CMP7202

Web Social Media Analytics and Visualization

REPORT TITLE:

Visualizing and Analyzing Twitter Trend

Student Name: Thi Tinh Lo

Student ID: 22236226

May 2023

Contents

List of Figures

List of Tables

1. Introduction

Social media has become an integral part of our daily lives, impacting various aspects such as social, businesses, political, and technological landscapes. With the rise of social media platforms like Twitter, Facebook, Instagram and YouTube, an enormous amount of data is being generated every second. This data offers valuable insights into how people communicate, exchange information and interact with each other. Analyzing these data has become a significant factor in various areas.

Twitter is a social media platform that allows users to share short messages, known as tweets, with their followers. The latest Twitter statistics also show that there are nearly 230 million daily active users (and growing), and the United Kingdom accounted for 23.2 million users (January, 2023). It has become a popular platform for public figures, celebrities, politicians, and businesses to connect with their audiences and share their messages. Twitter's impact on the world of social media and the wider world of communications is significant, and it is likely to remain a popular platform for years to come (Oberlo, 2023).

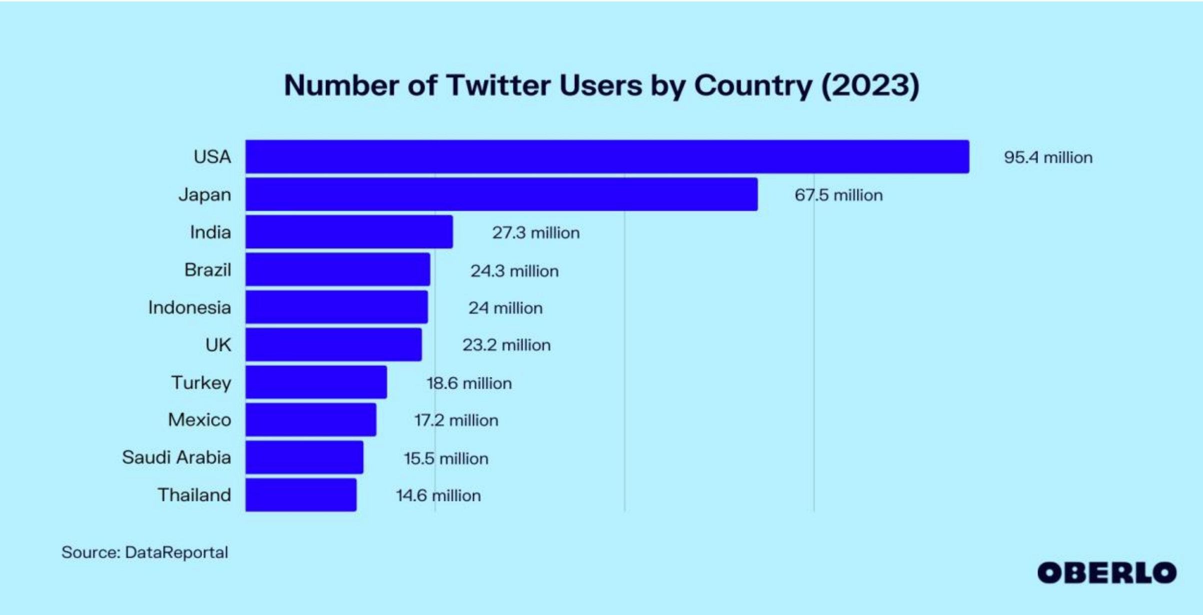

Figure 1: Number of Twitter Users by Country Q1 2023 (Oberlo, 2023)

This report focuses on visualizing and analyzing Twitter trends, specifically emphasizing the prevalent trends in the UK on March 24th, 2023. The project aims to showcase the process of preparing, manipulating, analyzing, and visualizing social media data by utilizing a range of social data analysis methods, including sentiment analysis, topic modeling, graph analysis, and artificial neural network algorithms. By doing so, the report demonstrate the effectiveness of web social media analytics and visualization in uncovering significant and useful insights from social media data, which can be useful in decision-making processes across different industries.

2. Data Collection and Preprocessing

2.1. Popular Trends on Twitter in The UK at 24th Match 2023

To retrieve data from Twitter, using the Twython library, which is a Python library that provides a convenient way to access Twitter data (twython, 2021). Then, extract the 10 most popular hashtags and topics in the UK on Friday, 24th Match 2023. These include hashtags such as “Macron”, “Year 1”, “Modi”, “Cristiano Ronaldo”, “Selena”, and so on.

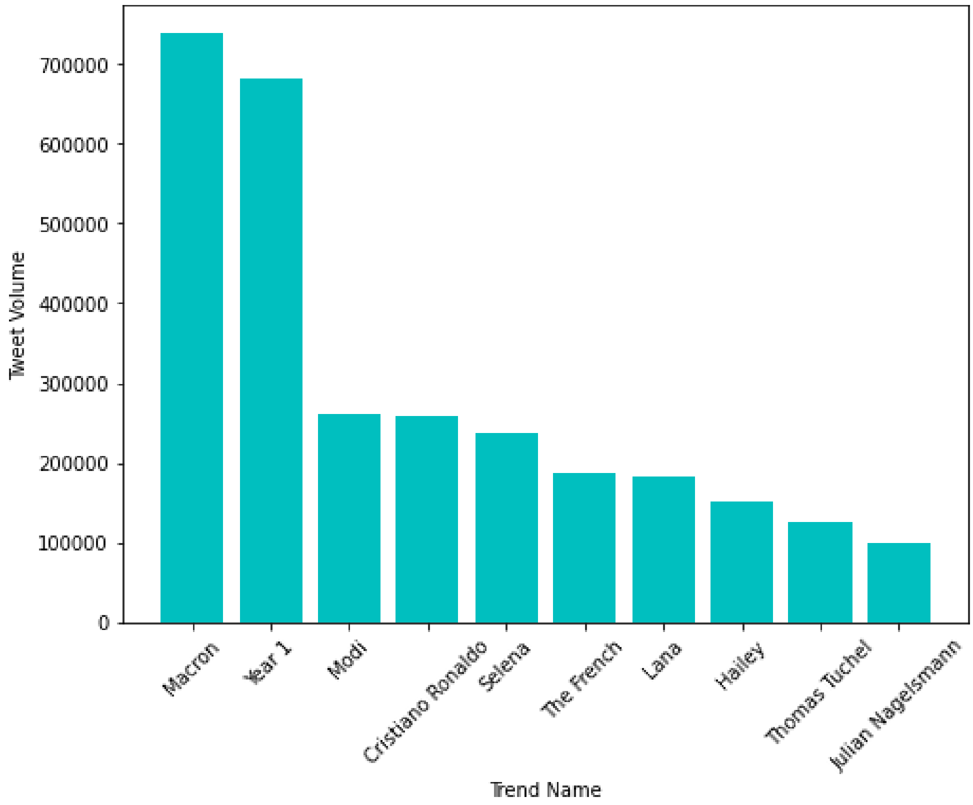

Figure 2: Popular Trends on Twitter in The UK at 24th Match 2023

2.2. Data Collection

I selected the hashtag "Macron" as my topic. To collect data from Twitter related to this topic, I will use the Twython library. The specific fields of data I am interested in are the user's name, date the tweet was created, the text of the tweet, the user's location, number of followers, hashtags used in the tweet, and the source of the tweet. I collect 2000 tweets on this topic to understand how people around the world, and in the UK, perceive and respond to French President Emmanuel Macron's policy of increasing the legal retirement age from 62 to 64. Over the past two months, protests related to this policy have attracted more than a million people. Overall, my aim is to gain insight into public on this issue.

2.3. Natural Language Processing (NLP)

Natural Language Processing (NLP), is broadly defined as the automatic manipulation of natural language, like speech and text, by software (Brownlee, 2017).

