Taylor's triumph: Social Media Analysis of the 2023 person of the year

Social Media Analytics - Q1 (2024)

Full Name	Student ID
Giulia Beccaria	851286
Lorenzo Lobosco	851289

Milan, January 15, 2024

Contents

1 Introduction **1**
1.1 Aim of the project 1

2 Data **2**
2.1 Reddit API 2
2.2 Data preprocessing 2

3 Social Network Analysis **4**
3.1 Metrics 4
3.2 Community Detection 5

4 Social Content Analysis **8**
4.1 Sentiment analysis 8
4.2 Emotion analysis 10
4.3 Named Entity Recognition 11

5 Conclusion **14**

6 References **15**

1 | Introduction

Taylor Swift is a famous American singer-songwriter who has won many awards and broken many records in the music industry. She is known for her catchy pop songs, her personal lyrics, and her loyal fan base. She started her career as a country singer at the age of 14, and later transitioned to pop music with her album 1989 in 2014. She has sold over 200 million records worldwide, making her one of the best-selling music artists of all time. She has also won 11 Grammy Awards, 28 American Music Awards, 32 Billboard Music Awards, and many other honors. She is one of the most influential and powerful celebrities in the world, with a net worth of over $400 million.
Time magazine named her the person of the year 2023 because of her achievements in the arts, her influence on culture, and her resilience in the face of challenges. She was the first woman to make a second appearance on the cover of Time, and the first person to be selected for her artistic contribution. Time praised her as an extraordinary storyteller who has the power to transform society.[5]
Some of the reasons why Time chose her as the person of the year are:

- She released two critically acclaimed and commercially successful albums in 2023: Red (Taylor's Version) and 1989 (Taylor's Version), which were re-recordings of her previous albums that she lost the rights to after a dispute with her former label. She reclaimed her artistic ownership and control, and inspired other artists to do the same.
- She embarked on a record-breaking global tour, The 1989 World Tour, which grossed over $1 billion and attracted over 10 million fans. She also performed at several major events, such as the Super Bowl halftime show, the Grammy Awards, and the Tokyo Olympics opening ceremony. She showcased her versatility, creativity, and charisma as a live performer.
- She used her platform and voice to advocate for various causes, such as women's rights, LGBTQ+ rights, voting rights, climate change, and mental health. She donated millions of dollars to various charities and organizations, and encouraged her fans to get involved and make a difference. She also spoke out against injustice, discrimination, and violence, and supported the victims of abuse and harassment. She was a role model and a leader for many people, especially young women.[1]

Taylor Swift is a remarkable artist and a remarkable person, who has shaped the world with her music, her words, and her actions.

1.1 | Aim of the project

In pur project, we aim to examine the reasons behind Taylor Swift's rise to becoming the TIME Person of 2023 and explore its impacts. Social media platforms, with their detailed analyses, serve as the optimal tools for this investigation. Specifically, we seek to understand how Taylor Swift's extensive fan community evolved, its structure, and the key topics associated with this iconic pop star.
We also plan to explore the sentiment and feelings that Taylor's supporters and others have expressed. It's crucial to note that Taylor Swift's recognition as the Person of 2023 isn't solely attributed to her musical achievements. Therefore, through this project, we aim to identify the other topics that are predominantly linked to her and determine which of these generate the most excitement within her community.

2 | Data

To achieve the goal of our project, we have chosen to utilize Reddit [6] as our primary source of analysis (please add references to the generic Reddit). Reddit is a widely open social platform, allowing for a more in-depth examination of the Taylor Swift community.
Reddit is an online platform for social news, entertainment, and forums, where registered users can share content in the form of text posts or hyperlinks (links). Additionally, users can express their approval or disapproval by giving a "up" or "down" vote (commonly referred to as "upvote" and "downvote") to the posted content ([8])
We specifically selected the main Taylor Swift subreddit (add link), which boasts around one and a half million users and millions of posts annually. Regarding data extraction from Reddit, we chose to extract the top posts of the subreddit in the year 2023, those with the highest number of positive votes. Additionally, we included the hottest posts of the subreddit in 2023, representing content that received a significant number of votes in a short period due to its innovative nature.[7]

2.1 | Reddit API

For data extraction from Reddit, we leveraged the platform's capability to provide data through its API. We requested a secret token that allowed us to extract data from the Taylor Swift subreddit using Python. Subsequently, we extracted the hottest and top posts from the subreddit in 2023. The extracted variables include:

- Subreddit: The name of the subreddit to which the post belongs.
- Title: The title of the post.
- Selftext: The text content of the post.
- Upvote_ratio: The ratio of positive votes to the total number of votes.
- Ups: The number of positive votes on the post.
- Downs: The number of negative votes on the post.
- Score: The difference between the total number of positive votes and negative votes.
- Author: The author of the post.
- Created: The date of post publication.
- Id: The post identifier.
- Link: The link to the post.

