

# **SENTIMENT & HATE SPEECH ANALYSIS: TESTATE GIORNALISTICHE SUI SOCIAL MEDIA**

**PROFESSORESSA  
DELFINA MALANDRINO**

**PROGETTO RETI GEOGRAFICHE**

**STUDENTI  
ALESSIO MADDALUNO  
ROBERTO PISCOPO**

# CONTENUTI

- 
- 01** INTRODUZIONE
  - 02** OBIETTIVO
  - 03** MOTIVAZIONI
  - 04** METODOLOGIA
  - 05** RACCOLTA COMMENTI
  - 06** CLASSIFICAZIONE DEI COMMENTI
  - 07** RISULTATI
  - 08** CONCLUSIONI E SVILUPPI FUTURI

# INTRODUZIONE

I social network hanno registrato un impressionante tasso di crescita esponenziale nel numero di utenti attivi mensili e nel guadagno finanziario.

Negli ultimi anni, sono diventati un terreno fertile per la diffusione di una varietà incredibile di comportamenti discutibili, soprattutto relazione ai discorsi d'odio.



# INTRODUZIONE

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 metologia.

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.



# OBIETTIVO

## Obiettivo n° 1

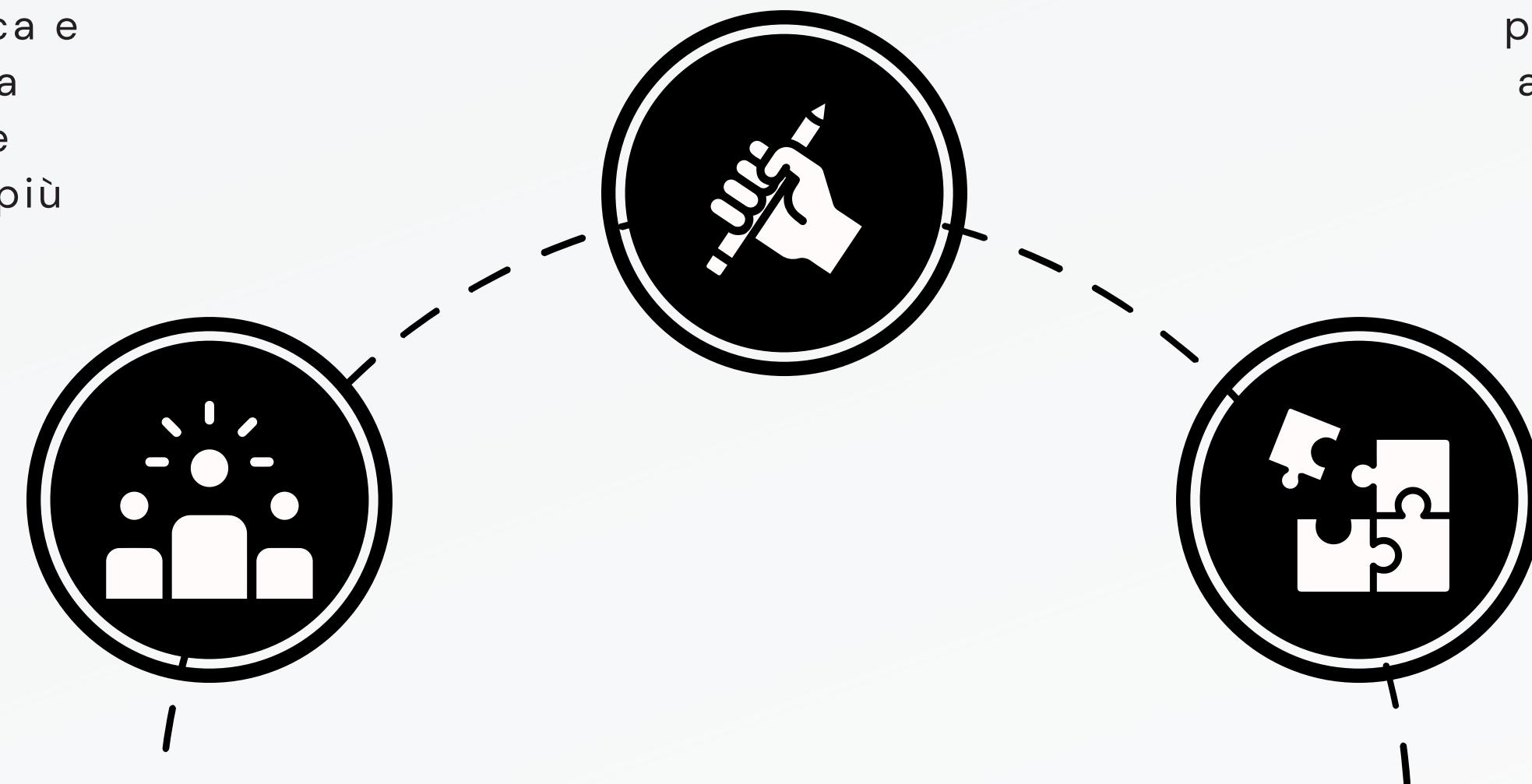
Analizzare i commenti associati a video e post di notizie su vari argomenti di Politica, Cronaca e Cronaca Nera delle testate giornalistiche più diffuse

## Obiettivo n° 2

Identificare quelle che generano un maggior numero di commenti di negativi e di odio (hate speech)

## Obiettivo n° 3

Analizzare le differenze le varie piattaforme per determinare quale presenta il livello più alto di hate speech.



# MOTIVAZIONI

01

Forte interesse nel comprendere il ruolo delle testate nel promuovere o contrastare l'odio online

02

Comprendere se il media influenza o meno la diffusione di commenti negativi e di odio.



# METODOLOGIA

**01**

SCELTA DEI SOCIAL

**02**

SCELTA DEI TOPIC

**03**

SCELTA DELLE TESTATE GIORNALISTICHE

**04**

SCELTA DELLE NOTIZIE

# SCELTA DEI SOCIAL



Facebook  
Post, Video,  
Immagini



Instagram  
Immagini, Video,  
Reels



Youtube  
Video, Shorts

# SCELTA DEI TOPIC



Il forte coinvolgimento emotivo istiga l'utente a reagire all'accaduto, commentandolo spesso senza filtri, soprattutto di fronte ad un caso d'omicidio.

**CRONACA NERA**



Le discussioni sulla politica sui social media variano da opinioni costruttive a manifestazioni di odio verso i politici. Le notizie sono spesso controverse e generano molti commenti.

**POLITICA**



La cronaca generalista offre un consistente bacino di notizie di varia natura, che invogliano diverse categorie di utenti a commentare

**CRONACA**

# **SCELTA DELLE TESTATE**

**fanpage.it**

***CORRIERE DELLA SERA***

**la Repubblica**

# SCELTA DELLE NOTIZIE

- *Abbiamo selezionato 4 notizie rappresentative per ciascun topic individuato.*
- *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.*

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

# RACCOLTA DEI COMMENTI

*I commenti sono stati estratti tramite l'utilizzo di Esuit, un insieme di estensioni per Chrome utilizzate principalmente per effettuare analisi da Social Media Managers*

## **DATASET CREATO:**

Campo	Descrizione
Social	Social da cui è stato estratto il commento
Giornale	Testata giornalistica (La Repubblica / Corriere della Sera /Fanpage.it)
Notizia	Titolo della notizia
Commento	Commento Estratto
Topic	Tipologia della notizia (Manualmente estratto, contiene i valori: POLITICA, CRONACA, CONACA NERA)

# DATA CLEANING

## EMOTICON

Rimozione di emoticon e simboli unicode, in quanto non sono efficaci per il modello utilizzato

## TAG @

Rimozione di riferimenti agli utenti preceduti dal simbolo "@" all'interno del testo

## LINK

Rimozione di eventuali URL o collegamenti web perchè potrebbero contenere parole che potrebbero influenzare il modello;

## CARATTERI SPECIALI

Rimozioni caratteri, come le virgolette doppie ("), e potrebbero essere considerati come rumore durante l'analisi

# CLASSIFICAZIONE SENTIMENT & HATE SPEECH

## MODELLI ML

Due modelli di machine learning basati sull'intelligenza Artificiale  
BERT di Google:

- Feel-it
- Hate-speech-it

Entrambi i modelli, disponibili su Hugging Face, sono pre-addestrati per compiti  
di Sentiment Analysis e Hate Speech detection.

I due modelli sono stati utilizzati come libreria di python.



**Hugging Face**



# MODELLI NEL DETTAGLIO

## FEEL-IT

classifica l'input in due categorie:

- positivo
- negativo.

## HATE-SPEECH-IT

classifica l'input in 4 categorie:

- assente
- inappropriato
- offensivo
- violento

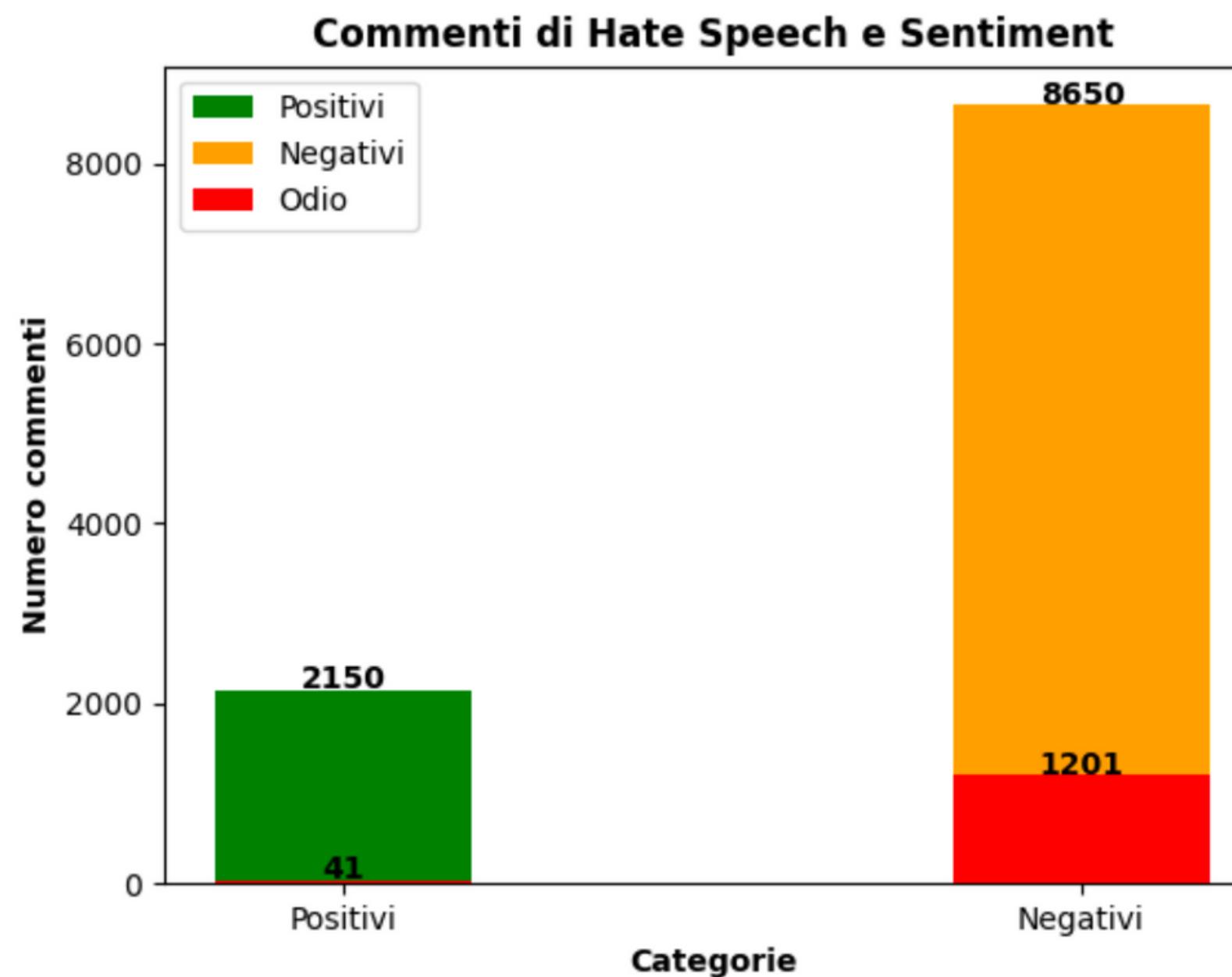
## CAMPI AGGIUNTI AL DATASET

Campo	Descrizione
Sentiment	Categorizzazione del sentimento del commento, ottenuto dal modello di machine learning.
Hate Speech	Categorizzazione del sentimento del commento, ottenuto dal modello di machine learning

