

# FERRAGNI BALOCCO

**Giorgia Costantini** 782261  
**Andrea Marcoccia** 775451  
**Filippo Navarra** 782571



# Sources



## INSTAGRAM & FACEBOOK

APIFY.com for Facebook.  
Exportgram.net for  
Instagram --> **limit of 100  
comments per post**



## REDDIT

PRAW library to access  
Reddit, fetching comments  
from specific posts  
identified by their IDs, then  
organizing this data into a  
DataFrame



## NEWS ARTICLES

Collected web articles via  
URLs and processed them  
with the *Newspaper3k* tool  
to isolate and analyze text  
content

# SENTIMENT ANALYSIS

## FEEL-IT

- FINE-TUNE OF BERT (ONE OF THE MOST POPULAR NEURAL ARCHITECTURES IN NLP)
- FINE-TUNED ON 2037 TWEETS IN ITALIAN
- **LOW ACCURACY**



## BERT-BASE-ITALIAN-CASE-SENTIMENT

- FINE-TUNE OF BERT
- FINE-TUNED ON 45K PREPROCESSED TWEETS
- **LOW ACCURACY**



LACK OF ACCURATE SENTIMENT ANALYSIS MODELS  
TO RECOGNIZE ITALIAN LANGUAGE FREELY  
AVAILABLE ON THE WEB

## MISTRAL API

- INPUT DATA IS CONVERTED IN A PROMPT SENT TO THE API
- THE MODEL IS ASKED TO CLASSIFY INPUT DATA INTO 'POSITIVE', 'NEUTRAL' OR 'NEGATIVE'
- **HIGH ACCURACY ON INSTAGRAM**
- **LOW ACCURACY ON REDDIT**



LLM (LARGE LANGUAGE MODEL) SO IT  
BETTER UNDERSTAND ITALIAN

# Difficulties

## **SCRAPE SOCIAL MEDIAS**

Instagram and Twitter have  
restrictive policies

## **LABEL DATA**

How can we be sure of the data  
sentiment labeling?

# FINAL GOALS

- NEWSPAPERS TITLE TRENDS
- MAIN TOPICS OF DISCUSSION (NEWSPAPERS)
- DIFFERENT SENTIMENTS FROM DIFFERENT SOURCES
- OVERALL SENTIMENT TO THE “SCANDAL”
- SENTIMENT OVER TIME OF FERRAGNI’S FANS (100 COMMENTS PER POST)