

Università Ca' Foscari
SOCIAL NETWORK ANALYSIS

Analisi comportamentale all'interno di
una piattaforma social

OTTOBRE/NOVEMBRE/DICEMBRE 2022

Scritto da:

Jinpeng Zhang

886854

Samuel Takyi Obeng

881431

INDICE

1 - Introduzione	3
- 1.1 Contesto	4
2 - Motivazione della ricerca	5
3 - Quesito	6
4 - Processo di analisi	7
- 4.1 Dati	7
- 4.2 Eventi analizzati	8
- 4.3 Analisi sentimentale	9
5 - Conclusioni	22

1. INTRODUZIONE

In questo report verrà riportato il comportamento degli utenti di una piattaforma dopo aver assistito alle azioni commesse da esponenti di un determinato movimento, per poi osservare le opinioni a riguardo.

Il movimento in questione lotta per divulgare e diffondere un messaggio di avvertenza a tutte le persone possibili che siano disposte ad ascoltarli.

Recentemente le attività sono aumentate a causa di un singolo avvenimento che ha portato alla diffusione di un comportamento analogo da parte di altri esponenti dello stesso movimento portando ad un incremento delle ricerche e discussioni online a riguardo.

Soprattutto in quest'ultimo periodo, in Italia, abbiamo sentito molto parlare di Just Stop Oil, un movimento nato da vario tempo in Gran Bretagna dove, degli attivisti del clima, imbrattano le opere d'arte dei musei per attirare l'attenzione di più persone possibili; l'obiettivo è quello di garantire che il governo si impegni a fermare tutte le autorizzazioni per lo sviluppo, l'esportazione e la produzione di combustibili fossili e i cambiamenti climatici.

Sono molti i pittori famosi in tutto il mondo, presi di mira dai giovani attivisti di gruppi che difendono l'ambiente come Just Stop Oil, Extinction Rebellion e Ultima Generazione (gruppo italiano). Questi gruppi di ragazzi mettono in scena atti dimostrativi in difesa dell'ambiente e del nostro pianeta. Chiedono che sia ascoltata la scienza e la parola degli scienziati impegnati negli studi per il clima.

La prima opera d'arte presa di mira con una torta, è stata La Gioconda di Leonardo da Vinci, ospite del museo parigino Il Louvre. Alla National Gallery della capitale inglese, gli attivisti si sono incollati al quadro The Hay Wain di John Constable, riportando in questo caso lievi danni all'opera d'arte. Successivamente, hanno preso di mira anche la Sala Botticelli degli Uffizi di Firenze, dove hanno appeso uno striscione con la scritta "Ultima generazione, No Gas No Carbon", per poi incollarsi al dipinto "La Primavera" di Botticelli...

Nella maggior parte dei casi, non si tratta di atti dimostrativi volti a danneggiare le opere d'arte, ma vogliono destabilizzare, creare disagio, suscitare delle reazioni per parlare di un tema così delicato come il cambiamento climatico.

Gli attivisti hanno capito che l'arte, per noi, rappresenta un punto debole per suscitare una nostra reazione essendo oggetti di valore inestimabile, nati dall'ingegno di artisti vissuti secoli fa.

1.1 Contesto

Dopo aver osservato questi eventi, causati da esponenti appartenenti a vari movimenti che lottano per il cambiamento climatico e soprattutto per diffondere un messaggio di avvertimento a tutto il pianeta, abbiamo notato una forte crescita delle ricerche riguardo tali avvenimenti. Analizzeremo come questi eventi abbiano avuto effetto sugli utenti di una piattaforma social attraverso l'analisi di Twitter tramite l'API (Application Programming Interfaces, ovvero le interfacce di programmazione delle applicazioni) che è stato fornito dall'azienda per determinare se hanno avuto un impatto positivo o negativo sul pubblico generale.

Gli eventi presi in considerazione per l'analisi sono dei reati commessi nei mesi di Ottobre/Novembre/Dicembre 2022 e i paesi presi in considerazione per l'analisi sono: Italia, Regno Unito, Germania, Francia, Spagna, Stati Uniti.

La protesta ambientalista infatti, è iniziata il 29 maggio 2022 al Louvre di **Parigi** dove è stata presa di mira la Gioconda di Leonardo da Vinci, colpita da un uomo travestito da signora anziana con una torta, egli è stato successivamente detenuto in un ospedale psichiatrico.

A luglio, una serie di musei inglesi entra nel mirino di Just Stop Oil, tra i quali la Courtauld Gallery di **Londra**, la Kelvingrove Art Gallery di Glasgow, la Manchester Art Gallery e la National Gallery della capitale dove i giovani attivisti si sono incollati ad un quadro di John Constable, riportando lievi danni.

Anche in **Italia**, nel mese di luglio si è verificata una protesta nella sala Botticelli degli Uffizi di Firenze, incollandosi al vetro che protegge il quadro che rappresenta La Primavera. A fine mese gli attivisti di Ultima Generazione attaccano il Museo del '900 di Milano e si incollano alla struttura che sostiene la scultura Forme uniche della continuità nello spazio di Umberto Boccioni.

A ottobre hanno preso di mira I girasoli di Vincent Van Gogh, custodita a **Londra**, con una zuppa di pomodoro. C'è stato anche il turno di Monet che è stato colpito da attivisti di Ultima Generazione presso il museo di Barberini in **Germania**.

Il 27 ottobre viene colpita l'opera di Vermeer, La ragazza con l'orecchino di perla, esposta nel museo Mauritshuis.

Due componenti di Futuro vegetal, al museo di **Madrid**, si sono incollati alle cornici dei dipinti di Francisco Goya.

Fortunatamente tali opere erano protette da un vetro perciò tutti i dipinti colpiti presentarono lievi oppure nessun tipo di danno.

2. Motivazione della ricerca

L'analisi sentimentale riguardo a individui che danneggiano opere d'arte è una ricerca importante per diversi motivi. In primo luogo, le opere d'arte sono un patrimonio culturale e storico che rappresenta una parte significativa della nostra identità e della nostra storia. Danneggiarle è un crimine che va contro il nostro patrimonio culturale e storico. In secondo luogo, l'analisi sentimentale può fornire informazioni utili per prevenire questi atti di vandalismo e proteggere le opere d'arte per le generazioni future. Ad esempio, l'analisi sentimentale può identificare i sentimenti e le opinioni degli individui sulla protezione delle opere d'arte, e queste informazioni possono essere utilizzate per formulare politiche più efficaci e migliorare l'educazione pubblica. Inoltre, l'analisi sentimentale può aiutare a capire le motivazioni alla base del comportamento vandalico e a trovare soluzioni per prevenirlo. In sintesi, l'analisi sentimentale riguardo a individui che danneggiano opere d'arte è una ricerca importante per proteggere e preservare il nostro patrimonio culturale e storico.

La motivazione dietro agli atti di vandalismo che vanno a danneggiare opere così importanti deve essere per loro di uguale importanza; infatti il motivo dietro alle loro azioni è la diffusione di politiche contro l'inquinamento dell'ambiente e la prevenzione di tale crisi.

La crisi climatica è un argomento molto attuale che avrà un impatto importante sulle generazioni future ma non è sempre al centro delle discussioni di piattaforme online, nonostante tutti sappiano quanto sia critica la situazione attuale.

Per delle ragioni psicologiche, non si parla molto del problema del riscaldamento globale anche se ne siamo a conoscenza.

Secondo lo psicologo Espen Stoknes, sono cinque le ragioni per cui la lotta al cambiamento climatico coinvolge poco le persone:

- la distanza temporale e spaziale delle conseguenze negative del cambiamento climatico;
- il destino, noi lo percepiamo come una cosa inevitabile;
- la dissonanza tra quello che facciamo tutti i giorni e quello che sappiamo dovremmo fare;
- rifiuto del problema, che ci viene spontaneo quando non vogliamo sentirci responsabili;
- l'identità culturale e i valori politici che nelle persone con orientamento conservatore spinge al rifiuto.

Oggi ci troviamo di fronte a fenomeni climatici sempre più estremi, frequenti e devastanti a causa delle attività umane, soprattutto nell'uso massiccio dei combustibili fossili.