Text data obtained from natural language is often messy and unorganized. Text preprocessing is the process of cleaning and standardizing the text data so that it can be effectively analyzed and used for further learning. This involves several steps, such as removing irrelevant words, removing duplicate, converting to lowercase, stemming, lemmatization, and removing emojis.

After undergoing this process, the "text" column will contain only words in their base or root form, free of any punctuation, tags, or extraneous characters. This makes the text ready for computers to understand, interpret, and visualize.

text
million french protest gather pari call remov ...
sum fear globalist technocrat happen right fra...
franc firefight follow polic begin stand choos...
macron oop moment macron realis he' wear expen...
million peopl street pari want dictat emmanuel...

Figure 3: Tweets' text after Preprocessing

3. Visualization

Instead of presenting data in multiple tables with numbers, visualization can be used to display the data in a more intuitive and graphical manner. This approach can help people better comprehend and interpret the data.

3.1. Word Cloud

A frequently used visualization method is the word cloud, which offers a rapid summary of the most commonly appearing words in a large body of text.

The most prevalent words in tweets are “remove”, “macron”, “protest”, “french”, “million”, ect.

Figure 4: Word Cloud Prevalent Words in Tweets

3.2. Location

In order to understand the geographical patterns of tweets, I utilize Twitter's geotagging feature. However, about 67% (1336/2000) of tweets are geotagged. The tweets predominantly come from European countries, such as Denmark, England, Sweden, France, as well as America. This may be due to the time zone difference, as I collected the data at 7 PM UK time. As a result, the majority of the data I obtained was from European and American countries.

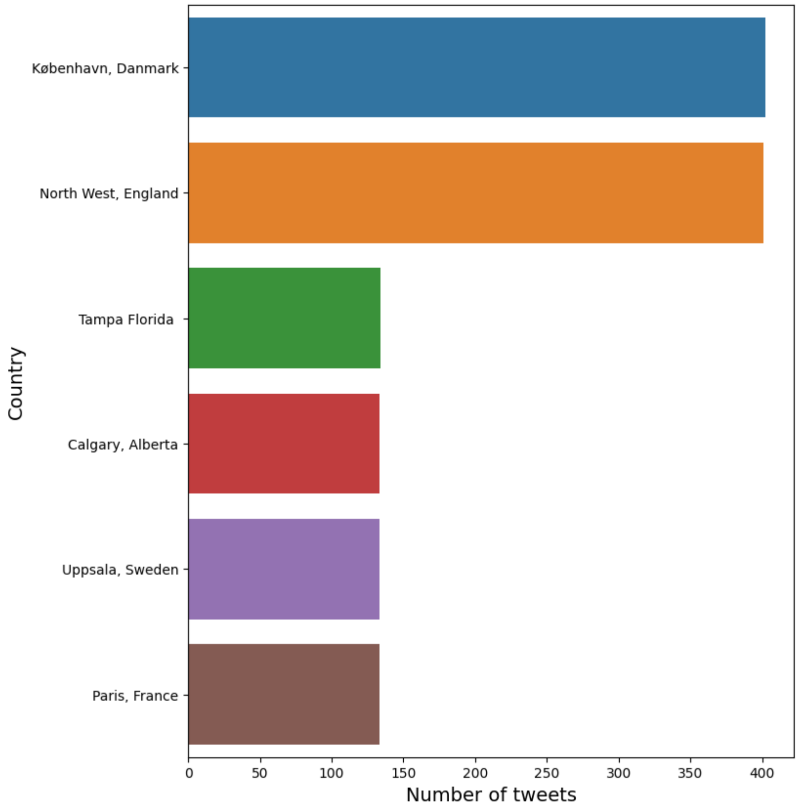

Figure 5: Countries with Number of Tweets

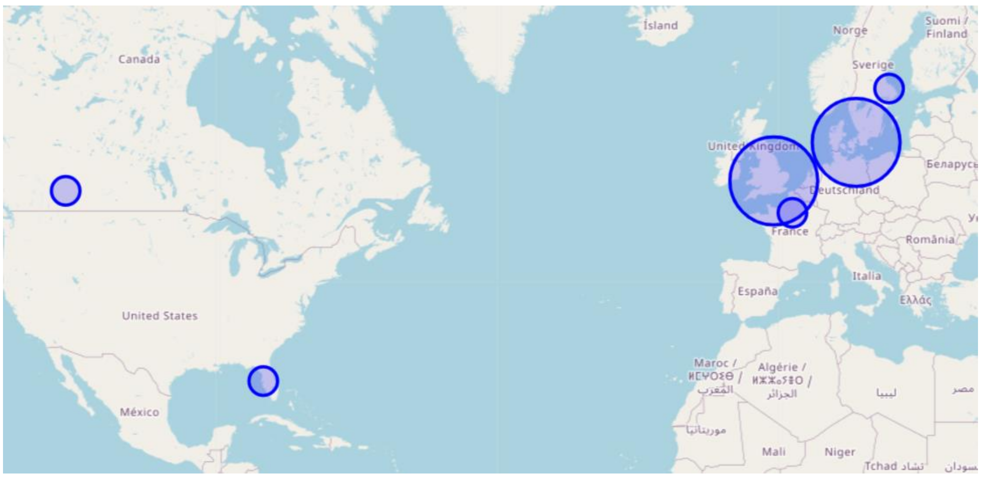

Figure 6: Map Distribution of Tweets

Through the visualization of tweet locations, analysts can gain valuable insights into diverse aspects of social media activity, including the geographic distribution of topics, sentiment, or user demographics in different regions.

3.3. Platform

In 2023, there are approximately 6.92 billion smartphone users globally, which is equivalent to 86.29% of the world's population who own a smartphone (Smartphones, n.d.). This widespread adoption of mobile phones has led to their increased use for browsing the internet and interacting with social media platforms.

Based on the data I obtained, only around 7% of people use the web, while the vast majority access these services through their mobile phones.

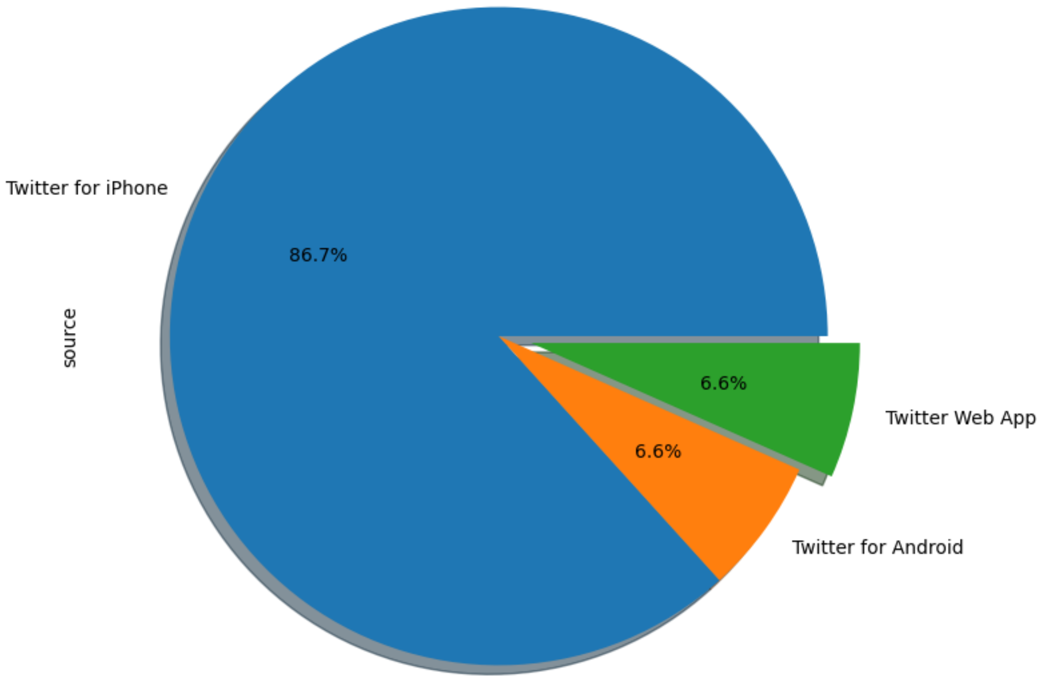

Figure 7: Platforms Ranked by Tweet Volume

By analyzing this trend, businesses and organizations can better understand the impact of mobile internet usage on their ability to engage with customers and audiences through social media and other digital channels.

