

# SENTIMENT ANALYSIS SUL CONCERTO DEL PRIMO MAGGIO

*A cura di Eric Spinelli 799827, Big Data in Behavioural Psychology*

## Introduzione

Il concerto del primo di maggio è un evento organizzato ogni anno dal 1990 a Roma per celebrare la Festa dei Lavoratori. Quest'anno, a causa della pandemia da coronavirus, la maratona musicale si tiene alla Cavea dell'auditorium Parco della Musica, senza pubblico.

Le esibizioni sono iniziate alle 16:35 e terminate alle 00:25, partecipandovi 41 artisti. A catalizzare l'attenzione mediatica durante questo evento è stato l'intervento di Fedez alle ore 21:25, in cui l'artista si è espresso esplicitamente contro affermazioni omofobe di rappresentanti del partito politico Lega Nord.

Lo scopo del progetto è quello di fare un'analisi della reazione degli utenti del social network Twitter, partendo dall'impatto numerico dei tweet legati all'intervento, andando gradualmente nello specifico illustrando i risultati della sentiment analysis di essi, studiando in particolare l'effetto emotivo del discorso e arrivando infine al dettaglio dell'analisi con considerazioni sulle parole più influenti che hanno caratterizzato l'evento.

## Panoramica metodologica

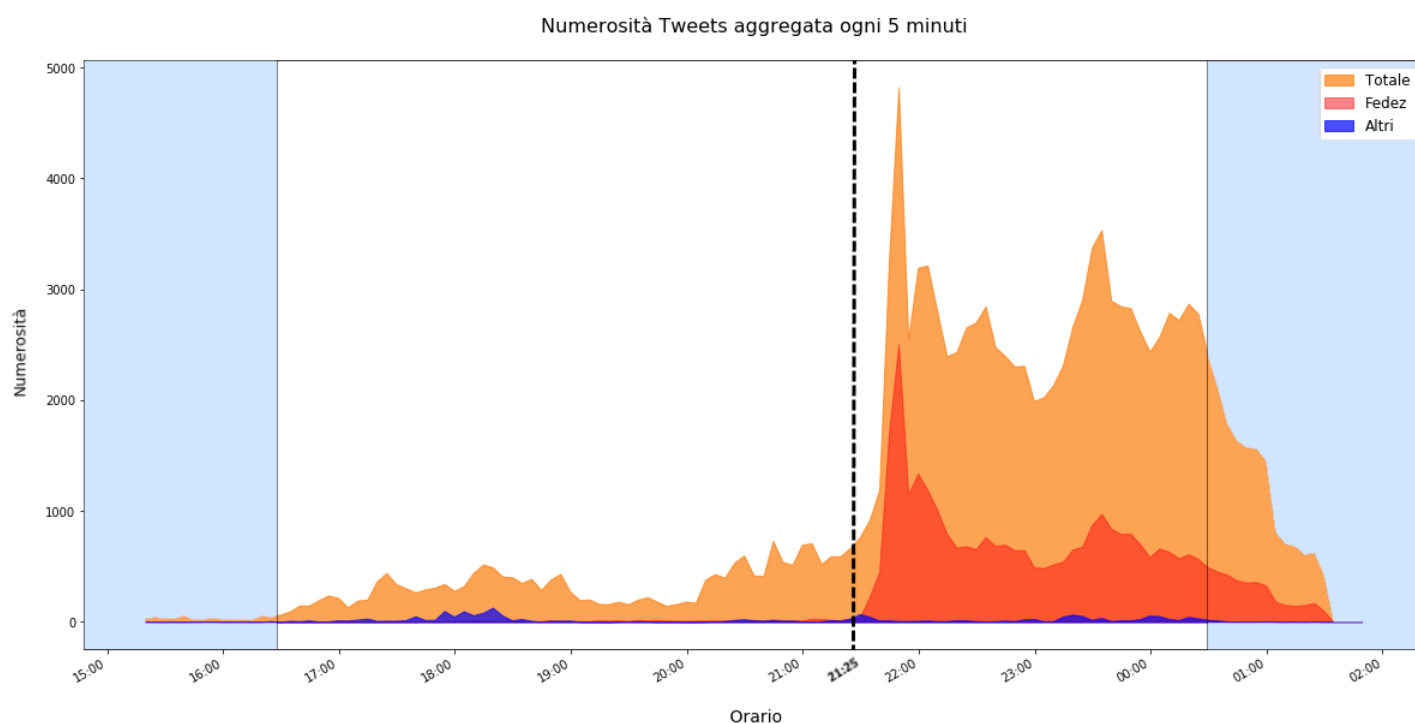
La prima fase del lavoro è consistita nell'effettuare una raccolta di dati in tempo reale (data streaming) sfruttando l'API di Twitter. Per rendere il processo efficiente si è scelto di utilizzare la tecnologia Kafka per gestire le code di dati, salvandoli in tempo reale nel database MongoDB. Si è scelto di mandare in parallelo due raccolte: la prima che comprendesse solo gli hashtag legati al concertone, di cui ne sono stati scelti 6 tra cui il più frequente #1M2021; la seconda relativa agli artisti, eligendo un unico hashtag per artista, di cui il più frequente #Fedez. La raccolta è iniziata alle 15:15 ed è terminata alla 1:30, in modo da cogliere l'effetto immediatamente prima e immediatamente dopo lo spettacolo. In totale sono stati raccolti 188152 tweet.