It's worth noting that the "Subreddit" variable is the same for all posts since we extracted data from a single subreddit. Additionally, the "Selftext" variable contains numerous NaN values, mainly because many posts only include the title and lack additional text. Following the raw data extraction phase, we proceeded to the cleaning and management of the data.

2.2 | Data preprocessing

The chapter begins by introducing the tools and libraries essential for the NLP tasks at hand. The Natural Language Toolkit (NLTK) is highlighted as a prominent platform for working with human language data in Python. Utilized the "TaylorSwift" subreddit in 2023 as the primary data source, capturing discussions from a vibrant community of Taylor Swift fans. Employed two datasets, "hot" and "top," focusing exclusively on the "top" dataset to prioritize posts with the highest vote counts. This strategic choice ensures a dataset enriched with content that has received significant community approval. The selected dataset serves as the foundation for the subsequent preprocessing phase, where text data is refined for in-depth analysis.
In the preliminary stages of our analysis, the dataset sourced from the "Taylortop.csv" file underwent a meticulous text preprocessing phase. This critical step was undertaken to refine and structure the textual content for subsequent advanced analyses. The preprocessing phase was facilitated by incorporating key libraries and linguistic resources. The pandas library was employed for efficient data manipulation,

while the Natural Language Toolkit (nltk) served as a robust tool for intricate language processing tasks. Specifically, nltk's WordNet was utilized for lemmatization, and a curated set of English stopwords was employed to enhance the quality of the text. The cornerstone of the preprocessing effort was the definition of the preprocess_text function. This function was crafted with precision to sequentially enhance the textual data:

- Numerical Removal: Numeric characters were systematically removed to maintain the linguistic integrity of the text.
- Lowercasing: All text was uniformly transformed to lowercase, ensuring consistency throughout the dataset.
- Lemmatization: Leveraging WordNetLemmatizer from nltk, the function normalized words to their base forms, capturing essential meanings.
- Punctuation Elimination: Punctuation was strategically eliminated to streamline subsequent processing steps.

The meticulous preprocess_text function was systematically applied to select columns within the dataset. Noteworthy columns, including 'subreddit,' 'title,' 'selftext,' 'author,' and 'link,' underwent the defined preprocessing steps. The outcomes of this process were then stored in new columns, denoted by the '_proc' suffix.
A final refinement step involved the selective removal of English stopwords from the processed text. This strategic elimination of common, non-substantive words aimed to further fine-tune the text for subsequent in-depth analyses.
After the text preprocessing, we leveraged nltk for word frequency calculation and WordCloud for graphical representation. .
From the processed 'selftext' column, word frequencies were computed using nltk's FreqDist. A visually appealing WordCloud was then crafted to highlight prominent terms within the dataset.
To enhance clarity, non-contributory terms like 'nan', 'wa', and 'taylor' were strategically removed from the WordCloud. This curation aimed at refining the visual representation for a more focused and insightful overview.
Matplotlib displayed the refined WordCloud, offering a concise snapshot of the dataset's salient terms. This visual tool synergized with text preprocessing, enriching our ability to extract meaningful insights from the linguistic nuances in the dataset.

Figure 2.1: Wordcloud of the dataset TaylorSwiftTop after the preprocessing

3 | Social Network Analysis

After the data preprocessing phase, we transitioned to implementing social network analysis.
Social Network Analysis (SNA) is a methodology that explores the relationships and interactions among elements within a system. In the realm of social network analysis, a "node" represents an actor, while an "edge" signifies the relationship between two actors. This approach is widely employed to comprehend the structure and dynamics of social networks.
For the social network analysis, we extensively utilized the Python library NetworkX. It is important to note that we had to slightly modify our data. To construct a graph, we considered only the latest 50 posts on the subreddit. For each post, we collected the authors of first and second-level comments. In graph theory, a graph is a collection of nodes (vertices) and edges that connect pairs of nodes. A graph can be directed or undirected.
Subsequently, we proceeded with the creation of an undirected graph, where edges were added based on the authors of first and second-level comments. The resulting graph visually illustrates the network's structure, with nodes representing users and edges representing relationships between them.

