

Social Media Analysis of

Taylor Swift

SMA project

Giulia Beccaria
Lorenzo Lobosco

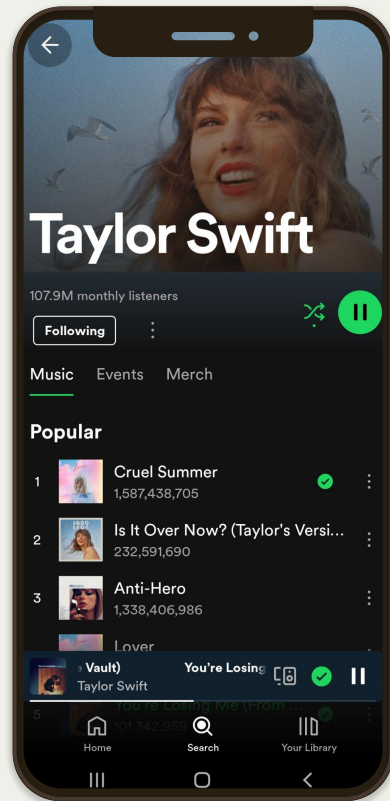


Table of contents

01

Introduction

Aim of the project

02

Data

Reddit API

Data preprocessing

03

SNA

Metrics

Community Detection

04

SCA

Sentiment Analysis

NER

05

Result & Conclusion

- Taylor Swift chosen as a case study for being the **Times' Person of 2023**.
- Significantly **impacted the American economy**, surpassing political figures
- Explored Taylor Swift's social media, particularly **Reddit**.
- Uncovered insights into public discourse and community development.
- Examined whether engagement extended beyond her music

Introduction





Data Collection

Reddit

Utilized the "TaylorSwift" subreddit in 2023 as the main data source, hosting a significant community of Taylor Swift fans.

Employed two datasets, "hot" and "top," with exclusive focus on the "top" dataset to prioritize posts with the highest vote counts.

Dataframe

Subreddit: Name of the subreddit

Title: Title of the post

Selftext: Text content of the post

upvote_ratio: Ratio of upvotes to total votes

Ups: Number of upvotes

Downs: Number of downvotes

Score: Overall score of the post

Author: Author of the post

Created: Date and time of post creation

ID: Unique identifier of the post

Link: Unique link to the post

Step 1

Stopwords removal

A stop word is a commonly used word

Step 2

Punctuation removal

And lowercase

Step 3

Stemming

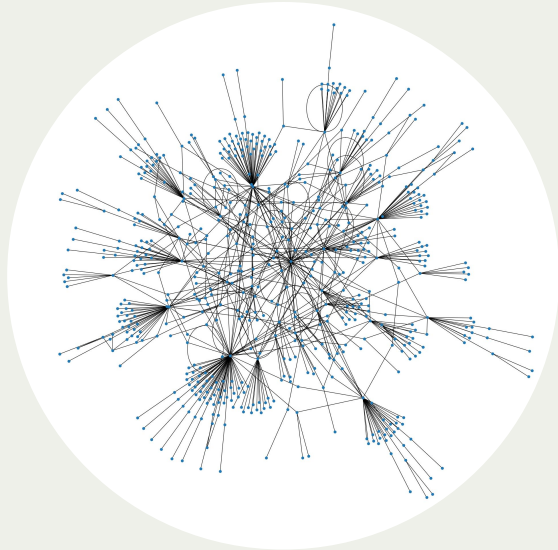
abbreviates words to their base or root form, often in writing

Preprocessing



Social **Network** Analysis

Metrics

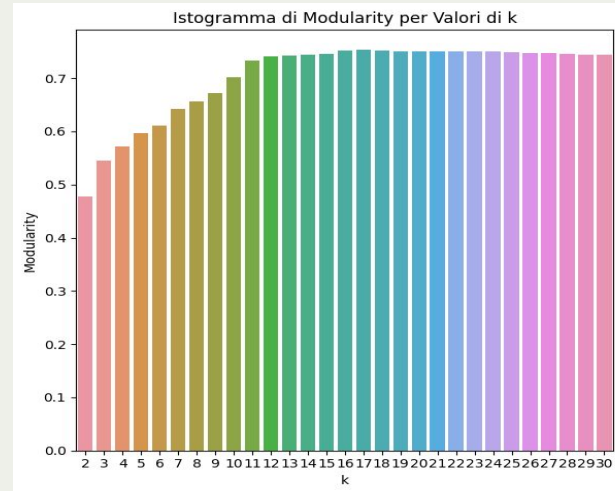
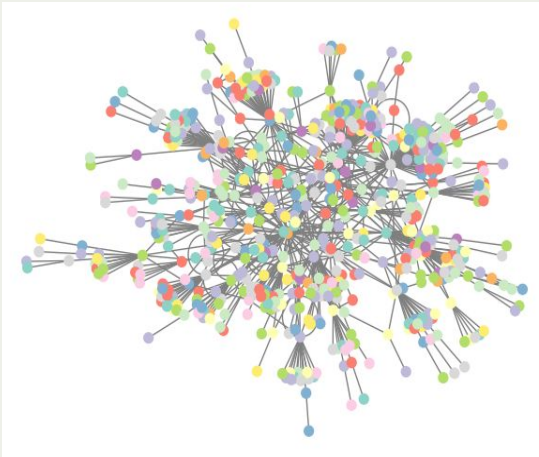


| Node | Max Betweenness Centrality | Max Closeness Centrality |
|-------------------------|----------------------------|--------------------------|
| AutoModerator | 0.12841091 | 0.47557252 |
| bubblecuffer13 | 0.10272873 | 0.37417417 |
| RandOmaccou ntpigeon | 0.07544141 | 0.36157864 |
| Mundane-Student7037 | 0.06741573 | 0.35217637 |
| Seachelle13o | 0.04975923 | 0.34960718 |

