si intende la divergenza delle preferenze politiche verso ideologie contrastanti. Il fenomeno trova un ambiente favorevole sia nei sistemi bipartitici che multipartitici, con una forma più elevata in ambito bipartitico in quanto le idee rispetto al movimento opposto sono definite in maniera molto più netta.

Sono stati individuati due tipi di polarizzazione politica:

- *Polarizzazione d'elite*, definita da un alto livello di distanza ideologica tra i partiti avversari e un alto livello di omogeneità all'interno del partito stesso (Druckman, Peterson e Slothuus 2013).
- *Polarizzazione di massa*, o polarizzazione popolare, che si verifica quando gli atteggiamenti di un elettorato nei confronti di questioni politiche, figure pubbliche o altri cittadini sono nettamente in contrasto con le linee di partito (McCarty, Poole e Rosenthal 2016). Con il termine "massa" si intende identificare tutto l'elettorato, cioè le persone non direttamente coinvolte nelle scelte attuate dal partito.

Per gli utilizzatori regolari di social media è necessario uno sforzo minimo per incontrare informazioni a supporto delle loro idee; addirittura potrebbe essere necessario uno sforzo maggiore per evitare tali informazioni. Questo conferma che l'utilizzo di questi strumenti facilita la frammentazione del pubblico in schieramenti diversi.

Poiché le piattaforme lasciano ampio spazio di pensiero e di opinione, oltre a connettere persone geograficamente distanti, amplificano in maniera esponenziale questi fenomeni, sui quali verranno discusse metodologie e metriche di verifica.

Echo-Chambers

Le Echo-Chambers, o casse di risonanza, sono ambienti virtuali nei quali si fruiscono o si trovano esclusivamente opinioni e convinzioni simili alle proprie, senza prendere in considerazione pensieri contrastanti[5].

Si presume che tale disconnessione e mancanza di prospettive alternative derivi da una combinazione di scelte individuali, come la selezione delle fonti di notizie da consultare o degli account dei social media da seguire. La definizione algoritmica di tali scelte, come portali di notizie, motori di ricerca e piattaforme social che evidenziano e raccomandano alcune fonti rispetto ad altre. Non è un fenomeno nuovo, ma con l'avvento dei social media ha avuto un'importanza ancora maggiore in quanto gli algoritmi delle piattaforme SNS imparano dalle scelte degli utilizzatori e gli utenti scelgono principalmente dalle opzioni promosse dagli algoritmi; si viene dunque a creare un ciclo di feedback auto-informativo che riduce gradualmente la scelta ad un insieme di opzioni sempre più ristrette ed omogenee (Bruns 2019).

[5] https://www.oxfordlearnersdictionaries.com/us/definition/english/echo-chamber.

L’obiettivo finale dei social è massimizzare il livello di soddisfazione dei fruitori e per fare ciò attuano un intricato sistema di personalizzazione dei contenuti, presentando ad ogni utente le stesse informazioni ma in modo differente per assecondare meglio le loro preferenze e rendere l’informazione più compatibile con il loro punto di vista (Tam e Ho 2006). A causa di queste personalizzazioni si verificano le cosiddette bolle di filtraggio[6], ossia il risultato delle ricerche su siti che registrano la storia del comportamento dell’utente. Questi sono in grado di utilizzare informazioni sui fruitori fornite in precedenza in maniera più o meno volontaria per scegliere selettivamente solo le risposte che l’utente stesso vorrà vedere. L’effetto è di escluderlo da informazioni che sono in contrasto con il suo punto di vista, isolandolo in tal modo nella sua bolla culturale o ideologica (Pariser 2011).

Sia *Sunstein* che *Pariser* hanno fatto previsioni disastrose su questo fenomeno, affermando che potrebbe portare ad ambienti di limitazione delle informazioni con impatti negativi sia individuali che sociali.

Pariser ha sottolineato gli effetti negativi dell'isolamento individuale nella creazione di bolle di filtraggio dove i punti di vista personali persistono, incontrastati e non testati. *Sunstein* ha sottolineato il potenziale della tecnologia di rafforzare la frammentazione su scala più ampia, in cui le persone non sono isolate individualmente, ma formano invece gruppi in cui individui con simili predilezioni ideologiche interagiscono esclusivamente tra loro. *Sunstein* ha sostenuto che le interazioni online possono rafforzare la segregazione ideologica e quindi facilitare pool di informazioni limitate che rafforzano i pregiudizi preesistenti, promuovono il pensiero di gruppo e incoraggiano l'adozione di punti di vista ancora più estremi. In questo modo, *Sunstein* era interessato alle interazioni tra gruppi omogenei che condividono un'identità sociale comune. In alternativa, *Pariser* si preoccupava dell'isolamento individuale e della mancanza di informazioni condivise sulla formazione. Dal punto di vista di *Pariser*, un individuo isolato nella propria bolla personalizzata può ancora soffrire degli impatti negativi di informazioni limitate, anche se questo isolamento li rende immuni dalle pressioni sociali che rafforzano la solidarietà di gruppo e generano gruppetti polarizzanti.

Nonostante queste differenze di prospettiva, sia *Pariser* che *Sunstein* hanno sostenuto che il rimedio per gli ambienti che limitano l'informazione è il consumo individuale di contenuti ideologicamente diversi. (Kitchens 2020)

[6] Chiamate anche Filter bubble.

Esempio di questo isolamento è la ricerca personalizzata di Google, che mostra per prime le informazioni che ritiene a noi più congeniali (Parramore 2010).
Diverse critiche sono state mosse contro le bolle di filtraggio: alcuni sostengono che gli algoritmi utilizzati dalle piattaforme online prendono decisioni per conto dell'utente, costringendolo e rendendolo inconsapevole delle scelte disponibili. Altri sostengono che i pregiudizi causati dagli algoritmi e dagli esseri umani stessi potrebbero diminuire la diversità dei punti di vista, il rispetto reciproco o permettere agli oppressori di prevalere a causa della mancanza di informazioni, in quanto si filtra solo una verità parziale (Bozdag 2015). Tutto ciò porta ad un dilemma etico: quanto pervasiva può essere la tecnologia nelle nostre vite e quanto ne siamo influenzati senza saperlo?

Dataset

L'analisi empirica è stata effettuata su un dataset di dati reali composto dai risultati di un sondaggio online già studiato in letteratura (Bail, et al. 2018). Il dataset è una collezione di risposte ricavate da un ampio campione di democratici e repubblicani che utilizzano la piattaforma Twitter almeno tre volte alla settimana. Il campione di utenti è stato intervistato su una serie di questioni di politica sociale. Al campione intervistato è stato assegnato, circa una settimana dopo l'intervista, un bot Twitter da seguire anche sotto la spinta di incentivi economici. Scopo del bot è stato quello di esporre l'utenza a messaggi diversificati per circa un mese. Tali contenuti rappresentavano una porzione di utenza con opinioni opposte alle ideologie politiche dell'utente interessato. Gli intervistati sono stati ripresi alla fine del mese per misurare l'effetto di questo trattamento e ad intervalli regolari durante il periodo di studio per monitorare la conformità del trattamento (Bail, et al. 2018). Oltre ai dati anagrafici e preferenze politiche agli intervistati è stato chiesto di rispondere alle seguenti domande:

- La regolamentazione governativa delle imprese è necessaria per tutelare l'interesse pubblico?
- La discriminazione razziale è la ragione principale per cui molte persone di colore non possono andare avanti in questi giorni?
- Gli immigrati oggi rafforzano il nostro paese grazie al loro lavoro e ai loro talenti?
- Le aziende fanno troppi profitti?
- L'omosessualità dovrebbe essere accettata dalla società?

- Leggi e regolamenti ambientali più severi costano troppi posti di lavoro e danneggiano l'economia?
- Il governo è quasi sempre inutile ed inefficiente?
- I poveri oggi hanno vita facile perché possono ottenere le indennità governative senza fare nulla in cambio?
- Il governo oggi non può permettersi di fare molto di più per aiutare i bisognosi?
- Il modo migliore per garantire la pace è attraverso l'intervento militare?

I seguenti dati sono stati successivamente processati per rendere il dataset binario[7], in modo da poter applicare le metriche.

Metriche

Le metriche utilizzate per l'analisi dei dati sono state ampiamente discusse in (Bramson, et al. 2016). Per una più semplice lettura, vengono riportate le metriche adottate.
Le misure fornite assumono le seguenti caratteristiche. Ci sono N agenti, partecipanti, o rispondenti: $a_1,a_2,...a_N$. Ogni agente a_i ha un valore di atteggiamento x_i su un intervallo normalizzato tra 0 e 1; questa è la posizione dell'atteggiamento di quell'agente lungo lo spettro. La distribuzione completa di tutti gli atteggiamenti degli agenti è X.

Media

Per un insieme di dati, la media, nota anche come media aritmetica, è un valore centrale di un insieme finito di numeri. In particolare, la somma dei valori è divisa per il numero di campioni. La media aritmetica di un insieme di numeri $x_1,x_2,...,x_n$ e tipicamente indicata da $\bar{X}$ è calcolata come segue:

$$\bar{X} = \frac{1}{N}\left(\sum_{i=1}^{N} x_i\right) = \frac{x_1 + x_2 + \ldots + x_N}{N}$$

In generale, questa formula fornisce una visione complessiva dell'andamento del dataset e indica approssimativamente il valore al quale tende la distribuzione dei campioni.

[7] Dataset binario, ovvero con risposte o uguali a 0 (In disaccordo) o uguali a 1 (Concordo).

Spread

Il concetto più semplice di polarizzazione è la diffusione degli atteggiamenti effettivamente rappresentati nel sistema. Senza tener conto della forma della distribuzione, o anche se vi è continuità tra gli estremi, più gli individui sono lontani più varie sono le loro idee. Se la differenza nelle opinioni più estreme detenute è ampia, le idee sono più polarizzate (in questo senso).

La polarizzazione nel senso di spread potrebbe essere misurata come il livello di atteggiamento dell'agente con il più alto valore (il massimo tra le risposte fornite) meno il livello di atteggiamento dell'agente con il valore più basso (il minimo). Questa misura è anche comunemente chiamata l'intervallo dei dati.

$$Spread \ = \ max_{xi} X - \ min_{xi} X$$

La misura appena descritta è volutamente comprensiva dell'intera popolazione. Ad esempio, lo spread potrebbe essere usato per confrontare i risultati attraverso gli anni, fra le affiliazioni politiche, o fra le nazioni.

Dispersione

Un'altra misura semplice e comune della polarizzazione è la dispersione statistica (o variazione statistica). A differenza dello spread, che considera solo gli estremi della popolazione, la dispersione considera la forma totale dei campioni raccolti. Il valore di questa metrica può aumentare quando i gruppi si allontanano, quando la distribuzione si appiattisce, o quando i dati degli agenti si allontanano dal centro verso le estremità opposte della distribuzione. Proprio come per lo spread, una maggiore dispersione implica livelli più elevati di polarizzazione. Ci sono molti metodi di calcolo che sono appropriati per questa metrica, tra cui la differenza media, la deviazione assoluta media, la deviazione standard, il coefficiente di variazione e l'entropia sono tutti candidati. Per semplicità utilizzeremo la deviazione media assoluta dalla media perché questa può essere calcolata dai dati aggregati del dataset. Nella nostra ricerca troviamo che le differenze di valore medie di coppia sono utili.

