



UNIVERSITÀ DEGLI STUDI DI SALERNO

Dipartimento di Informatica

Corso di Reti Geografiche: Struttura, Analisi e Prestazioni

Analisi manuale dei sotto-commenti relativi ai casi di femminicidio di Giulia Cecchettin e Giulia Tramontano

AUTORI: De Maio Dario, Santangelo Angelo

Indice

Elenco delle Figure	ii
1 Introduzione e Obiettivo	1
2 Metodologia	2
3 Risultati	4
3.1 Descrizione delle Motivazioni/Etichette	4
3.2 Report Complessivo	7
4 Conclusioni	8
4.1 Confronto finale	8
4.2 Considerazioni sull'utilizzo dell'Emotion Classifier	9
Bibliografia	10

Elenco delle figure

3.1	Andamento delle emozioni nell'arco temporale, riguardo l'omicidio di Giulia Cecchettin.	7
3.2	Andamento delle emozioni nell'arco temporale, riguardo l'omicidio di Giulia Tramontano.	7

CAPITOLO 1

Introduzione e Obiettivo

Nella fase precedente della ricerca, si è operato uno scarto dei commenti in risposta ad altri commenti poiché, dopo un'analisi campionaria, è emerso che molti di essi risultavano essere fuori contesto. Tuttavia, al fine di garantire un'analisi più accurata e una visione complessiva più completa, l'obiettivo è analizzare la frequenza dei sotto-commenti, mantenendo un costante monitoraggio dell'evoluzione temporale delle emozioni, tra cui rabbia, gioia, tristezza e paura. Subito dopo, sarà effettuata la selezione di un campione composto da 20 commenti per ciascuna delle emozioni menzionate, i quali saranno sottoposti a un'analisi manuale al fine di valutare se siano relativi al contesto o meno rispetto ai casi di femminicidio corrispondenti. Questo approccio è stato adottato per garantire una comprensione più approfondita delle dinamiche emozionali presenti nei commenti e per individuare eventuali pattern o anomalie che potrebbero essere emerse durante l'interazione. Attraverso questo metodo, si mira a ottenere una rappresentazione più accurata e dettagliata del contesto emotivo in cui sono inseriti i commenti, consentendo così una valutazione più precisa e significativa dei dati raccolti.

CAPITOLO 2

Metodologia

In questa sezione, seguendo il medesimo schema metodologico utilizzato nella ricerca precedente, l'approccio si articola in tre fasi fondamentali: raccolta dei dati, preparazione dei dati e, infine, classificazione dei dati. Come nella ricerca precedente, le tre sotto-fasi rimangono identiche. I commenti sono stati estratti da Reddit utilizzando la libreria PRAW di Python. Una volta ottenuti il testo dei commenti, precisamente 3529, insieme alla relativa data di pubblicazione, sono stati creati due dataset: uno per i commenti relativi al caso Cecchetti e un altro per i commenti relativi al caso Tramontano.

Successivamente, durante la fase di preparazione dei dati, sono state adottate le medesime procedure: identificazione ed eliminazione delle stopwords, rimozione dei commenti non composti esclusivamente da testo e conversione delle date dal formato UTF al formato Datetime.

Al termine di questa fase, i commenti degli utenti sono stati processati attraverso il classificatore di emozioni, utilizzando l'Emotion Classifier della libreria "Feel-it" [1]. Il cuore del progetto è stato raggiunto con la suddivisione del dataset etichettato in sotto-dataset, ciascuno corrispondente a una specifica emozione. Successivamente, sono stati selezionati i primi 20 commenti per ciascuna categoria emotiva.

A questi commenti sono state quindi assegnate delle etichette che indicano se sono

considerati pertinenti o meno al contesto. Questo passaggio è fondamentale per comprendere se i commenti estratti sono in linea con l'argomento trattato.