# RISULTATI

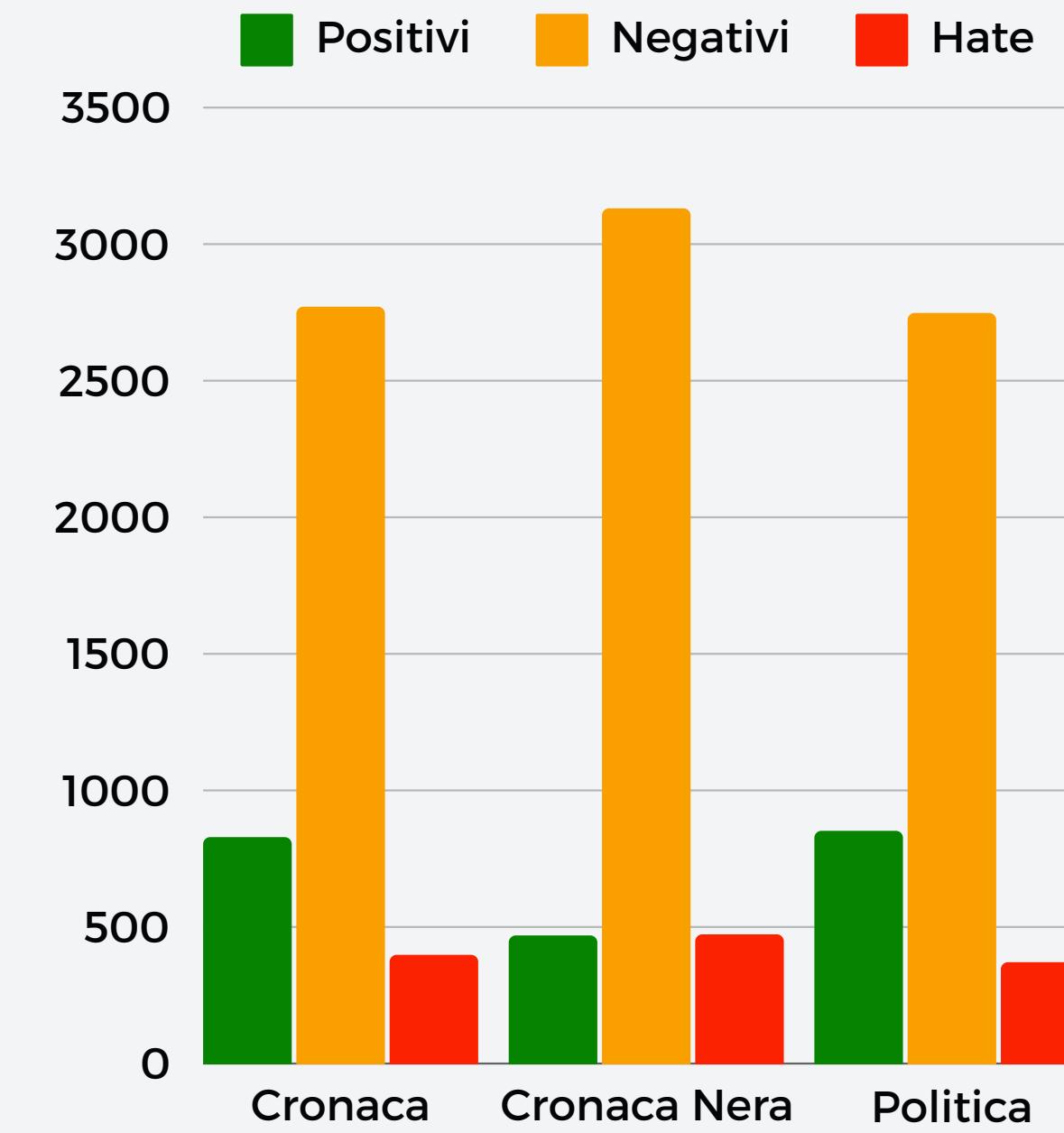
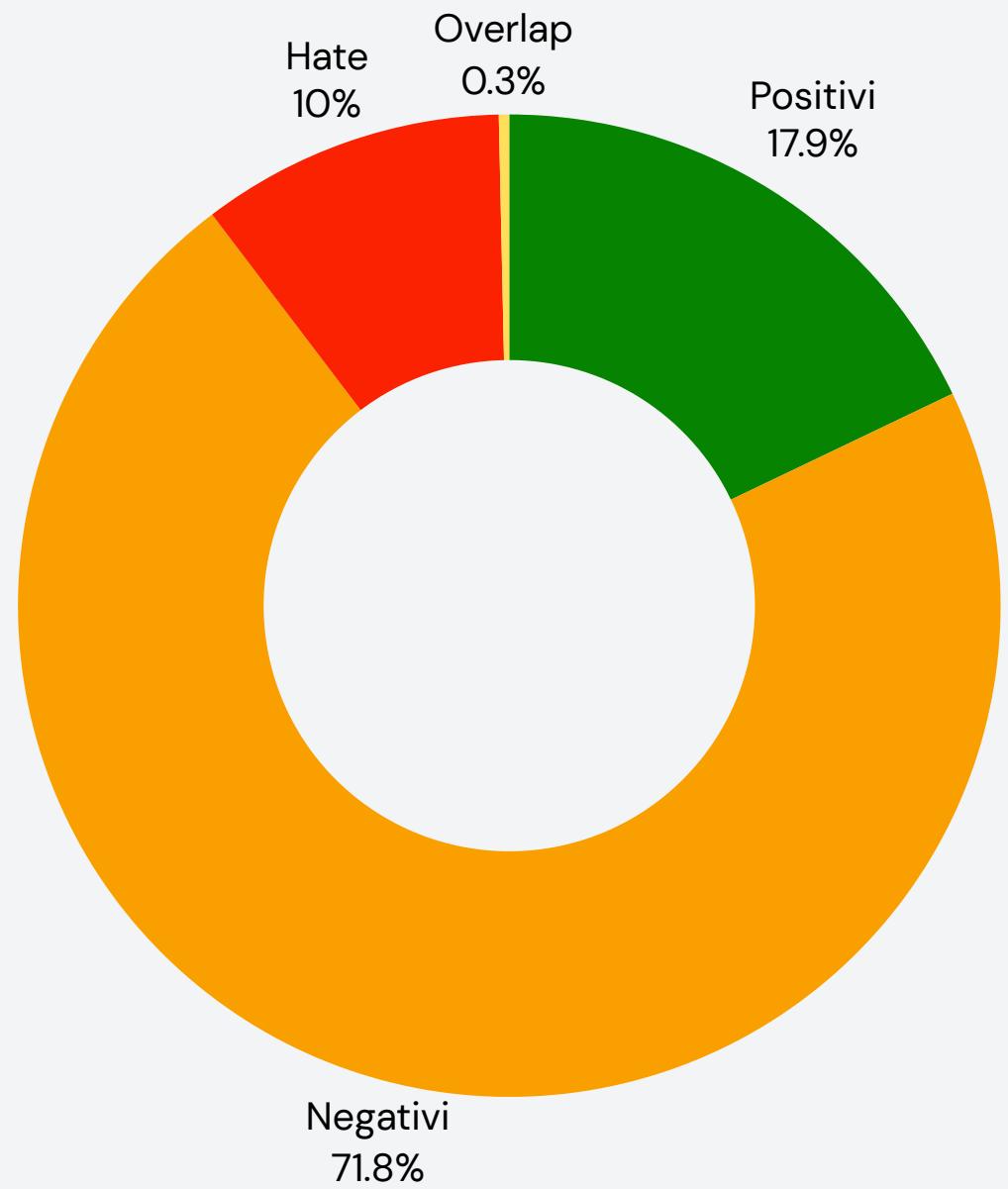