Per una popolazione con N individui la misura di deviazione assoluta (normalizzata tra zero e uno) diventa

$$dispersione \ = \frac{2}{N} \sum_{x_i}^{N} |x_i - \ \bar{\mathrm{X}} \ |$$

in cui $\bar{X}$ è il valore medio della distribuzione X.

Differenziabilità

Dopo aver identificato o definito i gruppi, ci si può chiedere quanto siano diverse le sottopopolazioni o sottodivisioni tra loro. Ciò che conta per la polarizzazione nel senso di distinzione è quanto bene possiamo distinguere i gruppi. Questi più possono essere visti come separati, più la popolazione totale è polarizzata, indipendentemente dalla distanza, dalle dimensioni e dai livelli di consenso interno tra i gruppi. Un altro modo di pensare a questo è quanto sia improbabile confondere una persona appartenente ad una fazione come appartenente ad un altra; meno è probabile la confusione, più distinti sono i gruppi. In molte applicazioni una misura semplice può essere sufficiente per catturare il livello di distinzione. Se i gruppi sono definiti esogeni[8] la sovrapposizione tra gruppi funziona come un confronto a coppie.

Una misura meno elegante della distinzione esogena di due gruppi (ma che è utile per i nostri dati previsti da indagini e modelli basati su agenti) è fornita da:

$$\frac{1}{|A| + |B|} 1 - \sum_{r \in R} min\{y_A(r), y_B(r)\}$$

Questa formula determina la proporzione della sovrapposizione di due distribuzioni normalizzando la somma dei valori ad ogni individuo di una qualsiasi distribuzione che abbia un valore più basso.

Matrice di Correlazione

La matrice di correlazione è un metodo statistico per mostrare la relazione tra due o più variabili e l'interrelazione nei loro movimenti. Aiuta a definire la relazione e la dipendenza tra le variabili.

Il coefficiente di correlazione di Pearson è una misura di come due istanze sono linearmente correlate (Kijsipongse, et al. 2011). Il valore di $r_{x,y}$ varia da -1 a 1. Vale zero se due istanze non sono legate. Quando il valore è positivo, X e Y sono correlate. Maggiore è il valore, maggiore è la correlazione. Se il valore di $r_{x,y}$ è negativo, allora X e Y sono correlati negativamente.

[8] Il numero di individui in gruppo A che abbina i valori di atteggiamento con gli individui del gruppo B.

La correlazione tra tutte le coppie di istanze può essere espressa come la matrice di correlazione in cui ogni elemento è il coefficiente di Pearson, $r_{x,y}$, delle diverse coppie di istanze (X,Y).

Supponiamo che i dati siano una matrice $n \times m$, dove n e il numero di istanze e m è il numero di attributi di un'istanza. Siano X e Y le istanze che contengono gli attributi m. Matematicamente, il coefficiente di correlazione di Pearson, $r_{x,y}$, tra due istanze X e Y è definito come:

$$r_{x,y} = \frac{\sum_{i=1}^{m}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{m}(X_i - \bar{X})^2}\sqrt{\sum_{i=1}^{m}(Y_i - \bar{Y})^2}}$$

dove $\bar{X}$ e $\bar{Y}$ sono i valori medi di X e Y.

Heatmap

Una heatmap (Wilkinson e Friendly 2009) è una tecnica di visualizzazione dei dati che mostra attraverso l'utilizzo del colore la grandezza di un fenomeno in due dimensioni. La variazione di colore può essere per tonalità o intensità. In una heatmap, le grandezze sono disposte in una matrice di dimensione fissa delle celle in cui le righe e le colonne sono fenomeni o categorie discreti e l'ordinamento è intenzionale, per riuscire a scovare cluster attraverso l'analisi statistica. La dimensione della cella è arbitraria ma abbastanza grande da essere chiaramente visibile.

Rete Neurale

Le reti neurali artificiali (Artificial Neural Network o ANN) sono dei sistemi computazionali che cercano di riprodurre il funzionamento dei processi del cervello umano. L'unita principale è il nodo, chiamato anche neurone o percettrone (N), che come il suo corrispettivo biologico, tramite dei collegamenti (L), comunica con gli altri nodi.

I vari nodi sono organizzati in diversi strati (layer). La struttura base prevede:

- Un layer di ingresso, il quale riceve gli input che devono essere elaborati dal sistema;
- Uno o più layer intermedi, detti anche nascosti (hidden). Questi estrapolano caratteristiche salienti nei dati di input, che vengono poi utilizzate per il calcolo dei valori ritornati;
- Un layer finale che fornisce gli output.

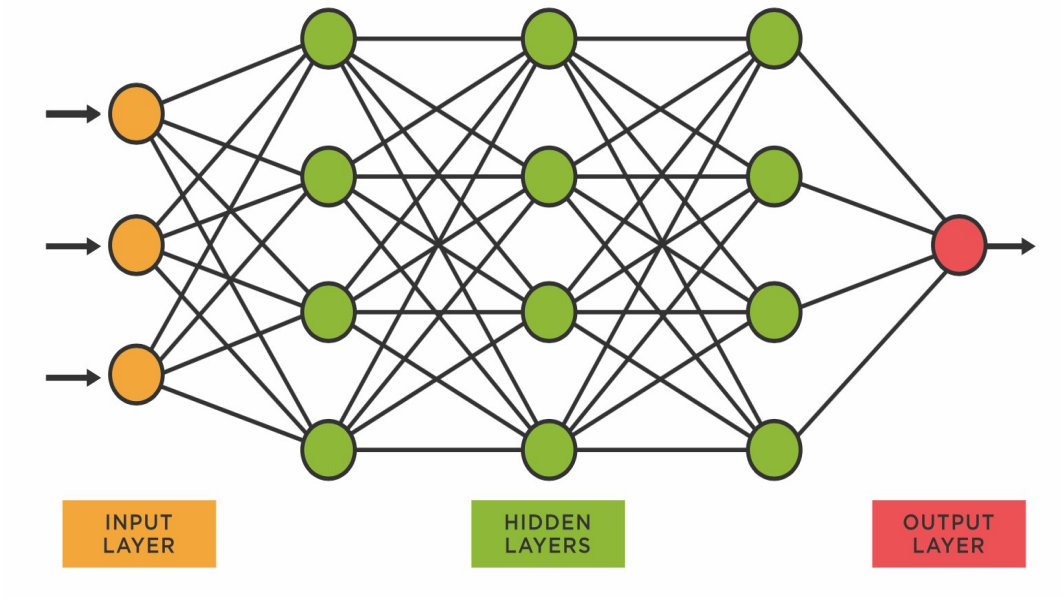

Figura 1: Esempio di rete neurale fully connected con un input layer con 3 neuroni, 3 hidden layer con 4 neuroni ciascuno e un output layer con un neurone

L'elemento base, il percettrone, è caratterizzato da tre elementi fondamentali:

- Un numero variabile di connessioni, ognuna caratterizzata da un proprio peso che può assumere sia valori positivi che negativi;
- Una funzione che calcola il risultato delle varie connessioni di input in base al loro peso, producendo come output un valore numerico;
- Una funzione di attivazione, che ha lo scopo di normalizzare i valori restituiti dal sommatore.

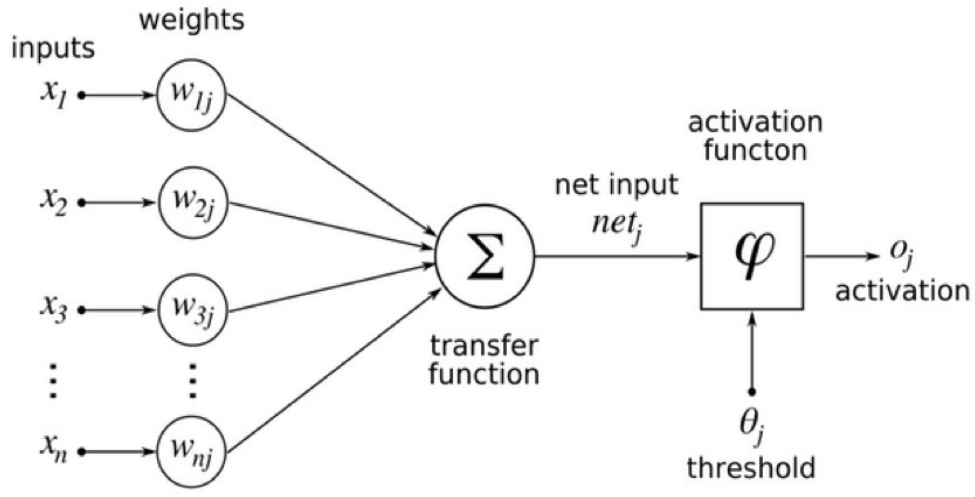

Figura 2: Esempio del funzionamento di un percettrone

Matematicamente, un neurone *k* può essere rappresentato dalle seguenti equazioni:

$$u_k = \sum_{j=1}^{m} w_{kj} \cdot x_j$$

$$y_k = \varphi(u_k + b_k)$$

dove:

- x_i è il peso associato al neurone;
- w_{kj} è il peso associato alla connessione tra i vari neuroni;
- u_k è la combinazione lineare degli input del neurone k;
- b_i è il peso del bias[9];
- $\varphi(x)$ è la funzione di attivazione;
- y_k è il valore di output della funzione di attivazione.

Ci possono essere diverse tipologie di ANN, che si differenziano per il tipo di connessioni presenti tra i diversi neuroni artificiali e per il tipo di problema che riescono a risolvere. In particolare, le reti attualmente più conosciute sono:

- *Feedforward neural network*, implementazione più semplice, nella quale i dati di input viaggiano in una sola direzione. I livelli nascosti possono essere presenti, ma non necessariamente, mentre i livelli di input e output sono necessari. Il numero di strati dipende dalla complessità della funzione da rappresentare.
  Ha una propagazione unidirezionale in avanti ma nessuna propagazione all'indietro. I pesi sono statici. Una funzione di attivazione è alimentata da input che vengono moltiplicati per i pesi. Per fare ciò, viene utilizzata la funzione di attivazione o la funzione di attivazione step. Sono abbastanza semplici da mantenere e sono utilizzate per trattare con dati che contengono un rumore elevato.
- *Multi-Layer Perceptron (MLP)*, rete neurale complessa dove i dati di input viaggiano attraverso vari strati di neuroni artificiali. Ogni singolo nodo è collegato a tutti i neuroni dello strato successivo; questo rende la rete completamente connessa.

[9] Un valore costate aggiunto al valore ottenuto dal sommatore

I livelli di input e output sono presenti ma possono esserci più hidden layers.

Ha una propagazione bidirezionale. Gli input sono moltiplicati con i pesi e alimentati alla funzione di attivazione e, nella backpropagation, sono modificati per ridurre la perdita; dunque, i pesi sono valori appresi dalla rete. Si auto-regolano a seconda della differenza tra uscite previste e input di allenamento.

