



Università degli Studi di Salerno
Dipartimento di Informatica

Corso di Laurea Magistrale in Informatica

Progetto Reti Geografiche: Sentiment e Hate Speech Analysis delle maggiori testate giornalistiche sui Social Media

Professoressa
Delfina Malandrino

Studenti
Roberto Piscopo
Alessio Maddaluno

Anno Accademico 2023/2024

Abstract

Con la crescente diffusione dei social media e l'espansione della loro utenza, si è osservato un incremento significativo dei commenti negativi e dell'hate speech online, specialmente nel contesto delle notizie e dell'informazione. Questo fenomeno rappresenta una sfida rilevante nell'ambito della comunicazione digitale, poiché tali commenti possono influenzare l'opinione pubblica, alimentare la polarizzazione e diffondere discorsi dannosi. L'aumento della visibilità e dell'accessibilità delle piattaforme digitali ha amplificato l'impatto di tali comportamenti, rendendo cruciale un'analisi approfondita della loro natura e diffusione. La rilevazione e la gestione di questi contenuti rappresentano una sfida complessa, data la vastità e la varietà delle informazioni scambiate quotidianamente online. In risposta a questa sfida, vengono utilizzati metodi avanzati di analisi del testo e di machine learning per identificare automaticamente i commenti negativi e l'hate speech nei social media. Questo elaborato si propone di esaminare e classificare, tramite modelli basati su Trasformer, i commenti presenti sui social media, concentrandosi sulle principali testate giornalistiche, al fine di comprendere meglio la diffusione e la natura dell'hate speech online e stabilire se la tipologia di media e di social influenzano in qualche modo la quantità di commenti negativi e di odio nei confronti di notizie di vario genere. L'obiettivo è individuare le fonti che sono associate al maggior numero di commenti negativi. Le motivazioni che guidano questo studio rispecchiano un forte interesse nel comprendere il ruolo delle testate nel promuovere o contrastare l'odio online. La ricerca mette in luce le sfide significative nell'identificare e classificare i commenti sui social media, sottolineando i limiti degli strumenti messi a disposizione per fare attività di scraping.

Indice

1	Introduzione	4
1.1	Obiettivo	4
1.2	Motivazioni	4
2	Concetti e definizioni	6
2.1	Stato dell'arte	6
2.1.1	Sentiment Analysis	6
2.1.2	Hate Speech	7
3	Metodologia	9
3.1	La scelta dei Social Network	9
3.2	Individuazione delle notizie	10
3.2.1	Scelta dei topic	10
3.2.2	Scelta delle testate giornalistiche	10
3.2.3	Scelta delle notizie	11
3.3	Tools e tecnologie utilizzate	11
3.4	Creazione del Dataset	12
3.5	Estrazione notizie	12
3.6	Data cleaning	13
3.7	Classificazione Sentiment e Hate Speech	14
3.7.1	Sentiment Analysis	14
3.7.2	Hate Speech Detection	14
4	Analisi	16
4.1	Distribuzione dei commenti	16
4.2	Analisi per Topic	17
4.2.1	Politica	17
4.2.2	Cronaca	19
4.2.3	Cronaca Nera	21
4.3	Analisi medie commenti SOCIAL delle testate	23
4.3.1	Tabella media commenti per topic	25
4.3.2	Risultati analisi delle Testate Giornalistiche	25
4.3.3	Risultati analisi dei Social	26

5	Considerazioni finali	27
5.1	Qualità del dataset	27
5.1.1	Collaborazione tra modelli	27
5.2	Strumenti di scraping massivi	28
5.2.1	Selezione delle notizie priva di bias	28
5.3	Sviluppi Futuri	29

Capitolo 1

Introduzione

I social network hanno assunto un ruolo centrale nelle nostre vite, sia personali che professionali, fornendo una varietà di contenuti che spaziano dai video divertenti ai documentari. Considerando il diffuso utilizzo dei social media da parte di milioni di persone, numerose aziende hanno ampliato la propria presenza in questo contesto. Le piattaforme di social media hanno radicalmente modificato il modo in cui le persone interagiscono e condividono informazioni, contribuendo anche alla proliferazione di contenuti inappropriati e discutibili. La rilevazione e la gestione di questi contenuti rappresentano una sfida complessa, data la vastità e la varietà delle informazioni scambiate quotidianamente online. In risposta a questa sfida, vengono utilizzati metodi avanzati di analisi del testo e di machine learning per identificare automaticamente i commenti negativi e l'hate speech nei social media (*Sentiment e hate speech analysis*). La maggior parte dei principali social media come Youtube, Facebook e Instagram hanno le loro politiche che regolano quali tipi di discorsi di odio sono consentiti sui loro siti. Questa ricerca si focalizzerà sulle testate giornalistiche più attive sui social media, analizzando i commenti per esplorare la natura del problema dell'incitamento all'odio.

1.1 Obiettivo

Il nostro obiettivo è analizzare i commenti associati a video, articoli e post di notizie su vari argomenti (*Politica, Cronaca, Cronaca Nera*) delle testate giornalistiche più diffuse. L'obiettivo è identificare quelle che generano un maggior numero di commenti negativi e di odio (*hate speech*). Faremo ciò, confrontando le stesse notizie tra le diverse testate e analizzando le differenze tra le testate sulle varie piattaforme per determinare quale presenta il livello più alto di *sentiment* negativo e di hate speech.

1.2 Motivazioni

I giornali hanno sempre cercato di attirare lettori con titoli allarmistici o che incitassero all'odio in modo tale da creare scalpore e far acquistare sempre più copie. Con l'avvento del digitale, le testate giornalistiche non hanno cambiato metodologia. Oggi titoli clickbait o titoli che fanno nascere sentimenti negativi sono all'ordine del giorno. Questo perché sono titoli che attirano click e guadagni relativamente facili. Le motivazioni che guidano questo studio rispecchiano un forte interesse nel

comprendere il ruolo delle testate nel promuovere o contrastare l'odio online e la diffusione di commenti negativi. Inoltre, andando a confrontare varie testate sugli stessi argomenti, ci permette di capire ad esempio che non è solo la notizia in se che genera Hate Speech ma, anche il modo in cui viene detto e attraverso quale strumento. Capire, ad esempio, se un video genera più odio di un post.

Capitolo 2

Concetti e definizioni

2.1 Stato dell'arte

2.1.1 Sentiment Analysis

La Sentiment Analysis (Analisi del sentimento), detta anche Opinion Mining, è l'attività mirata ad "ascoltare" il Web per conoscere quello che si dice in Rete della propria società e del brand. La sentiment analysis, in particolare, gioca un ruolo fondamentale nell'analizzare le emozioni e le opinioni espresse nei commenti online nei Social Network. Attraverso l'analisi del linguaggio naturale e l'identificazione di parole chiave e pattern linguistici, è possibile determinare il sentiment di un dato testo e classificarlo come positivo, negativo o neutro. Questa tecnica è cruciale per comprendere il tono e il contenuto emotivo dei commenti sui social media e può essere impiegata efficacemente per monitorare e mitigare la diffusione di contenuti dannosi.

La Sentiment Analysis si è rivelata uno strumento prezioso per valutare l'opinione pubblica in diverse discipline. È stata utilizzata con successo nella previsione dei mercati finanziari, nei problemi di salute, nell'analisi dei clienti, nella valutazione commerciale, nel brand marketing, nella politica, crime prediction e nella gestione delle emergenze.

Tecniche di Sentiment Analysis

Le tecniche di Sentiment Analysis possono essere classificate in approcci di machine learning, approcci basati sul lessico e metodi ibridi.

Le Sentiment Analysis più recenti utilizzano modelli di Deep learning ed elaborazione del linguaggio naturale (NLP, Natural Language Processing) per analizzare il testo scritto. Le metriche utilizzate sono progettate per rilevare se il *sentiment* complessivo di un testo è positivo, negativo o neutro.

Le Deep Neural Network (DNN) sono costituite da reti neurali artificiali con più strati nascosti tra lo strato di input e lo strato di output. Alcuni dei modelli di deep learning più popolari utilizzati nella Sentiment Analysis sono: CNN, Rec NN, RNN.

