CMP 7202

Web Social Media Analytics and Visualisation

Assessment 2

Statistical Analysis and Text Mining

Sachin

Student Id: 23235298

BIRMINGHAM CITY
University

Faculty of Computing, Engineering, and The Built Environment

School of Computing and Digital Technology

Table of Contents

Tables

Figures

1. Introduction

This report delves into the convergence of social media analytics and natural language processing (NLP). It begins by examining different methodologies and platforms utilized in social media analytics. Subsequently, it explores NLP techniques like topic modelling, sentiment analysis, and text summarization.

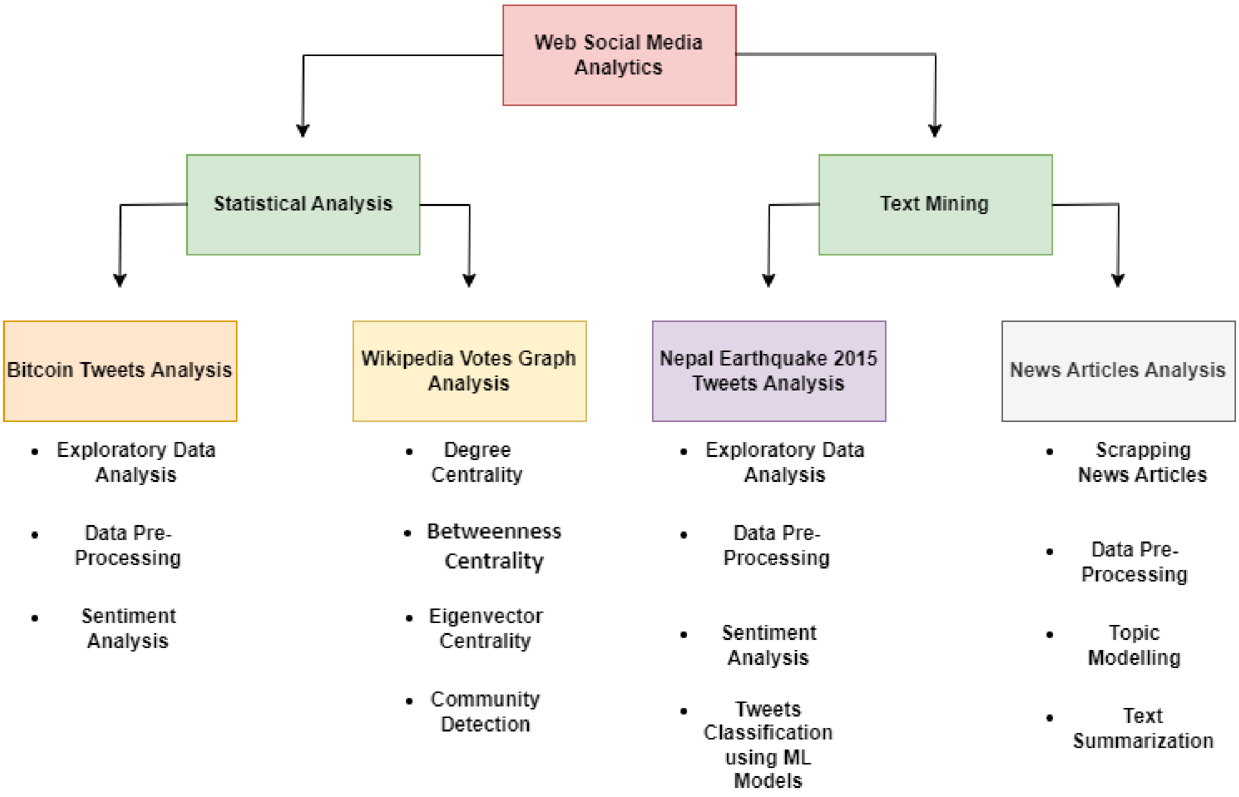


Figure 1: Project Flow Diagram

2. Twitter Trend Analysis

Social media platforms like Twitter host discussions on various topics, some gaining significant attention and becoming trends. Trends reflect the interests and attention of global and local communities, showcasing dynamic and transient topics that spread rapidly across interconnected networks, offering valuable insights into collective interests (Nilekar, 2020).