- *Convolutional Neural Network*, rete che contiene una disposizione tridimensionale dei neuroni, invece dell'array bidimensionale standard. Il primo strato è chiamato strato convoluzionale. Questo tipo di rete mostra risultati molto efficaci nel riconoscimento di immagini e video, nell'analisi semantica e nel rilevamento di parafrasi. Ogni neurone nello strato convoluzionale elabora solo le informazioni da una piccola parte del campo visivo. Le caratteristiche di input sono prese in batch come un filtro. La rete comprende le immagini in parti e può calcolare queste operazioni più volte per completare l'elaborazione.
- *Recurrent Neural Network,* tipo di rete neurale artificiale che utilizza dati sequenziali o dati di serie temporali. Questi algoritmi di apprendimento profondo sono comunemente usati per problemi ordinali o temporali, come la traduzione del linguaggio, l'elaborazione del linguaggio naturale, il riconoscimento vocale e la didascalia delle immagini. Come le reti neurali *feedforward* e *convoluzionali*, le reti neurali ricorrenti utilizzano i dati di addestramento per imparare.

  Si distinguono per la loro "memoria" in quanto prendono informazioni da input precedenti per influenzare l'input e l'output corrente.

  Mentre le reti neurali profonde tradizionali presuppongono che gli input e gli output siano indipendenti l'uno dall'altro, l'output delle reti neurali ricorrenti dipende dagli elementi precedenti all'interno della sequenza.

In questo articolo considereremo solo le reti *feedforward* con la proprietà di essere *fully connected*, cioè con i neuroni di ogni layer completamente connessi con quelli degli stati precedenti.

Una volta stabilite tutte le caratteristiche della rete, tra cui

- topologia;
- numero e tipo di neuroni;
- collegamenti;

bisogna inizializzare i pesi e addestrare la rete per arrivare ad ottenere buoni risultati.

L’addestramento non viene effettuato sull’intero campione di dati, il dataset viene diviso in due parti, il *training set* e il *test set*.

Durante la fase di allenamento della rete viene usato solo il training set, che permetterà alla rete di aggiornare il valore dei pesi interni per diminuire l’errore presente, al fine di migliorare l'accuratezza del risultato. Questo viene fatto minimizzando gli errori osservati. L'apprendimento è completo quando l'esame di osservazioni aggiuntive non riduce utilmente il tasso di errore. Anche dopo l'apprendimento, il tasso di errore in genere non raggiunge 0. Se dopo l'apprendimento, il tasso di errore è troppo alto, la rete in genere deve essere riprogettata.

Una volta terminato l’addestramento la rete viene testata sul campione di dati appartenenti al test set per valutare le sue performance finali.

La rete sviluppata avrà due implementazioni in base alla funzione di attivazione utilizzata in output. L’input layer avrà 21 neuroni, di cui uno di bias, mentre l’unico hidden layer presente sarà composto da 16 neuroni. Entrambi questi layer avranno come funzione di attivazione ReLU[10]. È presente un solo layer nascosto, in quanto il teorema di approssimazione universale per le reti neurali afferma che ogni funzione continua che mappa intervalli di numeri reali in qualche intervallo di output di numeri reali può essere approssimata da una rete MLP con un solo hidden layer. Questo risultato vale per un’ampia gamma di funzioni di attivazione, tra cui quelle utilizzate per l’implementazione.

[10] Rectified Linear Unit

Funzioni di Attivazione

Una funzione di attivazione trasforma i suoi ingressi in uscite che hanno un certo intervallo. Ci sono vari tipi di funzioni di attivazione, nei prossimi paragrafi introduciamo brevemente le più usate (Nwankpa e al. 2018) (Sharma, Sharma e Anidhya 2017) (Xu 2015).

ReLU

ReLU sta per unità di linea rettificata ed è una funzione di attivazione non lineare ampiamente utilizzata. Il vantaggio della funzione ReLU è che i neuroni non vengono attivati contemporaneamente. Ciò implica che un neurone sarà disattivato solo quando l'uscita della trasformazione lineare è zero. Matematicamente, si può definire come:

$$f(x) = max\ (0, x) = \begin{cases} x_i\ se\ \ x_i \geq 0 \\ 0\ se\ x_i < \ 0 \end{cases}$$

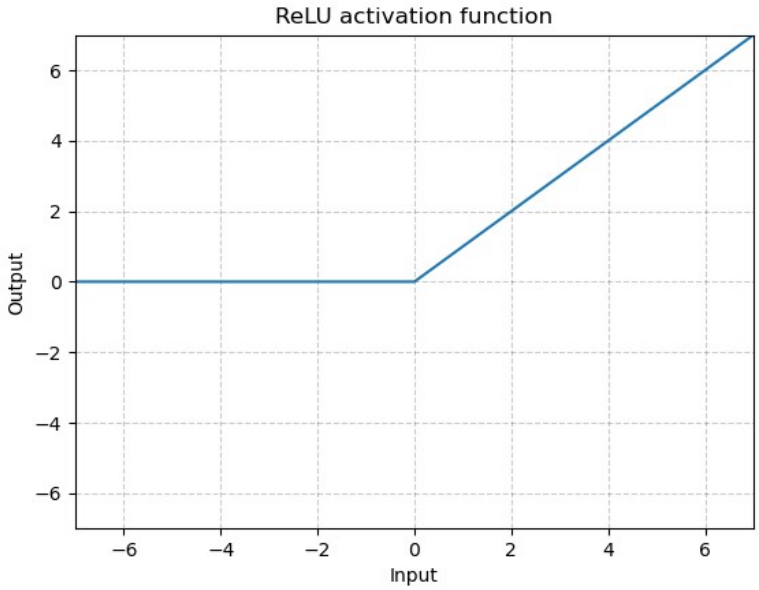

Figura 3: Funzione di attivazione ReLU

Il vantaggio principale dell'utilizzo di ReLU è che garantisce un calcolo più veloce rispetto alle altre funzioni di attivazione, in quanto non calcola funzioni complesse fornendo una velocità complessiva di calcolo migliore. Ha però il limite di essere particolarmente fragile: a causa della sua forma alcuni neuroni potrebbero non venire mai attivati durante l'addestramento, causando il non aggiornamento dei pesi.

Per limitare questa fragilità sono state proposte diverse varianti, che permettono di minimizzare gli effetti negativi presenti.

Le principali varianti utilizzate sono la Leaky ReLU[11] e la PReLU[12]. La differenza sostanziale rispetto alla ReLU originale è che queste funzioni possiedono una pendenza per valori negativi invece di una pendenza piana.

Matematicamente queste funzioni si presentano nella forma:

$$y_i = \begin{cases} x_i \; se \; x_i \geq 0 \\ \dfrac{x_i}{a_i} \; se \; x_i < 0 \end{cases}$$

dove a_i è un parametro fisso nell'intervallo $(1, +\infty)$.

La differenza tra le due funzioni sta nel fatto che la Leaky ReLU ha il coefficiente determinato prima del training, mentre nella PReLU a_i viene appreso durante l'addestramento tramite retro-propagazione.

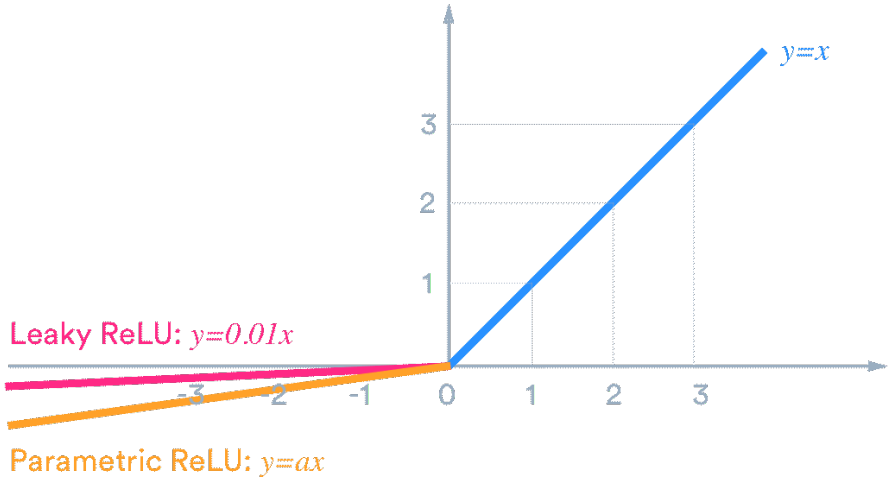

Figura 4: Funzione di attivazione ReLU con le sue varianti

È stato dimostrato che per certi tipi di dataset queste modifiche portano a prestazioni migliori della rete neurale. In particolare, è stato dimostrato che queste funzioni hanno effetto di prevenire lo spegnimento dei neuroni di cui soffre la ReLU.

Inoltre, diversi studi hanno evidenziato come sul training set, l'errore di PReLU è sempre il più basso mentre l'errore di Leaky ReLU è più alto della funzione originale.

Questo indica che PReLU soffre di un grave problema di overfitting nei dataset di piccole dimensioni.

[11] Leaky Rectified Linear Unit.

[12] Parametric Rectified Linear Unit.

Sigmoide

È una funzione non lineare che trasforma i valori nell'intervallo da 0 a 1. Matematicamente, e definita come:

$$f(x) = \frac{1}{1 + e^{-x}}$$

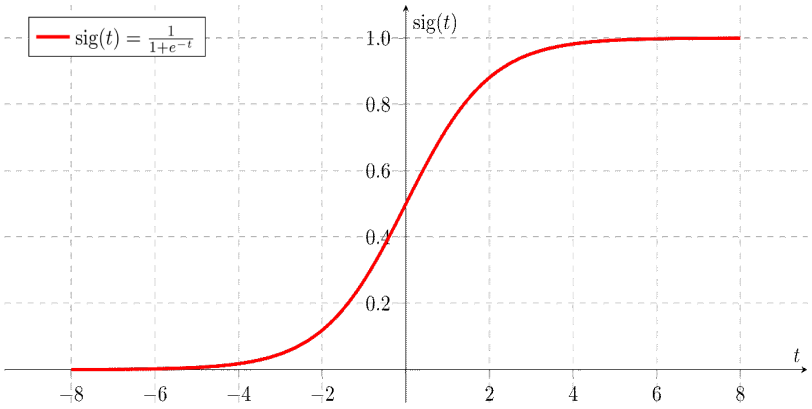

Figura 5: Funzione di attivazione sigmoide

La funzione sigmoide è continuamente differenziabile e ha l'importante proprietà che la sua derivata è uguale a:

$$f(x)' = 1 - sigmoid(x)$$

Viene evidenziato come la forma di questa derivata sia a campana.

Inoltre, non è simmetrica attorno allo 0, il che significa che il segno di tutti i valori di uscita dei neuroni saranno uguali. Questo problema può essere migliorato scalando la funzione sigmoide.

Tra i problemi principali che la contraddistinguono ci sono la saturazione del gradiente, la convergenza lenta e l'uscita non centrata a zero, causando così la propagazione degli aggiornamenti del gradiente in direzioni diverse.