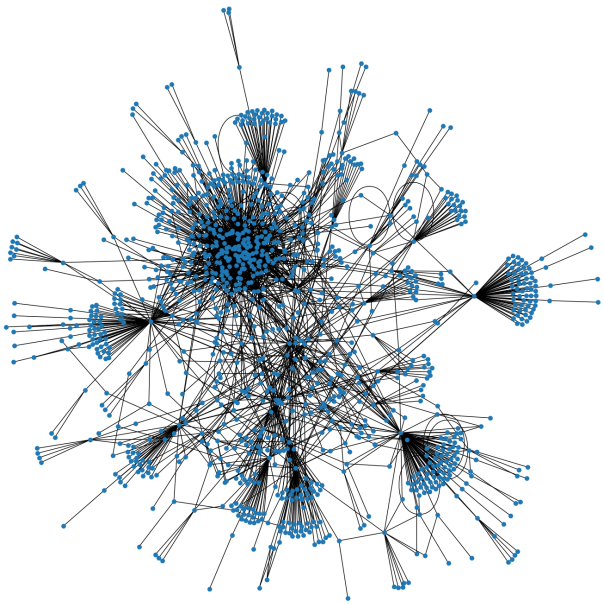

Figure 3.1: Graph of the Taylor Swift community

3.1 | Metrics

We conducted an analysis of centrality metrics, including degree centrality, closeness centrality, and betweenness centrality, to pinpoint key actors within the network. These metrics offer insights into the significance and influence of individual nodes in the social network.

1. **Degree Centrality:**
   Degree centrality quantifies the number of connections a node has. In a social network, a high degree centrality suggests a node with many connections.
   $C_D(v) = \frac{\text{Number of edges connected to node } v}{\text{Total number of nodes in the network}}$

   - **Top 5 nodes with the highest degree centrality:**
     - Node PassionateAsSin: 0.19755102040816327
     - Node ChampagneSnakes13: 0.08326530612244898
     - Node nfl: 0.06612244897959184
     - Node Mazerr44: 0.062040816326530614

- Node this_diamond_ring: 0.0489795918367347

2. **Closeness Centrality:**
   Closeness centrality measures how close a node is to all other nodes in the network. Nodes with high closeness centrality are positioned close to other nodes.
   $C_C(v) = \frac{1}{\sum_u \text{Shortest path from } v \text{ to } u}$

   - **Top 5 nodes with the highest closeness centrality:**
     - Node AutoModerator: 0.41287495786990225
     - Node PassionateAsSin: 0.38257339163023113
     - Node bubblecuffer13: 0.36654697785757034
     - Node None: 0.3641498216409037
     - Node icoulddance: 0.35944835680751175

3. **Betweenness Centrality:**
   Betweenness centrality identifies nodes that act as bridges along the shortest paths between other nodes. Nodes with high betweenness centrality play a crucial role in connecting different parts of the network.
   $C_B(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$

   - **Top 5 nodes with the highest betweenness centrality:**
     - Node AutoModerator: 0.3642134411978155
     - Node PassionateAsSin: 0.29280227158142985
     - Node ChampagneSnakes13: 0.184224971267353
     - Node nfl: 0.14416303481595036
     - Node Mazerr44: 0.1215781868195742

We also computed the diameter and radius of the graph:

- Diameter: The diameter of a graph is the length of the longest shortest path between any two nodes. It represents the maximum distance within the network.the resulting diameter is 7.
- The radius of a graph is the length of the shortest path from a central node to the farthest node in the network. It indicates the overall compactness of the network. The resulting radius is 4.

3.2 | Community Detection

Community detection, also known as network clustering or graph partitioning, is a crucial aspect of network analysis that involves identifying groups of nodes in a network that are more densely connected to each other than to nodes in other groups. These groups are referred to as communities or clusters.

- **Louvain Community Detection**
  At first we applied Louvain community detection algorithm using the 'community_louvain' module. Louvain is a popular method for finding communities in large networks. we used 'best_partition(G)' which identifies communities in the graph using the Louvain algorithm and assigns each node to a community, and we also used 'modularity(part, G)' that computes the modularity of the partition, indicating how well the network is divided into communities. Modularity obtained in this is equal to 0.606.