4. Sentiment Analysis use TextBlob

Another important question when analyzing text data is how people's feelings or emotions are expressed in the text? I use TextBlob in Python library to perform sentiment analysis. There is a higher ratio of negative opinions compared to positive opinions.

When examining text data, it is crucial to understand how people convey their emotions and feelings through their language. To accomplish this task, I utilized the TextBlob Python library, which allowed me to perform sentiment analysis on the text. Based on my analysis, the data showed a greater proportion of negative opinions expressed compared to positive opinions.

When analyzing text data, an important question to consider is how people convey their emotions or sentiments through the text. To accomplish this, I used the TextBlob library which is a Python library for processing textual data. It provides a simple API for diving into common natural language processing (NLP) tasks such as part-of-speech tagging, noun phrase extraction, sentiment analysis, classification, translation, and more (Loria, 2020).

TextBlob provides the ability to compute the polarity and subjectivity of tweets. It then classifies the tweet as 'positive' (if polarity > 0), 'negative' (if polarity < 0), or 'neutral' (if polarity = 0) based on the polarity value.

	text	subjectivity	polarity	sentiment
0	million french protest gather pari call remov ...	0.000000	0.000000	neutral
1	sum fear globalist technocrat happen right fra...	0.267857	0.142857	positive
2	franc firefight follow polic begin stand choos...	0.000000	0.000000	neutral
3	macron oop moment macron realis he' wear expen...	0.000000	0.000000	neutral
4	million peopl street pari want dictat emmanuel...	0.000000	0.000000	neutral
5	million protest call “macron dictator” remov m...	0.000000	0.000000	neutral

Figure 8: DataFrame Sentiment Scores and Types of Tweets.

Visualize the data, indicating that a number of negative opinions were expressed in comparison to positive ones.

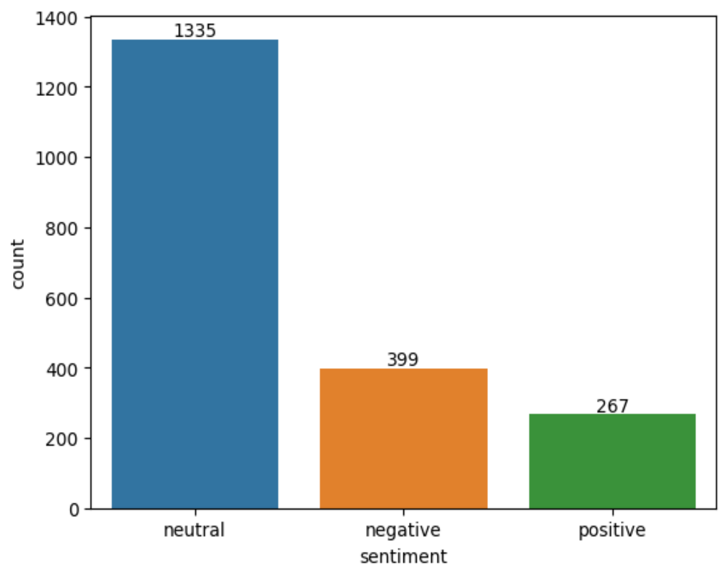

Figure 9: Distribution of Sentiment Categories (tweets)

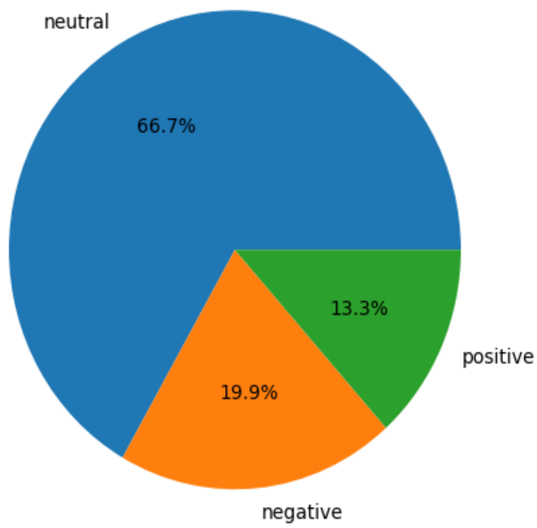

Figure 10: Distribution of Sentiment Categories (%)

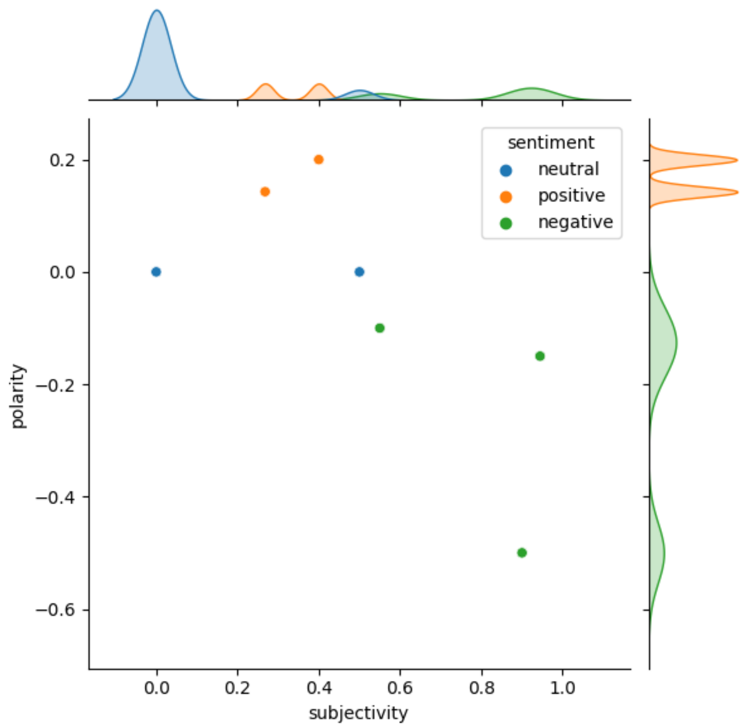

Figure 11: Scatter Plot of Distribution Sentiment by Categories

The most commonly occurring words in positive tweets are 'franc', 'happen', 'stand', 'globalist', 'firefight', etc.