Softmax

La funzione Softmax è una combinazione di molteplici funzioni sigmoidi. Poiché sappiamo che una funzione sigmoide restituisce valori nell'intervallo da 0 a 1, questi possono essere trattati come probabilità dei punti dati di una particolare classe. Questa a differenza delle funzioni sigmoidi utilizzate per la classificazione binaria, può essere utilizzata per problemi di classificazione multiclasse. La funzione restituisce la probabilità di ogni dato per singola classe; si riesce dunque ad avere una visione complessiva di quanto un certo punto sia appartenente ad una certa classe e quanto è correlato rispetto alle altre. Matematicamente può essere espressa come:

$$f(x_i) = \frac{e^{x_i}}{\sum_j x_j}$$

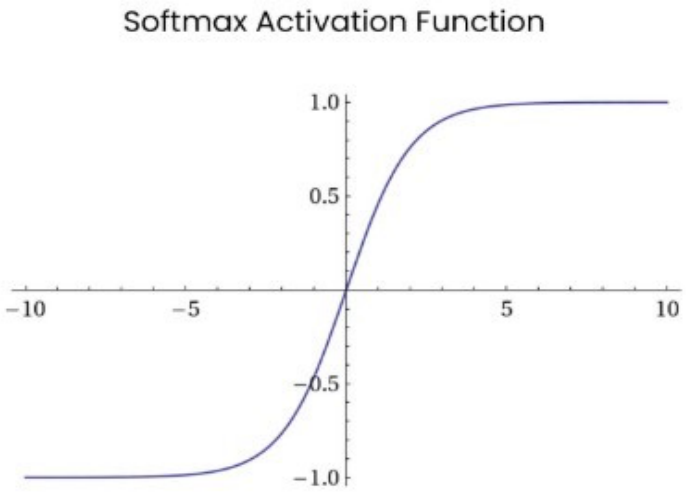

Figura 6: Funzione di attivazione softmax

Quando costruiamo una rete o un modello per la classificazione multiclasse, lo strato di output della rete avrà lo stesso numero di neuroni del numero di classi presenti.

Dropout

Per prevenire fenomeni di overfitting[13] è stata utilizzata la tecnica del *Dropout*. Questo consiste nel rimuovere unità (nascoste e visibili) in una rete neurale. Scaricando un neurone, intendiamo rimuoverlo temporaneamente dalla rete, insieme a tutte le sue connessioni in entrata e in uscita.

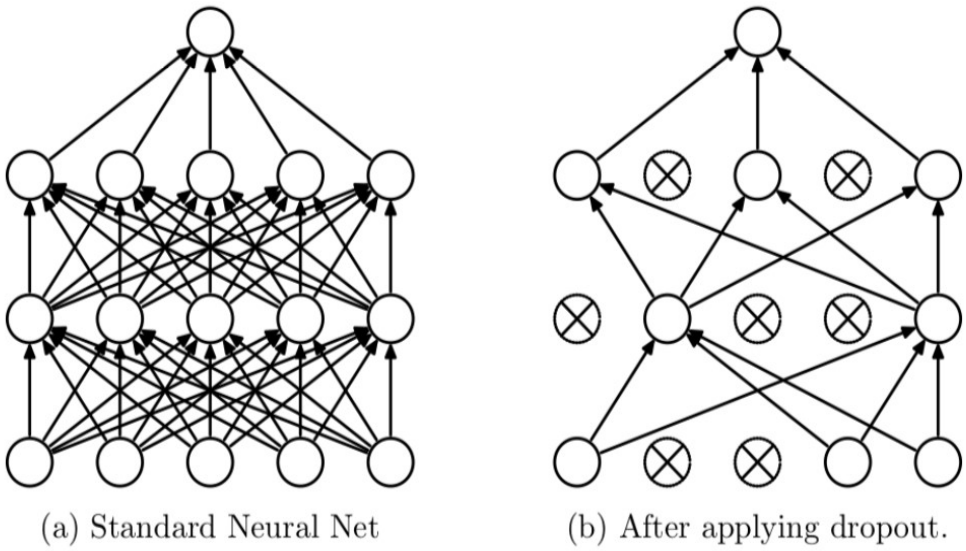

Figura 7: Modello Dropout su Rete Neurale. A sinistra: Una rete neurale standard con 2 hidden layer. A destra: Un esempio di una rete assottigliata prodotta applicando dropout sulla rete a sinistra. I neuroni segnati sono stati esclusi.

La scelta delle unità da eliminare è casuale. Nel caso più semplice, ogni unità è mantenuta con una probabilità fissa *p* indipendente da altre unità, dove *p* può essere scelto utilizzando un set di validazione o può semplicemente essere impostato a 0.5, che sembra essere vicino all'ottimale per una vasta gamma di reti e attività (Srivastava, et al. 2014).

Il *Dropout* può anche essere interpretato come un modo per regolarizzare una rete neurale aggiungendo rumore alle sue unità nascoste.

I benefici apportati da questo metodo sono molteplici: permette di ridurre il numero di pesi da calcolare e inoltre diminuisce il tempo necessario per effettuare il training della rete. Questo perché una rete neurale con *n* unità può essere vista come una collezione di 2^n possibili reti neurali alle quali il *Dropout* è stato applicato.

[13] La produzione di un'analisi che corrisponda troppo strettamente o esattamente a un particolare insieme di dati, e che possa quindi non essere in grado di inserire dati aggiuntivi o di prevedere in modo affidabile le osservazioni future.

Queste condividono tutti i pesi in modo che il numero totale di parametri sia ancora $O(n^2)$, complessità esponenziale che richiede un notevole sforzo computazionale per modelli con dimensioni elevate.

Per *n* elevati ogni volta che bisogna effettuare un training, è probabile utilizzare una nuova rete 'assottigliata'. Così il training di una rete neurale con *Dropout* può essere visto come l'addestramento di una raccolta di 2^n reti assottigliate con un enorme condivisione di pesi, dove ogni rete viene interamente addestrata molto raramente.

Quando il modello viene utilizzato per il test, non è possibile calcolare una media esplicita delle previsioni a partire dal numero esponenziale di modelli assottigliati. Tuttavia, un semplice metodo di media approssimata funziona bene. L'idea è quella di utilizzare una singola rete neurale, senza *Dropout*, al momento del test. I pesi di questa rete di test sono versioni scalate dei pesi delle reti utilizzate durante l'allenamento.

I pesi sono dimensionati in modo tale che per ogni dato input di un'unità nascosta l'output atteso (sotto la distribuzione utilizzata per eliminare le unità al momento dell'allenamento) sia lo stesso dell'output al momento della prova. Quindi, se un'unità viene mantenuta con probabilità *p*, questo equivale a moltiplicare i pesi in uscita di quell'unità per *p*.

Con questo metodo approssimativo, 2^n reti con pesi condivisi possono essere combinate in una singola rete neurale da utilizzare al momento della prova. La formazione di una rete con *Dropout* e l'utilizzo del metodo di media approssimata al momento del test porta a un errore significativamente inferiore su una vasta gamma di problemi di classificazione (Srivastava, Improving neural networks with dropout. 2013).

Risultati

In questa sezione riportiamo i risultati ottenuti applicando le metriche introdotte precedentemente al dataset descritto.

Questo permette di definire delle misure formali per la verifica dell'effettiva polarizzazione delle opinioni all'interno del questionario.

Prima di dividere i dati in sottogruppi, scelti secondo i parametri forniti di risposta al sondaggio, si è calcolata la media e la dispersione di tutto l'insieme.

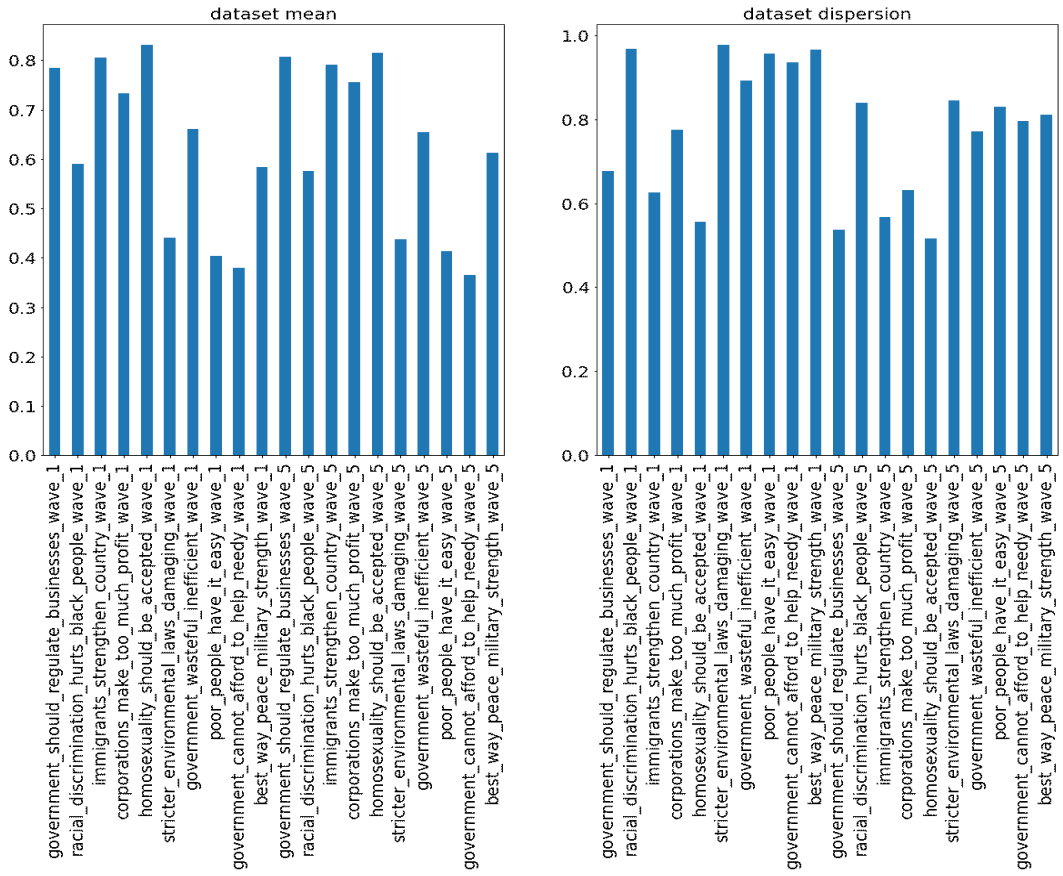

Figura 8: Risultati delle metriche applicate all'intero dataset, a sinistra la media mentre a destra la dispersione

Come si può notare dalla Figura 8 il valore della media oscilla da domanda a domanda: alcune non incrementano la polarizzazione del dataset in quanto il loro valore medio tende a 0 oppure a 1, vi è dunque una convergenza di opinioni da parte della popolazione intervistata. Per essere totalmente polarizzato la media dovrebbe essere vicino a 0.5, questo evidenzia che ci sono due gruppi distinti, con circa lo stesso numero di partecipanti, che hanno opinioni diametralmente opposte.

Le stesse considerazioni valgono anche per la dispersione: questa ha un valore massimo pari a 1, che indica che il dataset è completamente polarizzato. Viene osservato che alcune domande si avvicinano a questo limite, mettendo in evidenza come su alcuni argomenti il pensiero degli intervistati sia in aperta opposizione. La dispersione è maggiore nelle domande dove la media è circa pari a 0.5, che conferma la nostra ipotesi di dataset polarizzato, con alcune domande che hanno un peso maggiore nel determinare questo fenomeno.