2.1 Bitcoin Tweets Analysis

With the increasing prominence of cryptocurrencies in the rapidly evolving realms of technology and finance, it becomes crucial to assess their associated risks and rewards, particularly when considering sentiment and emotions. Sentiment analysis, a method employed to comprehend the emotions and opinions of individuals regarding a particular topic, proves invaluable in this regard (Kulcsar, 2023).

Figure 2 explains the flow of the Bitcoin tweet analysis.

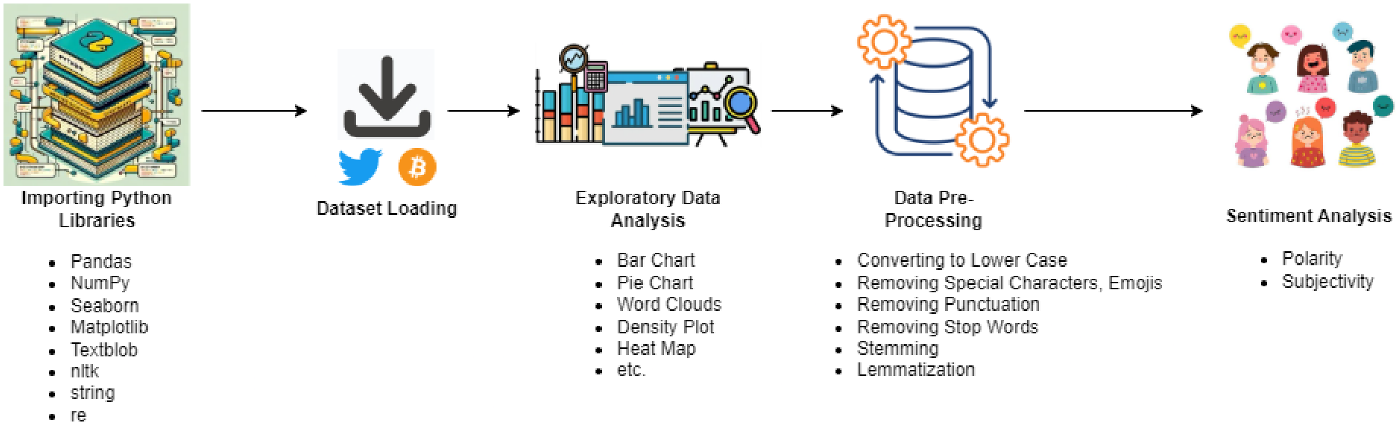


Figure 2: Flow of Bitcoin Analysis

2.2 Data Description

The dataset comprises over 100,000 records, each containing 13 columns. This information was gathered from Kaggle. Below is a description of these features-

Feature Name	Data Type	Description
user_name	String	The name of the user, as they've defined it
user_location	String	The user-defined location for this account's profile
user_description	String	The user-defined UTF-8 string describing their account
user_created	String	Time and date, when the account was created
user_followers	Integer	The number of followers an account currently has
user_friends	Integer	The number of friends an account currently has
user_favourites	Integer	The number of favorites an account currently has
user_verified	Boolean	When true, it indicates that the user has a verified account
date	String	UTC and date when the Tweet were created
text	String	The actual UTF-8 text of the Tweet
hashtags	String	All the hashtags posted in the tweets
source	String	Source of the tweet like web, mobile app, etc
is_retweet	Boolean	Tweet has been Retweeted by the authenticating user.

Table 1: Bitcoin Tweets Data Description

2.3 Identifying Missing Data

The heatmap analysis reveals missing values in the "user_location" and "user_description" columns. Addressing these gaps will enhance data completeness and analytical accuracy within the dataset.