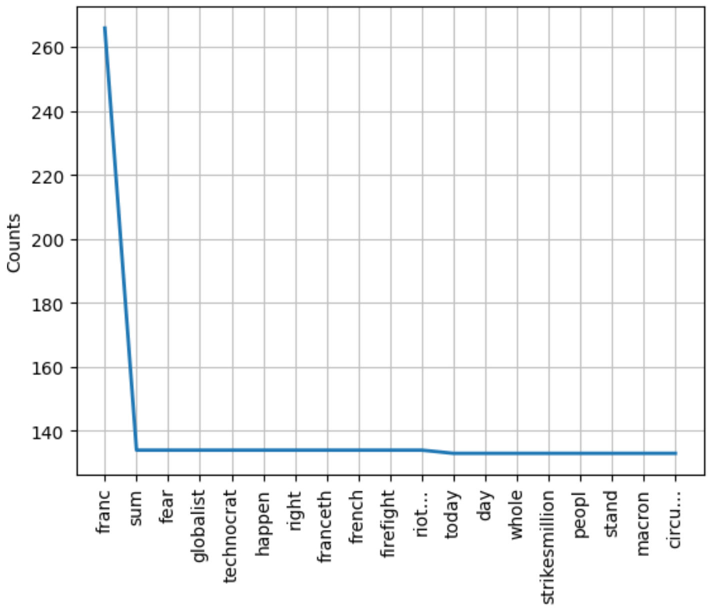

Figure 12: Words Frequency Distribution for Positive Tweets

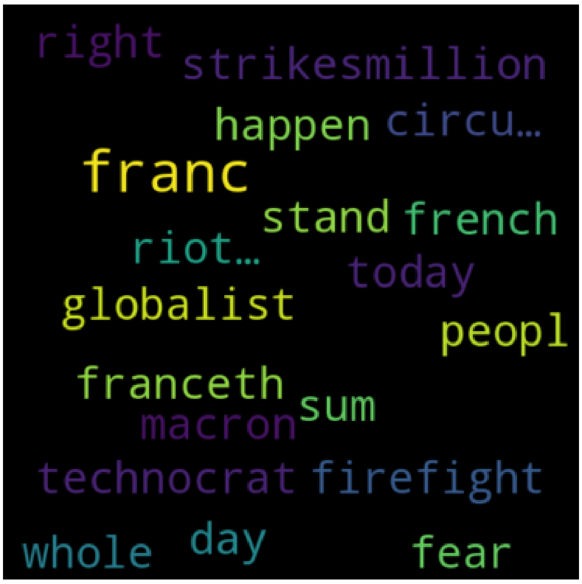

Figure 13: Word cloud for positive tweets

The most commonly occurring words in negative tweets are ‘macron’, ‘protest’, ‘million’, ‘people’, ‘remove’, ‘police’, ect.

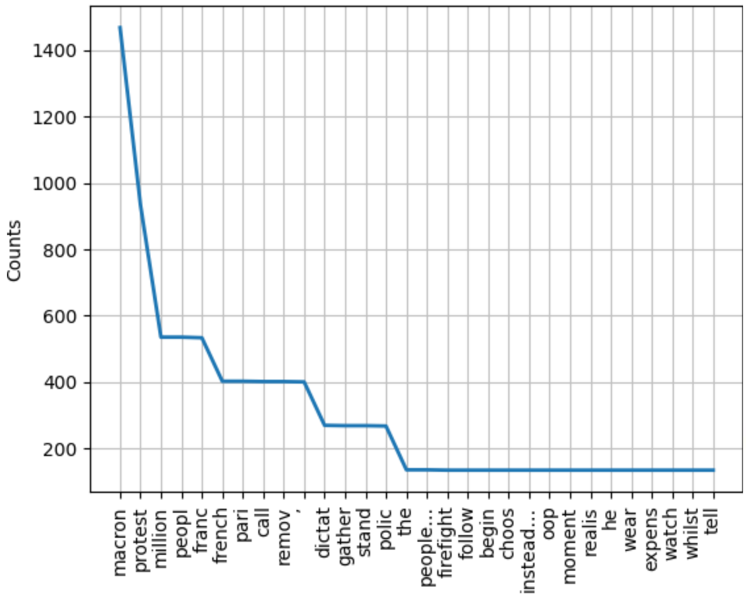

Figure 14: Words frequency Distribution for Negative Tweets

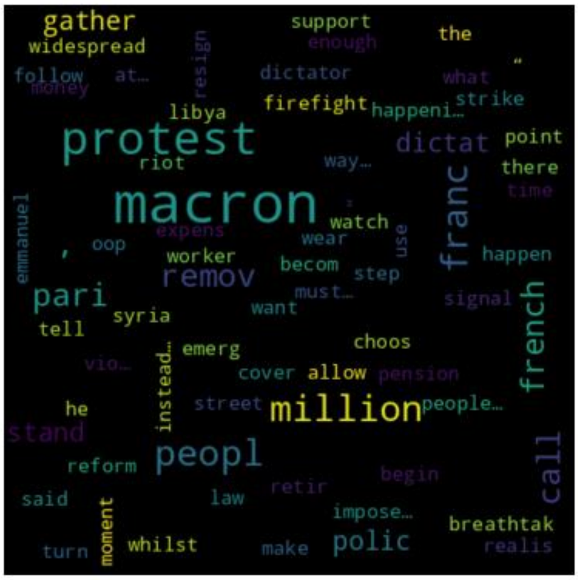

Figure 15: Word Cloud for Negative Tweets

By analyzing the tone, emotion, and attitude expressed in large volumes of social media data, customer feedback, or other online content, sentiment analysis can provide valuable information that can be used to make informed decisions, improve customer experience, manage brand reputation, and even detect potential crises or risks.

5. Topic Modelling

Approximately 70% of text data falls under the neutral category, indicating the absence of a distinct positive or negative sentiment. To gain valuable insights from such data, I employ topic modeling, which is a statistical model that reveals the underlying "topics" present in a set of documents. Topic modeling is a popular text-mining technique used to uncover hidden semantic structures in textual data (Topic model, n.d.). This technique help to identify common themes or topics that are frequently mentioned in the text.

I utilize Gensim, a Python library designed for handling large collections of text documents, specifically for natural language processing and information retrieval purposes. It offers efficient implementations of commonly used algorithms such as Latent Semantic Indexing (LSI) and Latent Dirichlet Allocation (LDA) for topic modeling (gensim, 2023). In this context, I used CoherenceModel to evaluate and compare the performance of LSI and LDA models.

```
LSI Coherence Score: 0.24813808451385244
LDA Coherence Score: 0.28127232391308415
```

Figure 16: Coherence Score of LSI & LDA Models

The results indicate that the LDA model outperforms the LSI model in terms of coherence, implying that the topics identified by the LDA model are more meaningful and logically connected than those identified by the LSI model

After implementing the LDA model, I obtained six topic models as illustrated below:

Figure 17: Six Topic Models using LDA

6. Neural Network

6.1. Build Sentiment Analysis Model use Artificial Neural Network

Artificial neural networks are computational models inspired by the nervous system of living beings. They have the ability to acquire and maintain knowledge (information based) and can be defined as a set of processing units, represented by artificial neurons, interlinked by a lot of interconnections (artificial synapses), implemented by vectors and matrices of synaptic weights (Spatti, 2017).

In this project, Artificial Neural Network (ANN) is be applied in sentiment analysis to classify text data into different sentiment categories (positive, negative and neutral tweets). Using TensorFlow which is an open-source software library for machine learning and artificial intelligence across a range of tasks but has a particular focus on training and inference of deep neural networks (TensorFlow, n.d.), to build the model.

Implementing training model for sentiment analysis using Tensorflow with the following parameters:

No.	Parameters	Values
1	Maximum length of the text sequences (max_len)	500

2	Training data – Testing data	70% - 30%
3	LSTM layer (output: dimensional hidden states)	64
4	Dense layer	256
5	Activation function	Sigmoid
6	Batch size (take 80 tweets in each iteration and train)	80
7	Epochs (the model will train on the data 10 times)	6

Table 1: Parameters for Training a TensorFlow Model

After final epoch of training (6th). On the training data, the value of the loss function of the model is 0.4835, the accuracy is 0.7653. On the validation data, the value of the loss function is 0.5287, the accuracy of the model is 0.7450.

```
Epoch 1/6
315/315 [==============================] - 154s 483ms/step - loss: 0.6214 - accuracy: 0.6379 - val_loss: 0.5393 - val_accuracy: 0.7211
Epoch 2/6
315/315 [==============================] - 147s 466ms/step - loss: 0.5147 - accuracy: 0.7469 - val_loss: 0.5159 - val_accuracy: 0.7446
Epoch 3/6
315/315 [==============================] - 147s 468ms/step - loss: 0.4942 - accuracy: 0.7617 - val_loss: 0.5240 - val_accuracy: 0.7343
Epoch 4/6
315/315 [==============================] - 150s 478ms/step - loss: 0.4837 - accuracy: 0.7687 - val_loss: 0.5203 - val_accuracy: 0.7443
Epoch 5/6
315/315 [==============================] - 146s 464ms/step - loss: 0.4764 - accuracy: 0.7731 - val_loss: 0.5246 - val_accuracy: 0.7450
Epoch 6/6
315/315 [==============================] - 151s 478ms/step - loss: 0.4835 - accuracy: 0.7653 - val_loss: 0.5287 - val_accuracy: 0.7450
Training finished !!
```

Figure 18: Loss and Accuracy Metrics on Training and Validation Data

The evaluation results of the model on the test data, the value of the loss function is 0.5237, and the accuracy of the model on the test data is 0.7470.

```
375/375 [==============================] - 31s 80ms/step - loss: 0.5237 - accuracy: 0.7470
```

Figure 19: Loss and Accuracy Metrics on Testing Data

To evaluate the performance of the model I use Confusion matrix (error matrix) which is a table with two rows and two columns that reports the number of true positives, false negatives, false positives, and true negatives. This allows more detailed analysis than simply observing the proportion of correct classifications (accuracy) (Confusion matrix, n.d.).