Il modello CNN più popolare per la classificazione del sentiment a livello di frase è il lavoro svolto da Kim [1]. L'autore ha condotto un esperimento con la CNN costruito su word2vec pre-addestrato. I

risultati sperimentali mostrano che i vettori pre-addestrati possono fungere da eccellente estrattore di feature per attività relative alla NLP utilizzando il deep learning.

Vari ricercatori hanno anche combinato varie tecniche di deep learning per la classificazione del sentiment, come nel lavoro di Huang et al. [2], dove hanno combinato due popolari reti di deep learning chiamate LSTM e CNN.

Lo stato dell'arte in letteratura vede l'utilizzo di modelli pre allenati sul linguaggio testuale, nello specifico : Bidirectional Encoder Representations from Transformers (BERTs) è una tecnica di apprendimento automatico basata su trasformatore per l'elaborazione del linguaggio naturale che è stata pre-addestrata e sviluppata da Google [3] sulla base della conoscenza estratta dal testo (vettori) utilizzando il testo circostante per stabilire il contesto. Bert è una tecnica basata su reti neurali per l'elaborazione del linguaggio. Questa tecnologia consente a chiunque di addestrare il proprio sistema per l'analisi del testo custom, effettuando Fine-Tuning. Il lavoro in [4] ha usato il modello italiano di BERT (UmBERTo [5]) ed ha effettuato il fine-tune per classificare le emozioni del testo. Hanno poi raggruppati le classi in 2 gruppi tramite la mappatura felicità nella classe positiva e rabbia, paura e tristezza nella classe negativa.

La Sentiment Analysis si basa non solo sulla polarità negativa o positiva di parole e concetti, ma anche sull'albero sintattico della frase analizzata. Il sistema cerca di leggere tra le righe, individuando espressioni idiomatiche o colloquiali, dando interpretazione alle negazioni, modificando la polarità delle parole in base ai relativi avverbi, aggettivi, congiunzioni o verbi, tenendo conto di specifici complementi logico-funzionali.

2.1.2 Hate Speech

Il termine "hate speech" comprende tutte le forme di espressione che *"diffondono, incitano, promuovono o giustificano odio razziale, xenofobia, antisemitismo o altre forme di odio basate sull'intolleranza, compresa l'intolleranza espressa dal nazionalismo aggressivo e dall'etnocentrismo, discriminazione e ostilità contro le minoranze, i migranti e le persone di origine immigrata"* - Consiglio d'Europa.

L'hate speech detection si concentra sull'identificazione di linguaggio offensivo, discriminatorio o violento nei commenti online. Gli impatti del hate speech sono già travolgenti a causa della diffusa adozione dei Social Media e dell'anonimato di cui godono gli utenti online. Utilizzando modelli di machine learning addestrati su grandi dataset di testi annotati, è possibile riconoscere automaticamente i commenti che violano le norme della community e promuovono odio e discriminazione. Questi sistemi di detection sono fondamentali per garantire un ambiente online sicuro e inclusivo, proteggendo gli utenti da contenuti dannosi e contribuendo a promuovere un dialogo civile e rispettoso sui social media.

Tecniche di Hate Speech Detection

Il primo approccio è quello che si basa su metodi di machine learning tradizionali di natura statistica o probabilistica con quali si sfruttano noti algoritmi di apprendimento automatico quali SVM (Support Vector Machine), Decision Tree e Naive-Bayes.

Un'altro approccio che è possibile utilizzare per realizzare un sistema di detection di hate speech prevede l'uso di tecnologie basate su deep learning. Il deep learning è una branca specifica del machine

learning che ha preso sempre più piede negli ultimi anni. Anche in questo caso l'utilizzo di Bert, succitato, è molto diffuso.

Nel lavoro svolto da [6] è stato usato, AlBERTo [7], un modello linguistico basato su BERT pre-addestrato su una raccolta di tweet in lingua italiana. Il testo viene classificato in 4 categorie: accettabile, inappropriato, offensivo e violento.

Capitolo 3

Metodologia

3.1 La scelta dei Social Network

Per identificare i social più idonei per condurre questo studio, abbiamo prima valutato il livello di coinvolgimento degli utenti su diverse piattaforme social, individuando quelli che, secondo la letteratura recente [8], sono più frequentati dagli utenti in cerca di informazioni di vario genere. Questo perché, nonostante la libertà di pubblicazione, le principali testate giornalistiche tendono a utilizzare i social più appropriati per diffondere le proprie notizie. Vedi figura 3.2.

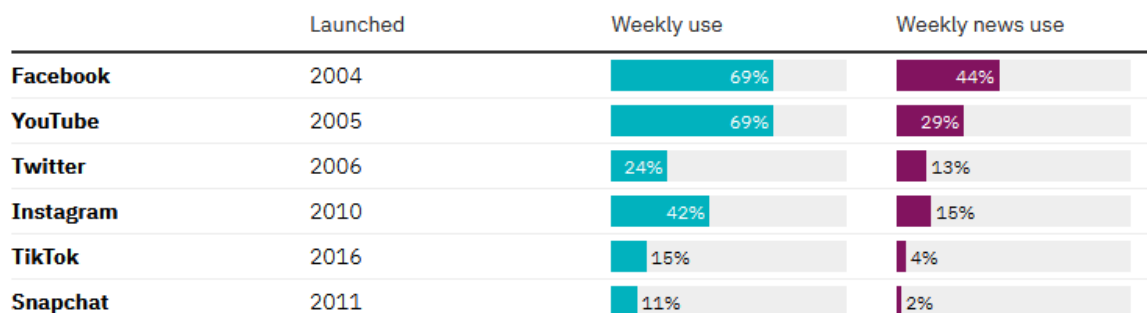


Figura 3.1: Profilazione dell'utilizzo dei social da parte degli utenti.

Dal grafico sembra evidente che i social con più appeal verso il mondo delle news e contemporaneamente più frequentati dagli utenti sono:

- Facebook
- Youtube
- X (ex Twitter)
- Instagram

Tuttavia, dopo il recente caso di acquisizione di Twitter (ora X), che ha causato una riduzione della partecipazione degli utenti a causa del cambio di brand e delle limitazioni della nuova piattaforma, abbiamo deciso di escluderlo da questo studio specifico, poiché non fornisce materiale sufficiente su cui basare un dataset convincente.

3.2 Individuazione delle notizie

3.2.1 Scelta dei topic

Nel processo di costruzione del dataset per la sentiment analysis, ci siamo trovati di fronte al primo ostacolo importante: individuare i post sui social media. Il primo passo è stato identificare dei temi (o generi di notizie) che potessero suscitare discussioni e commenti da parte degli utenti. Abbiamo scelto di focalizzarci su tre temi principali:

- **Cronaca Nera:** una parte significativa delle discussioni online si verifica in risposta a eventi che suscitano forte coinvolgimento emotivo, come ad esempio i casi di omicidio. Questi eventi spesso generano una forte reazione emotiva negli utenti, spingendoli a commentare e condividere le proprie opinioni.
- **Politica:** Analogamente alla cronaca nera, la politica è uno dei temi più discussi e controversi sui social media. Le discussioni riguardanti la politica possono variare da scambi di opinioni costruttive a manifestazioni di odio verso la classe politica.
- **Cronaca Generale:** Abbiamo scelto di includere la cronaca generale come terzo tema per garantire diversità nel dataset. Questa categoria comprende una vasta gamma di notizie che non rientrano esclusivamente nei temi della politica o della cronaca nera. Si tratta di notizie che possono riguardare argomenti come disastri naturali, la finanza o eventi di rilievo.

Questa selezione di temi ci ha permesso di coprire un'ampia gamma di argomenti e di raccogliere dati diversificati per l'analisi.

3.2.2 Scelta delle testate giornalistiche

Per le testate giornalistiche ci siamo avvalsi di un'analisi empirica dei profili social per individuare quelle meno legate al giornale fisico e che negli ultimi anni hanno investito più in ambiente social. Questo rende la possibilità di ottenere un engagement maggiore degli utenti e quindi un dataset più completo per la nostra analisi.