Da ciò nacquero molti movimenti e proteste per poter diffondere ed aumentare la consapevolezza di tale crisi climatica a tutte le persone possibili nella speranza di portare un cambiamento al nostro pianeta.

Di recente il movimento che fece più scalpore è la denominata “Last Generation”, i protagonisti degli avvenimenti sono esponenti appartenenti a tale movimento che hanno deciso di infliggere danni ad opere d’arte; i dipinti in questione erano tutti situati in zone molto frequentate per poter attirare l’attenzione di più persone possibili.

Il movimento è chiamato Last Generation per enfatizzare il cambiamento climatico e dovrebbe rappresentare l’ultima generazione in grado di portare cambiamenti ed introdurre politiche che riescano a mantenere il nostro pianeta abitabile.

3. Quesito

Dopo aver analizzato attentamente la situazioni ci siamo posti delle domande:

- Tali azioni erano davvero necessarie per attirare l’attenzione del pubblico generale?
- Commettere reati riesce davvero ad avere un impatto positivo sulla massa?
- Come possiamo aiutare il nostro pianeta per noi e per le generazioni future?
- Come è stato percepito tale messaggio dagli utenti della piattaforma social di twitter?
- Quali sono i sentimenti e le opinioni degli individui riguardo alla protezione delle opere d’arte?
- Quali sono le motivazioni alla base del comportamento vandalico?
- Può l’analisi sentimentale prevenire il vandalismo delle opere d’arte?
- Quali sono le politiche che si possono mettere in atto per proteggere le opere d’arte?

Per poter rispondere a tali domande è necessaria un’analisi molto dettagliata del comportamento di un largo gruppo di individui all’interno di una piattaforma, ossia effettuare un’analisi sentimentale e ricavarne i risultati.

L’analisi della piattaforma può offrire importanti informazioni e risposte a determinati quesiti, come ad esempio quelli che ci siamo posti precedentemente.

La raccolta e l’analisi dei dati su una piattaforma, come ad esempio Twitter, può fornire una vasta quantità di informazioni sui comportamenti, le opinioni e le preferenze degli utenti. Queste informazioni possono essere utilizzate per rispondere a molte domande.

L’analisi della piattaforma può fornire risposte a questi quesiti e molti altri utilizzando tecniche di text mining, sentiment analysis e network analysis. Queste tecniche possono essere utilizzate per analizzare i dati raccolti dalla piattaforma, come i tweet, ed estrarre informazioni rilevanti.

Nonostante questi limiti, l’analisi della piattaforma può comunque fornire informazioni preziose e risposte a determinati quesiti. Utilizzando metodi rigorosi e adeguati, i ricercatori possono ottenere risultati affidabili e significativi che possono essere utilizzati per capire l’atteggiamento degli utenti appartenenti alla determinata piattaforma.

4. Processo di Analisi

L'analisi è un processo che consiste nell'esaminare e interpretare dati o informazioni per poi trarre conclusioni. Il processo di analisi inizia con la raccolta e l'organizzazione dei dati, che possono essere in qualsiasi formato, nel nostro caso delle frasi oppure dei testi. Successivamente, questi dati vengono elaborati utilizzando metodi matematici, statistici o di machine learning per identificare modelli e tendenze. Infine, le informazioni ottenute vengono interpretate tramite la creazione di modelli grafici che rappresentano i dati e le relazioni tra di loro per poi utilizzarle successivamente al fine di formulare conclusioni.

Le informazioni di cui avevamo bisogno e con cui lavoreremo sono state prelevate da Twitter tramite l'API fornito dall'azienda; i dati verranno poi immagazzinati e salvati localmente all'interno di data frame per poi essere convertiti in file excel e csv. La raccolta è avvenuta periodicamente per evitare la perdita di possibili informazioni importanti, se si presentano invece informazioni ridondanti verranno eliminate per evitare eventuali errori nel processo di analisi.

Essendo l'api di twitter molto limitata la raccolta di tutte le informazioni necessarie non è stata possibile quindi l'analisi della piattaforma non risulterà completa, nonostante ciò abbiamo deciso di lavorare ed elaborare le informazioni a noi disponibili.

4.1 Dati

La raccolta periodica di dati da Twitter tramite l'API può essere utilizzata in un progetto di analisi dei social network per raccogliere informazioni sugli utenti, sui tweet e sulle interazioni tra di essi. Questi dati possono essere utilizzati per comprendere meglio come le conversazioni su Twitter evolvono nel tempo, quali argomenti sono più popolari e come le opinioni e le emozioni degli utenti variano in base a eventi specifici.

Ad esempio, utilizzando l'API di Twitter, è possibile raccogliere i tweet pubblicati su un determinato argomento, come una notizia o un evento, e analizzare le emozioni e le opinioni espresse in questi tweet. Questo può aiutare a comprendere meglio il sentimento pubblico riguardo a una questione specifica e a identificare eventuali tendenze o pattern.

Inoltre, la raccolta periodica di dati da Twitter tramite l'API può essere utilizzata per analizzare le interazioni tra gli utenti, come ad esempio i retweet, le risposte e le menzioni. Questo può fornire informazioni sulle reti sociali e sui cluster di conversazioni all'interno della piattaforma, permettendo di comprendere meglio come le informazioni e le opinioni si diffondono su Twitter.

Come per ogni progetto che utilizza dati da fonti esterne, è importante rispettare le politiche sulla privacy e i termini di utilizzo dell'API di Twitter e garantire che i dati raccolti siano trattati in modo etico e responsabile.

Detto ciò abbiamo deciso di raccogliere manualmente e periodicamente i tweet, successivamente i tweet raccolti verranno salvati e raggruppati in file excel attraverso il codice in R tramite l'api di Twitter per poi essere elaborati in python. Sfortunatamente la libreria da noi usata per raccogliere i tweet non contiene alcune informazioni molto importanti per ricavare grafi come l'id del retweet, la località del tweet e molte altre colonne che risultano NaN.

Dopo aver raccolto periodicamente un paio di volte al giorno i tweet abbiamo raggiunto una somma di circa 500mila tweet raccolti che contenessero parole chiave come: "Just Stop Oil", Climate Activist, Oil Activist e anche i nomi degli artisti i cui dipinti sono stati danneggiati.

4.2 Eventi analizzati

Gli eventi che abbiamo deciso di analizzare sono gli atti di vandalismo di cui abbiamo parlato nell'introduzione del report, tali avvenimenti sono accaduti nel periodo dell'anno 2022. La raccolta di queste informazioni è avvenuta verso la fine dell'anno nell'arco di 3 mesi; durante la raccolta sono state prese di mira altre due opere aumentando il numero di informazioni ottenibili in quell'arco di tempo.

Gli eventi presi in considerazione per l'analisi della piattaforma social includono:

1. Tendenze: I trend delle parole chiave e delle frasi sono uno dei fattori più importanti presi in considerazione durante l'analisi dei social media. Questi trend possono fornire informazioni sulle opinioni e le preferenze degli utenti riguardo a determinati argomenti.
2. Sentimento: Il sentimento è un altro fattore importante preso in considerazione durante l'analisi dei social media. Questo fattore cerca di identificare se un tweet o un post contiene emozioni positive, negative o neutre, aiutando a capire il punto di vista degli utenti su un determinato argomento.
3. Campagne pubblicitarie: Le campagne pubblicitarie sui social media possono avere un impatto significativo sulle opinioni degli utenti. L'analisi dei social media può aiutare a valutare l'efficacia di queste campagne e a capire se stanno raggiungendo il loro obiettivo.
4. Interazioni degli utenti: L'analisi dei social media può anche prendere in considerazione le interazioni degli utenti, come i like, i commenti e le condivisioni. Queste interazioni possono fornire informazioni sul grado di coinvolgimento degli utenti riguardo a un determinato argomento.

4.3 Sentiment Analysis

La prima analisi è stata effettuata su un file excel contenente tutti i tweet raccolti, dal file in questione sono state rimosse le righe contenenti tutte le informazioni uguali che potrebbero presentare delle ridondanze a causa della raccolta di tweet giornaliera.

La sentiment analysis è un processo di analisi che analizza per poi indicare e categorizzare gli atteggiamenti degli utenti verso un determinato argomento, tramite l'analisi del testo attraverso un algoritmo si riesce a determinare le opinioni di ciascuno degli utenti.