		Predicted condition	
	Total population	**Positive (PP)**	**Negative (PN)**
Actual condition	**Positive (P)**	True positive (TP) **0.73%**	False negative (FN) **0.27%**
	Negative (N)	False positive (FP) **0.23%**	True negative (TN) **0.77%**

Table 2: Confusion Matrix of Sentiment Analysis Model

The table provides a summary of the binary classification model's performance. The model accurately predicted the positive class 73% of the time and the negative class 77% of the time. The Type 1 error rate is 0.23, indicating that 23% of the time, the model mistakenly classified a negative review as positive.

6.2. Implementing a Sentiment Analysis Model on Twitter Datasets

After utilizing the sentiment analysis model on the tweets data, the following outcomes were obtained:

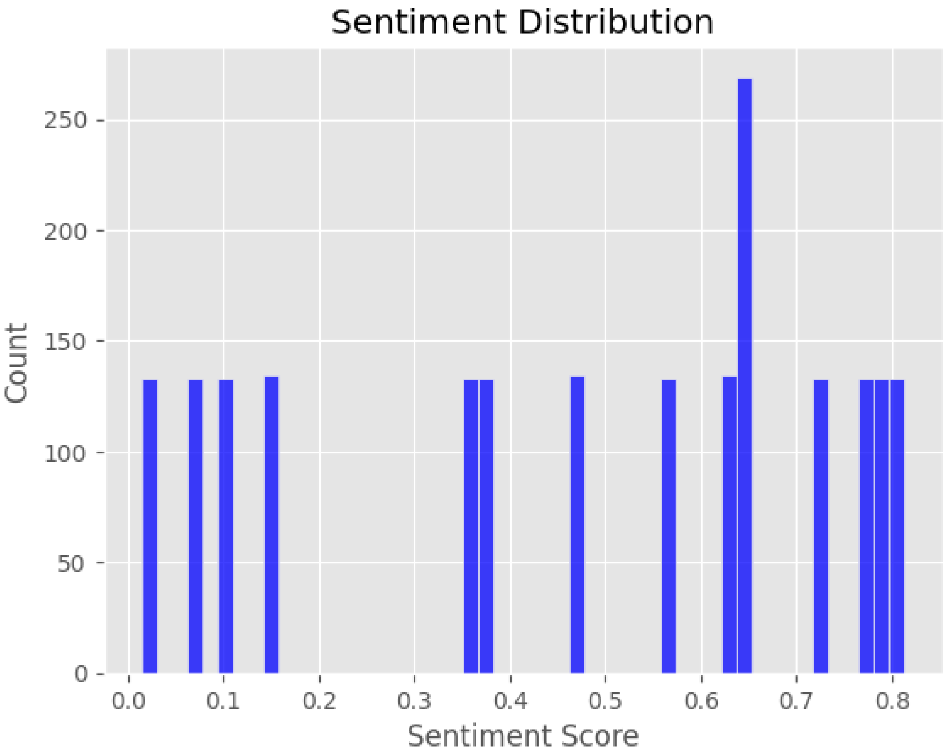

Figure 20: Sentiment Distribution using Tensorflow (scores)

	count	label
0	936	Neutral
-1	533	Negative
1	532	Positive

Table 3: Sentiment Values by Categories Using Tensorflow

Setting threshold: More than 0.7 - positive tweets(1), lower 0.3 - negative tweets(-1), between 0.3 and 0.7 - Neutral tweets (0)

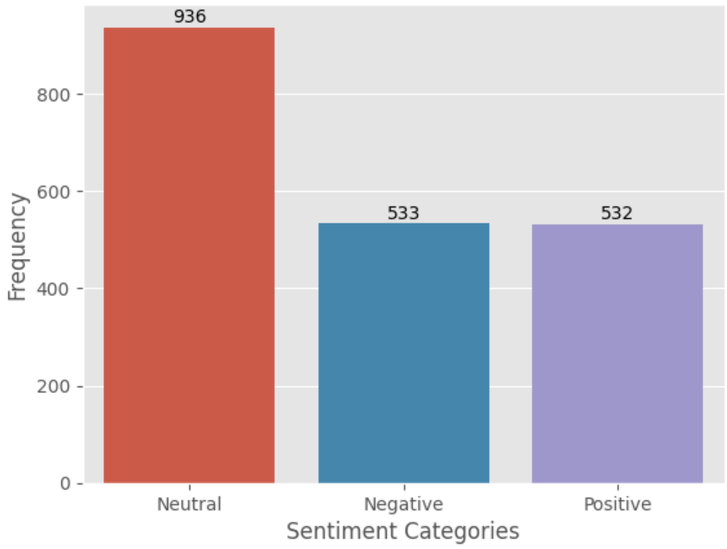

Figure 21: Distribution of Sentiment Categories using Tensorflow (tweets)

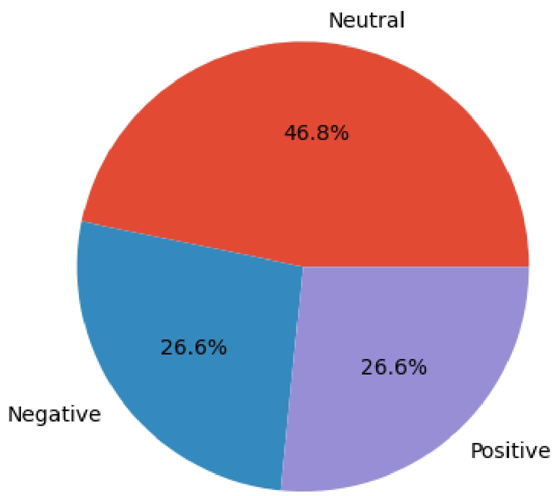

Figure 22: Distribution of Sentiment Categories using Tensorflow (%)

Comparing to sentiment analyses use TextBlob. There are a little difference between Negative and positive tweets. Using TextBlod Negative(399 tweets – 19.9%) over than Positive tweets (267 tweets – 13.3%), by using ANNs Negative (533 tweets) and Positive (532 tweets) almost similar.

6.3. Automatic summarization For Positive, Neutral, and Negative Tweets

Automatic summarization is the process of shortening a set of data computationally, to create a subset (a summary) that represents the most important or relevant information within the original content (Automatic summarization, n.d.). I use Automatic summarization technique to summarize the information from categories (positive, neutral, and negative tweets) that I got from Sentiment Analysis use Artificial Neural Network.