La seconda fase è stata la pulizia dei dati finalizzata all'ottimizzazione degli algoritmi di sentiment analysis. Per questo processo si sono aggregati i tweet delle due raccolte e si sono filtrati solo i tweet in italiano, che consistevano nell'oltre 90% dei tweet. La pulizia ha compreso: eliminazione di stopwords, rimozione della punteggiatura e caratteri speciali, trasformazione in lowercase, eliminazione hyperlinks, eliminazione di emoji (successivamente estratte per la emoji analysis), lemmatization del testo.

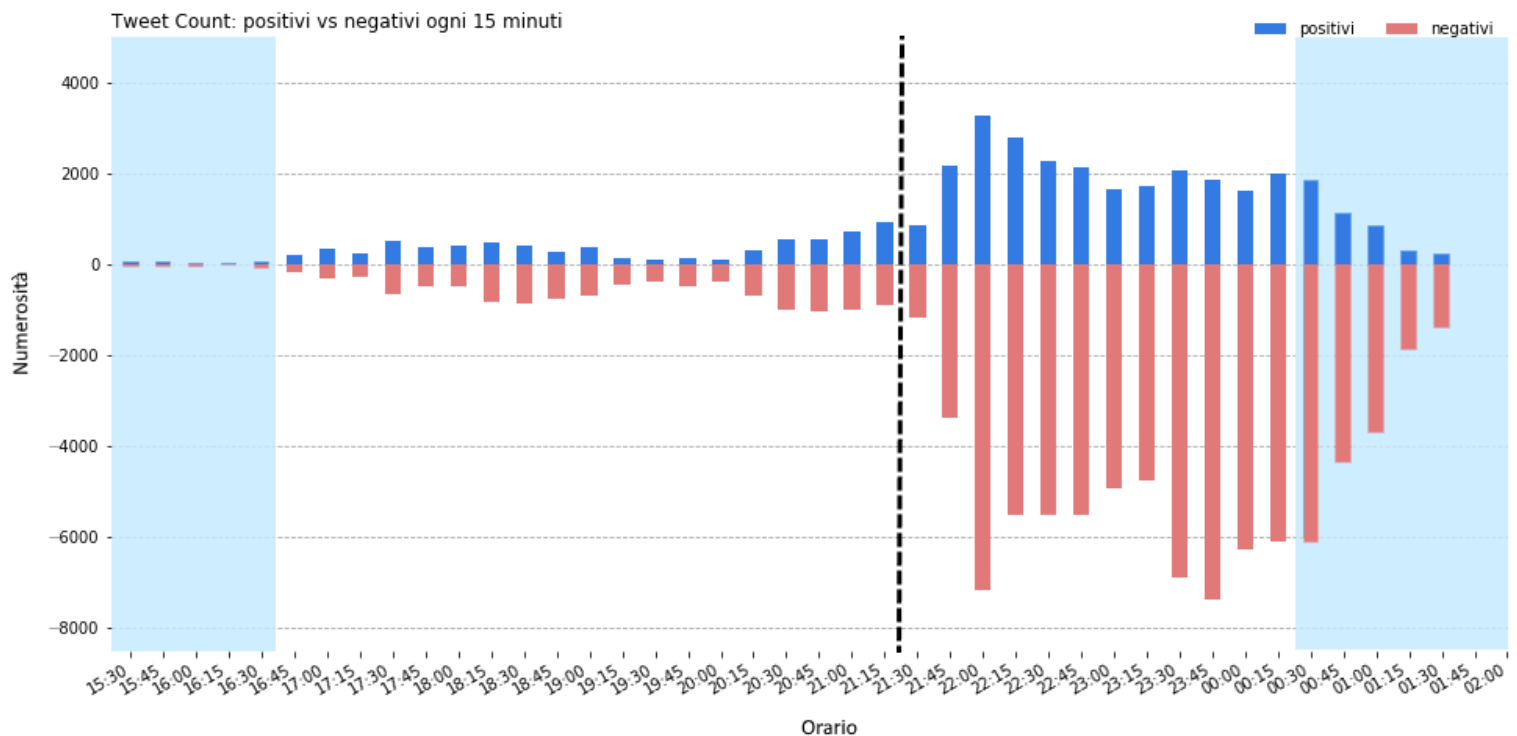
La terza fase ha consistito nell'applicazione degli algoritmi di sentiment analysis. In particolare, è stato scelto il pacchetto FEEL-IT, basato sull'algoritmo BERT (Bidirectional Encoder Representations from Transformers) allo stato dell'arte dei modelli machine learning usati per task di NLP. In particolare, FEEL-IT è ottimizzato per tweet italiani e mostra ottime performance di accuracy. Il risultato prodotto da questo algoritmo è una classificazione in base all'emozione catturata dal tweet, che può essere: rabbia, paura, gioia, tristezza. Sulla base di questo algoritmo si sono prodotti i grafici di sentiment. La emoji analysis è basata su un conteggio, sfruttando l'ottimizzazione del filtro grazie alla libreria "emoji". I worldcloud sono stati prodotti a partire da 4 subset: tweet positivi prima del discorso e dopo di esso; tweet negativi prima del discorso e dopo di esso.