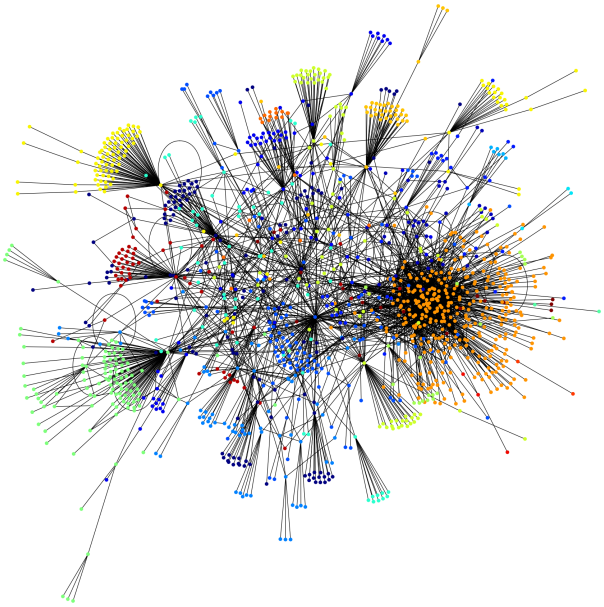

Figure 3.2: Community detected by Louvain algorithm

We considered also the also the Assortativity in our analysis. Assortativity measures the preference of a network's nodes to attach to others that are similar in some way. In this case the assortativity coefficient resulting is -0.2, which suggests a slight disassortative mixing, meaning nodes with different degrees tend to connect more often than nodes with similar degrees.

- **Greedy Modularity Communities**
  Then we applied the Greedy Modularity Communities algorithm, it Identifies communities based on maximizing modularity, which quantifies the quality of a partition. We obtained the following results:

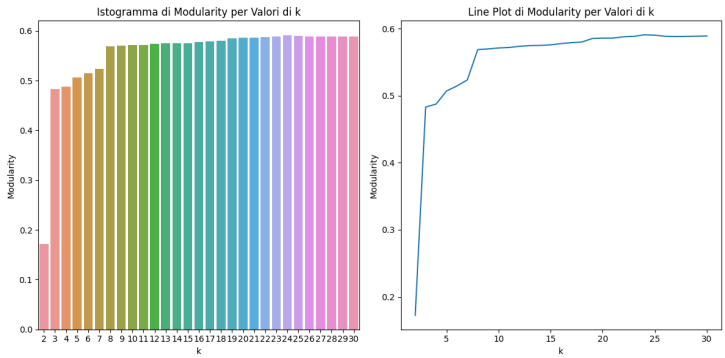

Figure 3.3: Community detectuon by Girvan-Newman algorithm

Analyzing the data, it appears that as the number of modules (represented by "k") increases, modularity tends to rise until a certain point, after which it stabilizes or slightly decreases. It seems that the maximum modularity value is reached when the number of modules is around 12-14. This could indicate an optimal point where the division into modules is well-balanced. Beyond this optimal point, there is a slight decrease in modularity, suggesting that excessive subdivision into modules beyond a certain threshold may lead to reduced cohesion between communities. In conclusion, the network seems to be highly divided, which is plausible given that a subreddit about Taylor Swift with many active users is likely to result in a highly fragmented network without few distinct and well-defined groups.

- **Crossing Edges**
  In our anaysis we included also the crossing edge analysis in community detection. Crossing edges are edges within a network that connect nodes belonging to different communities or clusters. In the context of community detection, identifying and understanding crossing edges is crucial for analyzing the interconnectivity between different groups of nodes. we obtained

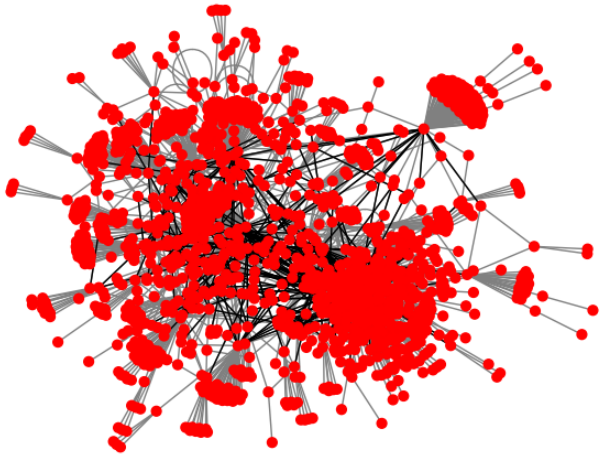

Figure 3.4: Crossing edges in community detection

As evident from this figure, the community structure appears highly subdivided with numerous small groups, each exhibiting relatively low density, yet interconnected by a multitude of links.