Come detto precedentemente il file che abbiamo preso in considerazione si tratta di un file excel che contiene tutti i tweet raccolti nell'arco di 3 mesi, sono approssimativamente 600 mila righe e per la prima analisi utilizzeremo un meccanismo di excel.

Il meccanismo da noi utilizzato è un add-in implementato su excel chiamato Azure Machine Learning che ha la capacità di determinare lo stato d'animo di una persona dal testo attraverso un algoritmo e grazie a ciò siamo riusciti a ricostruire un grafico che riporti il numero di tweet con sentimento negativo, neutro e positivo del dataframe contenente tutti i dati raccolti; l'algoritmo determina il sentimento di una persona e lo misura con un numero, ossia lo score, che varia da 0 a 1, se il valore risulta vicino a 0 allora il sentimento è negativo, se invece risulta vicino a 1 allora è positivo mentre se si aggira intorno al valore di 0.5 il sentimento sarà neutro.

I risultati ricavati da tale analisi sono i seguenti:

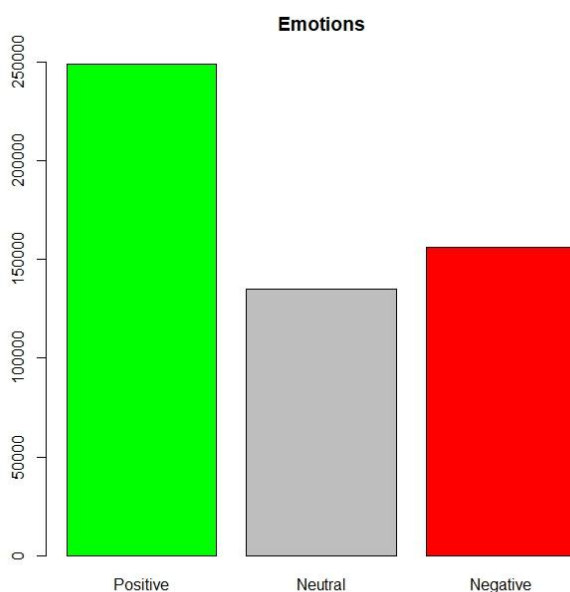


Fig. 1 Grafico che rappresenta il numero di tweet per ogni sentimento

Come possiamo notare dal grafico le emozioni positive ricavate sono notevolmente superiori rispetto a quelle negative.

Tale osservazione come possiamo notare però non è del tutto precisa dato che la raccolta di dati tramite keyword ha preso in considerazione anche tweet che non menzionano l'argomento da noi trattato; infatti discutono del cambiamento climatico in generale.

Detto ciò abbiamo deciso di diminuire ulteriormente il numero di informazioni a nostra disposizione per creare un grafico a barre più appropriato e mirato.

Dopo aver effettuato le apposite approssimazioni siamo riusciti a ricavare il seguente:

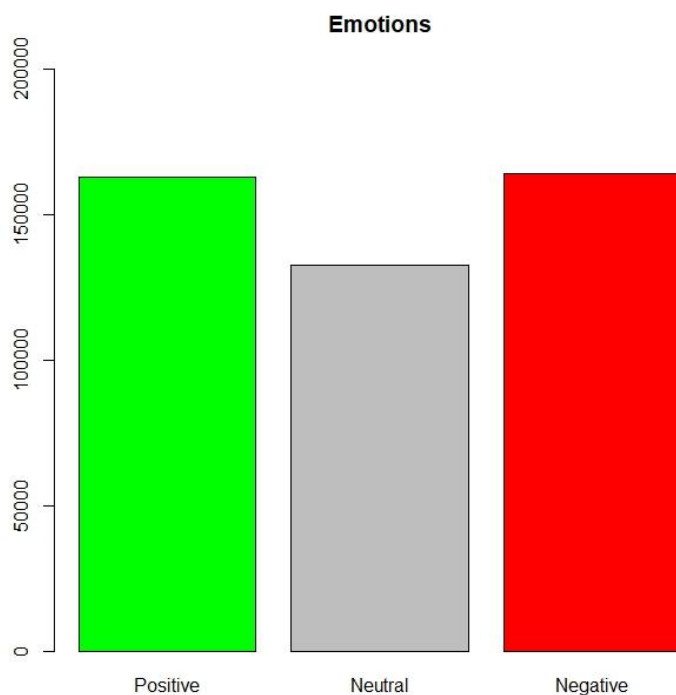


Fig. 2 Grafico a barre che prende di mira un gruppo più ristretto di utenti

Un altro fattore da prendere in considerazione è che l'algoritmo in questione non è del tutto affidabile dato che può commettere errori di valutazione se il testo del tweet risulta molto lungo, perciò abbiamo deciso di utilizzare la libreria textblob di python per effettuare un'analisi più precisa.

TextBlob è una libreria NLP (Natural Language Processing) di python che utilizza il kit NLTK (Natural Language Toolkit), tale kit è facile da usare e può anche essere utilizzato per analisi complesse; quando TextBlob riceve un input, ossia un dato con testo, mostrerà due output che rappresentano la soggettività e la polarità della frase.

Per ora osserveremo semplicemente le emozioni degli utenti tramite dei grafici a barre:

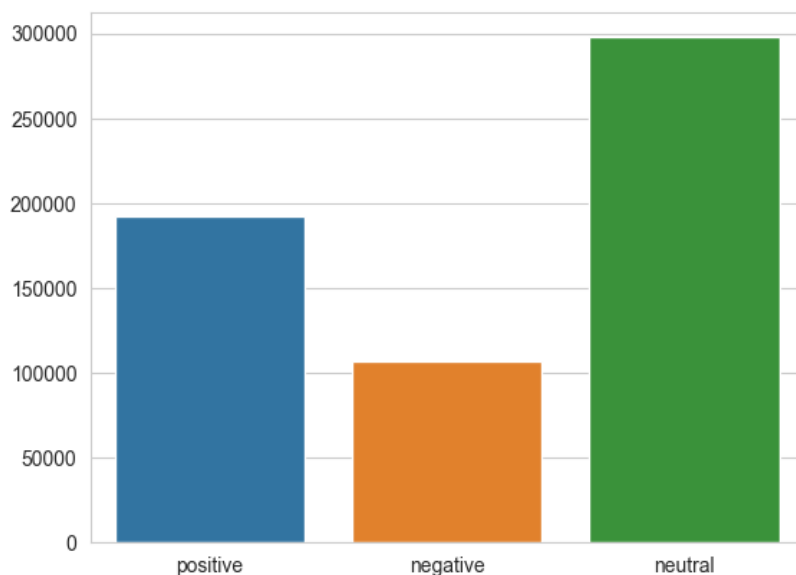


Fig. 3 Grafico a barre che rappresenta i sentimenti dei tweet totali ricavati tramite la libreria textblob di python

Come possiamo notare dall'immagine la differenza dei dati ricavati dal tool di excel e l'algoritmo all'interno della libreria di python è notevole, perciò lavoreremo con tali informazioni per evitare discontinuità nel processo di analisi.

Sorprendentemente il feedback positivo è ancora superiore a quello negativo, dato che le notizie più diffuse riguardo a tali avvenimenti durante quel periodo erano maggiormente negative, mentre le risposte neutrali sono nettamente superiori ad entrambi; questo perché i tweet presi in considerazione per questo grafico a barre sono tutti quelli che comprendono le keyword che sono presenti anche in tweet che non trattano l'argomento interessato.