## Analisi dei dati

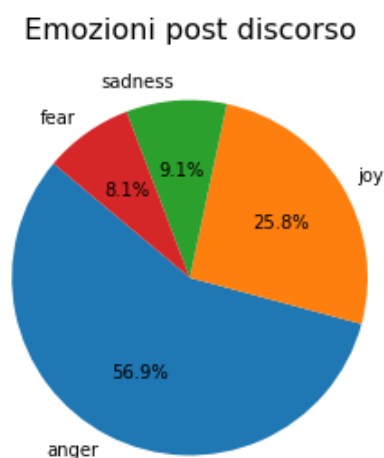
Il primo grafico proposto mette a confronto la numerosità dei tweet aggregata ogni 5 minuti. La zona viola consiste nella numerosità aggregata dei tweet legati agli artisti; la zona arancione scuro rappresenta i soli tweet relativi a Fedez; La zona arancione chiaro, che include le altre, indica la numerosità aggregata dei tweet relativi al concerto e agli artisti. Il momento del discorso di Fedez è denotato dalla linea di demarcazione tratteggiata. Commentando la parte del grafico precedente all'evento si può vedere che la numerosità raggiunge i 290 tweet ogni 5 minuti in media e che l'attenzione è focalizzata sull'evento del concerto in sé. Come si può notare, subito dopo il discorso di Fedez il picco di tweet si verifica dopo circa un quarto d'ora raggiungendo quasi i 5000 tweet in 5 minuti e Fedez, che fino a quel momento non ha destato alcuna particolare attenzione dai social, diventa il protagonista assoluto. Da quel momento fino alla fine del concerto la numerosità dei tweet in media è di 2305 tweet ogni 5 minuti.