La metrica dello spread non viene presa in considerazione: per come è stata definita produce esiti non vantaggiosi all'analisi. Prendendo solo il valore più alto e più basso dell'intero dataset ritorna valori inconcludenti; avendo a disposizione solo due possibilità di risposta (In Disaccordo/Concordo) vi saranno sempre almeno due risposte distinte in riscontro ad una particolare domanda. Non verrà più presentata questa metrica in quanto produce risultati infruttuosi, specifica che la polarizzazione è presente ma non è in grado di discernere la quantità.

Partendo da questi presupposti è stata effettuata un'analisi in sottogruppi, dividendo il dataset secondo le seguenti proprietà:

- Schieramento politico, divisione secondo il sistema americano diviso tra democratici e repubblicani;
- Genere, divisione in base al sesso dei partecipanti al sondaggio;
- Benessere sociale, divisione in base al reddito annuale medio degli intervistati, tra chi guadagna di più di 80000$ l'anno e chi ne guadagna meno;
- Livello di istruzione, divisione in base al massimo titolo di studio conseguito, la separazione è stata effettuata tra chi ha concluso la scuola superiore come titolo di studio più elevato e chi ha almeno conseguito un titolo universitario.

Come da aspettative si è notato che il livello più alto di polarizzazione si è riscontrato con il sottogruppo diviso da ideologia politica.

Secondo quanto riportato dai grafici in seguito l'opinione pubblica sente maggiormente l'influenza esterna riguardo ad alcuni temi specifici, tra i quali l'immigrazione, la corsa agli armamenti o la tassazione.

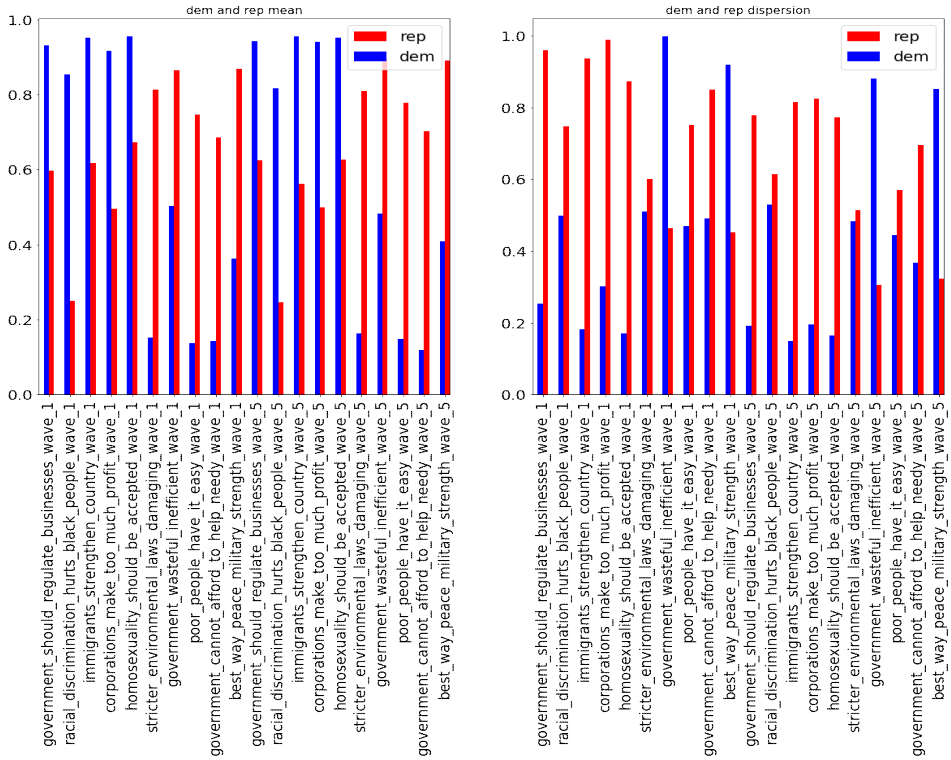

Figura 9: Risultati delle metriche applicate utilizzando come discriminate l'appartenenza ad uno dei due partiti politici americani, a sinistra la media mentre a destra la dispersione

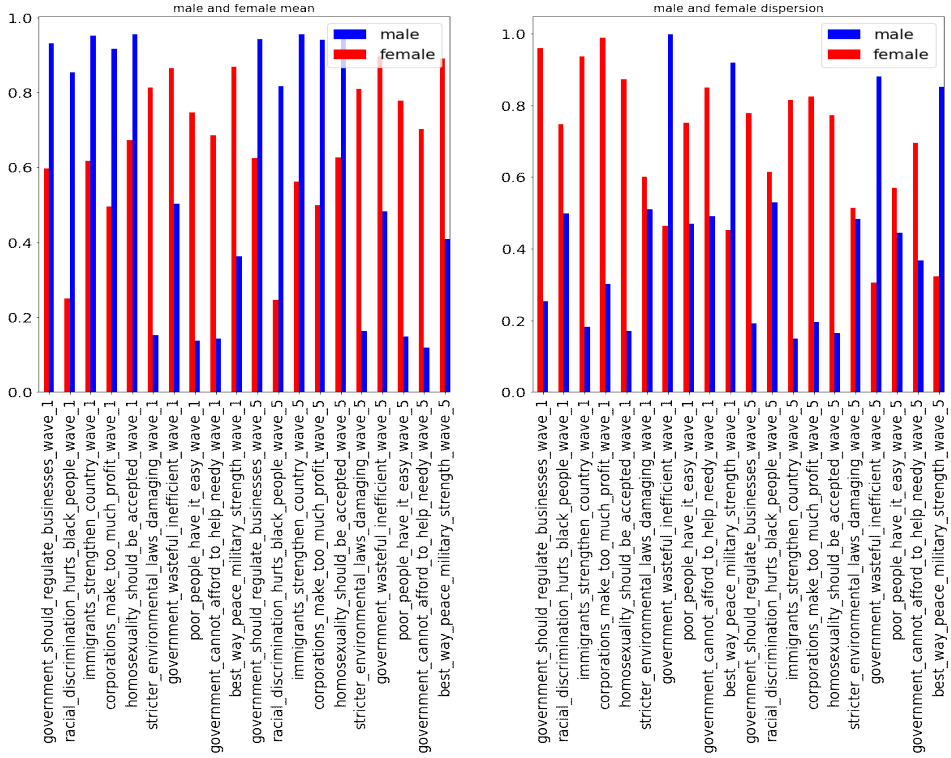

Figura 10: Risultati delle metriche applicate utilizzando come discriminate il genere degli intervistati, a sinistra la media mentre a destra la dispersione

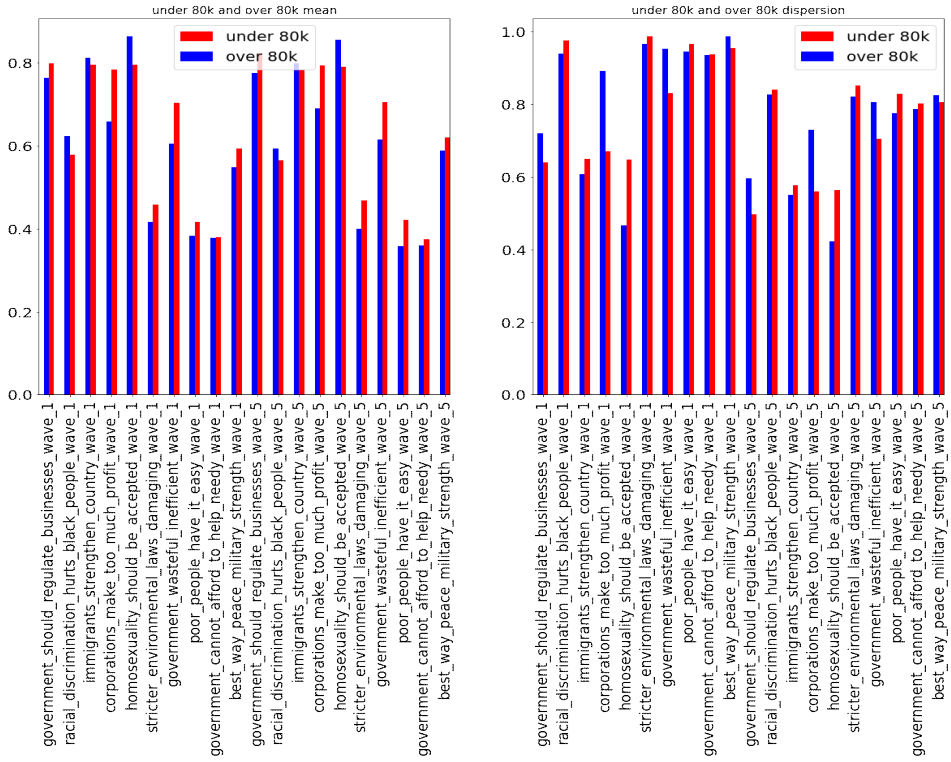

Figura 11: Risultati delle metriche applicate utilizzando come discriminate il reddito annuo degli intervistati, a sinistra la media mentre a destra la dispersione

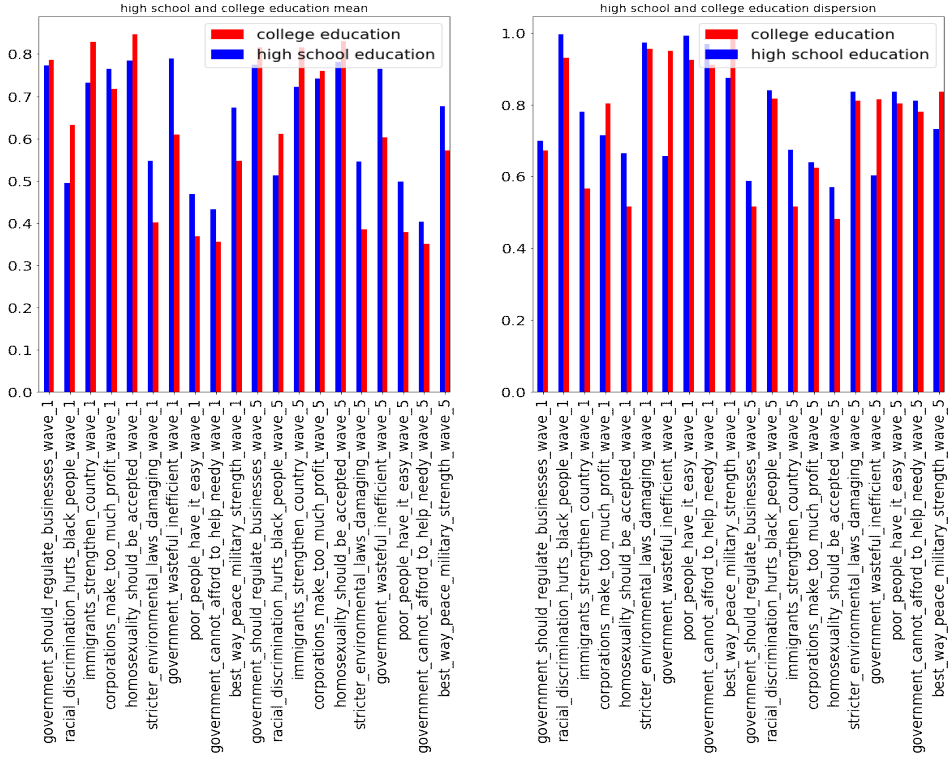

Figura 12: Risultati delle metriche applicate utilizzando come discriminate il massimo titolo di studio conseguito, a sinistra la media mentre a destra la dispersione

Tra i vari sottogruppi analizzati la polarizzazione maggiore si è riscontrata per opinione politica e genere. In questi due sottoinsiemi le influenze ambientali sono più forti, le idee della maggioranza sono ferme nei partecipanti. Anche in questo caso alcune domande generano più dissenso rispetto ad altre. Più la media tende a 0.5, e di conseguenza la dispersione tende a 1, più le risposte sono estremizzate verso due schieramenti opposti, generando un alto livello di polarizzazione.