NUMERO TOTALE DI COMMENTI SUDDIVISI PER CATEGORIA



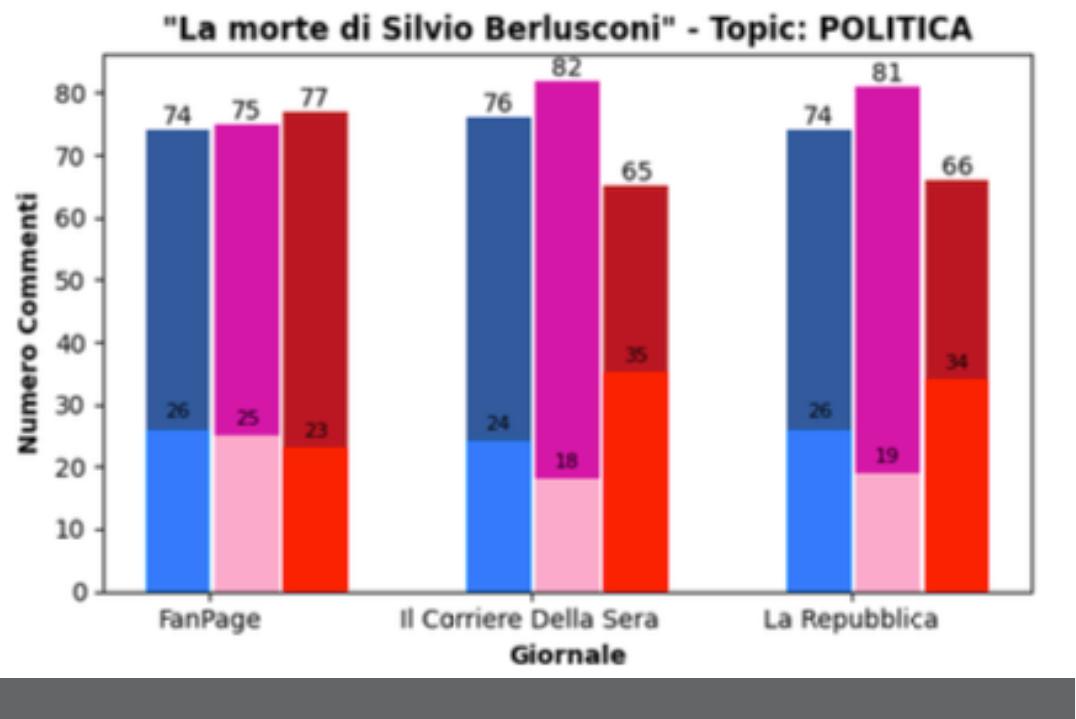
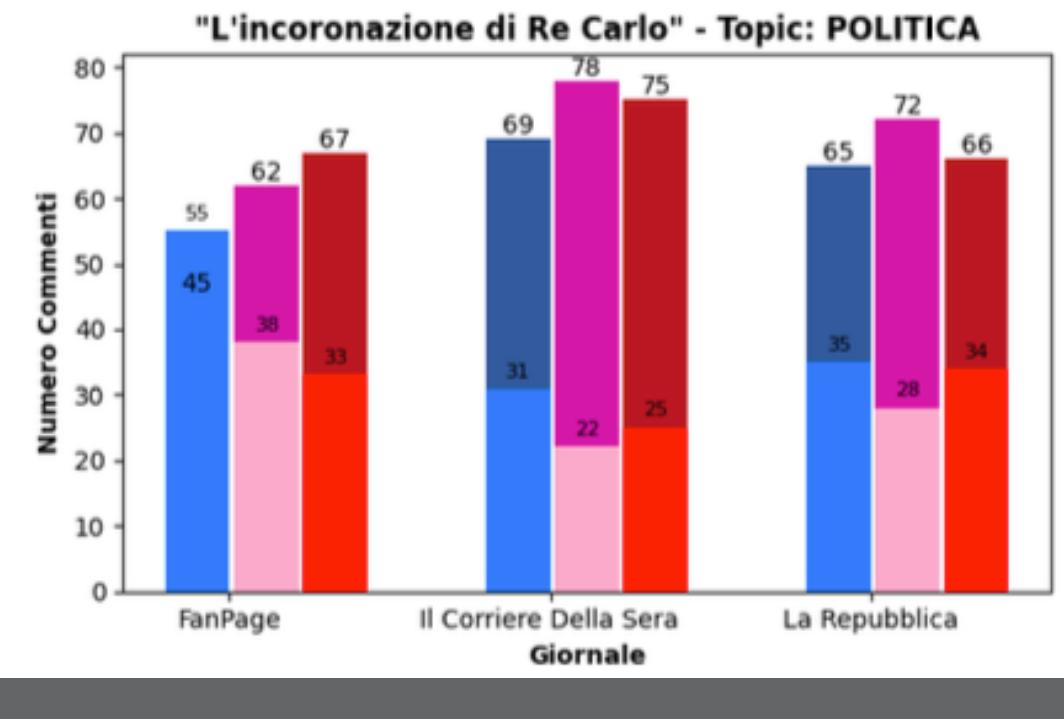
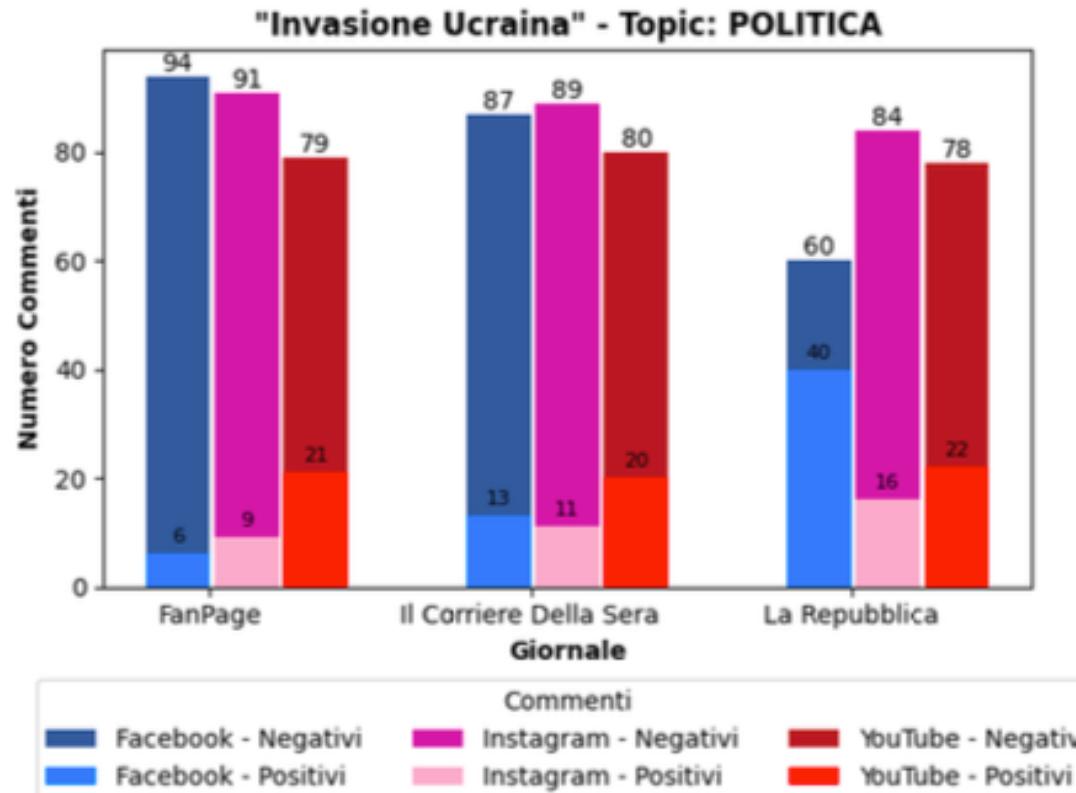
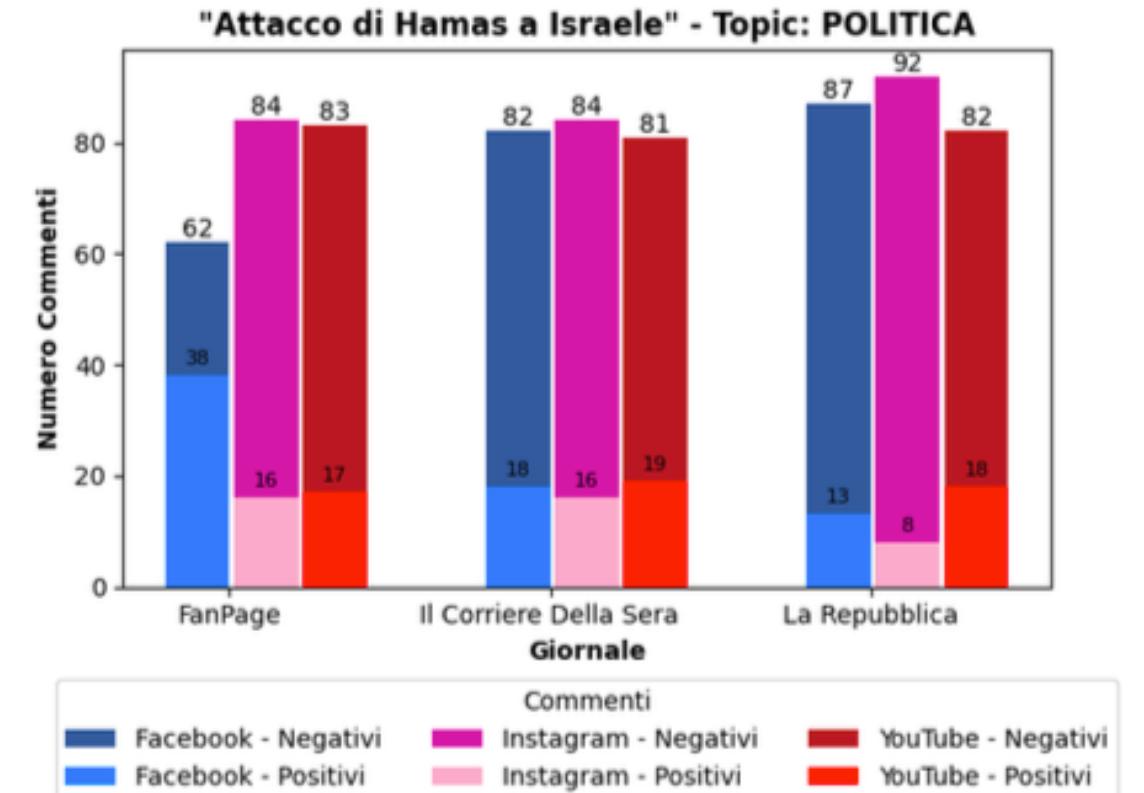
- 10800 COMMENTI ESTRATTI
  - 8650 COMMENTI **NEGATIVI** DI CUI 1201 DI **ODIO**
  - 2150 COMMENTI **POSITIVI** DI CUI 41 DI **ODIO**