- **Fluidc approach**
  At the end we analyzed the asynchronous fluid communities algorithm with multiple levels. Asynchronous fluid communities consider multiple levels of granularity, potentially capturing finer details in community structure compared to previous methods. We consider Two cases, when the cluster is 2, 3 and 8. 8 results the case with the higher modularity.

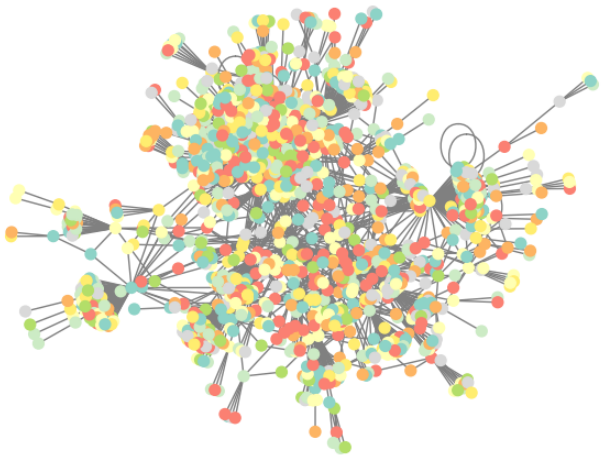

Figure 3.5: Fluidc community detection

4 | Social Content Analysis

stesssa cosa della SNA, si descrive cos'è come funziona, cosa abbamo fatto e calcolato noi. con qualche grafico e risultato.

4.1 | Sentiment analysis

In this phase, we delved into sentiment analysis, a process aimed at discerning the emotional tone behind textual content.
We undertook an exploration of the "title" column within our dataset, specifically sourced from the "TaylorSwift" subreddit in 2023. To anchor our sentiment analysis, we utilized the "upvote_ratio" column, representing the proportion of upvotes to total votes for each post. This metric served as a quantitative indicator of the community's response.
Given the nature of the Taylor Swift dataset, predominantly positive, we adopted a stringent approach, categorizing posts as negative only if they displayed a slight imbalance in dislikes compared to others. Hence, we opted to standardize the data and devised a function to group the "upvote_ratio" values, designating those close to zero as neutral.

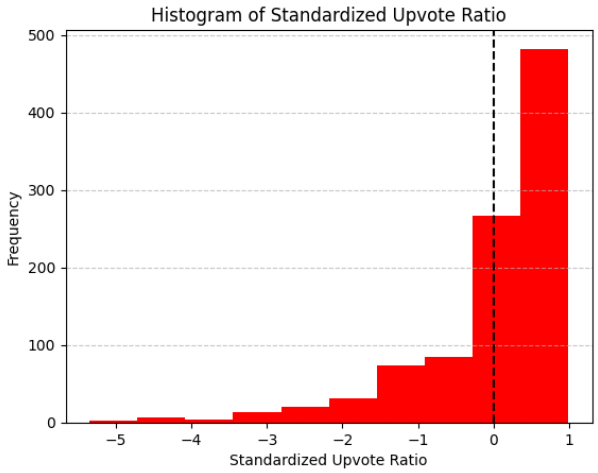

Figure 4.1: Histogram of upvote_ratio

Our approach involved classifying sentiment into three distinct categories:

- Negative (-1): Values less than -0.1
- Neutral (0): Values between -0.1 and 0.1
- Positive (1): Values greater than 0.1.
- **AFINN lexicon:**
  Utilizing the AFINN lexicon, a list of English terms manually rated for valence, we scored each title to ascertain sentiment intensity. Afinn utilizes a scoring system that assigns values from -5 to +5, where negative scores indicate a negative sentiment, positive scores denote a positive sentiment, and a score of 0 signifies neutrality. The magnitude of the score reflects the intensity of the sentiment. We applied Afinn's scoring mechanism to the "title" column, generating sentiment scores for each post. The resulting DataFrame included columns such as 'afinn', 'grouped_upvote_ratio', and 'title'. This approach provides a quick and accessible solution for sentiment analysis, particularly in scenarios where simplicity and ease of implementation are prioritized. Then, we grouped the values of the Afinn's scores to construct confusion matrices for both two and three sentiment groups:

	-1	0	1
-1	37	137	101
0	9	65	50
1	54	306	223

	-1	1
-1	37	101
1	54	223

Table 4.1: Confusion Matrices for three and two groups

We can calculate the accuracy by summing the values on the diagonal and obtain an accuracy of 0.331 for the three groups and 0.627 for the two groups.