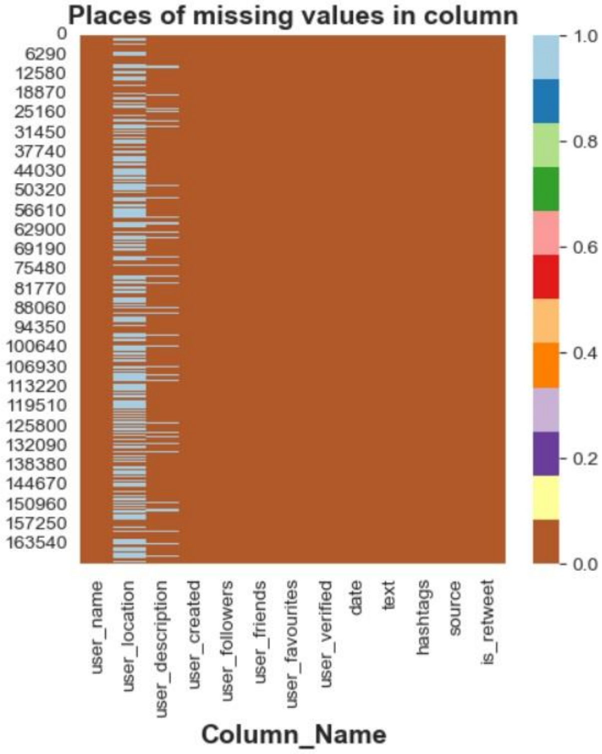


Figure 3: Missing Value Heat Map

2.4 What are the Prevalent Words in the Tweet?

Word clouds visually represent the most common terms in a dataset, offering a snapshot of its main themes. They simplify complex information, making it easier to identify key concepts, trends, and sentiments, aiding understanding and informing decision-making processes. Using a word cloud, prevalent words in the tweets were extracted to identify popular terms such as cryptocurrency, bitcoin, price, update, BTC, etc.

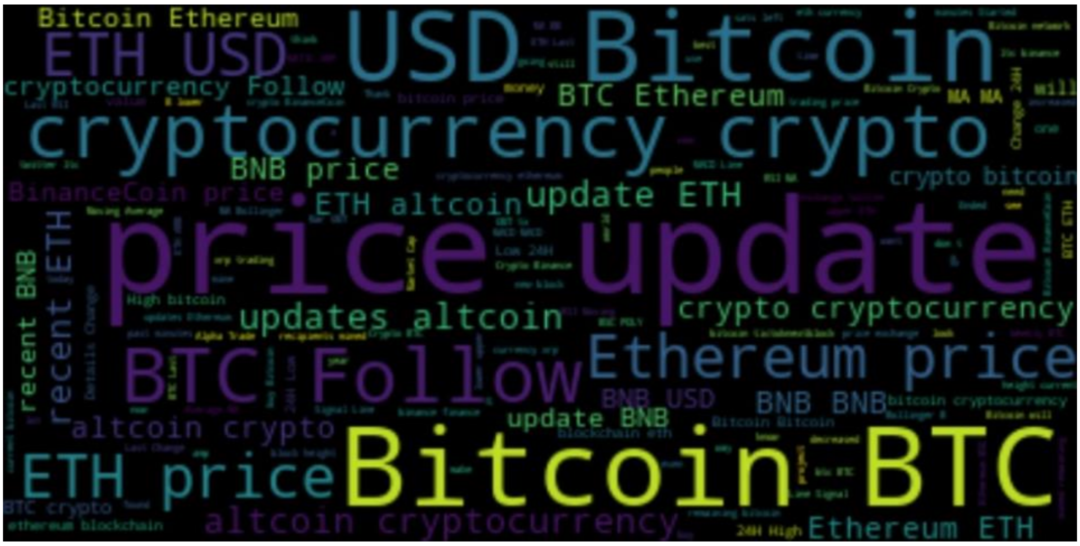


Figure 4: Prevalent Words

2.5 Which are the top countries that tweeted?

The bar graph displays tweeting activity on Bitcoin, with the United States leading with 4000+ tweets, followed by the UAE with 1000+ and London with 800+. It illuminates global interest in cryptocurrency, showcasing diverse geographical engagement and underscoring Bitcoin's significance in social media discourse.