Poiché questi due sottogruppi producono, tramite i risultati prodotti dalle metriche, un elevato livello del fenomeno abbiamo provato a verificare se, applicando entrambi i fattori di separazione, si è in grado di verificare se anche in questo caso i risultati sono correlati.

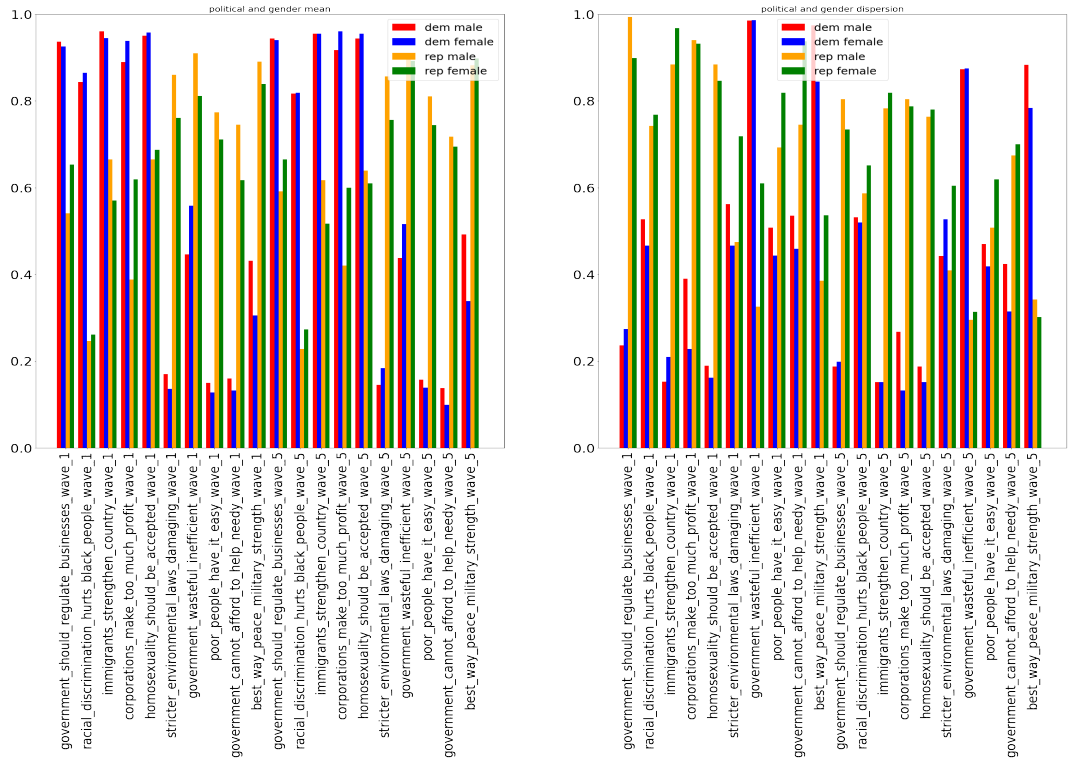

Figura 13: Risultati delle metriche applicate utilizzando come discriminate sia il partito politico di appartenenza che il genere dei partecipanti, a sinistra la media mentre a destra la dispersione

Le metriche calcolate sono le stesse delle analisi precedenti. In questo caso si nota come, quando vengono applicati entrambi i fattori, non si vengono a creare due gruppi distinti. La Figura 13 mette in evidenza come due persone appartenenti a due schieramenti opposti (ad esempio repubblicani e democratici maschi) hanno visioni simili su argomenti dove, senza la discriminante del sesso, avevano opinioni contrastanti. Questo ci porta ad

affermare che la polarizzazione deriva principalmente dal partito politico di appartenenza, e risulta essere il fattore principale di contrasto tra la popolazione analizzata.

Partendo da questo presupposto abbiamo calcolato la differenziabilità tra i due sottogruppi politici. Non è stata calcolata anche per gli altri fattori di divisione in quanto si è osservato che il livello del fenomeno non è così elevato. Questa metrica permette di constatare quanti individui appartenenti a un gruppo assumono atteggiamenti simili a quelli di un'altra fazione. Dunque, grazie a questa analisi, si è in grado di capire quante persone sono estremizzate nella loro idea e quante assumono invece atteggiamenti più moderati. Più il valore sarà basso più gli appartenenti ai due schieramenti saranno estremizzati nelle loro idee.

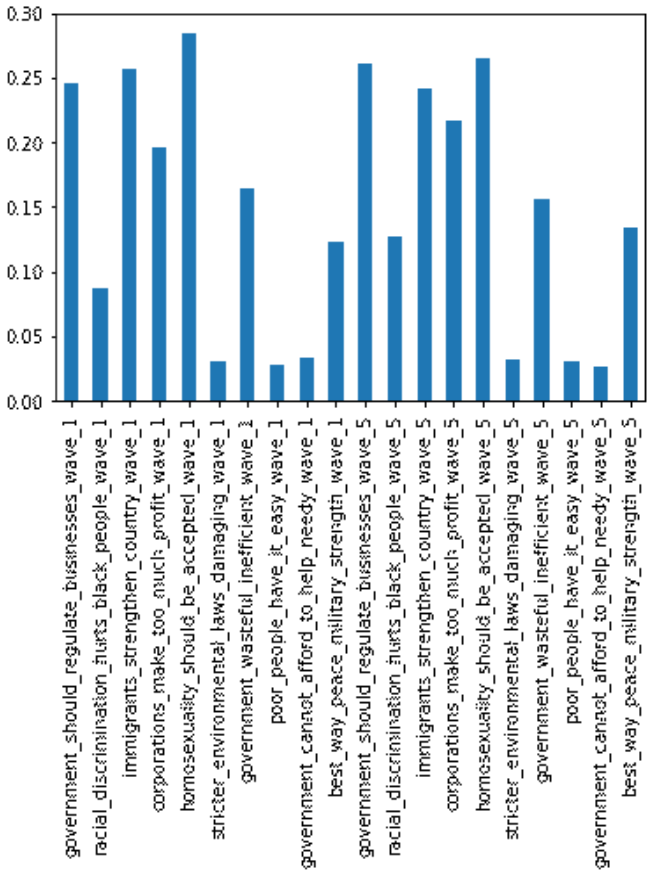

Figura 14: Differenziabilità tra i due sottogruppi divisi dalle loro opinioni politiche

Come si può notare la maggioranza dei valori assunti non è elevata, questo permette di affermare che i due gruppi sono distinti, dunque polarizzati. Viene evidenziato anche come alcune risposte generano più distinzione di altre, questo a causa degli atteggiamenti moderati che possono essere assunti all'interno di ciascun gruppo.

Tramite queste analisi possiamo giungere alla conclusione che i dati forniti in risposta al sondaggio sono polarizzati e considerando i vari sottogruppi possibili il fenomeno è molto più marcato quando si dividono le persone secondo le loro ideologie politiche.
Abbiamo notato come gli appartenenti ad una certa formazione politica prediligano una particolare linea di pensiero riguardo certi ambiti, cioè l'appartenenza ad un determinato schieramento influenza l'opinione di un individuo riguardo alcune tematiche trattate.
Le metriche evidenziano come i democratici siano portati ad avere idee più progressiste riguardo diritti civili ed ambiente, mentre i repubblicani siano più conservatori.
Sulla base di queste constatazioni abbiamo calcolato la matrice di correlazione, che evidenzia la connessione tra le varie risposte fornite dagli utenti.

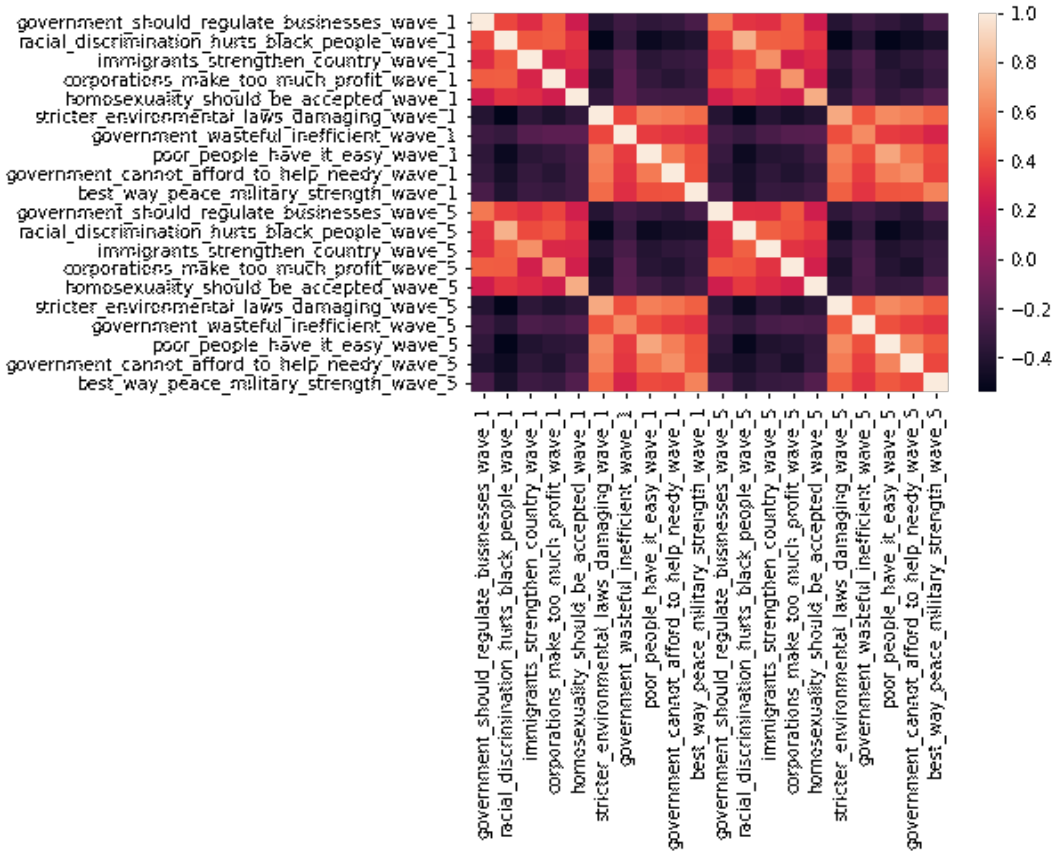

Figura 15: Matrice di correlazione

La Figura 15 evidenzia come le risposte fornite da alcuni utenti riguardo certe domande hanno una probabilità maggiore di portare una risposta simile riguardo certi argomenti. In particolare, si nota questo fatto in ambito sociale ed ambientale. La struttura a scacchi della matrice ci permette di constatare ulteriormente la polarizzazione dei dati: si sono sviluppate due linee di pensiero tra loro non correlate, che portano gli utenti ad un

contrasto ideologico. Questi due sottogruppi sono i due partiti politici utilizzati come discriminante, ovvero il partito democratico e quello repubblicano.