- **Leveraging the Opinion Lexicon:**
  The approach involved utilizing a lexicon-based method for sentiment analysis, specifically leveraging the Opinion Lexicon, which consists of positive and negative terms. The NLTK library was employed to implement sentiment analysis by counting the occurrences of positive and negative words in each title. The process began with downloading the opinion lexicon and creating sets of positive and negative words. Two functions, sentiment_sum and sentiment_neg, were then implemented for sentiment analysis. The sentiment_sum function assigned a sentiment score based on the overall count of positive and negative words, while the sentiment_neg function handled complex sentences by considering negations. These functions were applied to each title in the dataset, and the resulting sentiment scores were stored in the "op_lexicon" column of the DataFrame, providing a quantitative measure of sentiment for each title. Additionally, the scores were standardized and divided to obtain confusion matrices for the results.

	-1	0	1
-1	38	141	96
0	9	73	42
1	49	323	211

	-1	1
-1	40	72
1	59	194

Table 4.2: Confusion Matrices for three and two groups

- **VADER:**
  The final approach involved using VADER, a rule-based sentiment analysis tool that provides nuanced evaluation of sentiment intensity through compound scores. VADER is designed for social media sentiment analysis and is both lexicon and rule-based. The implementation utilized the VADER library, specifically the SentimentIntensityAnalyzer class from the vaderSentiment module. Two functions were created for sentiment assessment: the sentiment_scores function, which provided an overall sentiment dictionary with positive, negative, neutral, and compound scores for a given sentence, and the vader_score function, which calculated the compound sentiment score for each title in the dataset and stored the results in a new "vader" column. The final step involved standardizing and categorizing the new values into three groups based on sentiment scores: negative (-1) for scores less than -0.5, neutral (0) for scores between -0.5 and 0.5, and positive (1) for scores greater than 0.5. Confusion matrices were then obtained for both two and three sentiment groups.

	-1	0	1
-1	40	163	72
0	14	75	35
1	59	330	194

	-1	1
-1	40	72
1	59	194

Table 4.3: Confusion Matrices for three and two groups

FInally we can calculate the accuracy by summing the values on the diagonal and obtain an accuracy of 0.315 for the three groups and 0.641 for the two groups.

These accuracy scores indicate the model's success in correctly classifying instances into different sentiment groups. Notably, the VADER rule-based approach achieved the highest accuracy, with 64.1% accuracy for 2

sentiment groups. It suggests that VADER, designed for social media sentiments, performed comparatively well in capturing the nuanced sentiment expressed in the Taylor Swift subreddit during the specified timeframe.

- **Linear Classifier:**
  During this phase our focus shifted towards constructing a linear classifier for the dataset, leveraging the In this phase, the focus was on constructing a linear classifier for the dataset using the Logistic Regression model from the Scikit-learn library. The dataset was divided into training and testing sets using the train_test_split function. The text data was then transformed into a bag-of-words representation using the CountVectorizer from Scikit-learn, crucial for applying machine learning models.

  Next, a Logistic Regression model was created by instantiating the LogisticRegression() class and trained on the designated training data using the fit method. The trained model was applied to make predictions on the test data. To assess the model's performance, various evaluation metrics were computed. The accuracy metric, representing the model's overall correctness, was determined using the accuracy_score function. Precision, recall, F-measure, and support metrics were calculated through the precision_recall_fscore_support function, providing a comprehensive understanding of the model's proficiency in accurately categorizing instances into different sentiment groups.

  The resulting accuracy for this linear classifier was found to be 0.614, serving as a quantitative measure of the model's success in correctly classifying instances and offering valuable insights into its performance characteristics.

Metric	Value
Accuracy	0.614
Precision	0.375
Recall	0.404
F-score	0.384

Table 4.4: Performance measures of the model.