# DISTRIBUZIONE DEI COMMENTI



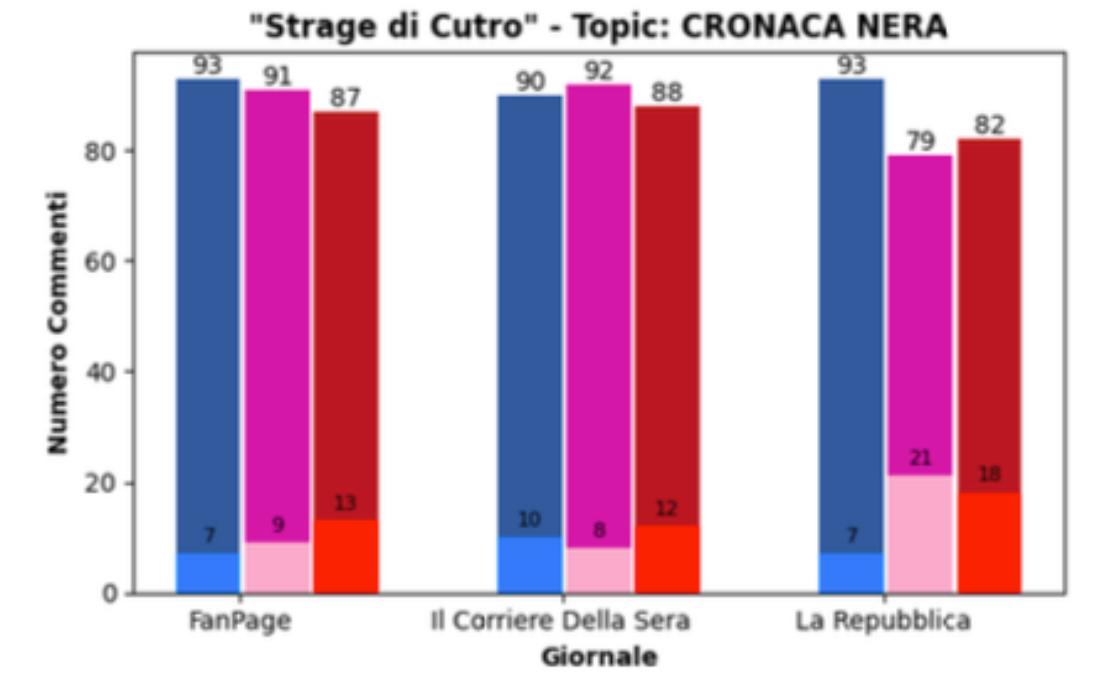
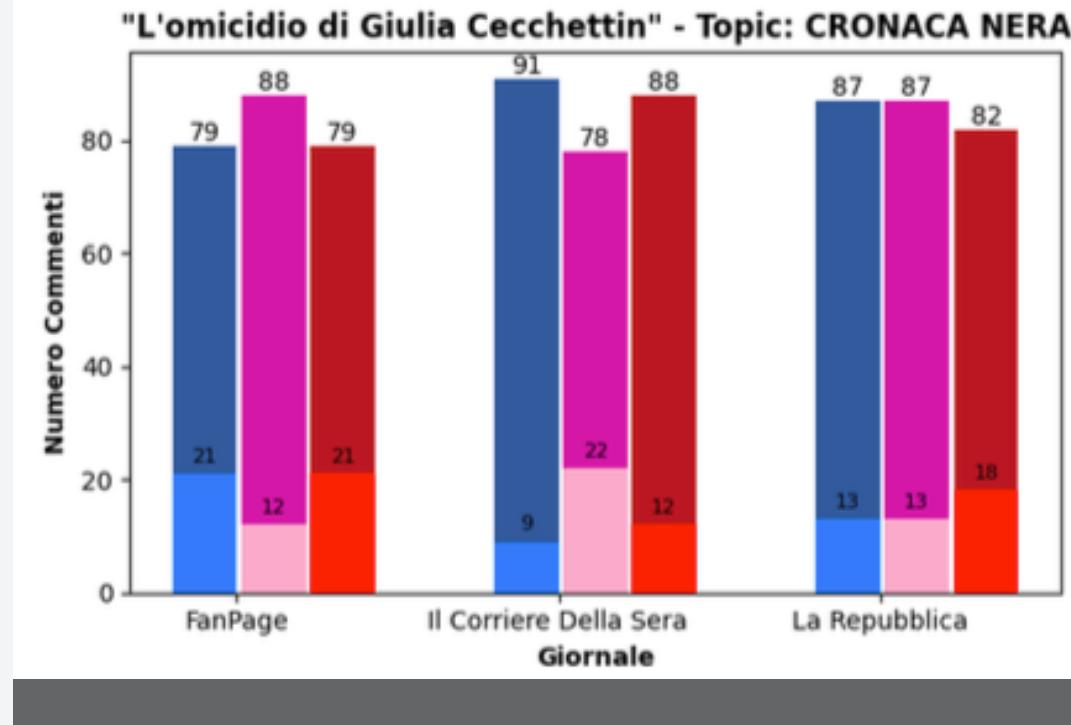
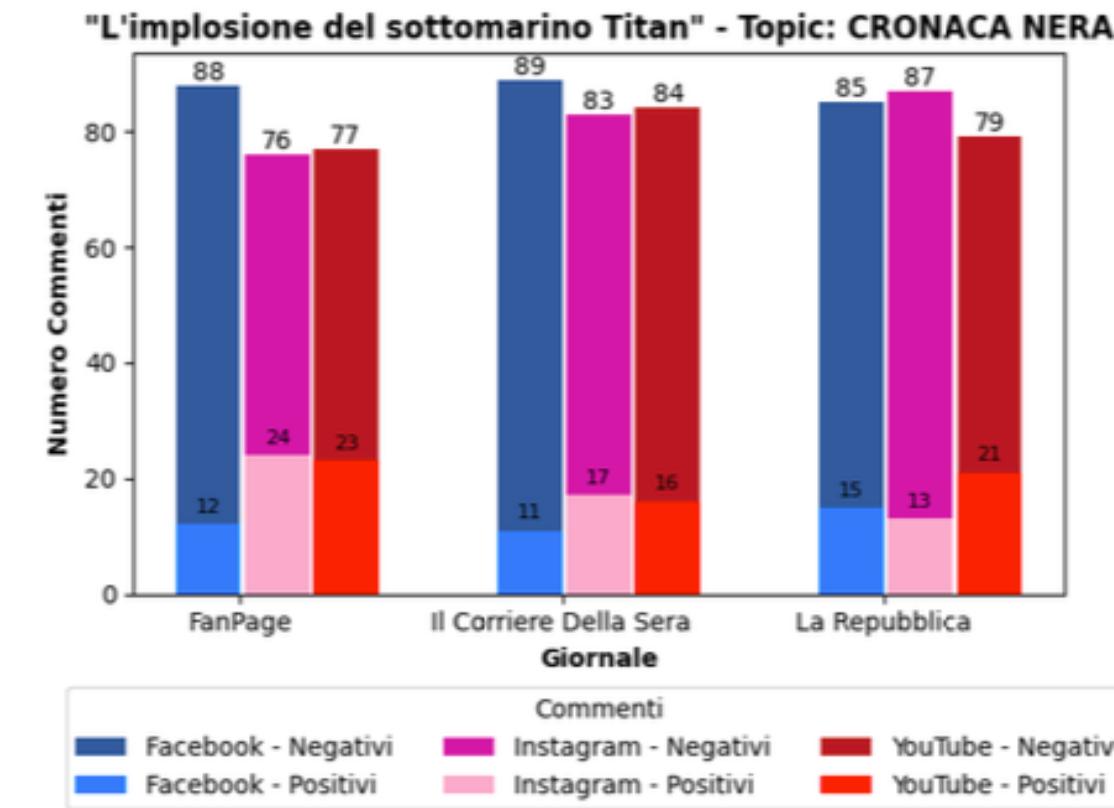
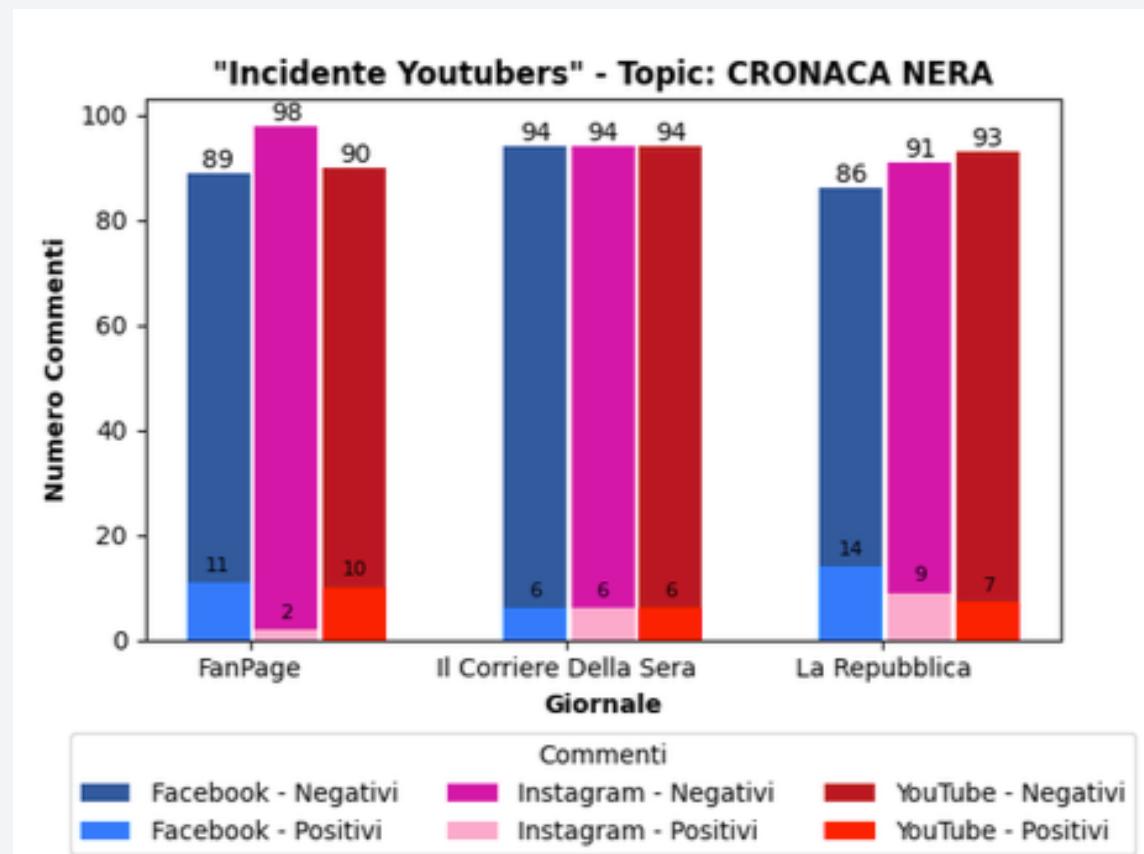
# ANALISI PER TOPIC

POLITICA



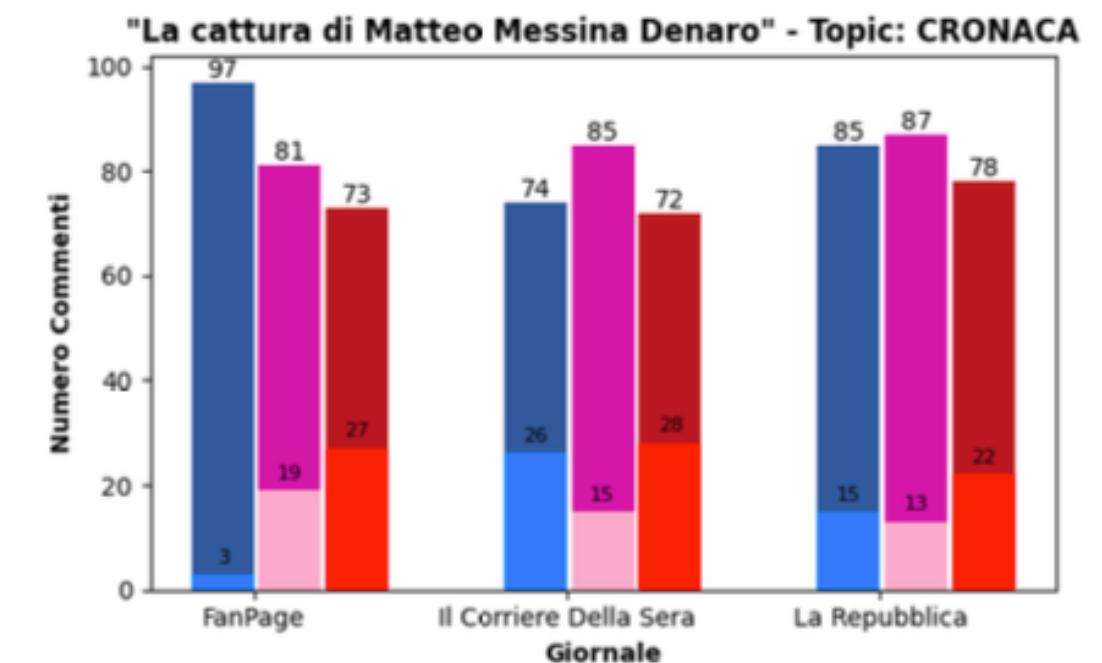
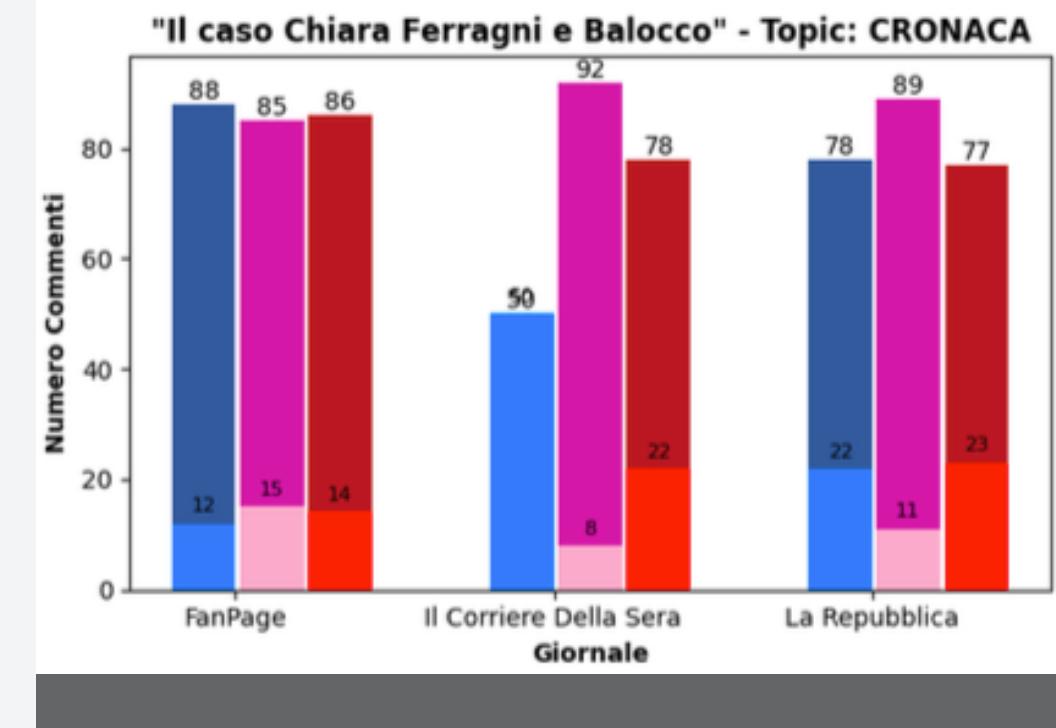
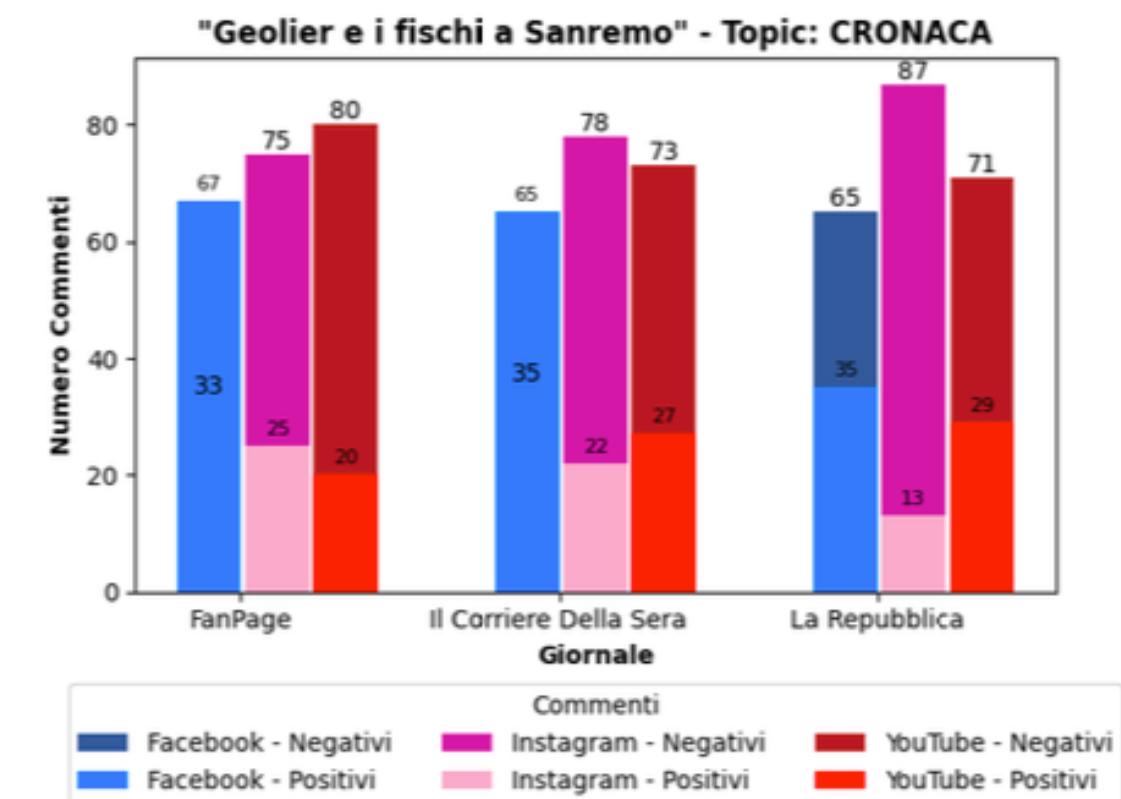
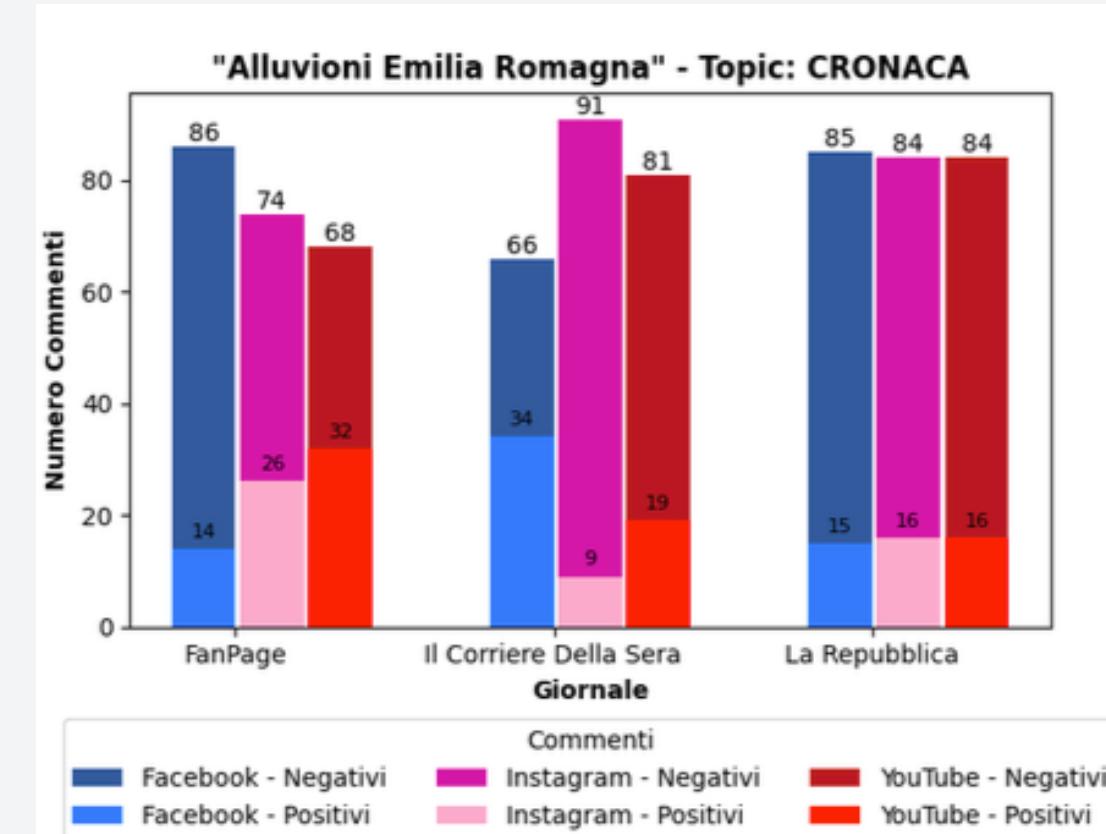
# ANALISI PER TOPIC