Prendendo, invece, in analisi i dati che contengono keyword e hashtags più specifici:

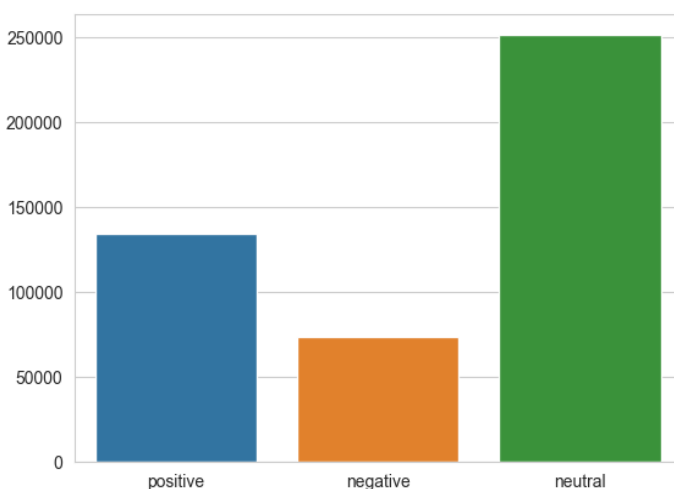


Fig. 4 Grafico a barre che rappresenta i sentimenti dei tweet mirati ricavati tramite la libreria textblob di python

Partendo da questi dati osserviamo come hanno reagito le piattaforme di diversi paesi, il grafico sottostante riporta i diversi feedback di diverse lingue:

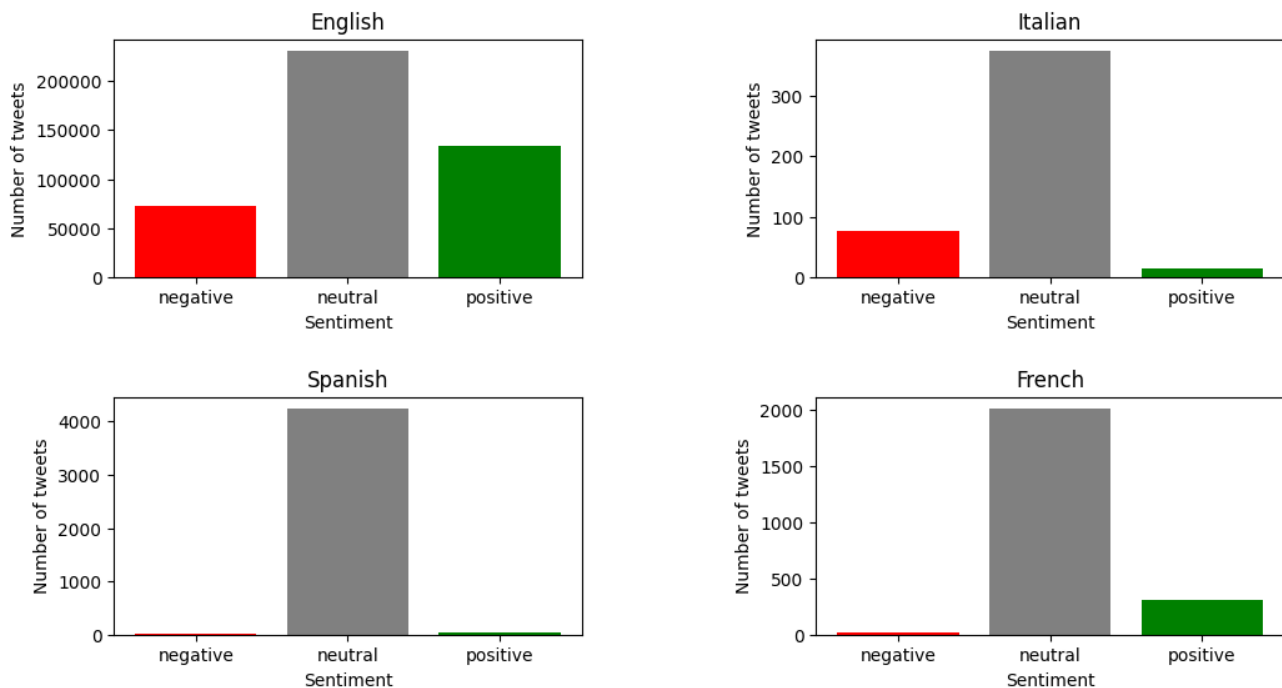
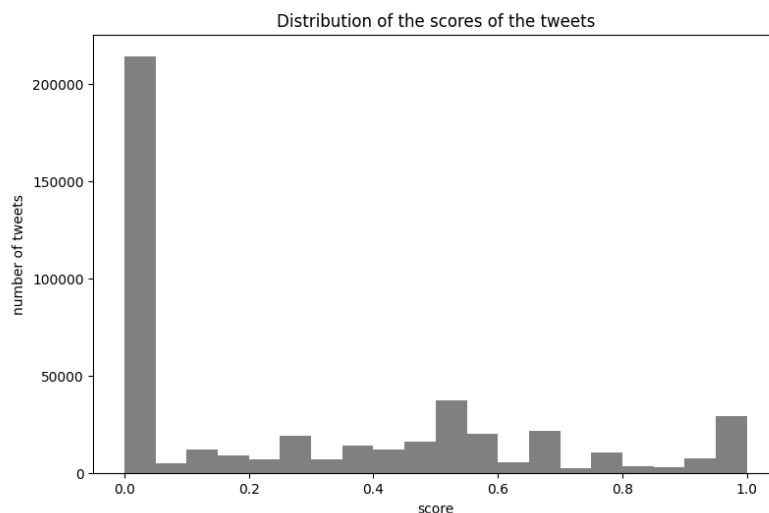


Fig. 5 Grafici a barre che rappresentano sentimenti per ogni lingua

I tweet riguardanti la lingua italiana risultano molto scarsi dato che le keyword e gli hashtag usati non erano molto diffusi all'interno della piattaforma italiana. Nonostante ciò possiamo notare come l'unica lingua dove i numeri di tweet negativi superano quelli positivi sia quella italiana, mentre in tutti gli altri i tweet con sentimenti positivi sono maggiori; ma ciò che veramente prevale sono i sentimenti neutri, fatto che è condiviso tra tutte le lingue prese in considerazione. Analizziamo tramite un istogramma quali sono i valori più ripetuti all'interno del dataframe:



Osserviamo anche il numero di tweet per giorno basandoci sui sentimenti che abbiamo ricavato dall'analisi precedente:

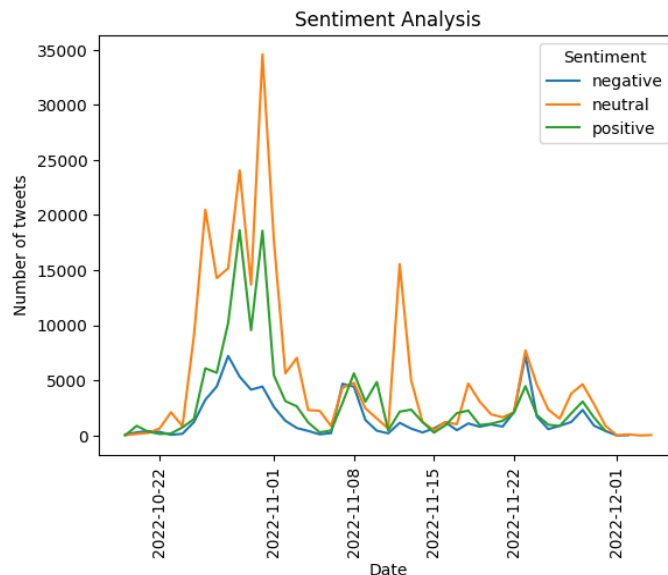


Fig. 6 Grafico che mostra l'andamento dei sentimenti

Come possiamo osservare dall'immagine i numeri di tweet sono diminuiti nel tempo a causa della perdita di interesse del pubblico generale, questo effetto si chiama "Recency Bias", ossia che gli eventi storici più recenti hanno più importanza. Da tenere in considerazione anche "Primacy Bias" che consiste nel ricordare maggiormente il primo fatto rispetto a quelli più recenti, infatti come possiamo notare dal grafico il picco è stato verso metà ottobre ossia lo stesso periodo in cui l'argomento più discusso era l'atto di vandalismo nei confronti del dipinto di Van Gogh nella galleria National Gallery situata a Londra. Rappresentiamo anche la percentuale dei sentimenti rispetto al numero di tweet con cui stiamo lavorando e notiamo che le opinioni neutrali sono poco più del 50% dell'intero dataset:

Sentiment Analysis Result of the entire dataset

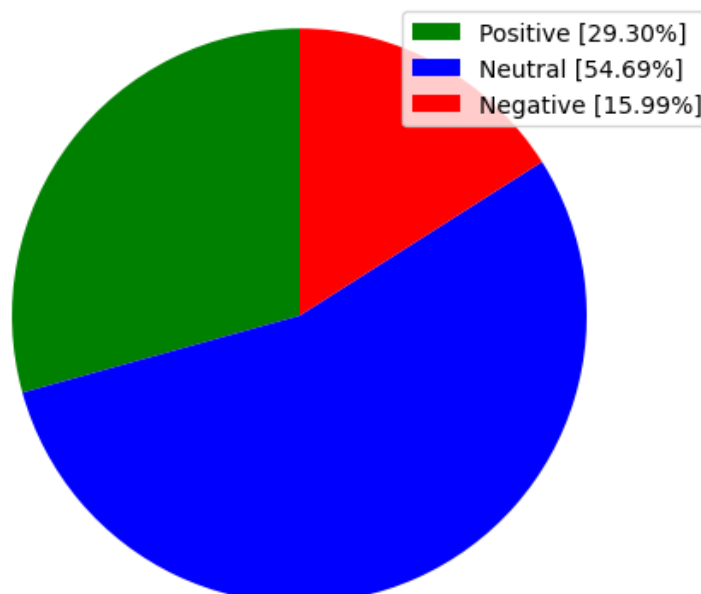
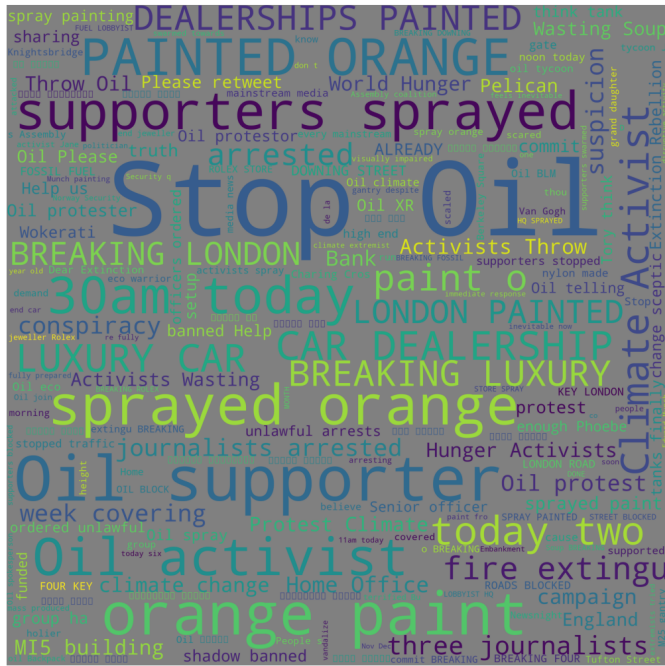


Fig. 7 Grafico a torta che rappresenta i sentimenti in percentuale

[illegible]

single organ
traffic
Gustav Klimt
Emergency Fund
nothing
billions
caused far
people left
one
young
Shepherds
orange
guess
group
London yesterday
hand
Met Police
building
protesters guilty
away
breakdown will
broken
silly liquid
Oil Pair
Motorists drag
Oil imagine
road
telling
XR
will lead
see
found guilty
arrest member
understand
climate crisis
Oil Failure
BREAKING
ROADS
seriously
slowly onto
need
paper thin
cost
Bush go
people
Britain
protestors
paint used
think
people getting
oil fanatics

Fig. 9 Parole dai tweet con sentimento negativo



Per appoggiare la teoria che abbiamo citato, ossia il bias del Recency e Primacy, la figura numero 11 sottostante mostra il numero di tweet raccolti con le stesse keyword e possiamo osservare come il numero di tweets decrementa con il passare del tempo, e come il picco si trovi verso ottobre.

Tramite i sentimenti che siamo riusciti a raccogliere e i valori di tali sentimenti abbiamo creato grafici scatter che rappresentano la relazione tra la “subjectivity” e la “polarity” di ogni utente, le quali abbiamo menzionato precedentemente. La “subjectivity” o soggettività rappresenta quanto, all’interno di un testo, un utente sia stato influenzato da dati di fatto oppure opinioni personali; con dati di fatto si intendono informazioni confermate tramite determinate procedure.

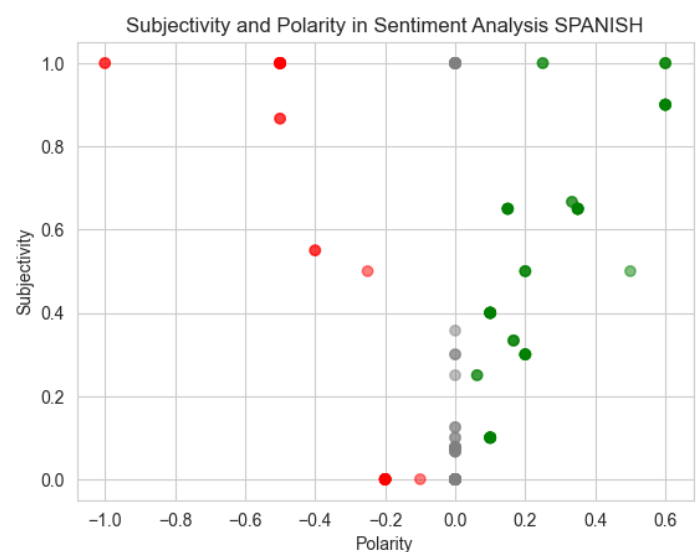
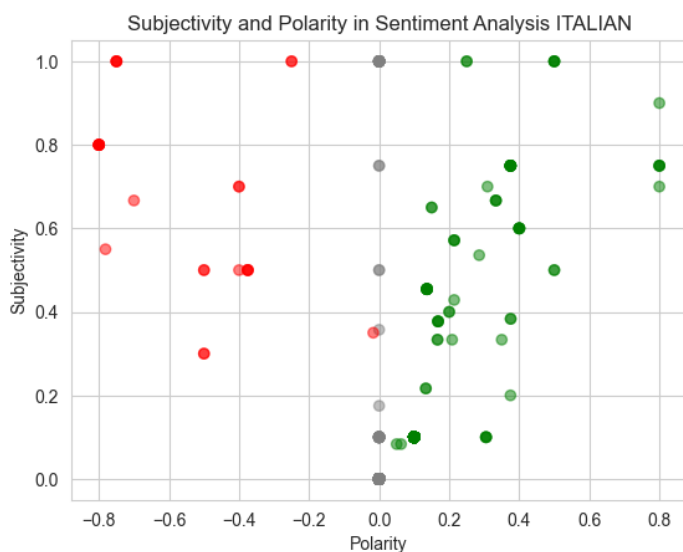
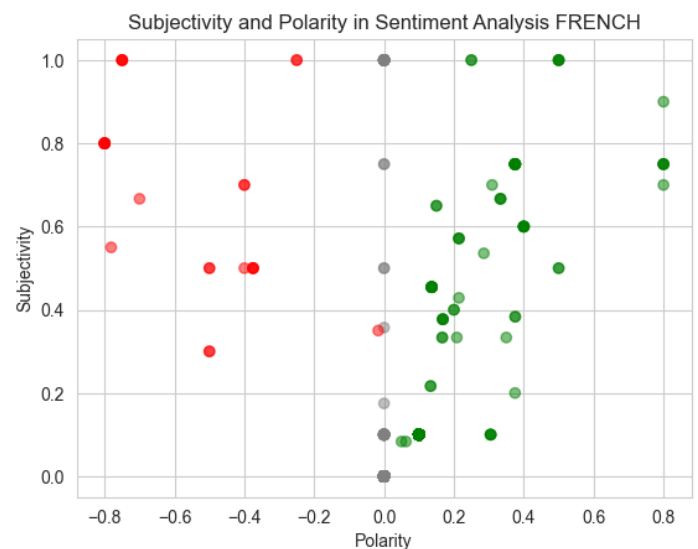
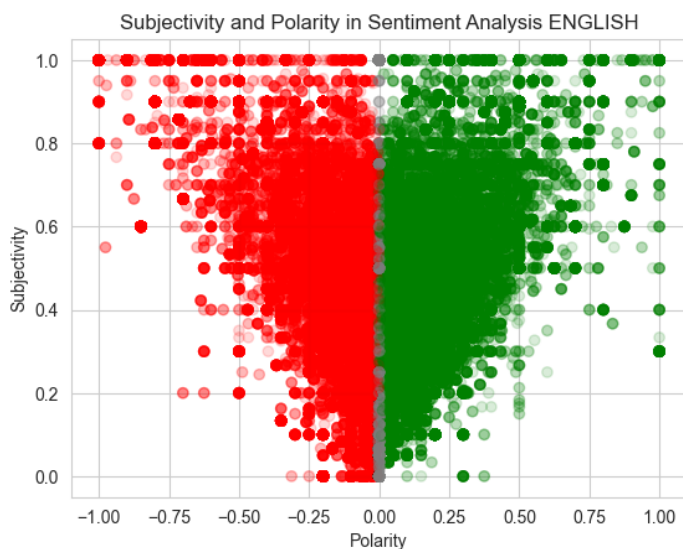
La “polarity” o polarità invece ha come scopo quello suddividere le opinioni in 2 emozioni, ossia positive e negative.