Le testate giornalistiche che abbiamo individuato sono:

- **Fanpage :** Una piattaforma giornalistica online italiana che si distingue per la sua presenza sui social media e la capacità di generare discussione su una vasta gamma di argomenti;
- **Il Corriere della Sera :** Uno dei principali quotidiani italiani con una lunga storia e una vasta diffusione nazionale, autorevole nell'informare il pubblico su eventi nazionali e internazionali, con particolare attenzione a politica, economia, cultura e attualità;

- **Repubblica** : Un quotidiano italiano noto per il suo approccio innovativo e progressista nell'informare il pubblico, con una copertura completa degli eventi nazionali e internazionali, stimolando il dibattito pubblico su politica, economia, cultura e attualità;

3.2.3 Scelta delle notizie

Abbiamo identificato 4 notizie rappresentative per ciascun topic individuato. Per ogni testata giornalistica scelta, abbiamo esaminato i profili ufficiali sui social media di Facebook, YouTube e Instagram al fine di individuare i post relativi alle notizie selezionate. Il topic di ogni notizia è stato classificato manualmente, insieme all'individuazione del post sui social, per garantire l'estrazione di commenti coerenti con il contesto. Nella tabella 3.3 è riportata la lista delle notizie con le relative categorie.

Tabella 3.1: Tabella delle notizie e dei relativi topic

Notizia	Topic
Omicidio Giulia Cecchettin	CRONACA NERA
Implosione del Titan (Sottomarino)	CRONACA NERA
Incidente Youtubers	CRONACA NERA
Strage di Cutro	CRONACA NERA
La cattura di Matteo Messina Denaro	CRONACA
Geolier e i fischi Sanremo	CRONACA
Alluvioni Emilia Romagna	CRONACA
Il caso Chiara Ferragni e Balocco	CRONACA
L'incoronamento di Re Carlo	POLITICA
Attacco di Hamas a Israele	POLITICA
Invasione Ucraina	POLITICA
La morte di Silvio Berlusconi	POLITICA

3.3 Tools e tecnologie utilizzate

- **Python**: è un linguaggio di programmazione versatile, con una sintassi chiara e leggibile, che supporta molteplici paradigmi di programmazione. È ampiamente utilizzato per lo sviluppo di una vasta gamma di applicazioni, inclusi script, applicazioni web, analisi dei dati, machine learning e molto altro;
- **Matplotlib**: è una libreria Python ampiamente utilizzata per la visualizzazione di dati. Con Matplotlib, è possibile creare una vasta gamma di grafici e visualizzazioni, tra cui grafici a linee, istogrammi, scatter plot, grafici a torta e molto altro ancora;
- **Pandas**: è una libreria Python progettata per facilitare la manipolazione e l'analisi dei dati. Introduce strutture dati potenti, come i DataFrame, che consentono di lavorare con dati tabulari in modo intuitivo e efficiente. Pandas offre una vasta gamma di funzioni per la pulizia, la trasformazione e l'analisi dei dati, rendendola una scelta popolare per gli analisti e i data scientist;

- **Hugging Face:** è una piattaforma open-source che offre modelli, dataset e strumenti per l'elaborazione del linguaggio naturale (NLP) e il machine learning. In particolare, è conosciuta per la sua vasta raccolta di modelli pre-addestrati per NLP, che possono essere utilizzati e adattati per una varietà di compiti, come la classificazione del testo, la generazione di testo e l'analisi del sentiment;
- **Esuit:** è un insieme di estensioni per Chrome utilizzate principalmente per effettuare analisi da Social Media Managers, che però nel contesto di questo progetto abbiamo utilizzato per estrarre i commenti, dal momento che alla stesura di questo documento le API non consentono di accedere gratuitamente a grosse quantità di dati. Abbiamo identificato in questa suite un workaround per accedere velocemente ai commenti di qualsiasi social;

3.4 Creazione del Dataset

Per ciascuna notizia pubblicata da una testata giornalistica (sotto forma di post o video) sono stati estratti fino a un massimo di 100 commenti. Questa limitazione deriva dall'impossibilità di accedere alle API della maggior parte dei social media e, dall'utilizzo conseguente, di software di terze parti (Esuit) che richiedono un pagamento per estrarre un numero maggiore di commenti.

3.5 Estrazione notizie

Per ottenere un dataset equilibrato sono state prese 4 notizie per ognuno dei topic elencati sopra. Ad ogni notizia è stato assegnato un titolo univoco, per uniformare le notizie tra i vari social. Ciascuna è stata cercata sui social (Facebook, Youtube, Instagram) delle tre testate giornalistiche scelte. Per ogni social è stato scritto un piccolo parser per uniformare il dataset relativamente ai social. In particolare ogni riga contiene:

Tabella 3.2: Struttura del dataset

Campo	Descrizione
Piattaforma Social	Social da cui è stato estratto il commento (Facebook / Instagram / Youtube)
Testata Giornalistica	Testata giornalistica (La Repubblica / Corriere della Sera / Fanpage.it)
Notizia	Titolo della notizia (Manualmente assegnato in modo da uniformare il dataset e estrarre facilmente i commenti da una notizia specifica)
Commento	Commento Estratto
Topic	Tipologia della notizia (Manualmente estratto, contiene i valori: POLITICA, CRONACA, CONACA NERA)
Sentiment	Categorizzazione del sentimento del commento, ottenuto dal modello di machine learning [6] . Identifica il sentiment secondo le categorie: positivo, negativo
Hate Speech	Categorizzazione del sentimento del commento, ottenuto dal modello di machine learning [4]. Identifica il sentiment secondo le categorie: assente, inappropriato, offensivo, violento

3.6 Data cleaning

La pulizia dei dati è una fase critica nel processo di analisi testuale, poiché influisce direttamente sulla qualità e sull'affidabilità dei risultati ottenuti dalla nostra analisi. Abbiamo dunque implementato una fase di data cleaning, garantendo che i dati in ingresso siano privi di elementi non utili o potenzialmente distorti, consentendo al modello di concentrarsi sui contenuti rilevanti e di fornire analisi accurate e informative. Gli elementi considerati in questa in questa pulizia sono:

- **Emoticon e simboli:** rimuovere una vasta gamma di emoticon e simboli unicode che potrebbero essere presenti nei dati. Questi dati possono indubbiamente contribuire nella manifestazione di un umore, tuttavia a seguito di alcuni test, abbiamo riscontrato che non sono particolarmente efficaci per il modello utilizzato pertanto abbiamo deciso di rimuoverle;
- **Rimozione dei riferimenti agli utenti (@):** identifichiamo e rimuoviamo i riferimenti agli utenti preceduti dal simbolo "@" all'interno del testo. Questi riferimenti sono spesso presenti nei commenti su piattaforme social come Instagram e potrebbero non essere rilevanti per le nostre analisi.
- **Eliminazione dei link:** individuiamo e rimuoviamo i link ipertestuali presenti nel testo. Questa operazione è utile per eliminare eventuali URL o collegamenti web perchè potrebbero contenere parole che potrebbero influenzare il modello;

- **Gestione dei caratteri speciali** : Alcuni caratteri, come le virgolette doppie ("), potrebbero essere presenti nei testi e potrebbero essere considerati come rumore durante l'analisi. Questi vengono rimossi sostituendoli con una stringa vuota;

3.7 Classificazione Sentiment e Hate Speech

Per l'operazione di classificazione sono stati individuati due modelli di machine learning basati sull'intelligenza Artificiale BERT di Google. Entrambi i modelli sono pre-addestrati e sono utilizzati *as a service* tramite la piattaforma Hugging Face, apprezzata per la sua facilità d'uso, la documentazione dettagliata e la comunità attiva che supporta lo sviluppo e l'utilizzo della libreria. Inoltre la presenza dei modelli in librerie python ci ha consentito di utilizzarli con estrema semplicità.

3.7.1 Sentiment Analysis

Per il campo del *sentiment* nel dataset, abbiamo impiegato il modello "feel-it" [4], il quale consente di effettuare la classificazione delle emozioni su un testo. Tale modello si occupa di classificare l'input in due categorie: positivo e negativo. È importante sottolineare che per "negativo" non si fa riferimento al "hate speech", ma piuttosto a un testo con connotazioni negative. Nel dataset abbiamo identificato un valore di threshold al di sotto del quale considerare il testo come negativo.