The sentiment analysis uncovered the emotional nuances within the Taylor Swift community, shedding light on prevailing sentiments. Comparing different lexicon-based and rule-based approaches allowed for a comprehensive understanding of the sentiment landscape. This analysis formed a pivotal step in comprehending user interactions and expressions within the Taylor Swift subreddit during the specified timeframe.

4.2 | Emotion analysis

We delved into a more sophisticated method of emotion detection, utilizing the NRC emotion lexicon. The objective was to discern the emotional tones embedded in the textual content of post titles within the Taylor Swift subreddit. Unlike sentiment analysis, which gauges the overall sentiment of a text, this approach aimed to identify specific emotions such as joy, anger, surprise, anticipation, trust, fear, disgust, and sadness.
To achieve this, we employed the NRCLex library, implementing a function named get_emotions(sentence) to calculate the raw emotion scores associated with each term in a sentence. This function was then applied to the "title" column of our dataset, generating a new column, "emotion," containing the raw emotion scores for each post.
One key aspect of our analysis was the categorization of posts into emotion categories: "Positive," "Negative," or "Neutral." This categorization was based on the predominant emotion detected in each post. For instance, posts exhibiting joy were labeled as "Positive," those containing anger were labeled as "Negative," and the rest were labeled as "Neutral."

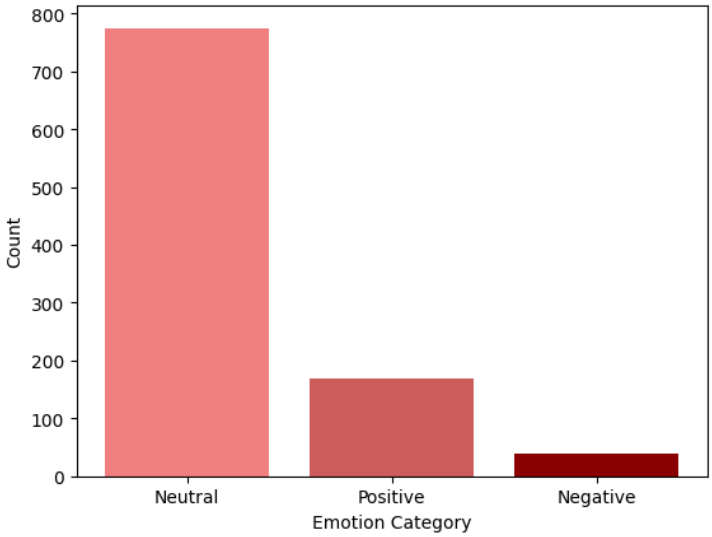

Figure 4.2: Emotion Category Distribution

The prevalence of the "Neutral" category is attributed to the classification of sentences with both positive and negative emotional elements as neutral. This approach captures nuanced instances where conflicting emotions coexist in a sentence, contributing to the higher count of the "Neutral" category.
Additionally, we explored the trust scores associated with different emotions, highlighting the emotional context and community response within the Taylor Swift subreddit. This nuanced approach to emotion detection contributes to a deeper understanding of the emotional nuances present in user-generated content, uncovering the diverse range of sentiments expressed within the community.

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

Table 4.5: Counts of Emotions

Looking at the emotion distribution in Taylor Swift subreddit titles, the counts reveal an interesting pattern. The category "Positive" dominates, indicating a prevalent expression of joy, positive sentiments, and anticipation. It's noteworthy that a substantial number fall into the "Neutral" category, likely stemming from titles containing both positive and negative emotions, along with genuinely neutral titles. On the other hand, "Negative" emotions, including anger, fear, disgust, and sadness, collectively form a relatively smaller portion, suggesting an overall positive and neutral tone in the subreddit titles.