Questo rapporto presenta i risultati di un'analisi della soggettività e della polarità effettuata su un insieme di dati, con lo scopo di identificare la percezione degli utenti e di comprendere la loro opinione sull'argomento in questione.

Tale rapporto verrà raffigurato nei grafici scatter sottostanti per rappresentare la relazione fra la soggettività di un individuo e la polarità per determinare quanto essi siano stati influenzati.

Prima analizzeremo le diverse lingue ufficiali dei paesi che abbiamo preso in considerazione.

Il grafici scatter ricavati sono i seguenti:



Come si nota dai grafici ricavati possiamo osservare la discrepanza dei numeri di tweet raccolti nelle varie lingue, a causa di un errore nella raccolta dei tweet nella quale abbiamo utilizzato hashtags e keywords prevalentemente diffusi nella lingua inglese.

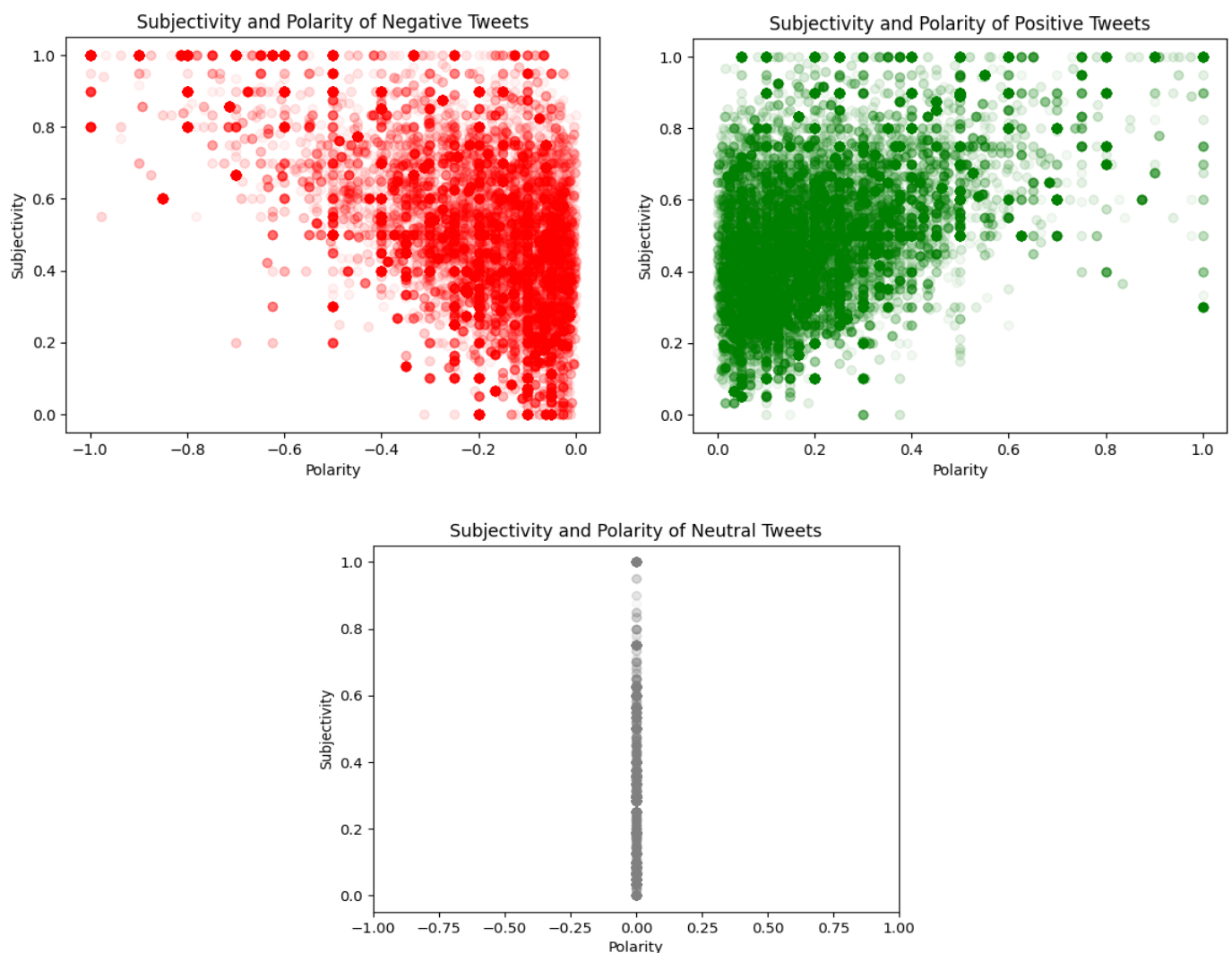
Ora osserveremo le diverse emozioni all'interno dell'intero dataframe raffigurate su diversi grafici scatter.

Lo scatter di colore rosso rappresenta i sentimenti negativi.

Lo scatter di colore verde rappresenta i sentimenti positivi.

Lo scatter di colore grigio rappresenta i sentimenti neutri.

Questi sono i scatter ricavati:



Unendo i 3 grafici ricaviamo:

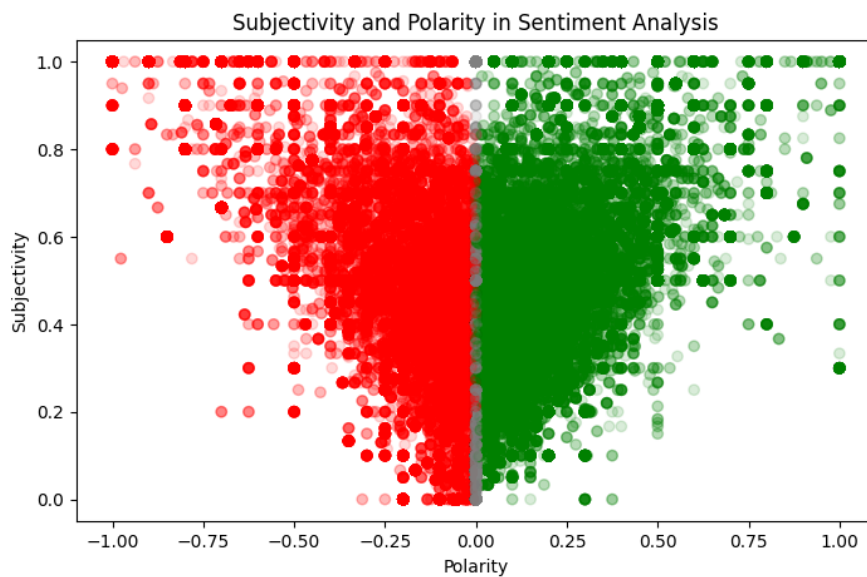


Fig. 12
Grafico
scatter che
mostra la
relazione tra
soggettività e
polarità di
tutti gli utenti

Come possiamo dedurre dai grafici i sentimenti neutri sono agglomerati sul fondo, ossia sono testi non influenzati dal sentimento e dall'opinione personale; quelli positivi e negativi invece sono maggiormente accumulati al centro del grafico. La maggioranza dei nodi ha polarità con valori vicino allo 0 (che rappresenta ovviamente la neutralità) e con una soggettività intorno ai 0.5, ossia sono stati influenzati, anche se minimamente, da opinioni personali.