Rete Neurale

Dopo questa analisi formale sulla polarizzazione effettuata utilizzando le metriche sopra riportate, abbiamo implementato una semplice rete neurale in due varianti: entrambe predicono il partito politico attuando due funzioni di attivazioni diverse nell'ultimo livello output, sigmoide e softmax.

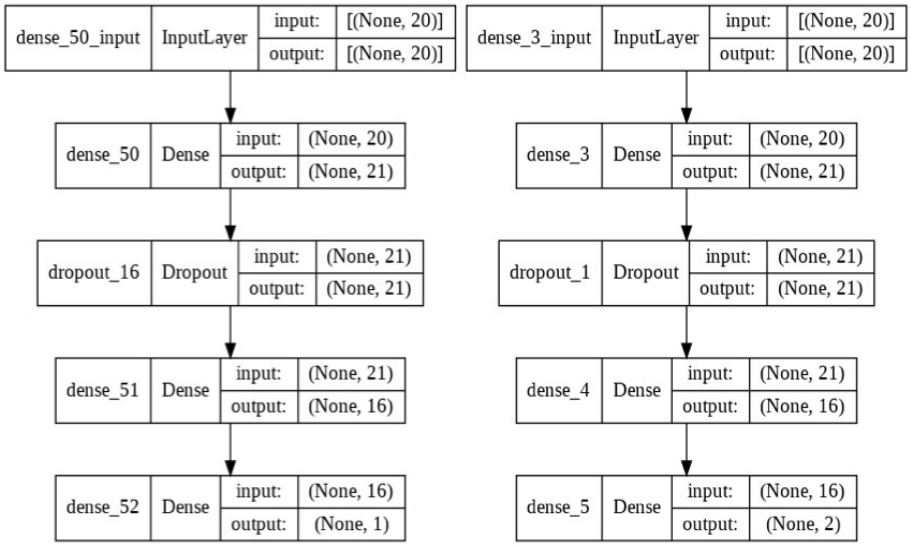

Figura 16: Riassunto delle reti sviluppate con il loro relativo design. A destra la rete con funzione di attivazione sigmoide, a sinistra la rete con funzione di attivazione softmax

Nel primo caso la rete è composta da 21 neuroni di input, che consistono nelle features che descrivono le risposte alle 20 domande fornite dagli utenti più un neurone di bias, 16 neuroni nell'hidden layer e un solo neurone di output, che riporta il partito predetto. Viene applicato il dropout con una percentuale di disattivazione dei neuroni pari a 0.3, per avere delle prestazioni migliori. I cicli di addestramento totali sono 10, analisi con un numero più elevato non portano benefici alla rete, questo è legato al numero di campioni ridotto che porta a situazioni di overfitting.

Una volta addestrata questa rete, si ottengono i seguenti risultati:

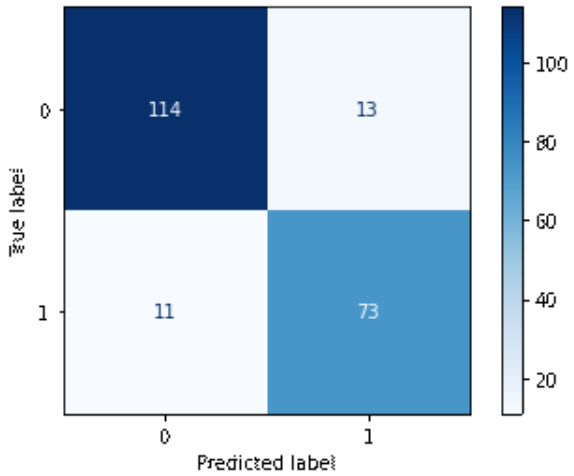

Figura 17: Riassunto dei risultati di predizione del partito politico prodotto dalla rete utilizzando la funzione di attivazione sigmoide, 0 corrisponde al partito democratico e 1 al partito repubblicano

Come si può notare la rete ha una buona percentuale di successo nella predizione del partito, passate come parametri le risposte al questionario: il tasso di successo è di circa 88%. Questo permette di avere sia buone prestazioni durante la fase di training che in validation (anche se leggermente inferiori a causa di un loss elevato).

La Figura 17 evidenzia come su un test con 211 campioni il modello è riuscito a classificare correttamente 114 utenti come democratici e 73 come repubblicani. Ci sono 24 elementi valutati in maniera errata; viene constatato che la classificazione delle persone appartenenti al partito repubblicano è più difficile in quanto gli intervistati sono presenti in maniera sbilanciata nel dataset.

Allo scopo di migliorare le prestazioni della rete abbiamo anche provato ad implementare il sistema usando come funzione di attivazione softmax in output. La rete presenta la stessa struttura sia nell'input layer che nell'hidden layer, come evidenzia la Figura 16. L'unica differenza si riscontra nell'output layer, dove sono presenti due neuroni complessivi. Questo permette di calcolare la percentuale di appartenenza ad entrambi i partiti, in modo da avere una visione d'insieme delle idee dell'individuo. Tramite questa analisi si è in grado di visualizzare il livello di polarizzazione all'interno del singolo individuo.

I risultati ritornati ci permettono di distinguere le figure moderate da quelle estremizzate. Tramite l'analisi delle risposte fornite viene calcolata la percentuale di appartenenza ai

partiti, si è in grado di visualizzare quanto un individuo condivida le idee del partito opposto. Questo ci permette di affermare se l'individuo è aperto al dialogo o se la sua opinione sia totalmente in linea con quella del suo schieramento politico.

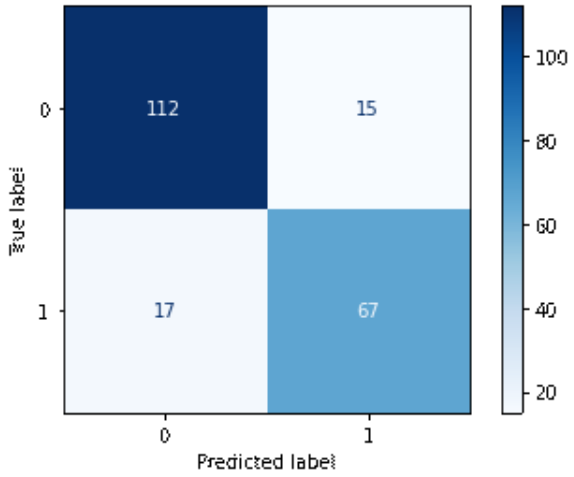

Figura 18: Riassunto dei risultati di predizione del partito politico prodotto dalla rete utilizzando la funzione di attivazione softmax, 0 corrisponde al partito democratico e 1 al partito repubblicano

Come si può constatare questa rete ottiene risultati simili a quelli precedenti, leggermente inferiori ma comunque accettabili. La differenza sostanziale rispetto al modello sviluppato in precedenza è che ora siamo in grado di conoscere la percentuale di consenso rispetto al proprio partito, siamo in grado di definire gli individui moderati da quelli completamente estremizzati.

Queste analisi inducono un dilemma etico: a partire da un questionario con domande che riflettono l'ambito ideologico, sociale e ambientale siamo in grado di conoscere in maniera affidabile, con una percentuale di successo pari a 85% circa, delle informazioni personali non fornite. Siamo quindi in grado di profilare gli utenti su questioni sensibili senza che questi abbiamo esplicitamente espresso opinioni a riguardo.

Problema Etico

L'analisi precedente ha portato alla luce un dilemma che sempre più spesso si interfaccia con gli utenti dei social. Quanti dati ed informazioni queste piattaforme salvano senza il nostro effettivo consenso?

Come è stato spiegato in precedenza gli algoritmi alla base di questi strumenti sono progettati per far visualizzare agli utenti i contenuti a loro più favorevoli, in modo da assecondare le preferenze degli utilizzatori. Per effettuare questo, tutte le interazioni con il sito vengono salvate dai fornitori di questi servizi. Tutto questo porta ad un'attività di profilazione non sempre consensuale da parte degli utenti (Contissa, et al. 2018).

Studiando in maniera continua le nostre abitudini, le piattaforme arrivano a conoscere le nostre preferenze in maniera sempre più accurata, riuscendo a prevedere le cose che vorremmo vedere e influenzandoci nelle nostre decisioni quotidiane.

Un caso eclatante di questo fenomeno è stato *Cambridge Analytica* (Schneble, Elger e Shawn 2018), società di consulenza britannica, che è riuscita a raccogliere dati di oltre 87 milioni di utenti di Facebook senza il loro consenso (ur Rehman 2019). L'azienda ha ottenuto l'accesso a 320.000 profili e dati dei loro amici attraverso l'app "thisisyourdigitallife"[14]. Per avere diritto all'utilizzo l'utente doveva avere un account Facebook ed essere un elettore degli Stati Uniti in modo che decine di milioni di profili potessero essere abbinati alle liste elettorali. Nel giro di pochi mesi Kogan e Cambridge Analytica hanno avuto un database di milioni di elettori statunitensi, che sono stati analizzati e dei quali è stata effettuata una profilazione politica.

In questo modo potevano decidere a chi indirizzare i loro messaggi e creare loro contenuti ad hoc in modo da attrarre gli individui e influenzare le loro azioni o pensieri, tecnica nota come micro-targeting. I partecipanti ai test avevano accettato di condividere i loro dati, e i termini predefiniti di Facebook permettevano di raccogliere le informazioni dei loro amici dall'app, a meno che questi non avessero modificato le loro impostazioni sulla privacy. Nessuno di loro, però, aveva accettato di vendere i loro dati ad aziende come Cambrdige Analytica che li hanno trasformati in uno strumento politico. Sebbene le informazioni siano state anonimizzate e aggregate, queste sono state poi utilizzate per modellare gli algoritmi di targeting per prevedere e influenzare il comportamento di voto individuale nelle elezioni presidenziali americane del 2016 (Hu 2020).

[14] Sviluppata dallo psicologo Alexandr Kogan dell'Universita di Cambridge, Regno Unito

Questo scandalo evidenzia come le applicazioni più popolari permettano l'accesso a informazioni. Questo mette a rischio la privacy di decine di milioni di utenti dei social.
Negli anni l'opinione pubblica ha richiesto un controllo più serrato sui contenuti che queste piattaforme possono salvare e far vedere ai loro utenti. La regolamentazione maggiore è stata attuata nell'Unione Europea con il Regolamento Generale sulla Protezione dei Dati (RGDP)[15]. Questo documento, relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati, è stato accolto come un passo progressivo verso la correzione dello sbilanciamento di potenza nella raccolta di dati digitali da parte di entità che sviluppano, mantengono e controllano l'accesso alle infrastrutture digitali. Il RGDP ha due obiettivi principali.