3.7.2 Hate Speech Detection

Per il campo *hate speech* del dataset, dopo aver testato la libreria *setfit* [9], che non classificava correttamente in molteplici casi, abbiamo deciso, dopo una lunga ricerca, di utilizzare la libreria *hate_speech_it* [6] che permette di avere anche il grado di odio suddiviso in 4 categorie (assente, inappropriato, offensivo, violento). In particolare, vengono utilizzate delle etichette numeriche (0, 1, 2, 3) così suddivise:

Tabella 3.3: Classificazione modello hate speech it

Etichetta	Classificazione	Descrizione
0	acceptable	Il contenuto è considerato appropriato e non solleva preoccupazioni di natura inappropriata o offensiva.
1	inappropriate	Il contenuto ha elementi che potrebbero essere considerati non adatti al contesto o alla situazione, ma non raggiunge il livello di offensività grave.
2	offensive	Il contenuto è offensivo o problematico e potrebbe causare disagio o fastidio a una parte del pubblico.
3	violent	Il contenuto è estremamente offensivo e potenzialmente dannoso, contenendo elementi di violenza verbale o visiva che possono essere traumatizzanti o dannosi per il pubblico.

Riassumendo, il processo di creazione e di classificazione di dati è il seguente:



Figura 3.2: Schema metodologia.

Capitolo 4

Analisi

4.1 Distribuzione dei commenti

Il dataset è composto da un totale di 10.800 commenti in totale, suddivisi per categoria come mostrato in figura 4.1 e per topic come mostra la figura 4.2.

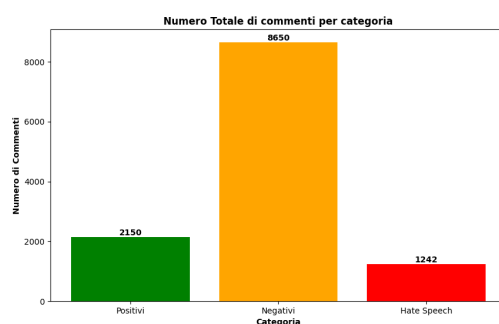


Figura 4.1: Distribuzione commenti per categoria

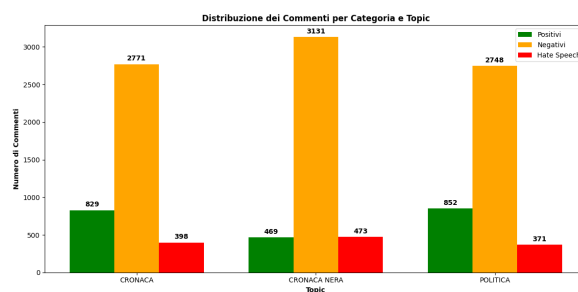


Figura 4.2: Distribuzione commenti per categoria e topic

4.2 Analisi per Topic

In questa sezione sono mostrati gli istogrammi per l'analisi dei commenti positivi e negativi delle varie testate giornalistiche sui tre social, suddivisi per topic.

4.2.1 Politica

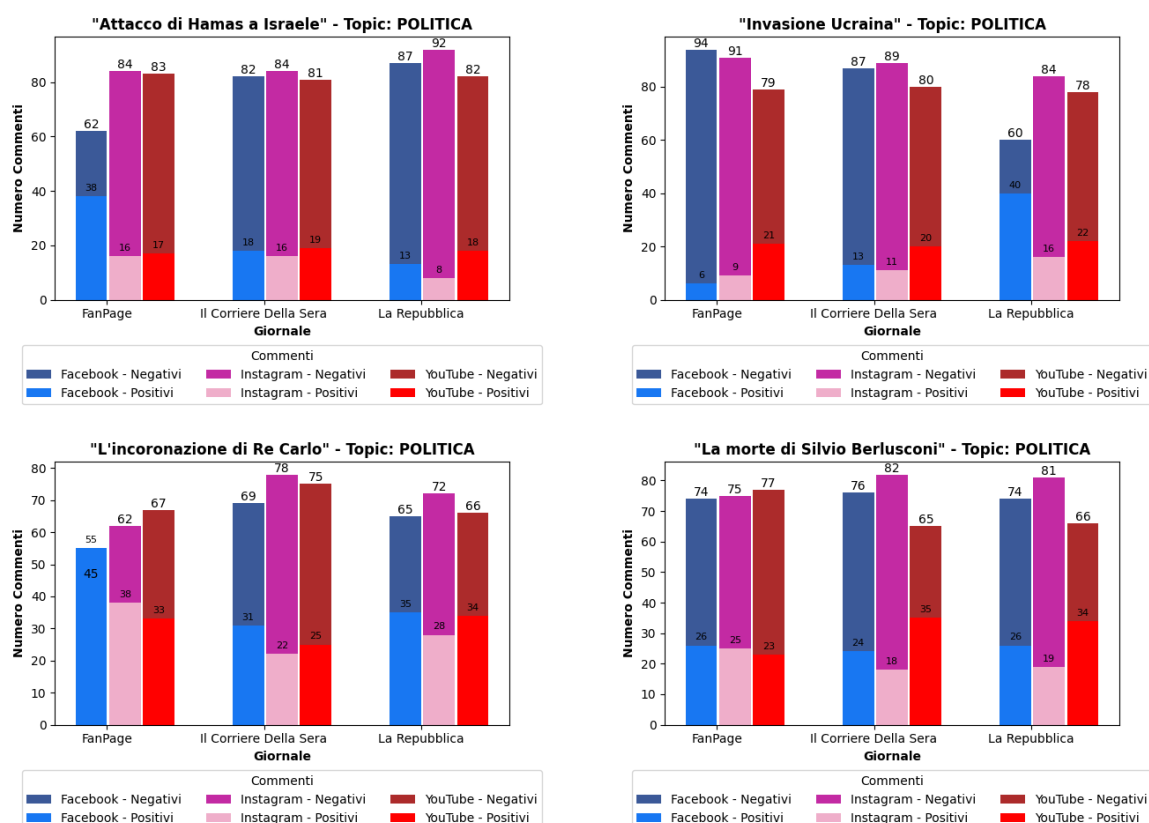


Figura 4.3: Numero di commenti positivi e negativi per ogni notizia per ogni social - POLITICA

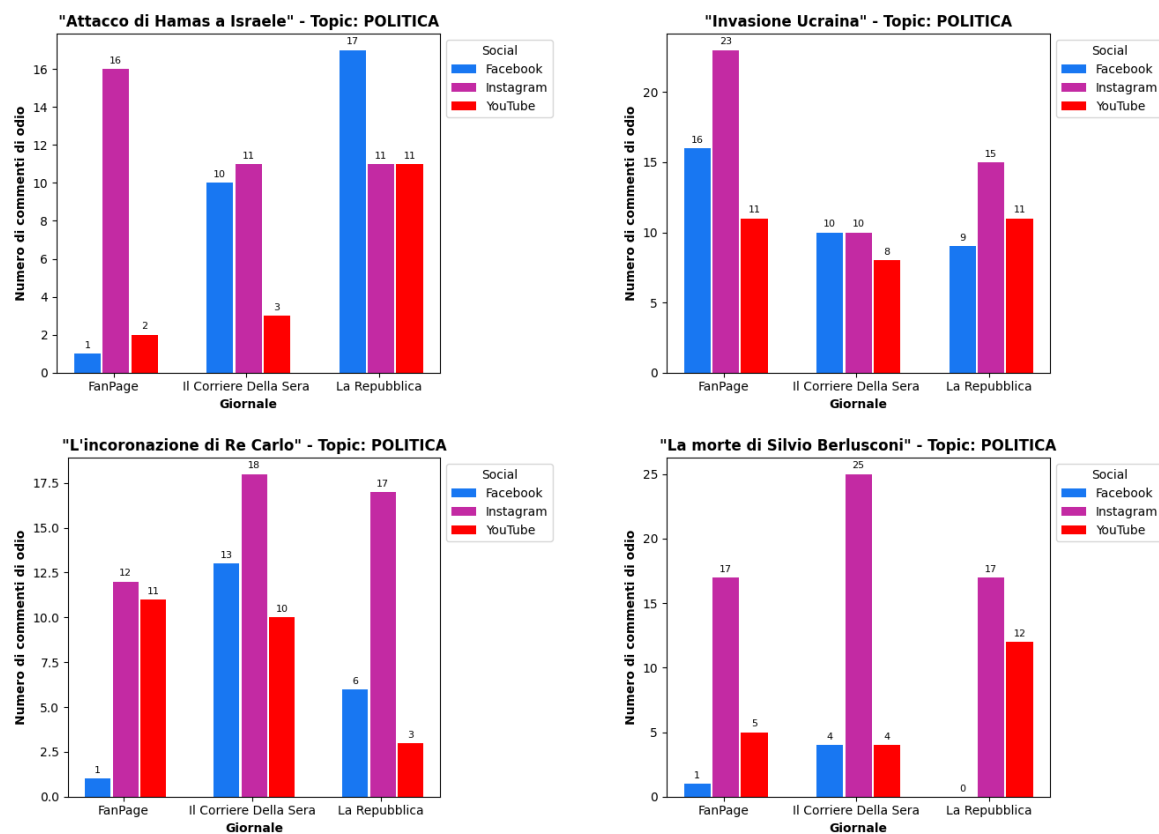


Figura 4.4: Numero di commenti di odio per ogni notizia per ogni social - POLITICA