CRONACA NERA



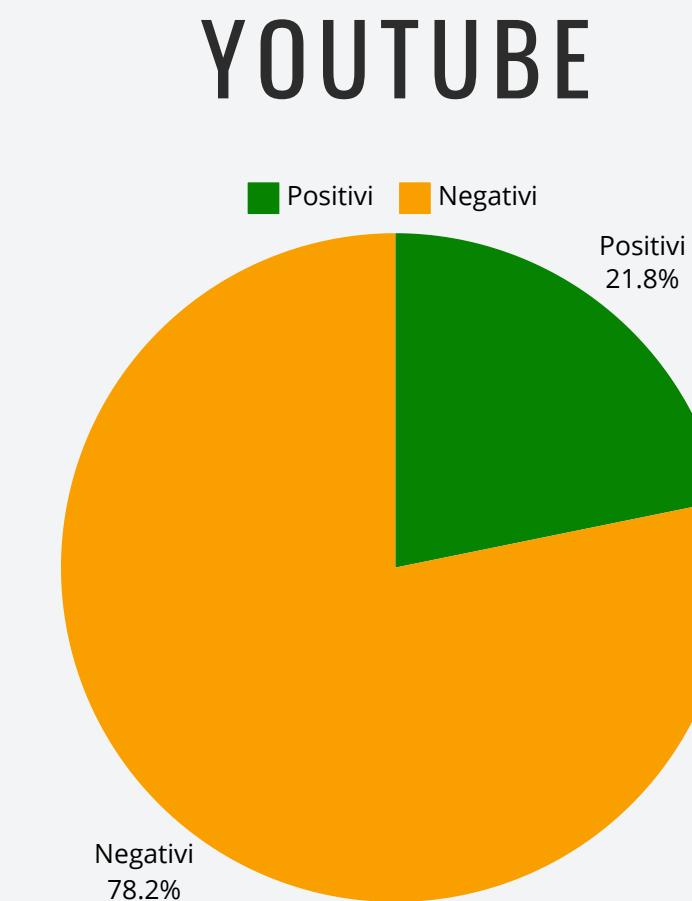
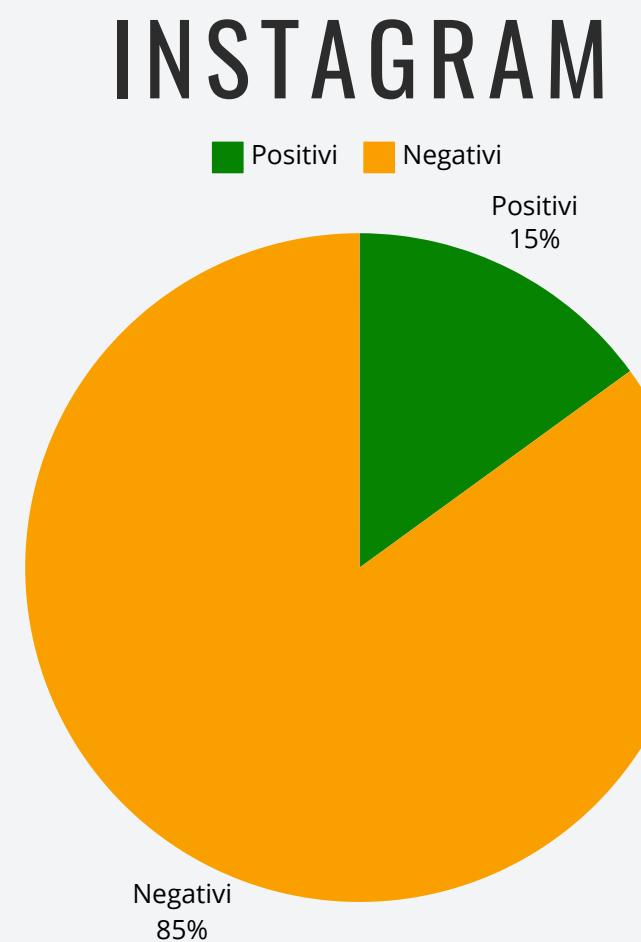
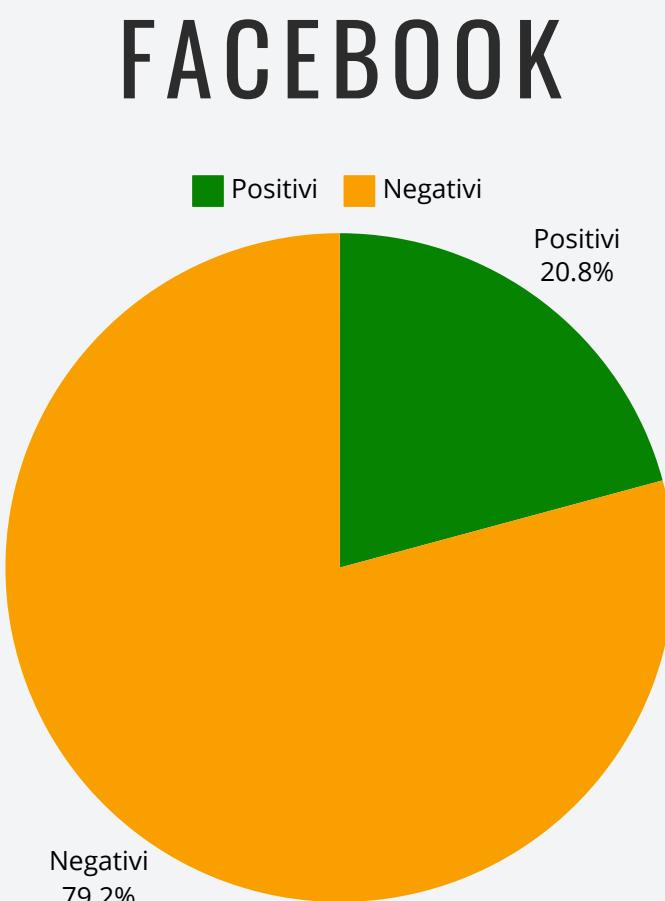
# ANALISI PER TOPIC