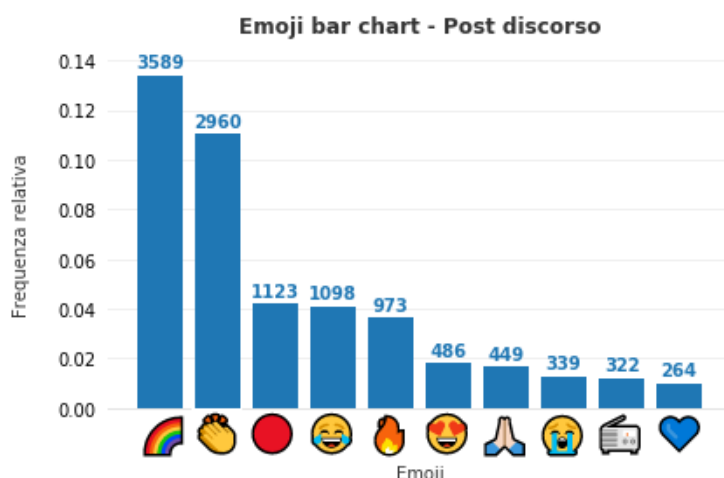
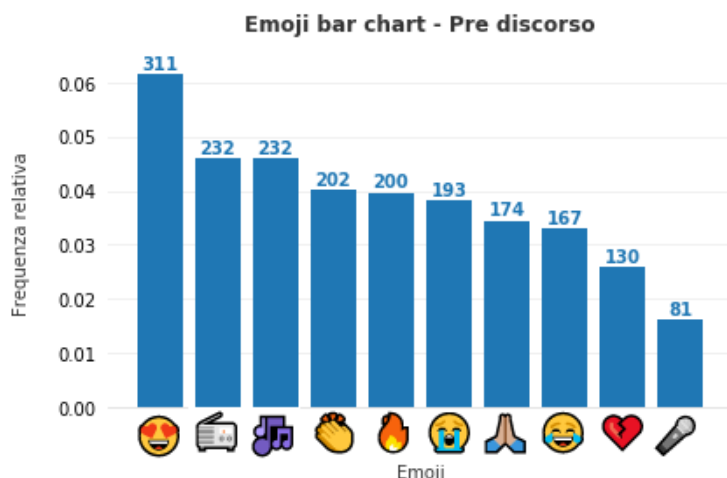
In una seconda fase si vuole mettere a confronto il sentimento positivo con quello negativo suscitato dai tweet in relazione alla loro numerosità. I tweet sono aggregati ogni 15 minuti. Si è scelto di rappresentare il conteggio dei tweet negativi con il rispettivo segno negativo. Possiamo notare che prima dell'intervento di Fedez i tweet positivi e quelli negativi tendono a bilanciarsi, pur essendo preponderante una nota negativa. Successivamente al discorso si nota come i tweet negativi diventano nettamente i più numerosi. Ben si sa che i social network in generale sono veicolo e amplificatore delle emozioni delle persone, e quelle negative tendono a diffondersi con più facilità. Questo periodo particolare porta con sé un gran malcontento rispetto a fenomeni politici, economici e sociali e l'evento del Concertone nella giornata della Festa dei Lavoratori è piazza ideale per queste tematiche. In questo contesto, il discorso di Fedez è stato la scintilla di questo gran fermento, creando scalpore e aprendo il dibattito sui temi menzionati. Si può così dare una spiegazione al pervadere di tanta negatività.



Si vuole quindi approfondire l'analisi focalizzando l'attenzione sulle emozioni suscitate dai tweet per capire in particolare come queste mutano con l'effetto del discorso. Per fare ciò si sceglie di considerare il periodo che precede il monologo di Fedez e confrontarlo con quello posteriore al discorso. Prima di procedere è bene specificare che i sentimenti di rabbia, tristezza, gioia e paura catturati dall'algoritmo possono essere frutto della classificazione di messaggi che comprendono le sfumature di queste emozioni, e che tali emozioni devono essere contestualizzate. Ad esempio, il sentimento di paura può essere legato a un certo grado di tensione riguardo alle conseguenze del discorso di Fedez. Il grafico a torta ci fa comprendere che se nel primo periodo i sentimenti di rabbia e gioia fossero bilanciati, e che il sentimento di paura fosse minimo, dopo il discorso la rabbia prende nettamente il sopravvento e il sentimento di paura pervade. Questo effetto denota chiaramente le reazioni del pubblico riguardo alle accuse mosse dall'artista.