4.2.2 Cronaca

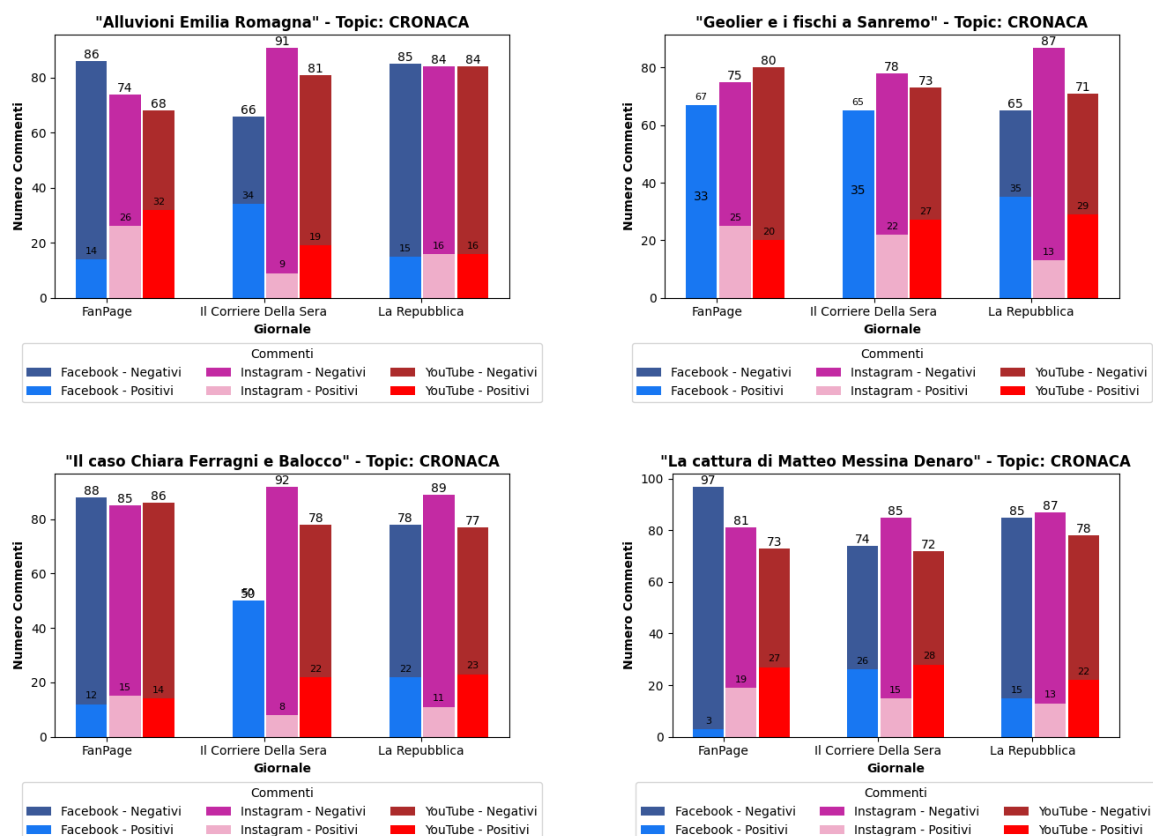


Figura 4.5: Numero di commenti positivi e negativi per ogni notizia per ogni social - CRONACA

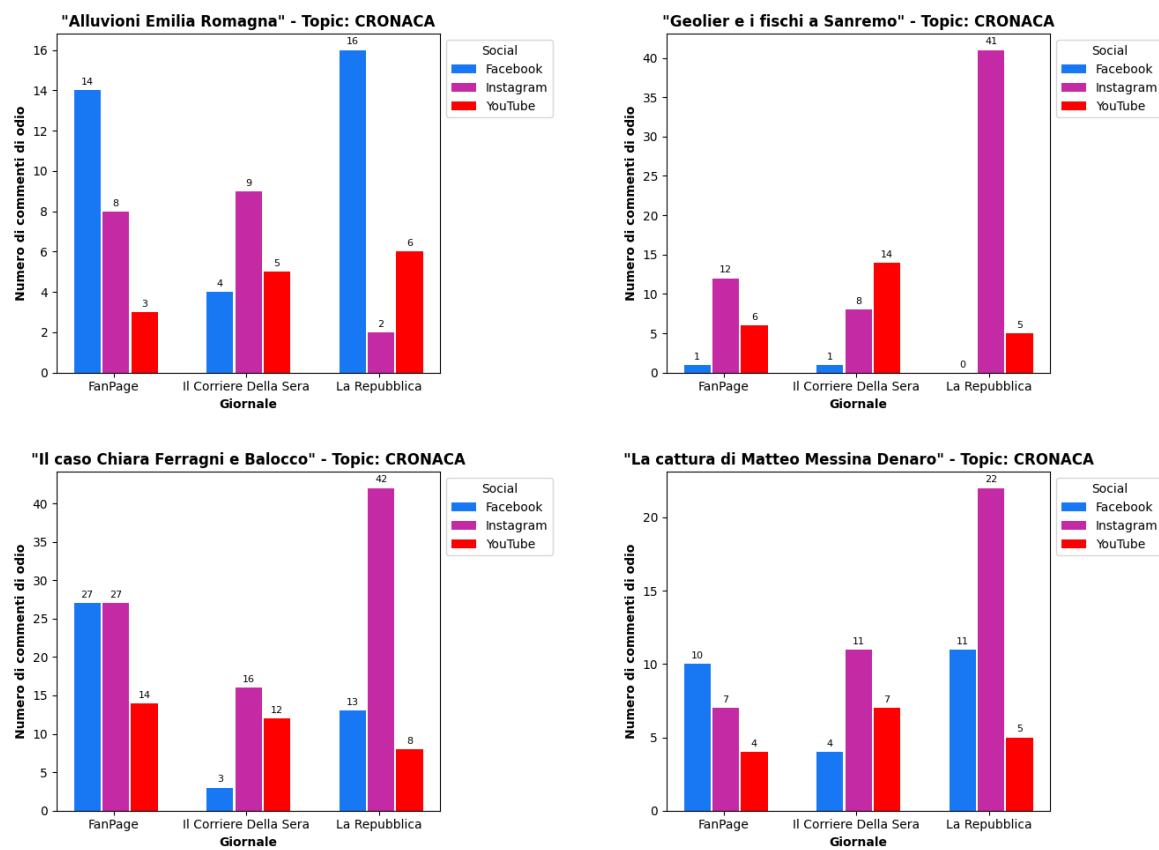


Figura 4.6: Numero di commenti di odio per ogni notizia per ogni social - CRONACA