CRONACA



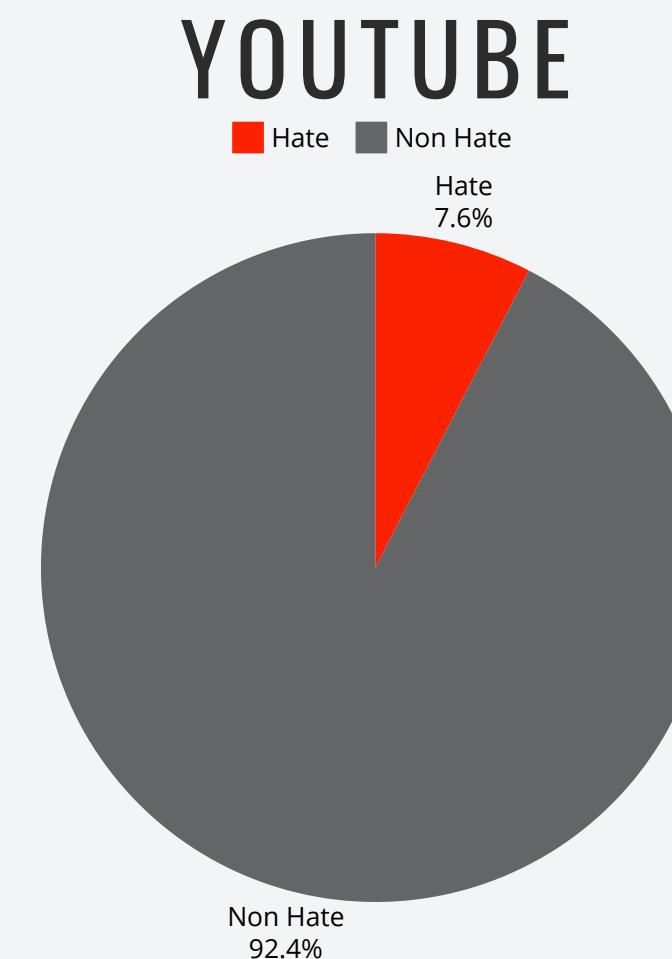
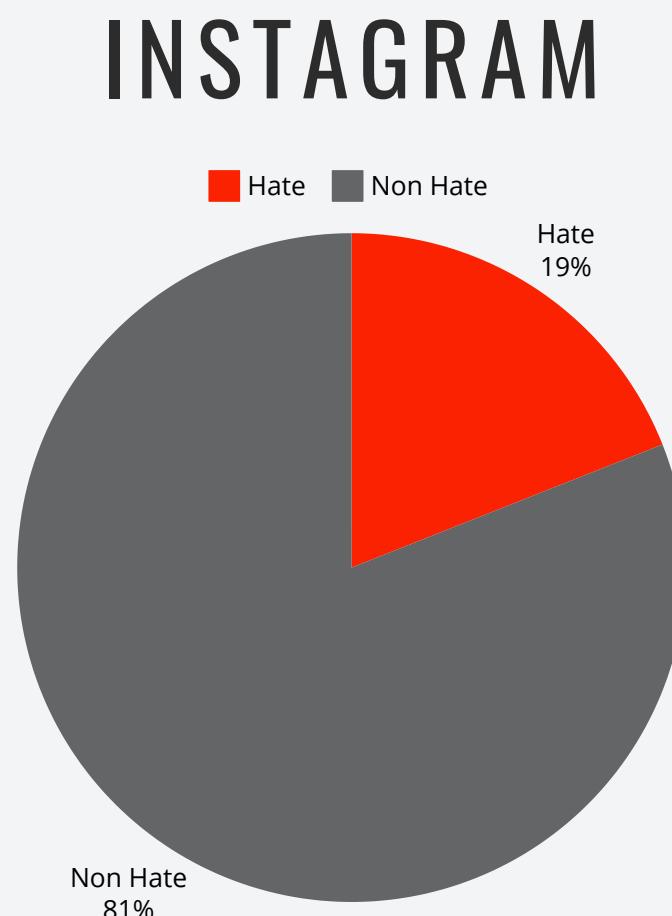
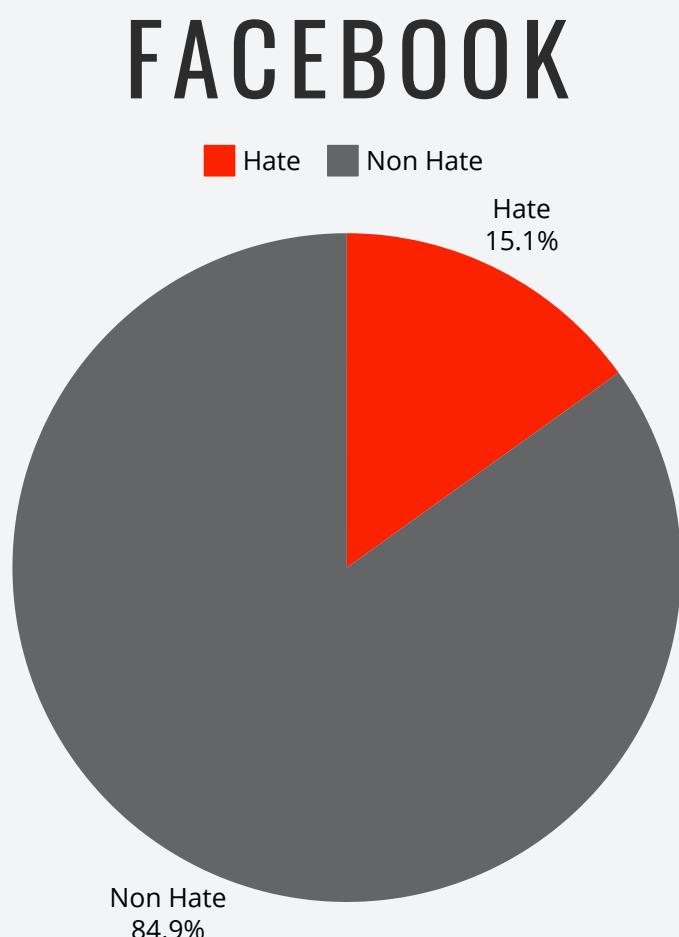
# MEDIE COMMENTI SOCIAL PER TESTATA