- Proteggere i diritti e le libertà fondamentali dei soggetti creando un regime di protezione per quanto riguarda il trattamento dei dati personali.
  Ciò è dovuto al fatto che le nuove tecnologie e i modelli organizzativi, sia nel settore privato che in quello pubblico, hanno reso facile la raccolta, l'utilizzo, la combinazione, l'aggregazione e l'elaborazione di un'ampia quantità di dati personali senza una supervisione sufficiente.
- Creare le condizioni ottimali affinché la libera circolazione dei dati personali, parallelamente alla libera circolazione delle merci e dei servizi, possa avvenire all'interno dell'UE, sostenendo la creazione del mercato unico europeo.

L'obiettivo del RGDP è quello di fornire un modo per raggiungere la libertà di accesso ai dati all'interno dell'UE, garantendo allo stesso tempo la protezione dei diritti e delle libertà fondamentali per gli individui (Kotsios 2019).
Diverse critiche sono state mosse a questa regolamentazione, tra le quali il fatto che l'attuazione di queste norme è molto costosa; dunque, causa alle piccole/medie aziende dei costi che non sono sostenibili. Questo favorisce i grandi colossi che hanno budget più elevati per pagare gli aggiornamenti software e professionisti della privacy.
Tutto questo mostra come la regolamentazione dei social, nonostante sia presente, sia ancora inadeguata (Loreggia e Sartor 2020).
La tecnologia sviluppata ha mostrato come sia facile riuscire ad acquisire in tempo reale informazioni riservate senza richiederle formalmente: è infatti possibile raccogliere

[15] Regolamento (UE) n. 2016/679.

informazioni e conoscenza comportamentale attraverso questionari che permettono di inferire tali informazioni a latere. Tramite queste si possono implementare politiche di targetting che possono portare alla creazione e al rafforzamento della divisione sociale, lasciando agli utenti solo la visione di contenuti che rispecchiano le loro idee. Questi diventano così assuefatti da informazioni simili che li porterà a credere che solo il loro pensiero sia quello che rispecchia la realtà.
Pubblicando apertamente su piattaforme private con politiche di privacy più o meno restrittive, gli utenti mettono a disposizione un enorme quantitativo di dati che possono essere utilizzati per la loro profilazione da parte di società terze.
Queste predizioni possono poi essere usate per scopi pubblicitari, per influenzare o rafforzare le idee pubblicate, senza affrontare mai un dibattito costruttivo (Quintarelli, et al. 2019).

Sviluppi Futuri

Il sistema proposto è solo una semplice analisi che fornisce misure di calcolo della polarizzazione e predizione del partito politico di interesse basandosi su dati che seguono il sistema istituzionale americano.
Diverse migliorie possono essere sviluppate:

- Incrementare il numero di campioni presenti nel dataset in modo da poter analizzare al meglio le prestazioni del modello predittivo;
- Studiare e sviluppare nuove metriche di misura della polarizzazione ed effettuare un confronto con quelle attuali, in modo da visualizzare quali forniscono risultati più accurati;
- I dati ottenuti si riferiscono al sistema istituzionale americano, sistema bipartitico con una netta divisione di idee. È possibile effettuare lo stesso studio anche su stati con istituzioni politiche differenti, sarà però necessario raccogliere dati specifici.

Conclusioni

L'analisi riportata in questo lavoro ha evidenziato come, partendo dalle risorse condivise sui social network, sia possibile distinguere gli utilizzatori in due classi diverse con visioni diametralmente opposte. Partendo dai risultati di un semplice sondaggio online si riescono a ricavare diverse valutazioni che evidenziano come i dati siano polarizzati.

È stato possibile identificare due gruppi contrapposti e mapparli nei due principali schieramenti politici.

Questo ha evidenziato come la polarizzazione sia legata in qualche modo al pensiero politico degli utenti, che li porta ad avere idee conflittuali con gli appartenenti al gruppo opposto.

Abbiamo mostrato come utilizzando metriche già presenti in letteratura sia possibile dimensionare il fenomeno in un particolare contesto. In particolare, abbiamo evidenziato che lo spread non è una metrica efficiente per la verifica in quanto non considera tutta la popolazione, ma soltanto gli estremi. La media, invece, è una buona misura per avere una visione d'insieme della totalità dei dati, così come la dispersione. Entrambe rendono possibile esprimere il grado di polarizzazione presente. La differenziabilità aiuta a visualizzare quanto estremizzati sono i dati analizzati, fornendoci una misura di quanti individui sono moderati e quanti estremizzati verso le proprie opinioni. Abbiamo verificato quale correlazione lega le diverse variabili implicate e sulla base di questa evidenza abbiamo sviluppato una rete neurale in grado di predire l'appartenenza ad un movimento politico con una buona affidabilità (circa 85% di accuratezza).

Questa analisi ci ha portato ad evidenziare quanto le piattaforme digitali siano in grado di estrarre informazioni personali ed utilizzarle per altri scopi, eludendo in molti casi i controlli e il nostro consenso.

Bibliografia

Bail, C. A., L. P. Argyle, T. W. Brown, J. P. Bumpus, H. Chen, M. F. Hunzaker, e A. Volfovsky. 2018. «Exposure to opposing views on social media can increase political polarization.» . *Proceedings of the National Academy of Sciences.* 9216-9221.

Barrett-Maitland, Nadine, e Jenice Lynch. 2019. «Social media, ethics and the privacy paradox. Security and privacy from a legal, ethical, and technical perspective.»

Boyd, D., e N. Ellison. 2008. «Social Network Sites: Definition, History, and Scholarship.» *Journal of Computer-Mediated Communication* 13: 210-230.

Bozdag, VE. 2015. «Bursting the filter bubble: Democracy, design, and ethics.» TU Delft.

Bramson, Aaron, Patrick Grim, Daniel J. Singer, Steven Fisher, William Berger, Graham Sack, e Carissa Flocken. 2016. «Disambiguation of social polarization concepts and measures.» *The Journal of Mathematical Sociology* 40 (2): 80-111.

Bruns, Axel. 2019. *It's not the technology, stupid: How the 'Echo Chamber'and 'Filter Bubble'metaphors have failed us.* International Association for Media and Communication Research.

Contissa, Giuseppe, Francesca Lagioia, Marco Lippi, Hans-Wolfgang Micklitz, Przemyslaw Palka, Sartor, e Paolo Torroni. 2018. «Towards consumer-empowering artificial intelligence.» *In International Joint Conference on Artificial Intelligence.* 5150-5157.

Druckman, J., E. Peterson, e R. Slothuus. 2013. «How Elite Partisan Polarization Affects Public Opinion Formation.» *American Political Science Review* 107 (1): 57-79.

Guerra, P., Meira Jr., W., C. Cardie, e R. Kleinberg. 2021. «A Measure of Polarization on Social Media Networks Based on Community Boundaries.» *Proceedings of the International AAAI Conference on Web and Social Media.* 215-224.

Hu, Margaret. 2020. «Cambridge Analytica's black box.» *Big Data & Society 7.2 (2020): 2053951720938091. APA* 7 (2).

Kijsipongse, E., S. U-ruekolan, C. Ngamphiw, e S. Tongsima. 2011. « Efficient large Pearson correlation matrix computing using hybrid MPI/CUDA.» *Eighth International Joint Conference on Computer Science and Software Engineering (JCSSE).* 237-241.

Kitchens, Brent, Steven L. Johnson, and Peter Gray. 2020. «Understanding Echo Chambers and Filter Bubbles: The Impact of Social Media on Diversification and Partisan Shifts in News Consumption.» *MIS Quarterly* .

Kotsios, Andreas, et al. 2019. «An analysis of the consequences of the general data protection regulation on social network research.» *ACM Transactions on Social Computing 2.3* 1-22.

Levy, Ro'ee. 2021. «Social media, news consumption, and polarization: Evidence from a field experiment. .» *American economic review* 111 (3): 831-70.

Loreggia, A., e G. Sartor. 2020. *The impact of algorithms for online content filtering or moderation - Upload filters.* European Parliament.

McCarty, Nolan, Keith T. Poole, e Howard Rosenthal. 2016. *Polarized America: The dance of ideology and unequal riches. mit Press, 2016.* Cambridge: MIT Press.

Myers, D. G., e H. Lamm. 1975. «The polarizing effect of group discussion: the discovery that discussion tends to enhance the average pre discussion tendency has stimulated new insights about the nature of group influence.» *American Scientist* 63 (3): 297-303.

Nguyen, N. P., T. N. Dinh, Y. Xuan, e M. T. Thai. 2011. «Adaptive algorithms for detecting community structure in dynamic social networks.» *Proceedings IEEE INFOCOM.* 2282-2290.

Nwankpa, Chigozie, e et al. 2018. «Activation functions: Comparison of trends in practice and research for deep learning.» . *arXiv preprint arXiv:1811.03378.*

Pariser, E. 2011. *The filter bubble: What the Internet is hiding from you.* Penguin, UK.

Parramore, Lynn, intervista di The Atlantic. 2010. *The Filter Bubble* (10 10).

Quintarelli, S., F. Corea, F. Fossa, A. Loreggia, e S. Sapienza. 2019. «AI: profili etici. una prospettiva etica sull'intelligenza artificiale: princıpi, diritti e raccomandazioni.» *BioLaw Journal-Rivista di BioDiritto* 18 (3): 183-204.

Schneble, Christophe Olivier, Bernice Simone Elger, e David Shawn. 2018. *The Cambridge Analytica affair and Internet-mediated research.* EMBO reports.

Sharma, Sagar, Simone Sharma, e Athaiya Anidhya. 2017. «Activation functions in neural networks.» *Towards data science* 6 (12): 310-316.

Sîrbu, A., D. Pedreschi, F. Giannotti, e J Kertész. 2019. «Algorithmic bias amplifies opinion fragmentation and polarization: A bounded confidence model.» *PLoS ONE* .

Srivastava, Nitish. 2013. «Improving neural networks with dropout.» (University of Toronto).

Srivastava, Nitish, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, e Ruslan Salakhutdinov. 2014. «Dropout: a simple way to prevent neural networks from overfitting.» *The journal of machine learning research* 15 (1): 1929-1958.

Sunstein, C. R. 2002. «The Law of Group Polarization.» *Journal of Political Philosophy* 10: 175-195.

Tam, Kar, e Shuk Ho. 2006. «Understanding the Impact of Web Personalization on User Information Processing and Decision Outcomes.» *MIS Quarterly*.

ur Rehman, Ikhlaq. 2019. «Facebook-Cambridge Analytica data harvesting: What you need to know.» *Library Philosophy and Practice* 1-11.

Wilkinson, Leland, e Michael Friendly. 2009. «The history of the cluster heat map.» *The American Statistician* 63. (2): 179-184.

Xu, Bing, et al. 2015. «Empirical evaluation of rectified activations in convolutional network.»

Indice delle Figure