4.2.3 Cronaca Nera

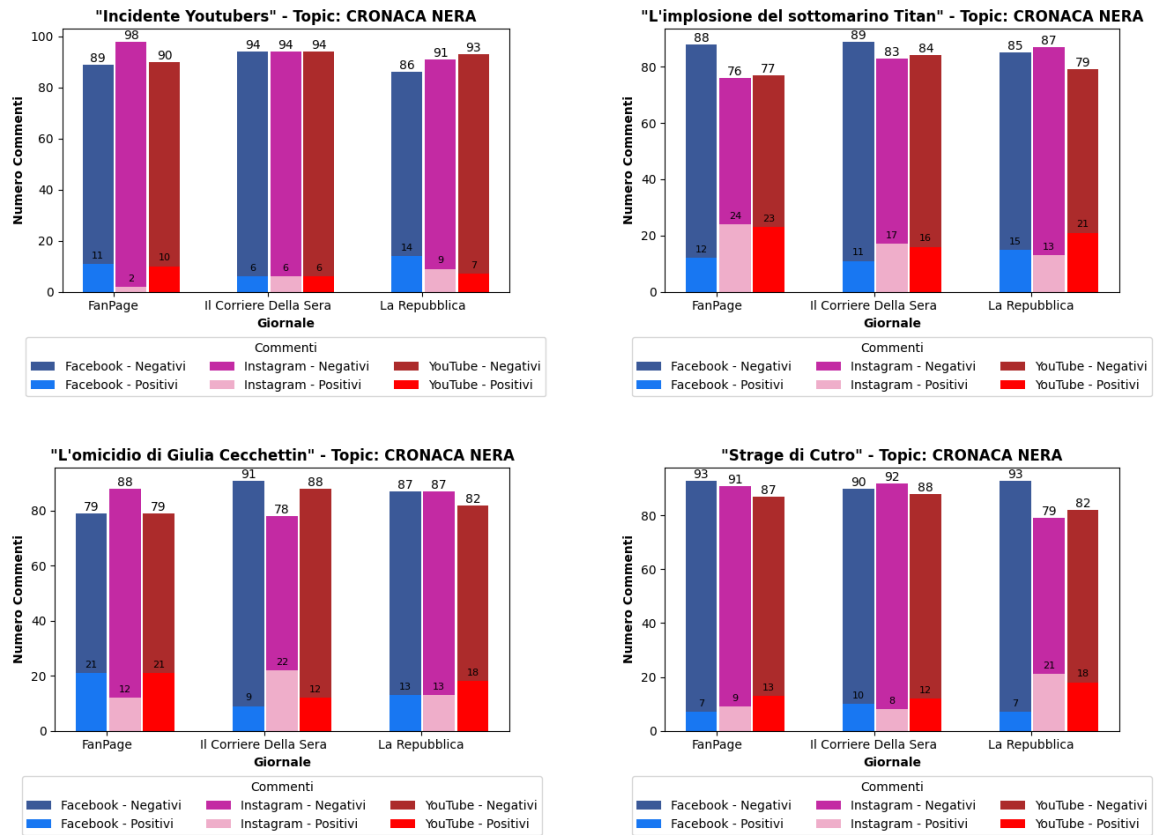


Figura 4.7: Numero di commenti positivi e negativi per ogni notizia per ogni social - CRONACA NERA

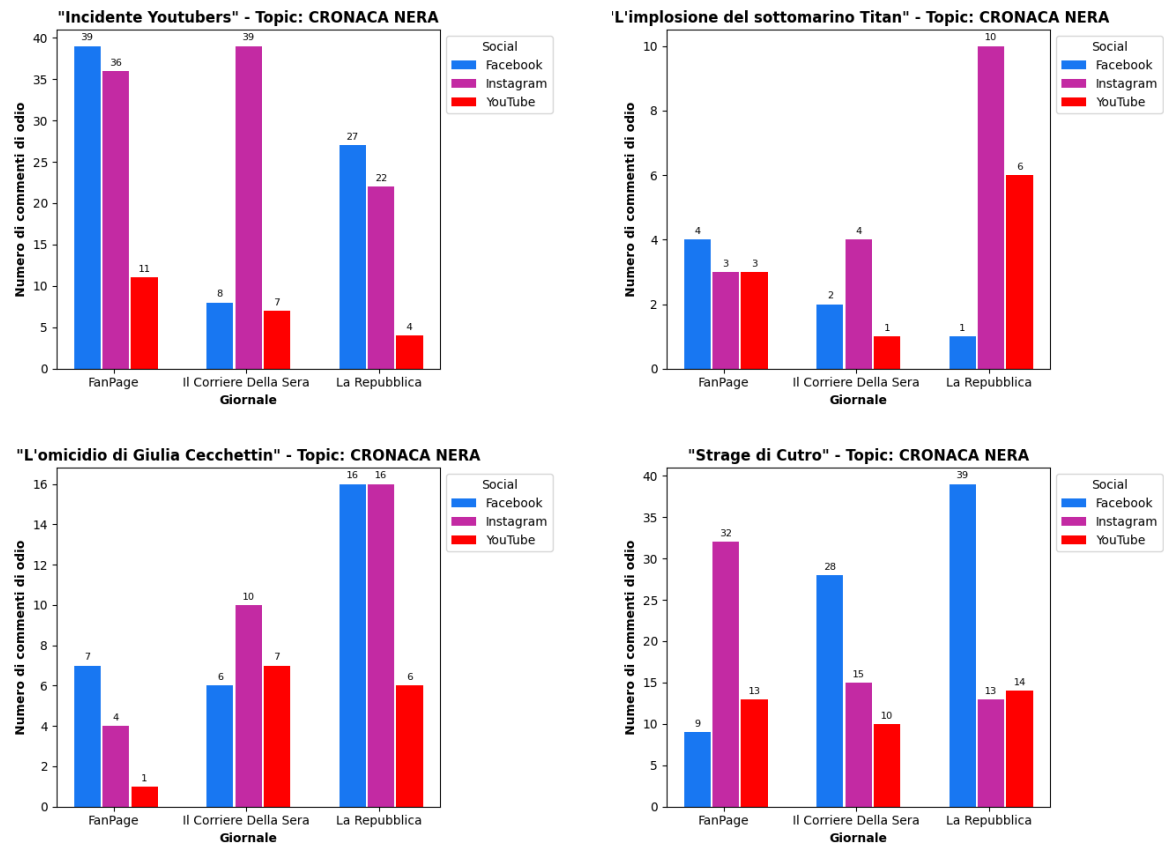


Figura 4.8: Numero di commenti di odio per ogni notizia per ogni social - CRONACA NERA

4.3 Analisi medie commenti SOCIAL delle testate

In questa sezione sono mostrate le medie di commenti positivi, negativi, e di odio per ogni testata giornalistica