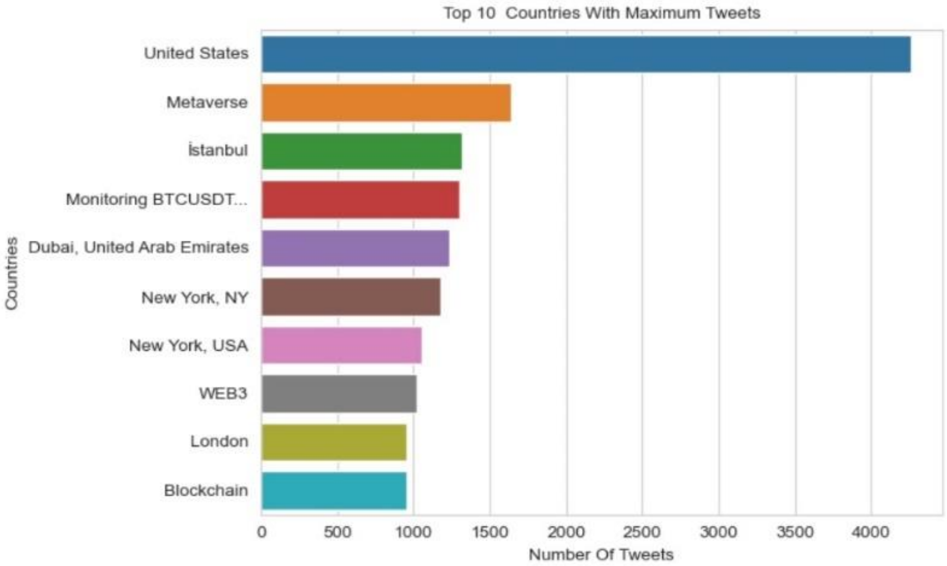


Figure 5: Top 10 Countries with Maximum Tweets

2.6 User Verification Disparity

The bar plot illustrates user verification status, revealing that a vast majority, 99.2%, are unverified, while only a minimal 0.8% are verified. This stark contrast underscores the prevalence of unverified users within the dataset, highlighting the importance of verifying user identities for credibility and authenticity.

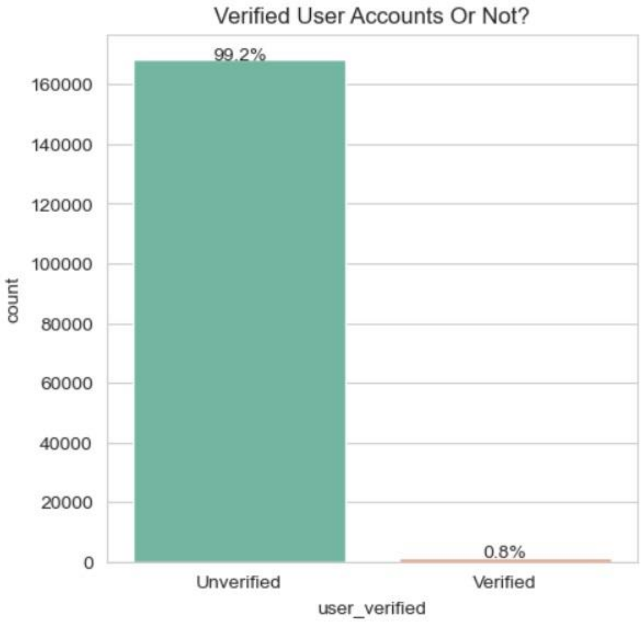


Figure 6: User Verified or Not

2.7 Tweet Distribution by Platform

The bar plot displays tweet distribution by platform, revealing that the Twitter Web app leads with 44.3%, followed by Android at 23.8%, and iPhone at 22.7%. This visualization underscores the dominance of the Twitter Web app in tweet generation, with notable contributions from Android and iPhone platforms as well.