## la Repubblica



# MEDIE COMMENTI SOCIAL PER TESTATA

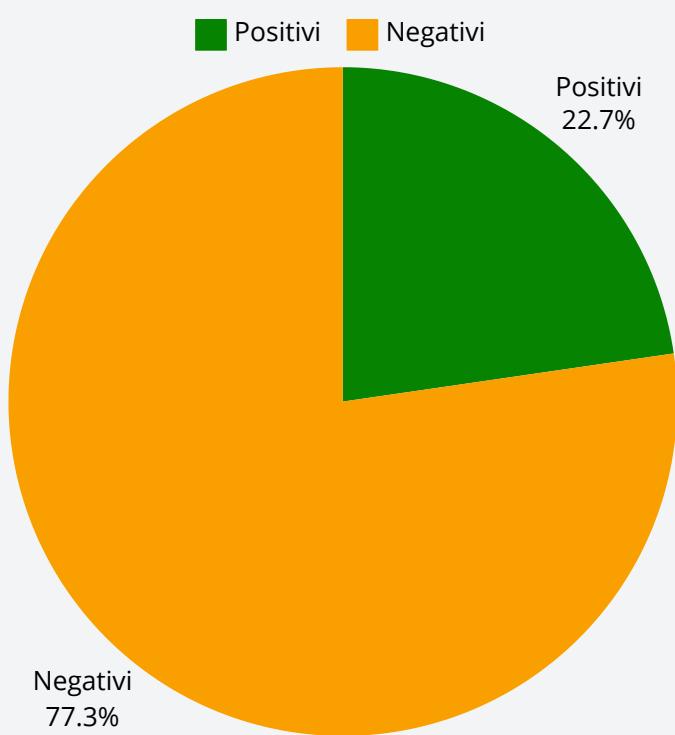
## la Repubblica



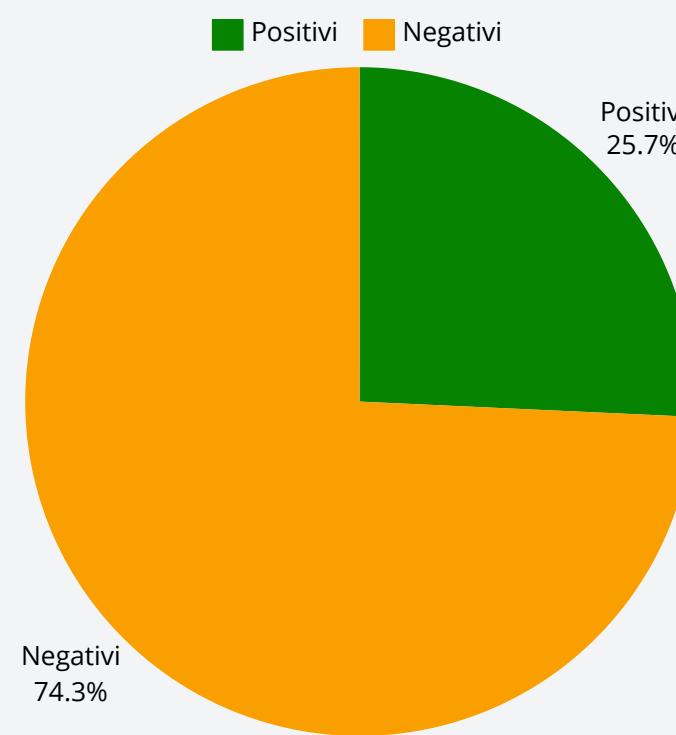
# MEDIE COMMENTI SOCIAL PER TESTATA

## fanpage.it

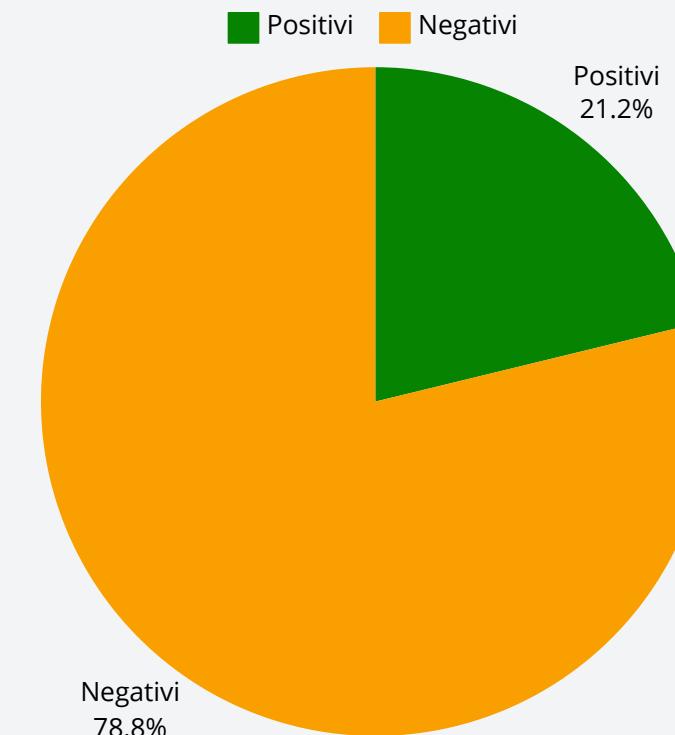
### FACEBOOK



### INSTAGRAM

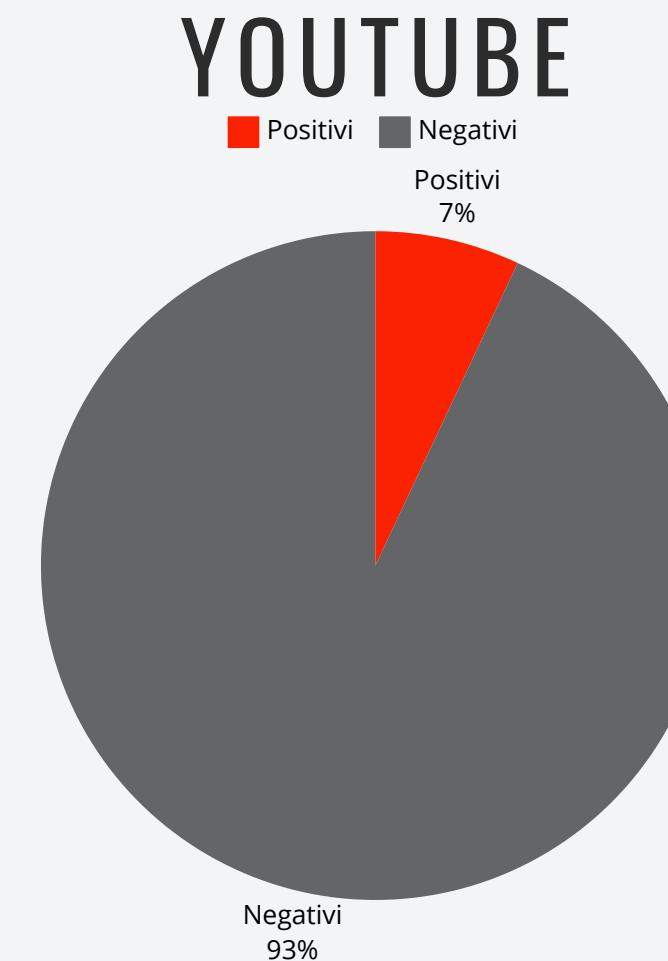
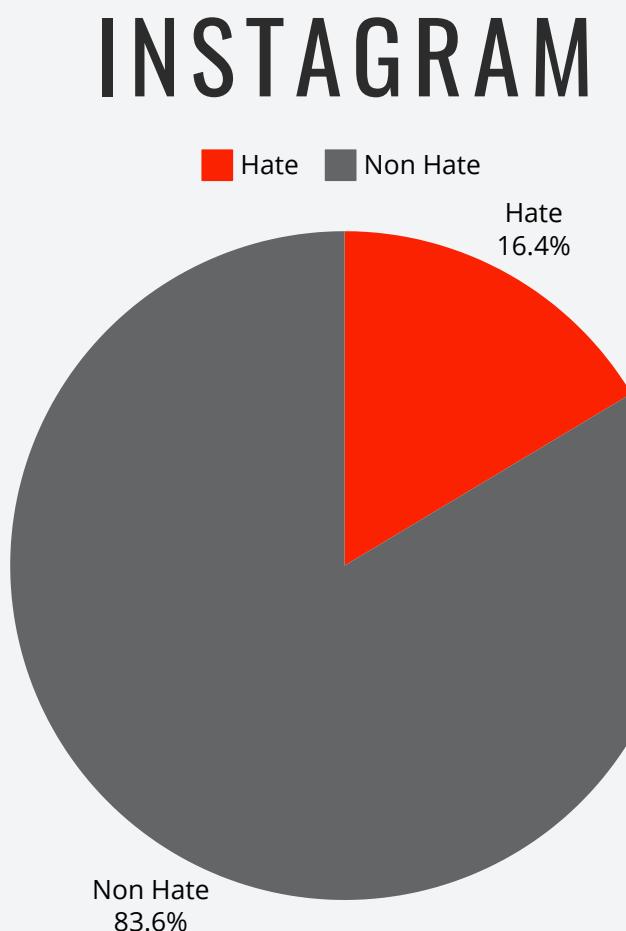
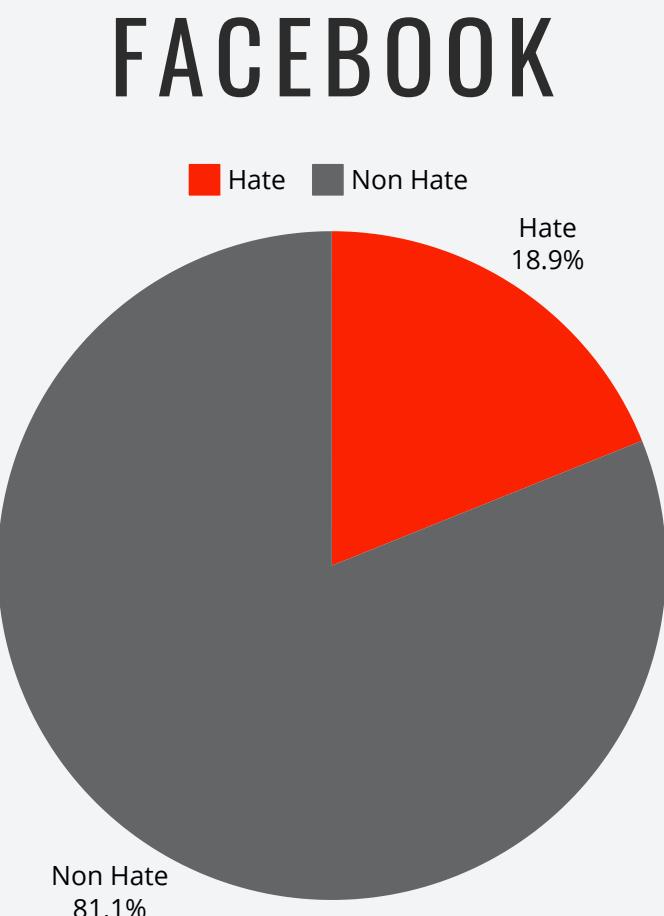