Extractive Summarization for Positive tweets

```
'macron slowly surely turning france violent dictatorship. riot police signal support protes
ters.. 9 million people wrong! france covered protests strikes macron said there's money all
ow workers retire at…. macron slowly surely turning france violent dictatorship. riot police
signal support protesters.. 9 million people wrong! france covered protests strikes macron s
aid there's money allow workers retire at…. macron slowly surely turning france violent dict
atorship. riot police signal support protesters.. 9 million people wrong! france covered pro
tests strikes macron said there's money allow workers retire at…. macron slowly surely turni
ng france violent dictatorship. riot police signal support protesters.. 9 million people wro
ng! france covered protests strikes macron said there's money allow workers retire at…. macr
on slowly surely turning france violent dictatorship. riot police signal support protester
s.. 9 million people wrong! france covered protests strikes macron said there's mo...'
```

Figure 23: Extractive Summarization for Positive Tweets

Extractive Summarization for Neutral tweets

```
'million french protesters gather paris calling removal macron dictator. sum fears globalist
technocrats happening right france.the french firefighters riot…. france - firefighters foll
ow police begin stand down, choosing stand people instead.…. million french protesters gathe
r paris calling removal macron.. syria libya -this france. million french protesters gather
paris calling removal macron dictator. sum fears globalist technocrats happening right franc
e.the french firefighters riot…. france - firefighters follow police begin stand down, choos
ing stand people instead.…. million french protesters gather paris calling removal macron..
syria libya -this france. million french protesters gather paris calling removal macron dict
ator. sum fears globalist technocrats happening right france.the french firefighters riot….
france - firefighters follow police begin stand down, choosing stand people instead.…. milli
on french protesters gather paris calling removal macron.. syria libya -this fra...'
```

Figure 24: Extractive Summarization for Neutral Tweets

Extractive Summarization for Negative tweets

```
'macron - oops moment macron realises he’s wearing expensive watch whilst telling french peo
ple must…. france - today day whole france strikes.millions people stand macron, circu…. who
le france behind anti-reform movement… macron lost control country. macron - oops moment mac
ron realises he’s wearing expensive watch whilst telling french people must…. france - today
day whole france strikes.millions people stand macron, circu…. whole france behind anti-refo
rm movement… macron lost control country. macron - oops moment macron realises he’s wearing
expensive watch whilst telling french people must…. france - today day whole france strikes.
millions people stand macron, circu…. whole france behind anti-reform movement… macron lost
control country. macron - oops moment macron realises he’s wearing expensive watch whilst te
lling french people must…. france - today day whole france strikes.millions people stand mac
ron, circu…. whole france behind anti-reform movement… macron lost control count...'
```

Figure 25: Extractive Summarization for Negative Tweets

7. Network Analysis

Network analysis is a method of studying the relationships between entities, such as individuals, organizations, or other types of nodes, and the interactions between them. It involves the use of mathematical and statistical models to analyze the structure of networks, including measures of centrality, density, clustering, and community detection (Carrington P. J., Models and Methods in Social Network Analysis_ page 8).

7.1. Data Collection

I utilized tweepy, a Python library that provides access to the Twitter API, to retrieve a list of users (Python – API, 2020). Specifically, I retrieved the followers of the Twitter account belonging to Glenn Greenwald, an influential journalist and political commentator who is recognized for his opinions on government surveillance and civil liberties. His tweets cover various political and social topics, and he has over 2 million followers.

The dataset that I obtained includes 5030 nodes and 5063 edges, representing the relationships between the users who follow Glenn Greenwald.

7.2. Network Analysis

Network analysis involves understanding the relationships between nodes within a network. There are three common methods for conducting network analysis. The first involves calculating network centrality, which identifies the most important nodes in the network. The second method involves calculating network density, which measures the overall connectedness and shape of the network. The third method is calculating network modularity, which assesses whether nodes cluster into distinct communities within the network.

7.2.1. Network Centrality

In network analysis, one way to measure the most important nodes is through network centrality. There are four types of centrality measures: degree centrality, closeness centrality, betweenness centrality, and eigenvector centrality.

- **Degree centrality** measures the total number of connections of each node (weighted degree) (Carrington P. J., Models and Methods in Social Network Analysis_ page 59).
    - *Weighted degree*: combining edge weight with degree
    - *Directed degrees*: combining edge direction with degree (in-degree and out-degree)
- **Closeness centrality** measures the average length of the shortest path between a given node and all other nodes in the network (Carrington P. J., Models and Methods in Social Network Analysis_ page 61).
- **Betweenness centrality** measures the number of shortest paths across the network (Carrington P. J., Models and Methods in Social Network Analysis_ page 62).
- **Eigenvector centrality** measures each node's connection to other highly connected nodes (Carrington P. J., Models and Methods in Social Network Analysis_ page 70).

Id	Weighted Degree
16076032	5000.0
2278187948	563.0
1.48789E+18	202.0
1.45478E+18	184.0
1.5868E+18	61.0
1.60714E+18	16.0
1.63794E+18	11.0
1.63762E+18	10.0
1.63747E+18	10.0
1.63752E+18	10.0

Table 5: Degree Centrality (Weighted Degree), Top 10 Nodes

Id	Degree
16076032	4013
2278187948	561
1.48789E+18	202
1.45478E+18	183
1.5868E+18	60
1.60714E+18	16
1.58797E+18	10
9.64455E+17	7
1.35914E+18	6
1.6395E+18	5

Table 4: Degree Centrality (Directed Degree), Top 10 Nodes

Id	Closeness Centrality
1.6395E+18	1.0
1.5868E+18	1.0
2278187948	1.0
1.48789E+18	1.0
1.45478E+18	1.0
1.60714E+18	1.0
1.58797E+18	1.0
9.64455E+17	1.0
1.35914E+18	1.0
1.62757E+18	1.0

Table 6: Closeness Centrality, Top 10 Nodes

Id	Betweenness Centrality
2278187948	550.0
1.48789E+18	197.0
1.45478E+18	171.5
1.5868E+18	53.0
1.60714E+18	14.0
1.58797E+18	8.5
9.64455E+17	6.0
1.35914E+18	5.0
1.62757E+18	4.0
1.58823E+18	4.0

Table 7: Betweenness Centrality, Top 10 Nodes

Id	Eigenvector Centrality
1.57115E+18	1.0
1.58812E+18	0.950549
1.57812E+18	0.950549
1.24419E+18	0.950549
1.27266E+18	0.524725
1.54325E+18	0.524725
1.59731E+18	0.524725
1.63905E+18	0.524725
1.5912E+18	0.524725
1.59741E+18	0.524725

Table 8: Eigenventor Centrality, Top 10 Nodes

No.	Statistics	Value
1	Network Interpretation	Directed
2	Average Degree	1.007
3	Average Weighted Degree	1.203
4	Average Path length	1.167
5	Number of Weakly Connected Components	1
6	Number of Strongly Connected Components	5030

Figure 26: Centrality Statistics Values

An average degree of 1.007 means that on average, each vertex in the graph is connected to slightly more than one other vertex. This indicates a relatively sparse graph, where the majority of vertices have only a small number of edges incident to them. The average path length of 1.167 indicates that it takes less than two steps to reach any node in the network from any other node. And there is only one weakly connected component, indicating that all nodes are connected to each other in some way.

In general, the graph has direct and tightly connected nodes with short paths between them. There are strong connections with a directed path from any node.