I 20 commenti, ognuno per ogni categoria, sono stati memorizzati, con la relativa etichetta, all'interno di file Excel separati. Questi file presentano 3 colonne:

1. Testo del commento;
2. Etichetta: il commento è IN o FUORI contesto;
3. Motivazione: rappresenta la motivazione per cui un commento è stato etichettato in un determinato modo;

Nella sezione relativa alla colonna *Motivazione*, sono state identificate diverse categorie che descrivono le ragioni dietro la scelta di una specifica etichetta. Le varie categorie saranno dettagliate nella sezione successiva, con particolare enfasi sull'analisi condotta.

3.1 Descrizione delle Motivazioni/Etichette

In questa sezione verranno illustrate le etichette utilizzate nel progetto, spiegandone il significato e la relativa associazione. Le prime 5 etichette indicano l'esclusione dei commenti, mentre le ultime 3 rappresentano la loro selezione e inclusione.

Discussione Tutti i commenti etichettati in questo modo sono stati esclusi poiché trattavano argomenti del tutto estranei al tema principale.

Molti di questi commenti presentavano una varietà di tematiche, tra cui dispute generate da punti di vista contrastanti, critiche alle istituzioni e reazioni provocate da incomprensioni.

Questi sottotemi hanno contribuito a un aumento delle frequenze di diverse emozioni, tra cui la rabbia, evidenziando la complessità e la diversità delle risposte emotive suscitate dall'argomento trattato.

Non contestualizzata Numerosi commenti sono stati identificati come 'non contestualizzati'.

Questa etichetta è stata assegnata per escludere tutti quei contributi che non solo non

affrontavano il tema principale, ma neanche lo sfioravano in minima parte. Molto spesso, gli utenti condividevano aneddoti personali che risultavano completamente estranei al contesto della discussione.

Allo stesso modo, alcuni commenti erano del tutto privi di significato, non apportando alcun contributo alla conversazione. Pertanto, è stato deciso di escluderli dall'esperimento al fine di mantenere l'analisi focalizzata sui commenti rilevanti e pertinenti.

Ironia decontestualizzata I commenti associati a questa motivazione sono stati ritenuti fuori contesto in quanto consistevano in espressioni ironiche e sarcastiche riguardanti argomenti diversi da quelli presi in considerazione nell'analisi.

Bot Si è valutato che tutti i commenti contrassegnati con questa specifica etichetta risultassero fuori contesto, in quanto riconducibili a messaggi generati da bot sulla piattaforma Reddit.

Espressione I commenti associati a questa motivazione sono stati considerati fuori contesto in quanto consistevano principalmente in espressioni insignificanti come "ehhh" e "ahia". In tali casi, è stato ritenuto che l'Emotion Classifier potesse commettere errori di classificazione, dato il carattere poco significativo di tali interazioni.

Riflessione sull'accaduto I commenti associati a questa motivazione sono stati ritenuti in contesto poiché rappresentavano opinioni significative relative ai femminicidi presi in esame. Essi comprendevano considerazioni sulla necessità di adottare precauzioni per prevenire ulteriori casi simili, come la frequenza di corsi di arti marziali o di autodifesa. Inoltre, si esprimevano pronostici sulla sequenza degli eventi successivi alla scomparsa di entrambi i corpi delle ragazze, manifestando rabbia nei confronti degli assassini e auspici di non vederli mai più liberi alla luce del sole. Alcuni commentatori esprimevano anche previsioni riguardo allo stato emotivo dei genitori, con particolare attenzione al padre di Giulia Cecchettin, colpito anche dalla

perdita della moglie (madre di Giulia) nello stesso anno. Questi elementi conferivano ai commenti un legame intrinseco con il contesto dell'analisi.

Spiegazione approfondita dell'accaduto I commenti contrassegnati con questa motivazione sono stati considerati in contesto poiché rappresentavano le riflessioni di alcuni utenti che, ad esempio, approfondivano dettagli relativi a eventi accaduti prima della scomparsa di Giulia Cecchettin insieme al fidanzato. Tali utenti cercavano di ricostruire scenari e individuare possibili cause che potrebbero aver portato Giulia Cecchettin a allontanarsi con il fidanzato. Allo stesso tempo, attraverso questi aneddoti, molti utenti esprimevano preoccupazioni riguardo al rischio di femminicidio. Pertanto, tali commenti sono stati considerati rilevanti e contestualmente connessi al tema oggetto dell'analisi.