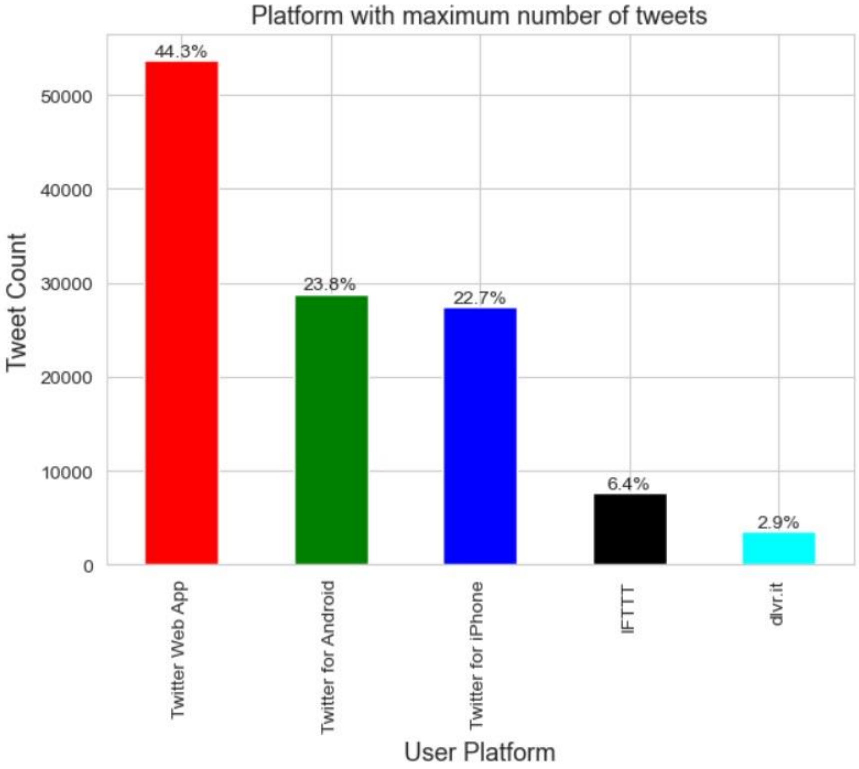


Figure 7: Platform used for maximum number of tweets

2.8 Top Tweet Sources: India, United States and United Kingdom

The pie charts illustrate the predominant tweet sources in India, the United States, and the United Kingdom.

In India, the majority of tweets originate from Android devices, accounting for 39.2%, followed by IFTTT at 29.9%, and the Twitter Web app at 20.1%.

For the United States, Microsoft Power Platform stands out as the leading tweet source with 36.6%, trailed by the Twitter Web app at 24%, and iPhones at 17%.

In the United Kingdom, iPhones contribute the most to tweet generation, representing 35.3%, followed closely by Android devices at 33.8%, and the Twitter Web app at 20.6%.

These visualizations offer insights into the preferred platforms for tweeting in each country, reflecting the diverse technological preferences and usage patterns across different regions.