4.3 | Named Entity Recognition

At the conclusion of our Social Content Analysis, we implemented Named Entity Recognition (NER), a natural language processing (NLP) technique that focuses on identifying and categorizing entities within text into predefined categories. Entities encompass specific pieces of information, such as names of individuals, organizations, locations, dates, numerical values, and more. NER aids in extracting structured information from unstructured text, facilitating analysis and comprehension.
Entity recognition represents just one phase; following this, we must process the data, employ various approaches to recognize entities, establish connections between identified entities and real-world entities,

and finally, engage in disambiguation. During disambiguation, we rectify potential errors in entity detection and ensure accurate linking to reality. Throughout our analysis, we explored various approaches:

- **NLTK-Based NER**
  NLTK, which stands for Natural Language Toolkit, is a powerful tool for working with language in Python. In the code provided, NLTK is used for basic Named Entity Recognition (NER). This means it identifies and categorizes specific things in the text based on their roles in the sentence and the context around them. This method is fundamental and might not work well in every situation, but sometimes it performs better. The steps applied are: Tokenization and Part-of-Speech Tagging; Named Entity Chunking; Entity Extraction. We obtained these entities as the most occured:

Table 4.6: Top 10 Most Common Entities

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

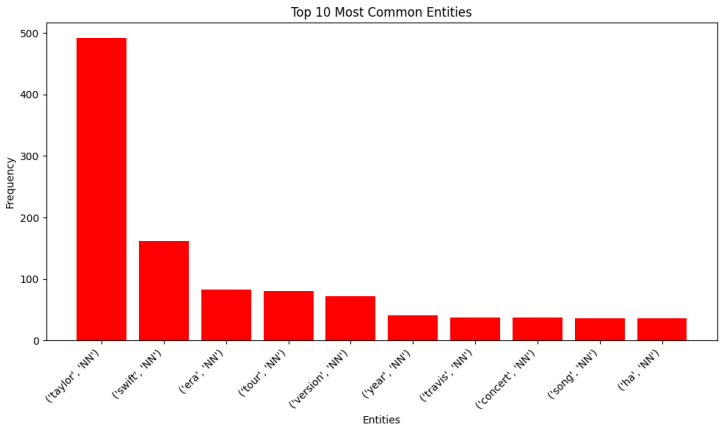

Figure 4.3: Top 10 Most Common Entities

- **Spacy-Based NER**
  Spacy is a more advanced NLP library that offers state-of-the-art tools for various NLP tasks, including Named Entity Recognition. Spacy is employed to recognize entities and assign them specific labels such as 'PERSON' or 'DATE. Spacy has an initial phase where it needs to be initialized, and you have to choose the reference language. Unfortunately, Spacy doesn't perform well with our dataframe. In fact, it only detected 1056 entities. Here, we present the most frequent entities:

Table 4.7: Entity Statistics

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

- **Dandelion, Babelfy, and Spotlight APIs**
  in conclusion, we implemented an alternative approach for entity recognition. Specifically, Dandelion [3], Babelfy [2], and Spotlight are external APIs that offer Named Entity Recognition services. These APIs employ diverse algorithms and databases to identify and categorize entities within text. Among them, the Babelfy API proved to be the most useful algorithm, yielding superior results for our dataset. It successfully identified entities and assigned them links, referencing its own website. The links provided by the entities were associated with DBpedia [4] and BabelNet, which served as points of reference for the categorization process.

5 | Conclusion

The social media analysis of Taylor Swift's online presence provided valuable insights into her popularity and impact on social media. The project utilized various techniques such as social network analysis, sentiment analysis, emotion analysis, and named entity recognition to analyze the data collected from Reddit.

The social network analysis revealed the structure and dynamics of Taylor Swift's online community, highlighting the key players and the dissemination of content within the community. The content analysis showcased the prevailing sentiments and emotions within the subreddit, reflecting the overall positive and neutral tone in the subreddit titles.

The findings showed that Taylor Swift had a strong and positive presence on social media, with a large and engaged fan base. The project also highlighted the importance of data preprocessing and the use of appropriate metrics and tools for accurate analysis.

Overall, the project was successful in achieving its aim of analyzing Taylor Swift's social media presence and providing insights into her impact on social media. The results of this analysis can be used to inform marketing and social media strategies for Taylor Swift and other public figures. Possible future developments could include expanding the analysis to other social media platforms and incorporating machine learning techniques for more advanced analysis.

6 | References

[1] Mack Ashworth. Is taylor swift time's person of the year 2023?, 2023. Published on December 6, 2023.

[2] BabelNet. Babelnet, 2024.

[3] Dandelion API. Dandelion - semantic text analytics as a service, 2024.

[4] DBpedia. Dbpedia, 2024.

[5] Sam Lansky. 2023 person of the year: Taylor swift, 2023. Published on December 6, 2023 on Time.

[6] Reddit. Reddit, 2024.

[7] Reddit. r/taylorswift, 2024.

[8] Wikipedia contributors. Reddit, 2024.