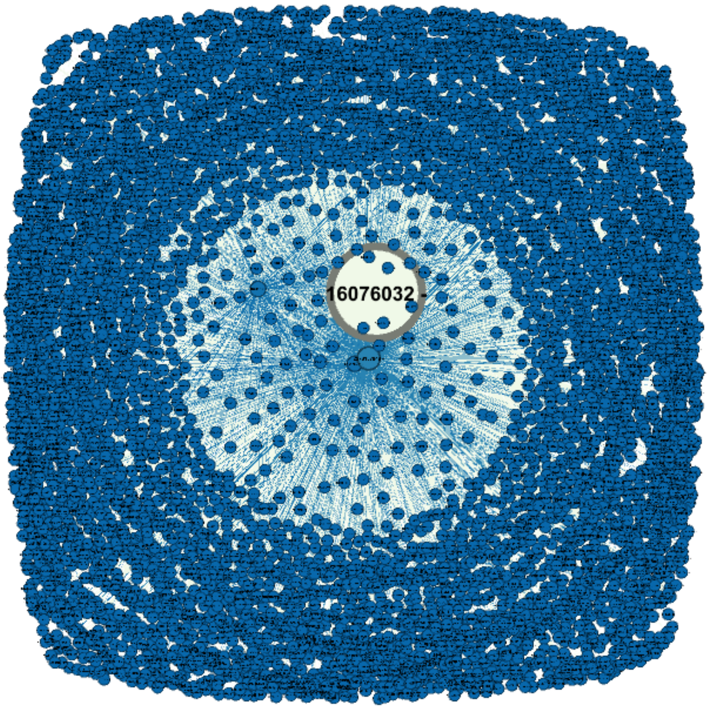

Figure 27: Degree Centrality Graph Visualization, Size & Color = Degree

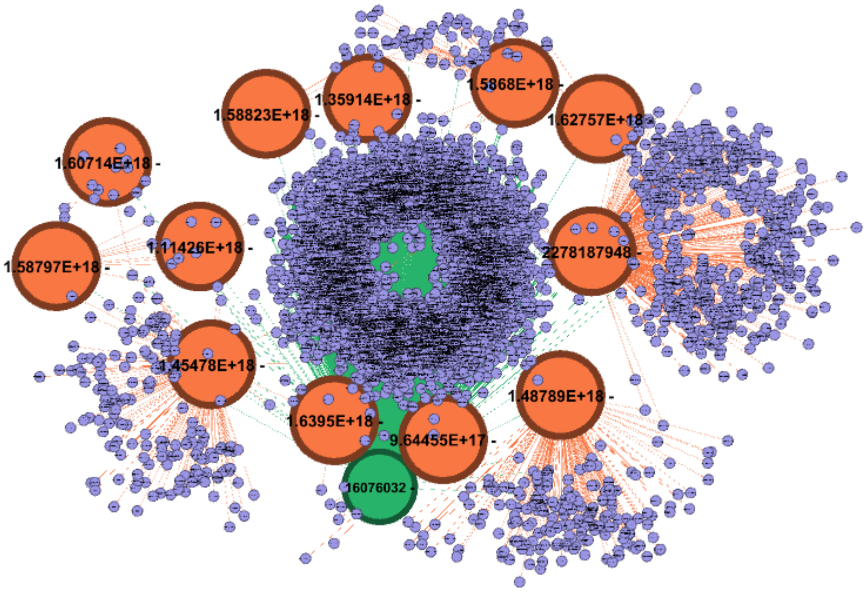

Figure 28: Closeness Centrality Graph Visualization, Size & Color = Closeness Centrality

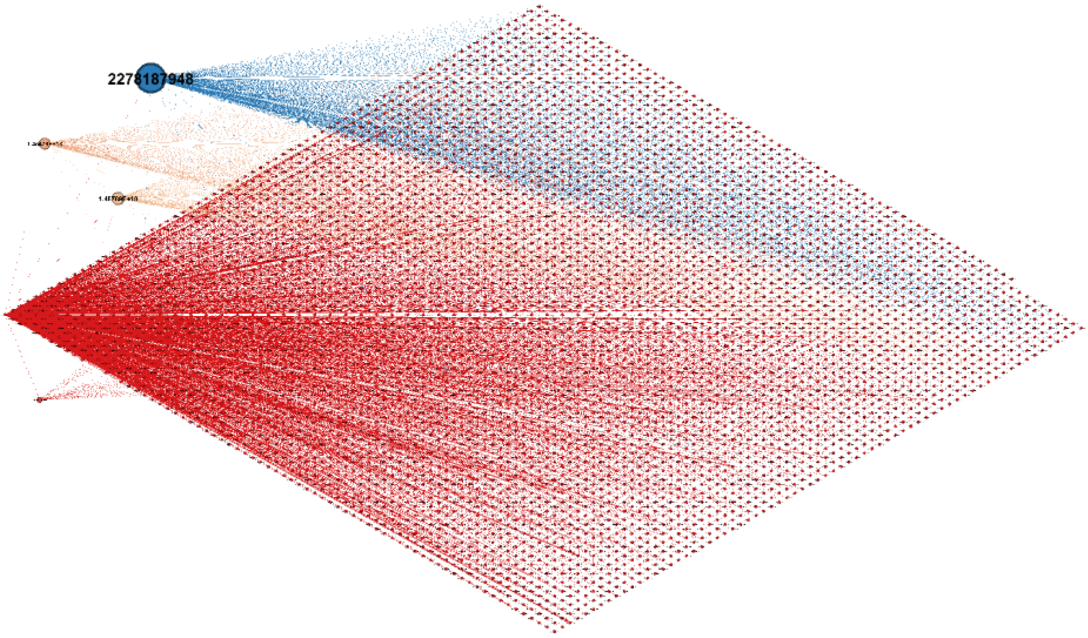

Figure 29: Betweenness Centrality Graph Visualization, Size & Color = Betweenness Centrality

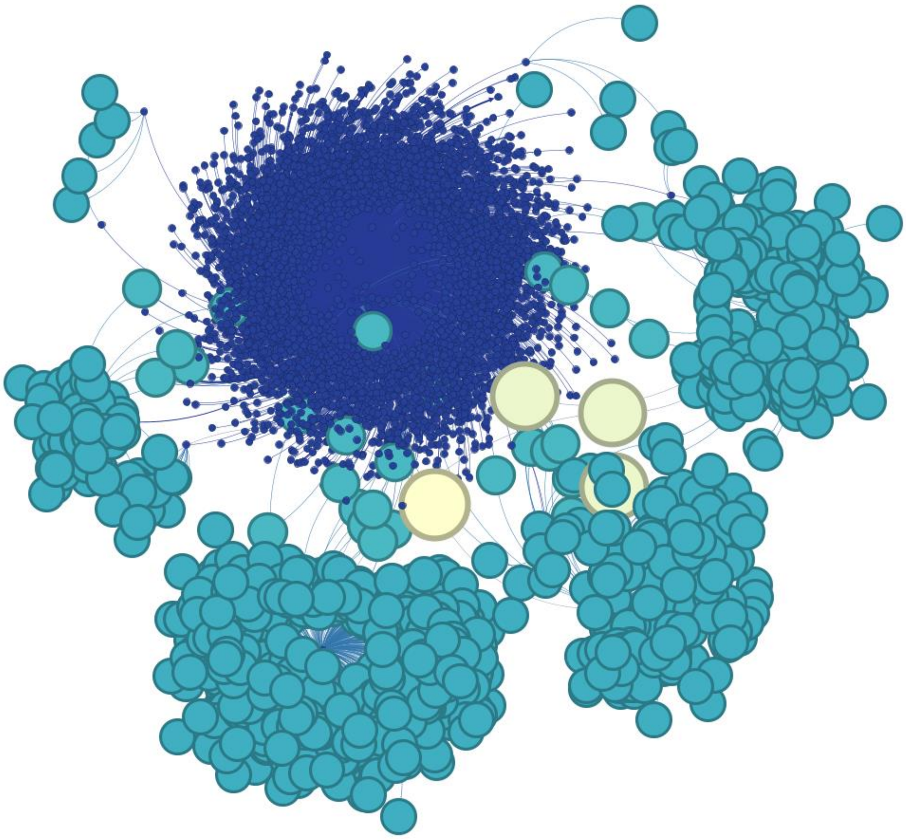

Figure 30: Eigenvector Centrality Graph Visualization, Size & Color = Eigenvector Centrality

7.2.2. Network Modularity:

Measure of the structure of networks or graphs which measures the strength of division of a network into modules (also called groups, clusters or communities*)* (Modularity (networks), n.d.).

The network has been clustered into 12 distinct communities.