Social **Network** Analysis

Community Detection

- Graph creation and clustering detection
- Several different approaches such as **Louvain Community Detection** or **Greedy Modularity**
- High Modularity for an high number of Groups, for each cases
- Network seems to be highly divided, without a solid creation of clusters well defined within the community



Social **Content** Analysis

Sentiment Analysis

Focus on the **upvote_ratio value**, categorizing it into three groups:

- - 1 for values less than -0.1 → Negative
- 0 for values between -0.1 and 0.1 → Neutral
- 1 for values greater than 0.1 → Positive

Then we try different approaches:

- Lexicon-based approach: **Afinn** and **Opinion Lexicon**
- **VADER** (Valence Aware Dictionary and sentiment Reasoner)

We will be using the **Logistic Regression** class from *Scikit-learn*

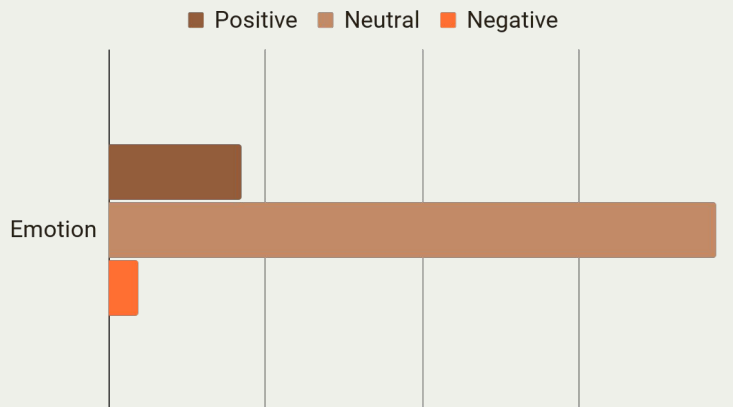
| Approach | Accuracy 3 Classes | Accuracy 2 Classes |
|-----------------|-----------------------|-----------------------|
| Afinn | 0.330957 | 0.626506 |
| Opinion Lexicon | 0.327902 | 0.631979 |
| VADER | 0.314664 | 0.641096 |

| Accuracy | Precision | Recall | F Score |
|----------|-----------|-----------|----------|
| 0.614213 | 0.375375 | 0.4047203 | 0.384321 |

Social **Content** Analysis

Emotion Analysis

We use the **Dictionary-based** emotion detection



We employed the NRC emotion lexicon using the *NRC*Lex function

| | |
|--------------|-----|
| trust | 187 |
| joy | 169 |
| positive | 283 |
| surprise | 86 |
| anticipation | 170 |
| anger | 49 |
| fear | 59 |
| disgust | 36 |
| sadness | 72 |
| negative | 104 |

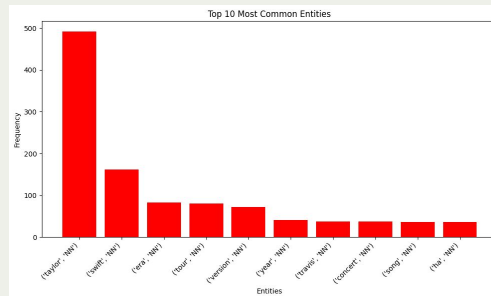
Social **Content** Analysis

Named Entity Recognition

Identifying and categorizing entities within text into predefined categories.

Several different approaches:

- **NLTK**-Based
Simply and basic method
- **Spacy**-Based
More sophisticated and powerful
- **Dandelion**, **Babelify**, and **Spotlight** APIs
This are based on external APIs that offer NER services



| Entity | Count |
|--------------------------|-------|
| taylor | 500 |
| tonight | 30 |
| today | 26 |
| one | 23 |
| first | 20 |
| midnight | 17 |
| last night | 16 |
| night | 9 |
| year | 9 |
| travis kelce | 8 |
| Total number of entities | 1056 |

| Entity | Frequency |
|-------------------|-----------|
| ('taylor', 'NN') | 492 |
| ('swift', 'NN') | 162 |
| ('era', 'NN') | 83 |
| ('tour', 'NN') | 81 |
| ('version', 'NN') | 72 |
| ('year', 'NN') | 41 |
| ('travis', 'NN') | 37 |
| ('concert', 'NN') | 37 |
| ('song', 'NN') | 36 |
| ('ha', 'NN') | 36 |



Conclusion

Community Structure & Content Dynamics

Social network analysis unveiled the structure and dynamics of Taylor Swift's online community, emphasizing its open and cohesive nature.

Sentiments and Emotions in Subreddit

Content analysis showcased prevailing sentiments and emotions in the subreddit, reflecting an overall positive and neutral tone in titles.

Strong and Positive Social Media Presence

Findings revealed Taylor Swift's strong and positive presence on social media, supported by a large and engaged fan base.

Success, Importance of Preprocessing, and Future Directions

The project successfully achieved its aim, highlighting the importance of data preprocessing. The results can inform marketing strategies, and potential future developments include expanding to other platforms and incorporating machine learning for advanced analysis.