### YOUTUBE



# MEDIE COMMENTI SOCIAL PER TESTATA

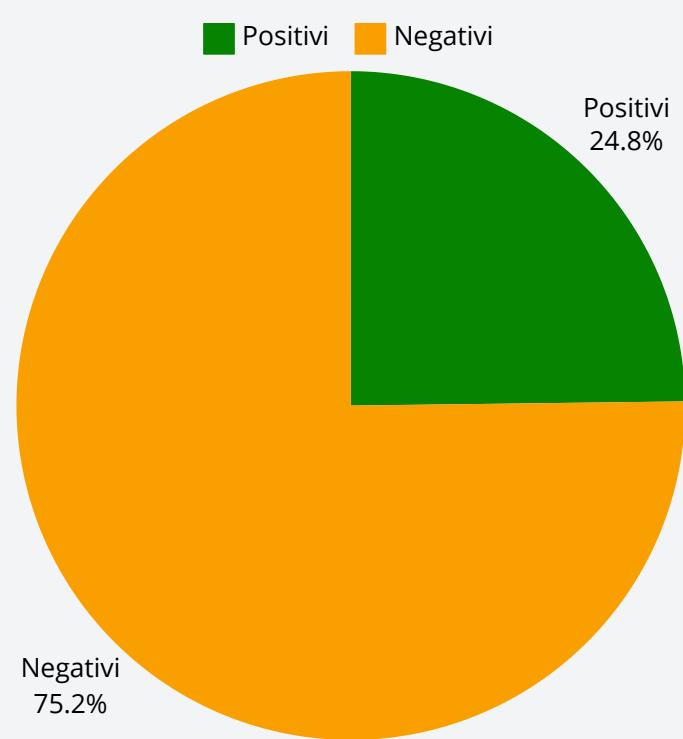
## fanpage.it



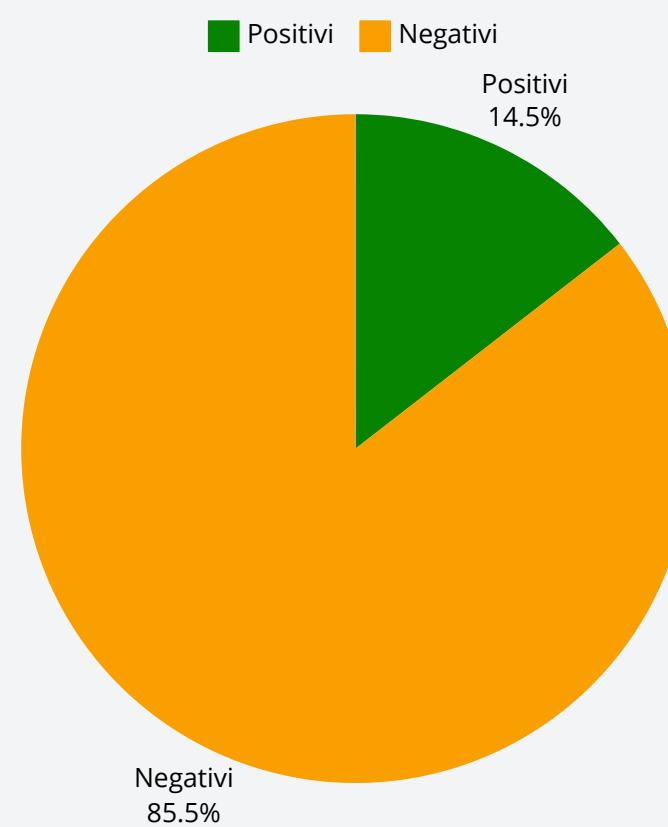
# MEDIE COMMENTI SOCIAL PER TESTATA

## ***CORRIERE DELLA SERA***

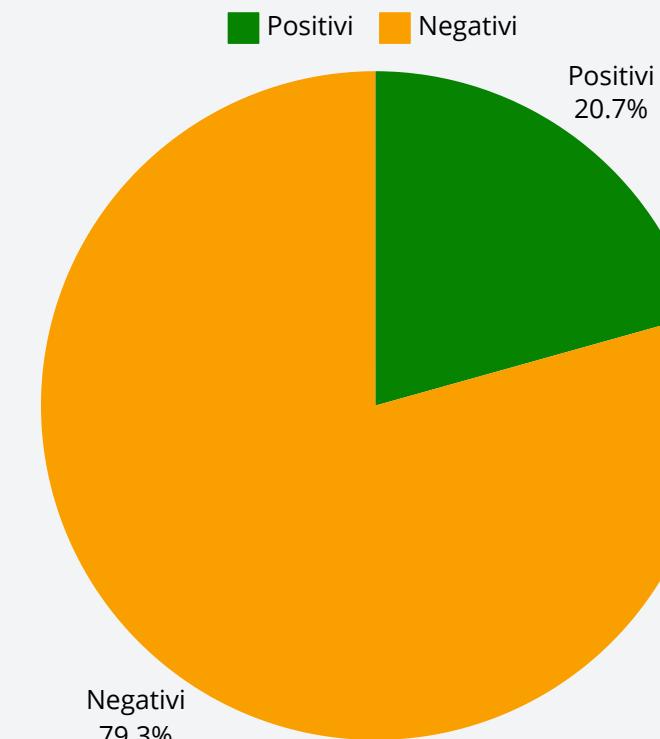
### FACEBOOK



### INSTAGRAM

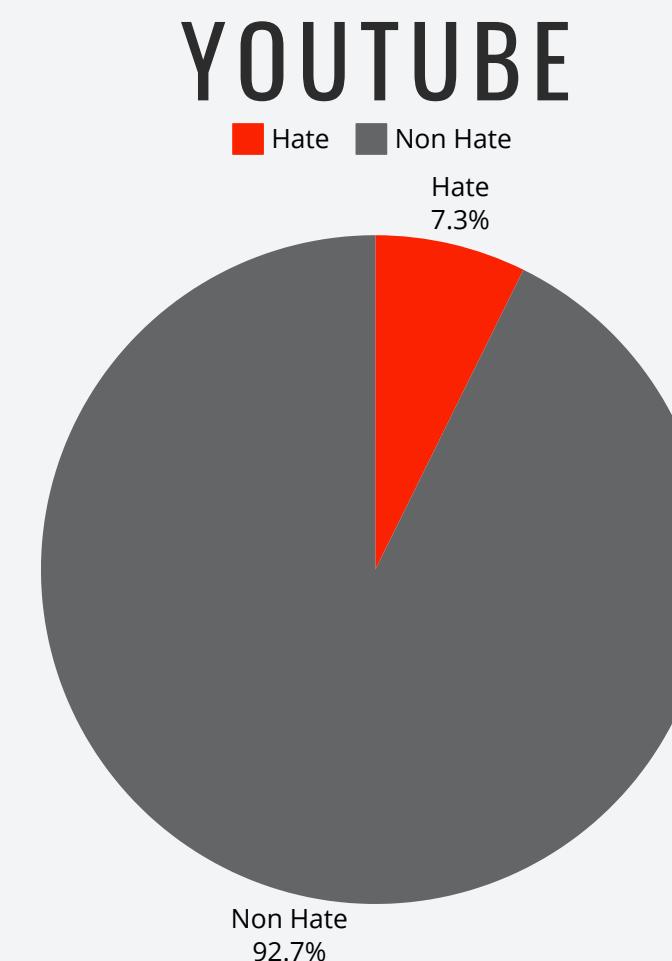
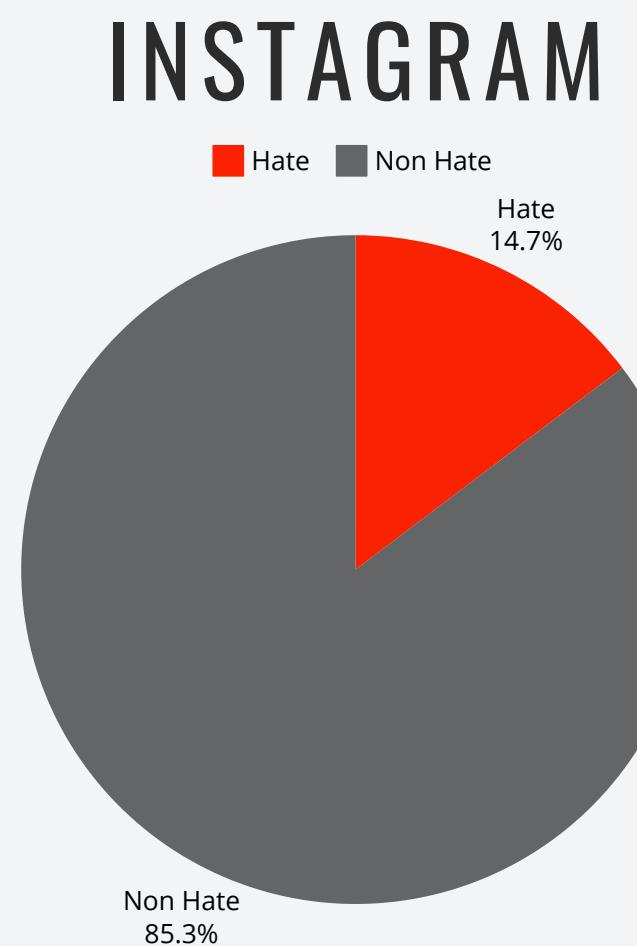
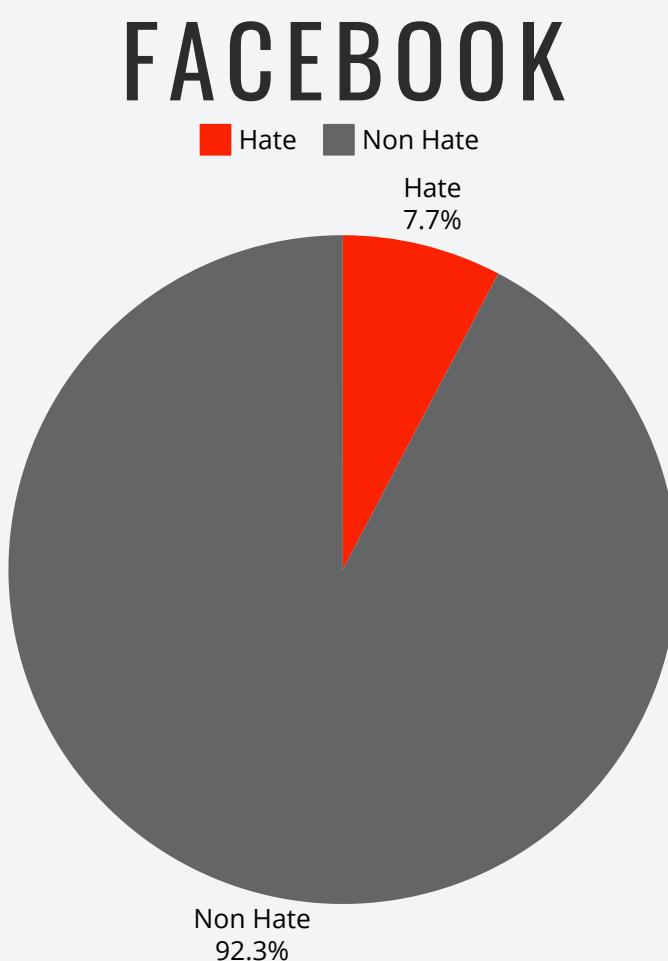


### YOUTUBE



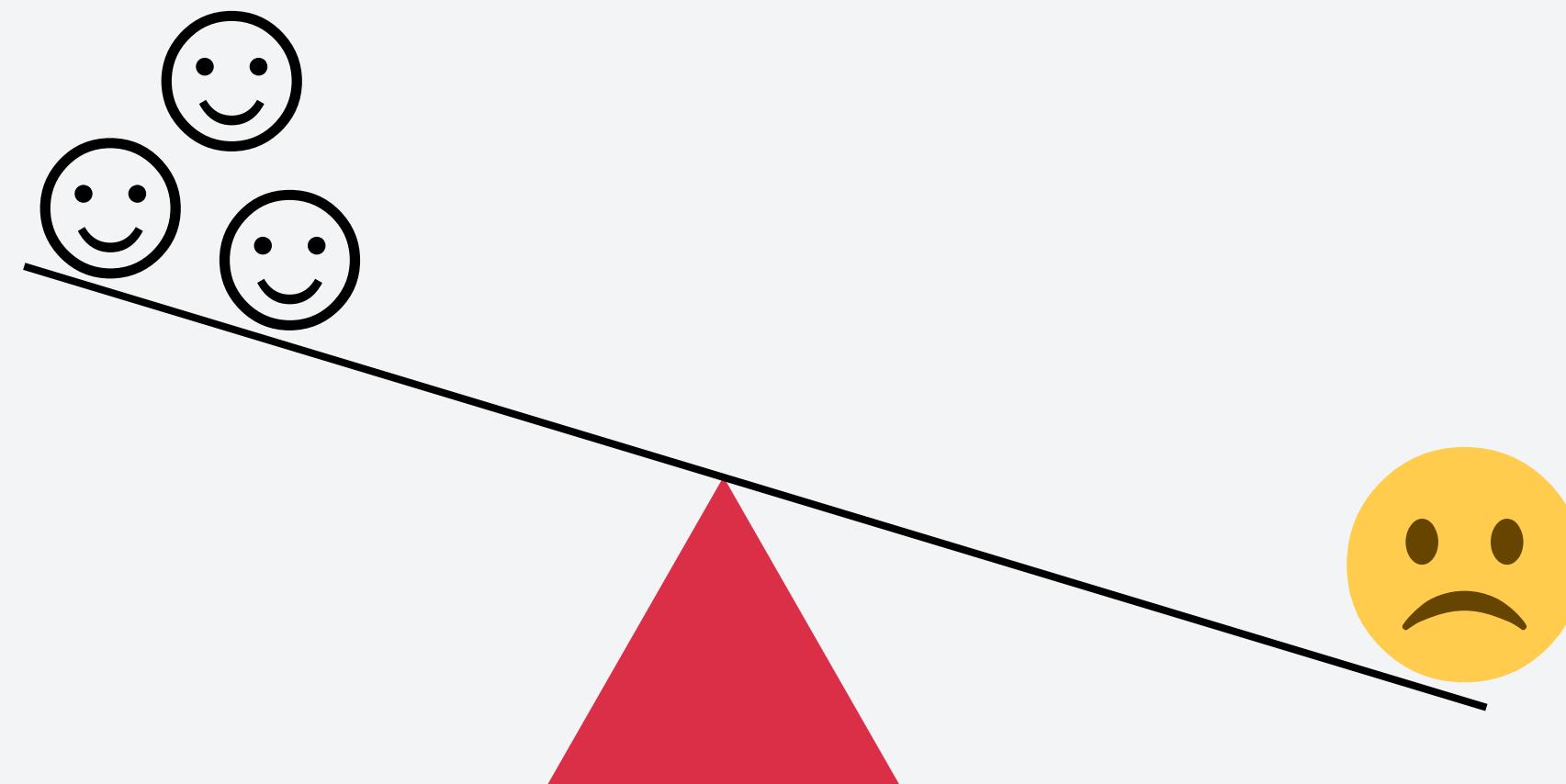
# MEDIE COMMENTI SOCIAL PER TESTATA

**CORRIERE DELLA SERA**



# NEGATIVITY BIAS

*Fenomeno psicologico che spinge all'esternazione del proprio coinvolgimento emotivo solo in senso negativo, dando più importanza agli eventi negativi rispetto a quelli positivi.*



# OVERLAPPING DEI MODELLI



Lo 0.3% del dataset soffre di un overlapping incoerente tra i modelli, ovvero il modello di sentiment classifica il commento come “positivo” e l’hate speech detector lo considera commento di odio! 😬

# OVERLAPPING DEI MODELLI: ESEMPI

Notizia	Commento	Sentiment	Hate Speech
Attacco di Hamas a Israele	I fasci sono tutti appesi	<b>POSITIVO</b>	<b>VIOLENTO</b>
Incidente autostradale Youtubers	Schifo	<b>POSITIVO</b>	<b>INAPPROPRIATO</b>
Geolier e i fischi a Sanremo	Giornalai!	<b>POSITIVO</b>	<b>OFFENSIVO</b>
L'incoronazione di Re Carlo	CHISSENEFREGAAAAAA!!!!	<b>POSITIVO</b>	<b>INAPPROPRIATO</b>

# CONCLUSIONI

Topic-Social che generano più commenti **negativi**:

1°		CRONACA NERA (88.7%)
2°		CRONACA NERA (87.0%)
3°		CRONACA NERA (85.2%)
4°		CRONACA (84%)
5°		CRONACA (76.8%)

Topic-Social che generano più commenti di **odio**:

1°		CRONACA (17.1%)
2°		CRONACA NERA (17.0%)
3°		POLITICA (16.0%)
4°		CRONACA NERA (15.5%)
5°		CRONACA (7.4%)

# CONCLUSIONI

Giornale-Topic-Social che generano più commenti di negativi:

1°		IL CORRIERE DELLA SERA CRONACA NERA	364
2°		IL CORRIERE DELLA SERA CRONACA NERA	354
3°		FANPAGE CRONACA NERA	353

Giornale-Topic-Social che generano più commenti di odio:

1°		LA REPUBBLICA CRONACA	107
2°		LA REPUBBLICA CRONACA NERA	83
3°		FANPAGE CRONACA NERA	75

# CONCLUSIONI

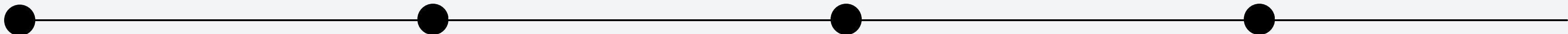
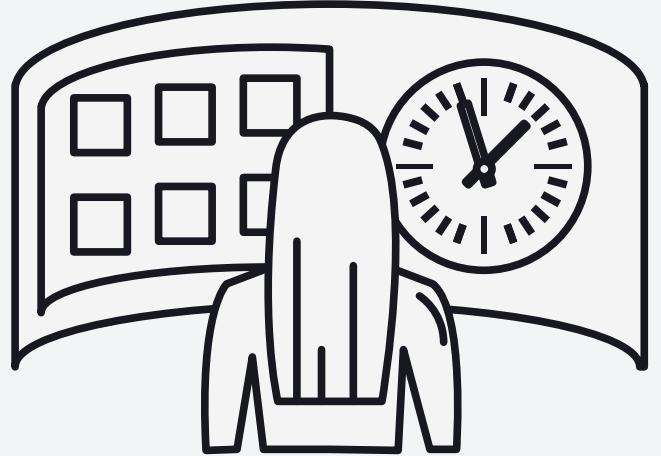
Social che generano più commenti **negativi**:

1°		3026
2°		2843
3°		2781

Social che generano più commenti di **odio**:

1°		601
2°		378
3°		263

# SVILUPPI FUTURI



## Analisi del media

Identificare gli effetti sulla tipologia di media nei confronti di una notizia e la relazione al sentimento dello stesso

## Previsione Hate Speech

Basandoci sul titolo della notizia, riuscire a predire la quantità di hate speech presente in un determinato social.

## 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 sentimento che l'hate speech, in modo da evitare errori di discrepanza

# GRAZIE



UNISA



GITHUB REPOSITORY