No.	Statistics	Value
1	Modularity	0.303
2	Modularity with resolution	0.303
3	Number of Communities	12

Table 9: Community Statistic Values

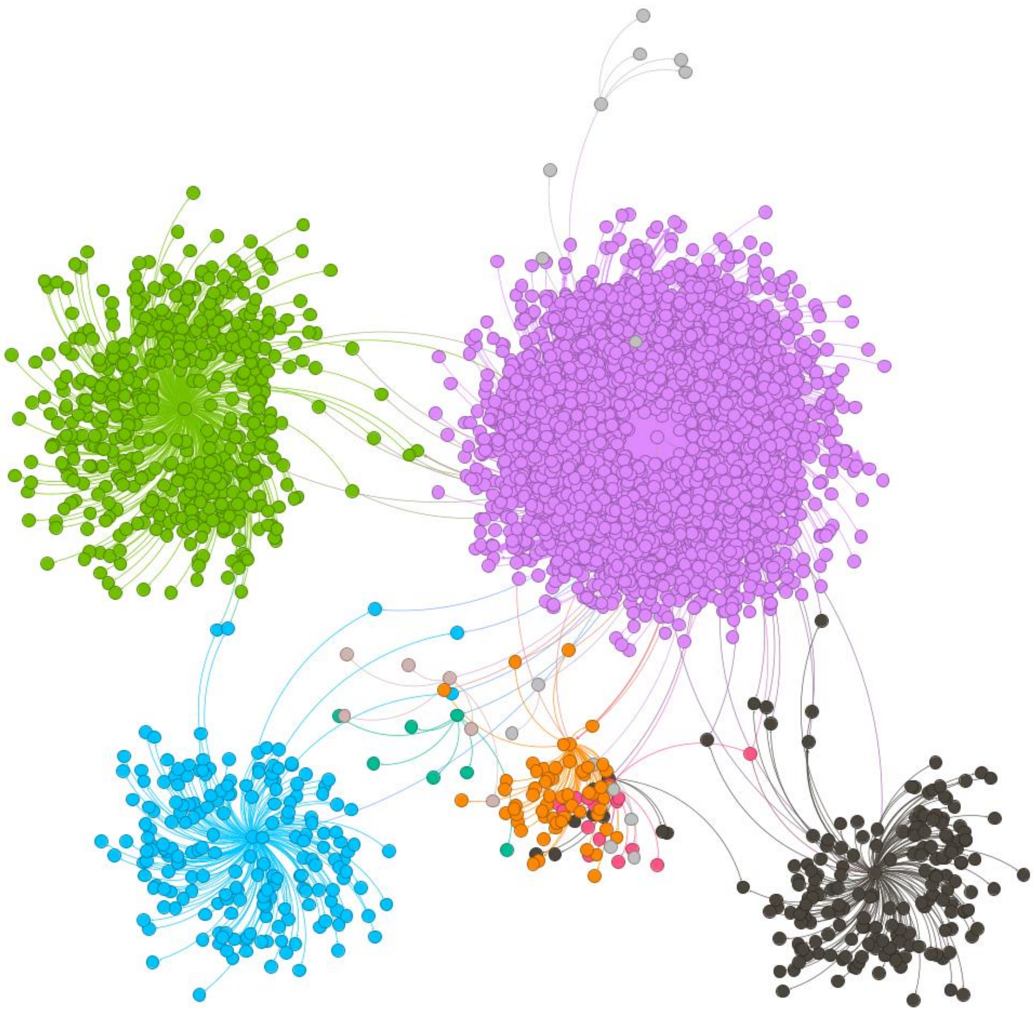

Figure 31: Community Graph Visualization, Color = Modularity Class

7.2.3. Network Density:

Measurement indicating the percentage of potential connections that actually exist among nodes in a network (Network science, n.d.).

Network-level measurement:

$$Network\ density = \frac{\text{Number of edges in network}}{\text{Number of possible edges in network}}$$

In this context, the network density is close to 0. This is because that the data is composed of Twitter followers for a particular account identified ID 16076032. There are minimal connections or edges among the followers themselves, with most connections originating from each follower towards the account with ID 16076032. This results in a lack of interconnectivity between the nodes, leading to a low score for network density.

8. Conclusion

In conclusion, this project has demonstrated the value of social media analytics in understanding Twitter trends. By applying various techniques such as visualization, sentiment analysis, topic modeling, and graph analysis, we were able to extract insights and identify patterns in the data. The application of a community detection algorithm allowed us to discover communities within the graph and analyze their characteristics. The results obtained from this project can have implications in a range of fields such as marketing, politics, and social sciences. It highlights the importance of data analytics in understanding social media trends, and the potential of social media platforms in providing insights into public opinion, behavior, and preferences.

Overall, this project has provided a practical and hands-on experience in the application of social media analytics techniques, and the importance of a multidisciplinary approach in analyzing complex social data. As social media continues to play an increasingly important role in our society, the techniques and methods applied in this project will continue to be relevant in understanding and analyzing social media trends.

References

Automatic summarization. (n.d.). Retrieved from Wikipedia_AS: https://en.wikipedia.org/wiki/Automatic_summarization

Brownlee, J. (2017). *Deep Learning for Natural Language Processing (page 2).*

Carrington, P. J. (n.d.). *Models and Methods in Social Network Analysis – page 70.*

Carrington, P. J. (n.d.). *Models and Methods in Social Network Analysis (page 61).*

Carrington, P. J. (n.d.). *Models and Methods in Social Network Analysis_ page 59.*

Carrington, P. J. (n.d.). *Models and Methods in Social Network Analysis_ page 59.*

Carrington, P. J. (n.d.). Models and Methods in Social Network Analysis_ page 61.

Carrington, P. J. (n.d.). *Models and Methods in Social Network Analysis_ page 62.*

Carrington, P. J. (n.d.). *Models and Methods in Social Network Analysis_ page 70.*

Carrington, P. J. (n.d.). Models and Methods in Social Network Analysis_ page 8.

Confusion matrix. (n.d.). Retrieved from Wikipedia_CM: https://en.wikipedia.org/wiki/Confusion_matrix

gensim. (2023, Mar 10). Retrieved from https://pypi.org/project/gensim/

Loria, S. (2020). *TextBlob*. Retrieved from TextBlob: Simplified Text Processing: https://textblob.readthedocs.io/en/dev/

Modularity (networks). (n.d.). Retrieved from Wikipedia: https://en.wikipedia.org/wiki/Modularity_(networks)

Network science. (n.d.). Retrieved from Wikipedia: https://en.wikipedia.org/wiki/Network_science

Oberlo. (2023). Retrieved from Number of Twitter Users by Country: https://www.oberlo.com/statistics/number-of-twitter-users-by-country

Python – API. (2020, Jun 05). Retrieved from Python – API.followers() in Tweepy: https://www.geeksforgeeks.org/python-api-followers-in-tweepy/

Smartphones. (n.d.). Retrieved from bankmycell: https://www.bankmycell.com/blog/how-many-phones-are-in-the-world

Spatti, I. N. (2017). *Artificial Neural Networks (page 5).*

TensorFlow. (n.d.). Retrieved from Wikipedia: https://en.wikipedia.org/wiki/TensorFlow

Topic model. (n.d.). Retrieved from Wikipedia: https://en.wikipedia.org/wiki/Topic_model

twython. (2021, Jul 16). Retrieved from Twython: https://pypi.org/project/twython/

Appendix A: Code Python

No1_get_tweet_trends_UK.ipynb – Get the current trends in the UK (id=23424975) on 2023/03/24

No2_get_data_hashtag_Macron.ipynb – Gathering tweets from Twitter using Twython API

No3_Preprocessing & Sentiment Analysis use TextBlob & Topic Models.ipynb – Preprocessing & Analysing sentiment by using TextBlob

Build_model_tweets_sentiment_analyses_Tensorflow.ipynb – Create a sentiment analysis model for tweets using Tensorflow and save the file name 'sentiment_model.h5'

No4_Sentiment Analyses use Tensorflow & Extractive_Summarization .ipynb – Analysing sentiment by using ANNs and Extractive_Summarization for sentiment categories.

No5_get_followers_graph.ipynb – Get followers of twitter account ID #16076032