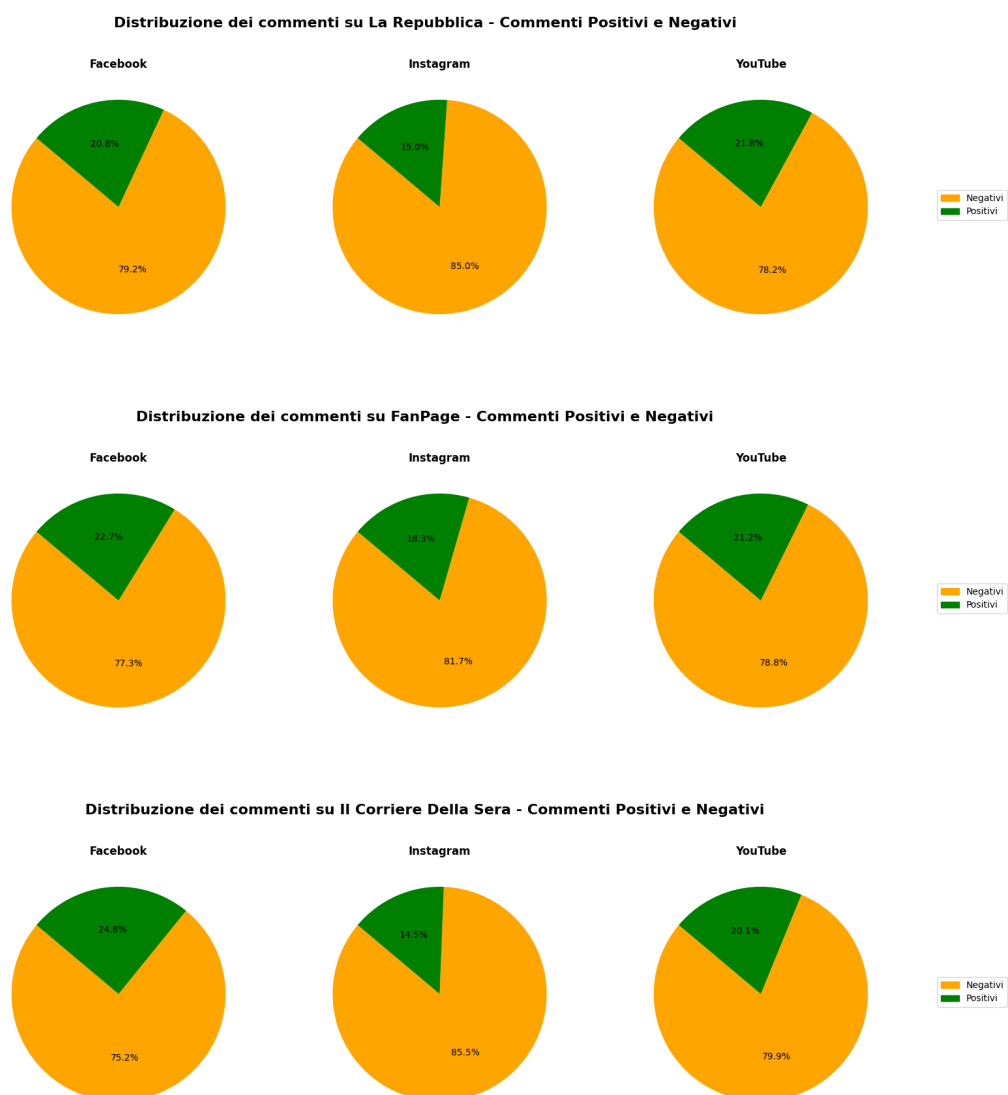


Figura 4.9: Media dei commenti positivi e negativi

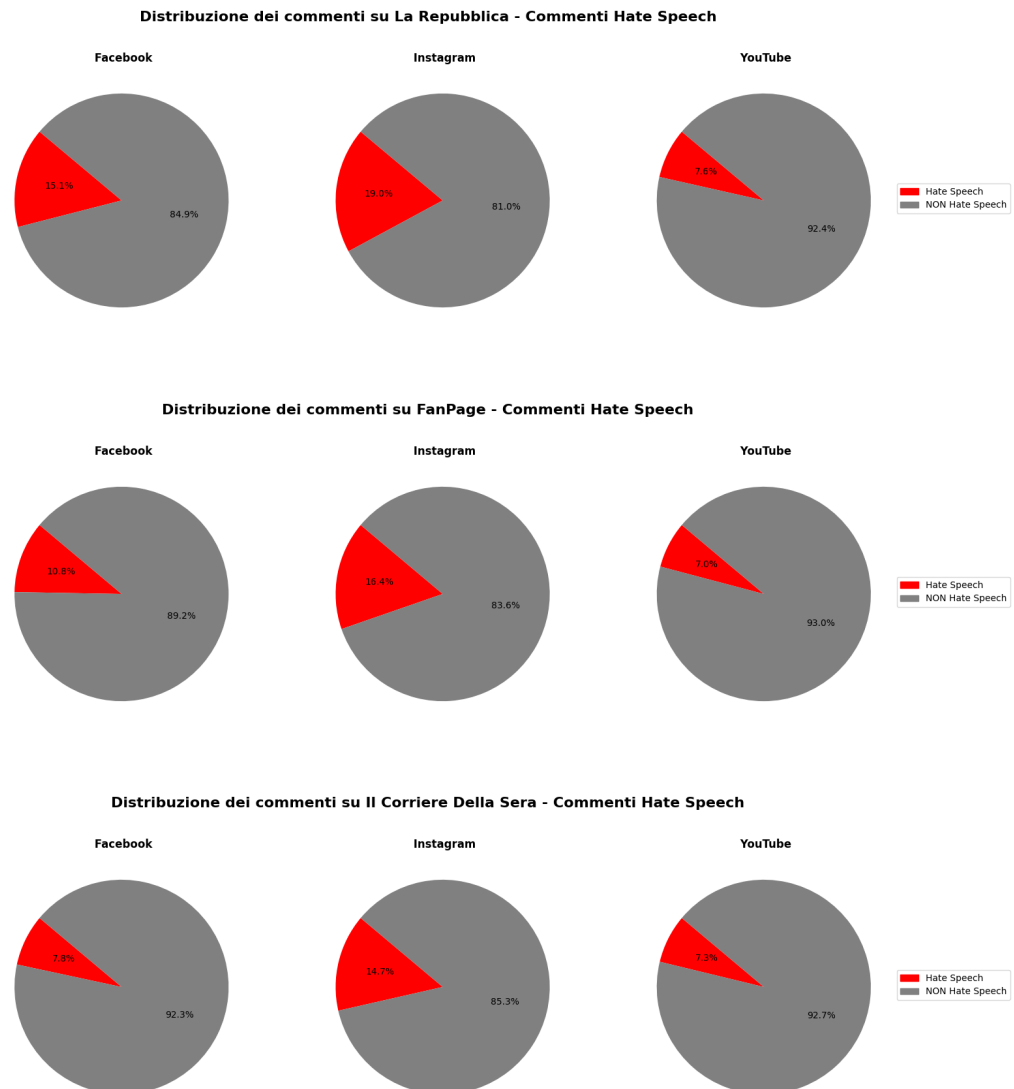


Figura 4.10: Media dei commenti hate speech e non

4.3.1 Tabella media commenti per topic

In questa sezione sono mostrate le medie di commenti positivi, negativi, e di odio per ogni accoppiata Topic-Social

Topic	Social	Totale	Negativi	Positivi	Hate
CRONACA	Facebook	1200	842 (70.2%)	358 (29.8%)	104 (8.7%)
CRONACA	Instagram	1200	1008 (84.0%)	192 (16.0%)	205 (17.1%)
CRONACA	Youtube	1200	921 (76.8%)	279 (23.2%)	89 (7.4%)
CRONACA NERA	Facebook	1200	1064 (88.7)	136 (11.3%)	186 (15.5%)
CRONACA NERA	Instagram	1200	1044 (87.0%)	156 (13.0%)	204 (17.0%)
CRONACA NERA	Youtube	1200	1023 (85.2%)	177 (14.8%)	83 (6.9%)
POLITICA	Facebook	1200	875 (72.9%)	325 (27.1%)	88 (7.3%)
POLITICA	Instagram	1200	974 (81.2%)	226 (18.8%)	192 (16.0%)
POLITICA	Youtube	1200	899 (74.9%)	301(25.1%)	91 (7.6%)

Tabella 4.1: Topic-social con le rispettive percentuali di commenti

4.3.2 Risultati analisi delle Testate Giornalistiche

In questa sezione sono mostrate le top terne di Giornale-Topic-Social con il numero di commenti negativi e di odio

Giornale	Topic	Social	Numero commenti negativi
Il Corriere Della Sera	CRONACA NERA	Facebook	364
Il Corriere Della Sera	CRONACA NERA	YouTube	354
FanPage	CRONACA NERA	Instagram	353
La Repubblica	CRONACA NERA	Facebook	351
FanPage	CRONACA NERA	Facebook	349
La Repubblica	CRONACA	Instagram	347

Tabella 4.2: Testate Giornalistiche con più commenti negativi per topic e social

Giornale	Topic	Social	Numero commenti di odio
La Repubblica	CRONACA	Instagram	107
La Repubblica	CRONACA NERA	Facebook	83
FanPage	CRONACA NERA	Instagram	75
Il Corriere Della Sera	CRONACA NERA	Instagram	68
FanPage	POLITICA	Instagram	68
Il Corriere Della Sera	POLITICA	Instagram	64

Tabella 4.3: Testate Giornalistiche con più commenti di odio per topic e social

4.3.3 Risultati analisi dei Social

In questa sezione sono mostrate le top 3 dei Social in base rispettivamente al numero di commenti negativi e di odio

Social	Numero commenti negativi
Instagram	3026
YouTube	2843
Facebook	2781

Tabella 4.4: Social con più commenti negativi

Social	Numero commenti di odio
Instagram	601
Facebook	378
YouTube	263

Tabella 4.5: Social con più commenti di odio

Dai dati presentati nella tabella 4.2, emerge che il mezzo di comunicazione utilizzato (video/post) non sembra avere un impatto significativo sul numero di commenti negativi, poiché tale numero risulta essere simile tra le piattaforme Facebook, YouTube e Instagram. Invece, per i commenti di odio lo strumento utilizzato ha un impatto, come confermato dalla tabella 4.5.