Ironia contestualizzata I commenti associati a questa motivazione sono stati ritenuti in contesto in quanto consistevano in espressioni ironiche e sarcastiche riguardanti gli argomenti presi in considerazione nell'analisi.

3.2 Report Complessivo

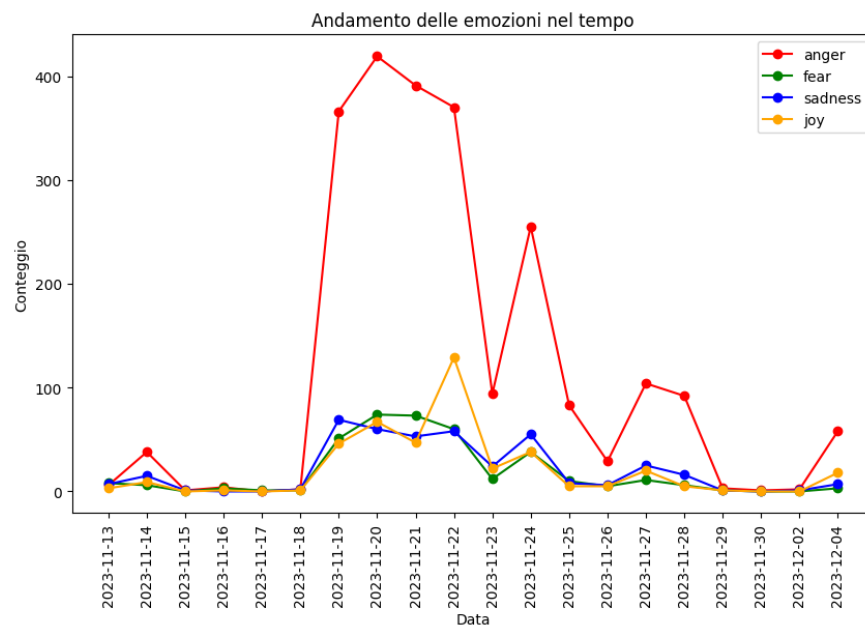


Figura 3.1: Andamento delle emozioni nell’arco temporale, riguardo l’omicidio di Giulia Cecchettin.

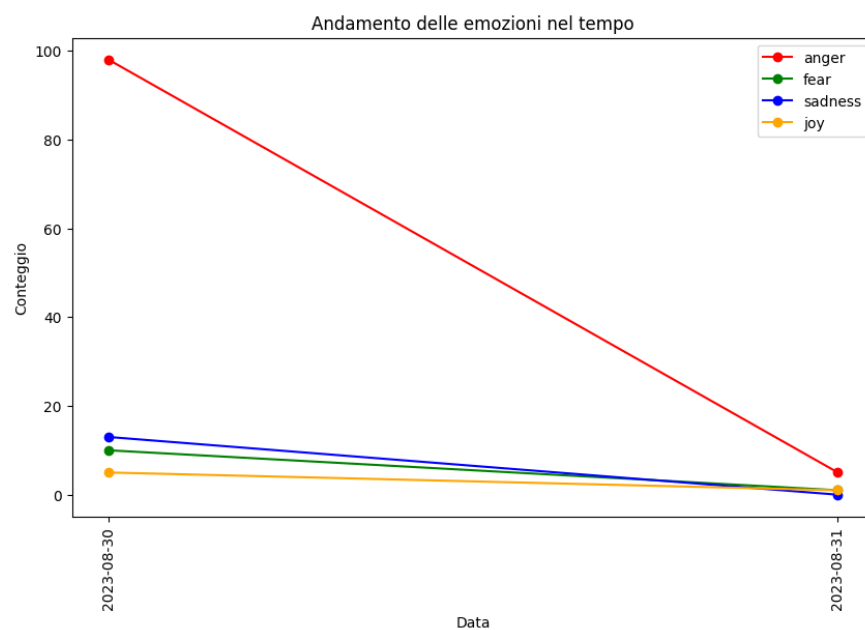


Figura 3.2: Andamento delle emozioni nell’arco temporale, riguardo l’omicidio di Giulia Tramontano.