Da tutto ciò possiamo ricavare la proiezione della rete bipartita attraverso un grafico. Una rete bipartita è un tipo di grafo in cui i nodi sono divisi in due gruppi distinti e non esistono relazioni all'interno dello stesso gruppo. Invece, ogni nodo del primo gruppo è connesso solo a nodi del secondo gruppo e viceversa. Questo tipo di struttura è utile per rappresentare relazioni binarie tra elementi appartenenti a due insiemi distinti. Ad esempio, in una rete bipartita può essere rappresentata la relazione tra gli utenti e i loro sentimenti.

Qua osserviamo un esempio di un piccolo grafo bipartito:

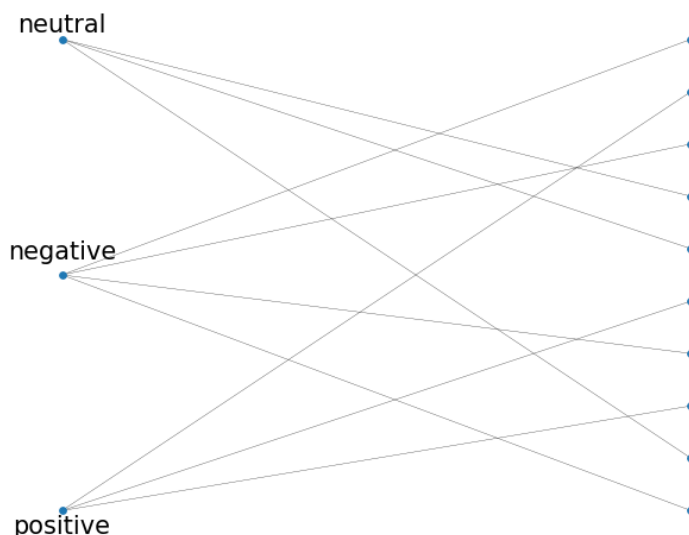


Fig 13 Rappresentazione di una rete
bipartita delle prime 10 righe del
dataframe

Un grafo bipartito è un tipo di grafo in cui i nodi sono divisi in due gruppi distinti, chiamati partizioni. Le relazioni tra i nodi sono rappresentate da archi che possono connettersi solo tra nodi di partizioni diverse.

I nodi a destra rappresentano gli utenti mentre a sinistra ci sono i nodi dei sentimenti; degli utenti non mostreremo gli ID dato che stiamo lavorando con un file di moderate dimensioni.

Il grafo bipartito che ricaviamo dall'intero dataframe è il seguente:

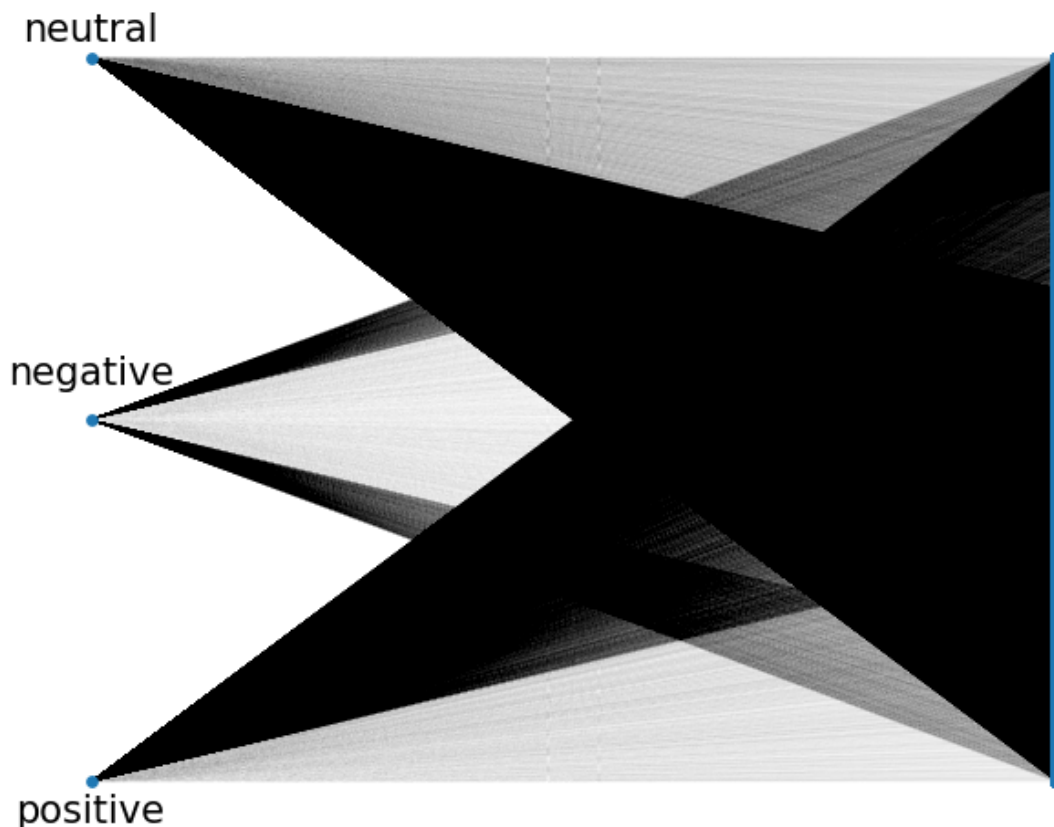


Fig. 14 Grafo bipartito che rappresenta le relazioni tra utenti e sentimenti in tutto il dataframe

Come possiamo notare dal grafico bipartito ottenuto la lettura dei dati non è molto comprensibile dato l'elevato numero di dati con cui stiamo lavorando, un modo per ridurre la dimensione del grafo ed il numero di dati è attraverso il clustering con PCA e K-Means ossia un processo di apprendimento non supervisionato.

Il processo di apprendimento non supervisionato è un tipo di apprendimento automatico in cui l'algoritmo viene addestrato su un dataset che non ha etichette o categorie predefinite. In altre parole, l'obiettivo dell'apprendimento non supervisionato è di scoprire i pattern o i modelli nascosti all'interno dei dati. Nell'apprendimento non supervisionato, l'algoritmo è libero di esplorare i dati e formare gruppi o cluster basati sulle somiglianze tra i dati. Questi gruppi possono quindi essere utilizzati per fare previsioni o per visualizzare la struttura dei dati.

Gli algoritmi di apprendimento non supervisionato includono il clustering, la riduzione della dimensionalità e la visualizzazione delle relazioni tra i dati.

Questo tipo di analisi è spesso utilizzato quando non sono presenti etichette disponibili o quando l'obiettivo è di scoprire i modelli nascosti all'interno dei dati.

Tuttavia, poiché l'algoritmo non viene direttamente guidato da etichette o categorie, i risultati possono essere meno precisi.

Il Clustering con PCA e K-Means è un processo di apprendimento non supervisionato che ha come obiettivo quello di raggruppare oggetti simili in un dataset.

K-Means Clustering è un metodo per trovare i cluster, ossia i gruppi di oggetti simili, all'interno di un dataset. L'algoritmo definisce per prima il numero di cluster che chiameremo K, e successivamente inizializza casualmente un numero di centroidi equivalente allo stesso numero di clusters. Gli oggetti nel dataset vengono quindi assegnati al centroide più vicino e i centroidi vengono aggiornati in base alla media degli oggetti assegnati a loro. Questo processo viene ripetuto fino a quando i centroidi non si stabilizzano.

PCA (Principal Component Analysis) è una tecnica per ridurre la dimensionalità di un dataset mantenendo la maggior parte della varianza possibile. Funziona trasformando le variabili originali in un nuovo insieme di variabili chiamate componenti principali, che sono ortogonali tra di loro e classificate in base alla quantità di varianza che spiegano. Questo metodo viene spesso utilizzato prima del K-Means Clustering per ridurre la dimensionalità del dataset, rendendo il processo di clustering più efficiente, efficace e veloce.

Il Clustering con PCA e K-Means viene spesso utilizzato in molteplici applicazioni, come l'analisi del testo ecc... Fornisce un modo semplice ed efficiente per raggruppare oggetti simili in un dataset.