Approfondendo l'analisi si propone una emoji analysis in grado di catturare l'espressione di quella parte di utenti che hanno voluto identificare una certa emozione che le parole non sono state sufficienti a spiegare, con un simbolo che la rappresenti. Si vuole sottolineare che l'uso delle emoji avviene da parte di una piccola percentuale della popolazione totale, per cui il campione è soggetto ad errore statistico. Anche in questo caso si può vedere che nella prima parte dello spettacolo le emoji fanno riferimento alle emozioni del concerto e si distribuiscono in modo abbastanza uniforme tra quelle più usate. Dopo il discorso prevalgono in modo netto i simboli legati alla diversità di genere o in modo più o meno esplicito alla comunità LGBT (circa 13.5% del totale delle emoji), simbolo di applauso quindi di approvazione (circa 11%) e curiosamente appare un simbolo di bollino rosso che può essere associato ai contenuti espliciti e incensurati dell'artista.



In ultimo si vuole entrare nel dettaglio delle parole più frequentemente utilizzate nei testi dei tweet, in modo da esplicitare i temi più ridondanti cercando di spiegare il motivo delle emozioni che questo concerto ha suscitato. Il primo confronto avviene tra il wordcloud ricavato dai tweet positivi con quello dei tweet negativi rispetto alla prima parte del concerto. Il primo mette in risalto parole come “Bello” e “Ciao” ricollegandosi probabilmente all'inno del partigiano. Le altre parole più usate sono legate al concerto: “canzone”, “cantare”, “palco” e vari nomi di artisti. Quello negativo, in nero, riflette temi legati al denaro come “euro”, “milione”, alla politica e soprattutto al lavoro.



Osservando i wordcloud successivi al discorso di Fedez si nota che al centro di tutte le tematiche vi è l'artista e le parole legate al discorso in sè e al fatto di aver dette determinate cose. Quindi risaltano parole come: "parola", "dire", "discorrere", "combattere", "persona" e una vera e propria nuvola di parole di elogio come "bravo", "coraggio", "amare", "grande". Il cloud negativo si focalizza sui dibattiti suscitati dal discorso e alla rabbia suscitata da esso, al cui centro resta sempre l'artista Fedez. La politica torna al centro dell'attenzione con "lega", la Rai è tra i temi più dibattuti in assoluto per il tentativo di censura del discorso, risaltano le parole "nomi" e "cognomi", in riferimento a quelli dei politici menzionati nel discorso. Parole come "ddl" e "zan" emergono in relazione al decreto mosso contro la discriminazione. Fa da sfondo il dibattito che ha coinvolto la presentatrice Ambra Angioini e la cantautrice Grazia Di Michele che ha contestato la disparità di genere degli artisti coinvolti nel concerto, di cui risalta la parola "grazia", le cui reazioni sono state certamente amplificate a seguito del discorso di Fedez.



## Conclusioni

Con questa analisi si è voluto studiare l'effetto sul comportamento delle persone a seguito dell'intervento di un influencer con un discorso inatteso per la sua denuncia esplicita contro dei personaggi politici per le loro affermazioni omofobe, dimostrando come questo sia in grado di catalizzare le emozioni del pubblico. Il social network Twitter può essere visto come campione rappresentativo dei sentimenti mossi dalla popolazione a questo merito. Si è potuto osservare l'effetto numerico sui tweet che il discorso ha immediatamente sollevato; si sono studiate le emozioni di rabbia che contrassegnano questo periodo in generale sui temi sociali, politici ed economici e quanto un discorso che ha come oggetto l'omofobia da parte di soggetti o partiti politici provochi indignazione e rabbia da parte delle persone. Si è infine cercato un filo rosso tra i termini più menzionati in base al contesto anteriore e posteriore al discorso, dando un senso a quelle emozioni precedentemente espresse solo in modo quantitativo. Per dare un senso all'analisi bisogna conoscere il contesto in cui viviamo e l'attuale periodo storico, in modo da comprendere questo sentimento di rabbia diffuso che rappresenta una conseguenza a difesa della libertà e della partecipazione.