Inoltre dalla tabella 4.2 è evidente la predominanza del topic "Cronaca Nera" come una delle principali fonti di generazione dei commenti negativi.

Una situazione simile si riscontra anche per i commenti di odio, dove il topic "Cronaca Nera", insieme al topic "Cronaca", emerge come uno dei principali temi di sviluppo delle discussioni di odio. Vedi tabella 4.3.

Per quanto riguarda le testate giornalistiche ritroviamo più volte nelle prime posizioni Il Corriere Della Sera e La Repubblica per i commenti negativi e di odio su topic differenti e su social diversi. Mentre FanPage presenta un ambiente più tranquillo solo nei topic non presenti nelle tabelle 4.2 e 4.3.

Mentre per i Social, Instagram è quello che genera più commenti negati e di odio. Vedi tabelle 4.4 e 4.5.

Capitolo 5

Considerazioni finali

Nella realizzazione di questo elaborato, si sono presentate una serie di problematiche oggetto di discussione e analisi. Alcune di esse non sono state puntualmente risolte ma potrebbe essere interessante una riflessione anche in relazione a possibili sviluppi futuri.

5.1 Qualità del dataset

L'operazione di data cleaning effettuata, per quanto efficace, è particolarmente rudimentale. La rimozione di elementi che causano inutile entropia è sicuramente un buon approccio ma vanno fatte una serie di considerazioni del linguaggio social:

- **L'utilizzo di emoji e stickers:** abbiamo rilevato una grossa presenza di emoji che inevitabilmente causano l'espressione di un sentimento, di un umore. Nella nostra analisi le abbiamo eliminate per rendere il testo compatibile col modello, tuttavia non è scontata una grossa perdita d'informazione che potrebbe essere interessante considerare;
- **La presenza di bot e spam:** non è stata intrapresa alcun tipo di azione per pulire il dataset da bot che commentano i post o comunque spam indesiderato. Questo decisamente va a influire nell'analisi dei dati ma risulta ostico interpretare o meno questa tipologia di comportamento. I social, di fatto, implementano già un sistema anti-spam che nasconde i commenti. Tuttavia potrebbe non essere sufficiente.

5.1.1 Collaborazione tra modelli

Durante l'analisi dei commenti sui social, è emerso un particolare errore relativo alla classificazione degli stessi. Nonostante la classificazione del sentiment abbia evidenziato che alcuni commenti classificati come positivi, figuravano anche come commenti di odio. Questa discrepanza è mostrata nella figura 5.1.

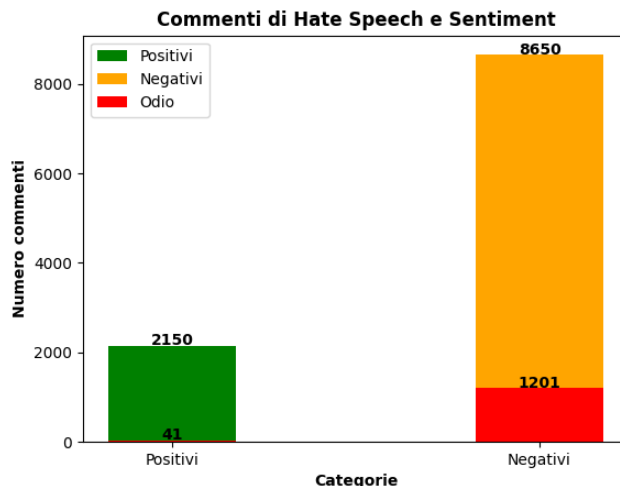


Figura 5.1: Enter Caption

È importante sottolineare che la classificazione del sentiment e dell'hate speech è stata eseguita attraverso l'impiego di modelli distinti. Questo errore sottolinea la complessità e la sfida nel valutare il contenuto testuale sui social media, dove la presenza di sottigliezze linguistiche e contestuali può rendere difficile una classificazione accurata. La necessità di affinare e migliorare i modelli di analisi del testo è cruciale per garantire una comprensione più completa e precisa dei contenuti pubblicati sui social media.

5.2 Strumenti di scraping massivi

Per questa nostra analisi abbiamo estratto i commenti con dei tool gratuiti di scraping. Questa tipologia di operazione però, può comportare il ban da parte dei social e, soprattutto, non consente una massiva cattura di commenti per creare un dataset più corposo. Diverse sono le soluzioni, con abbonamento o comunque dietro un pagamento, che potrebbero essere utilizzate. L'implementazione di uno scraper manuale sarebbe relegato solo all'analisi di una specifica finestra temporale in quanto, la struttura UI dei social insieme ad altri meccanismi (implementati dagli stessi proprio per evitare questo genere di operazioni) renderebbe necessario un aggiornamento continuo.

5.2.1 Selezione delle notizie priva di bias

La selezione manuale della notizia introduce inevitabilmente dei bias che vanno ad influenzare diversi aspetti dell'analisi. Ipotizzando di cercare i post social su una notizia, l'operazione di riportare per ogni social di ciascun testata giornalistica un post che tratti specificatamente lo stesso argomento è qualcosa che difficilmente risulta realizzabile dal momento che spesso sono le testate giornalistiche a dare una lettura diversa per la stessa notizia.

5.3 Sviluppi Futuri

Con una maggior qualità di dati e strumenti, ci sono diversi possibili sviluppi futuri di questo elaborato. Risolvendo il problema del bias della selezione notizie potrebbe essere interessante l'idea di espandere questo studio in varie direzioni:

- **Analisi del media** : Riuscire ad identificare gli effetti sulla tipologia di media (post, video, audio) nei confronti di una notizia e la relazione al sentiment dello stesso;
- **Previsione dell'hate speech** : Basandoci sul titolo della notizia, riuscire a predire la quantità di hate speech presente in un determinato social. *Feature* come tipologia di media, testata giornalistica e topic potrebbero essere relazionate a tale scopo;
- **Data Cleaning accurato**: Implementazione di un sistema di data cleaning del dataset, dando un peso all'emoji ed eliminando commenti off-topic e spam;
- **Singolo modello NLP**: Effettuare il Fine-Tuning di un unico modello in grado di classificare sia il sentiment che l'hate speech, in modo da evitare errori di discrepanza.

Bibliografia

- [1] Yoon Kim. “Convolutional Neural Networks for Sentence Classification”. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics.
- [2] Qiongxia Huang et al. “Deep Sentiment Representation Based on CNN and LSTM”. *2017 International Conference on Green Informatics (ICGI)* (2017), pp. 30–33. URL: <https://api.semanticscholar.org/CorpusID:36744331>.
- [3] Jacob Devlin et al. *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. 2019. arXiv: 1810.04805 [cs.CL].
- [4] Federico Bianchi, Debora Nozza e Dirk Hovy. “FEEL-IT: Emotion and Sentiment Classification for the Italian Language”. *Proceedings of the 11th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*. Association for Computational Linguistics, 2021.
- [5] Musixmatch Research. *Umberto: Musixmatch Research’s Hubert - a BERT-like model for Natural Language Understanding tasks*. <https://github.com/musixmatchresearch/umberto>.
- [6] Petra Kralj Novak et al. “Handling Disagreement in Hate Speech Modelling”. *Communications in Computer and Information Science*. Springer International Publishing, 2022. ISBN: 9783031089749. DOI: 10.1007/978-3-031-08974-9_54. URL: http://dx.doi.org/10.1007/978-3-031-08974-9_54.
- [7] Marco Polignano et al. “ALBERTo: Italian BERT Language Understanding Model for NLP Challenging Tasks Based on Tweets”. *Italian Conference on Computational Linguistics*. 2019. URL: <https://api.semanticscholar.org/CorpusID:204914950>.
- [8] Dr Simge Andi. *How and why do consumers access news on social media?* <https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2021/how-and-why-do-consumers-access-news-social-media>. 2021.
- [9] Lewis Tunstall et al. “Efficient Few-Shot Learning Without Prompts” (2022). DOI: 10.48550/ARXIV.2209.11055.