All'interno di questi grafici troveremo anche i centroidi; nel clustering, i centroidi sono utilizzati per rappresentare i cluster, ovvero gruppi di dati simili tra loro. Ad esempio, in un'analisi di clustering basata sulla somiglianza dei dati, i centroidi sarebbero calcolati come la media delle caratteristiche dei dati all'interno di un cluster. I dati verrebbero quindi assegnati ai cluster in base alla loro vicinanza ai centroidi.

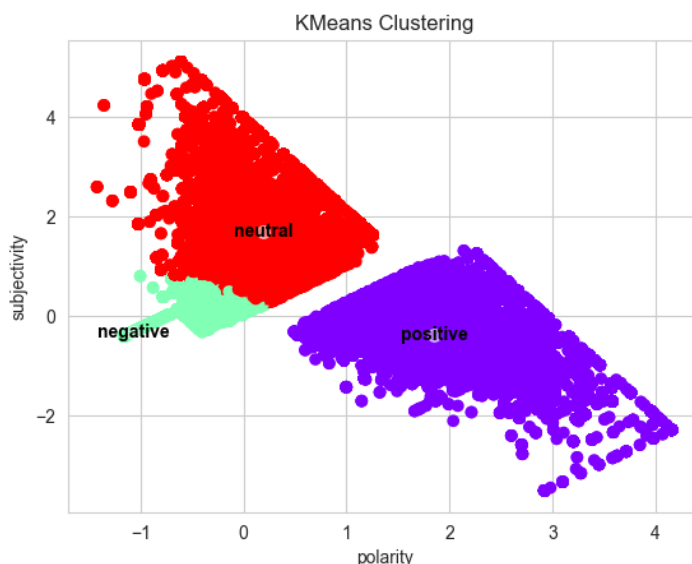


Fig.15 Clustering con soggettività e polarità

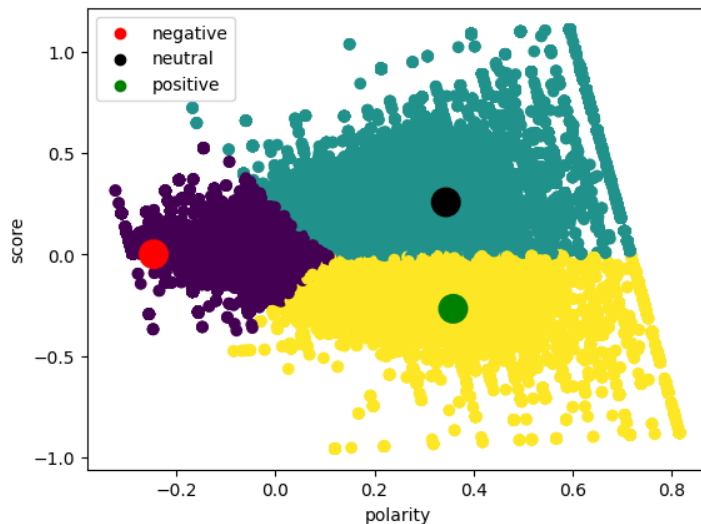


Fig.16 Clustering con score e polarit 

L'apprendimento non supervisionato per  porta alcuni svantaggi importanti:

1. Difficolt  nell'interpretazione dei risultati: L'apprendimento non supervisionato non fornisce risultati facilmente interpretabili, poich  non viene fornita alcuna etichetta o categoria predefinita per i dati. Questo pu  rendere difficile comprendere il significato dei risultati dell'apprendimento.
2. Assenza di informazione sulle relazioni tra i dati: L'apprendimento non supervisionato si concentra sulla identificazione di cluster o modelli all'interno dei dati, ma non fornisce informazioni sulle relazioni tra i dati.
3. Bias nella selezione dei modelli: L'apprendimento non supervisionato   spesso influenzato dalla selezione dei modelli, che pu  a sua volta influire sulla interpretazione dei risultati.
4. Problemi di sovrapposizione: L'apprendimento non supervisionato pu  essere soggetto a problemi di sovrapposizione, in cui i dati vengono assegnati a pi  di un cluster.
5. Difficolt  nella valutazione delle prestazioni: La mancanza di etichette o categorie predefinite rende difficile valutare le prestazioni dell'apprendimento non supervisionato.

L'apprendimento non supervisionato   stato utilizzato data la scarsa quantit  di etichette presenti all'interno dei dati raccolti e inoltre il dataframe ottenuto non presenta una struttura concreta da cui partire per la costruzione di grafi che ne rappresentino le relazioni.

Sfortunatamente, come menzionato all'inizio del report, durante il processo di analisi dei dati abbiamo riscontrato vari problemi, soprattutto nella raccolta dei tweet tramite il nostro api la quale restituisce valori non validi per gli id dei retweet o like, rendendo impossibile la creazione di un grafo composto da nodi e archi come la proiezione del grafo bipartita che si basa sul collegamento che hanno due utenti fra di loro.

L'unico elemento che risultava condiviso all'interno del dataframe tra gli utenti erano le emozioni, dalla quale siamo riusciti a creare collegamenti su cui abbiamo basato i nostri grafi bipartiti e cluster.

5. Conclusioni

I dati con cui abbiamo lavorato hanno mostrato come la maggior parte delle persone che hanno avuto la possibilità di esprimere la propria opinione riguardo a tali eventi abbiano lasciato un feedback nè positivo nè negativo, ciò ci ha sorpreso molto dato che stiamo lavorando con dati ottenuti dalla piattaforma di Twitter.

Questo significa che la maggior parte degli utenti ha deciso di essere più oggettivi nei confronti degli attivisti, tralasciando le proprie emozioni; questo fenomeno però potrebbe essere causato dal fatto che le opere non sono state realmente danneggiate, essendo protette da un vetro, e anche se lo fossero possono essere ricoverate facilmente dati i lievi danni ricevuti.

Un'altra giustificazione potrebbe essere l'importanza dell'argomento trasmesso tramite i messaggi che gli attivisti hanno tentato di diffondere, ossia la crisi climatica e le possibili conseguenze che potrebbe portare.

Il cambiamento climatico è un argomento di grande importanza e riconosciuto come una delle sfide più grandi che la società attuale deve affrontare, la sua importanza è confermata dalla vasta comunità scientifica che ha prodotto prove inequivocabili del suo impatto sull'ecosistema del pianeta e sull'impatto che potrebbe avere sulla vita umana.

Il cambiamento climatico è causato da attività umane come l'utilizzo di combustibili fossili, l'agricoltura e l'industria, che rilasciano grandi quantità di gas serra nell'atmosfera, causando un aumento della temperatura globale, tale aumento sta causando cambiamenti climatici e ambientali, come l'innalzamento del livello del mare, la riduzione delle risorse idriche, la distruzione di ecosistemi naturali e la riduzione della biodiversità.

Per questo motivo, il cambiamento climatico rappresenta una minaccia per la sicurezza alimentare, la salute pubblica e l'equità sociale, ed è diventato una priorità per i governi di tutto il mondo. Molti paesi hanno adottato politiche e programmi per ridurre le emissioni di gas serra e promuovere fonti di energia rinnovabile, ma ancora molto deve essere fatto per poter affrontare questa sfida globale.

In sintesi il cambiamento climatico è ormai ritenuto importante dalla maggior parte delle persone ed eventuali atti di vandalismo che coinvolgono il danno di opere d'arte non cambierà la visione delle persone a riguardo; come abbiamo dimostrato tramite i nostri grafici il sentimento più prevalente è quello neutro, ossia tali eventi non sono riusciti a suscitare la reazione che speravano dal pubblico generale.

Tramite atti momentanei che causano solo reazioni temporanee non raggiungeremo mai un cambiamento sostanziale per prevenire la crisi climatica, esistono molte possibili soluzioni e danneggiare opere, bloccare il traffico, vandalizzare edifici con vernice non sono tra quelle; per poter raggiungere tale obiettivo bisogna sensibilizzare i governi e le aziende riguardo ai loro effetti, portando all'attuazione di politiche e regolamentazioni climatiche, incentivando l'utilizzo di energie rinnovabili e l'investimento in tecnologie sostenibili.