4.1 Confronto finale

La comparazione dei due casi evidenzia ancora notevoli differenze, soprattutto per quanto riguarda la frequenza dei commenti. I risultati evidenziano notevoli disparità nella frequenza dei commenti tra i due casi, ma sotto il profilo delle emozioni, emerge una similitudine significativa, poiché in entrambe le situazioni prevale la rabbia. È interessante notare che, rispetto alla fase precedente dell'analisi, anche in questo contesto persiste un predominio della rabbia. Ciò è dovuto soprattutto da una presenza molto massiccia di commenti fuori contesto, riguardanti principalmente discussioni accese tra gli utenti della piattaforma anche su tematiche differenti a quelle considerate.

In effetti, l'analisi ha rivelato una prevalenza di commenti giudicati fuori contesto rispetto a quelli considerati in contesto, ognuno con motivazioni specifiche. La scelta di non selezionare i sotto-commenti nella prima parte dell'analisi è risultata strategica in quanto il classificatore avrebbe potuto non completamente capace di comprendere il contesto, rendendo la sentiment analysis meno efficiente. Tuttavia, è importante sottolineare che questa analisi risulta intrinsecamente riduttiva, poiché ci si è limitati a esaminare solo i primi 20 commenti per ciascuna emozione, pur avendo a disposi-

zione un campione ampio di circa 3500 commenti. Tuttavia, l'ampiezza del campione rimasto offre spunti interessanti per ulteriori approfondimenti e per comprendere meglio le dinamiche emozionali all'interno del contesto analizzato.

4.2 Considerazioni sull'utilizzo dell'Emotion Classifier

Dopo un'analisi dettagliata dei risultati ottenuti dai file Excel, emerge chiaramente che il classificatore utilizzato non sempre categorizza correttamente i commenti, nonostante i buoni risultati mostrati all'interno del paper ufficiale del modello[1]. Questo scenario solleva interrogativi significativi sulla sua affidabilità e applicabilità in contesti reali.

In effetti, è evidente che, soprattutto per l'emozione di gioia, il numero di classificazioni erranee risulta significativo. Ad esempio, viene riportato un caso emblematico di un commento associato erroneamente all'emozione di gioia. Nel caso specifico di "Giulia Cecchettin", il commento **"Finalmente. Spero non veda mai più direttamente la luce del sole."** è stato erroneamente classificato come gioioso, mentre è chiaro che esprime un sentimento di rabbia. Questo esempio illustra in modo tangibile la criticità delle classificazioni effettuate dal modello.

Di conseguenza, questi errori compromettono la precisione complessiva del processo di classificazione, rendendo fondamentale un'analisi critica e un miglioramento del sistema. La presenza di tali errori solleva dubbi sulla validità delle conclusioni tratte dai dati classificati e sottolinea la necessità di esaminare attentamente il funzionamento del classificatore.

Pertanto, sarebbe opportuno considerare l'impiego di modelli alternativi mirati e condurre un'analisi su un campione più ampio di commenti rispetto ai 20 (per ogni emozione e per ognuno dei due casi di femminicidio) considerati in questa ricerca. L'aumento della dimensione del campione non solo potrebbe contribuire a ridurre gli errori di classificazione generale, ma potrebbe anche conferire maggiore affidabilità all'analisi condotta in questa ricerca. In particolare, ciò sarebbe cruciale per l'analisi manuale dei commenti fuori contesto o contestualizzati, consentendo una valutazione più accurata delle emozioni espresse.

Bibliografia

- [1] F. Bianchi, D. Nozza, and D. Hovy, ""FEEL-IT: Emotion and Sentiment Classification for the Italian Language",," in *Proceedings of the 11th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*. Association for Computational Linguistics, 2021. (Citato alle pagine 2 e 9)