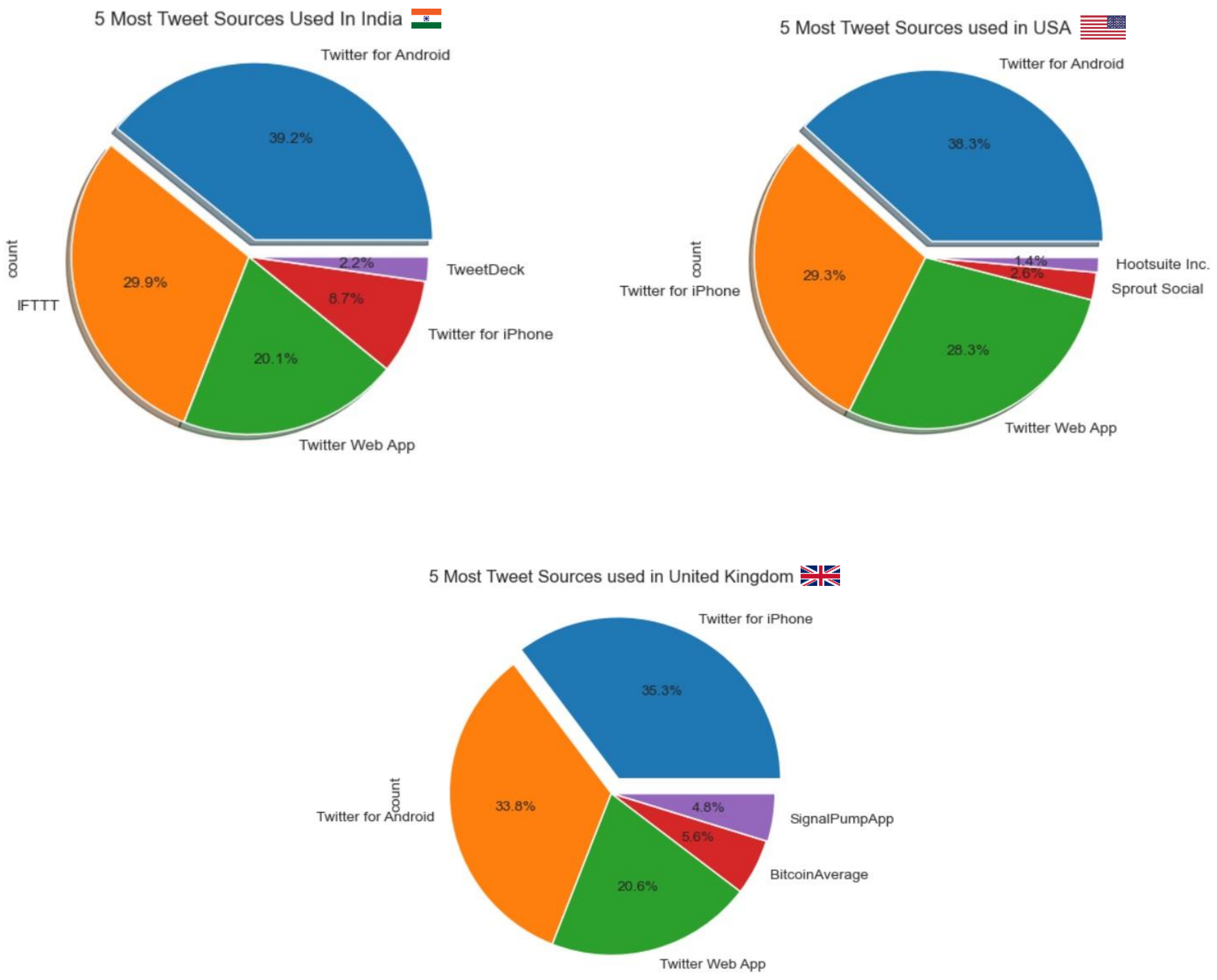


Figure 8: Tweet Sources in India, USA, and UK

2.9 Sentiment Analysis

Twitter sentiment analysis, employing NLP and ML, determines whether tweets convey negative, positive, or neutral emotions. Also called opinion mining, it categorizes sentiments in text, providing insights into public opinion on diverse social media topics

(Shahzad, 2021).

Polarity: Polarity indicates the emotional stance conveyed in analyzed text, typically categorized as positive, negative, or neutral. In sentiment analysis, the objective is to ascertain the overall sentiment conveyed, whether it leans towards positivity, negativity, or neutrality.

Subjectivity: Subjectivity, conversely, denotes the extent to which a statement reflects personal sentiments, beliefs, or viewpoints rather than factual data. Subjective statements express individual judgments or emotions, while objective statements are grounded in observable facts and remain unaffected by personal emotions.

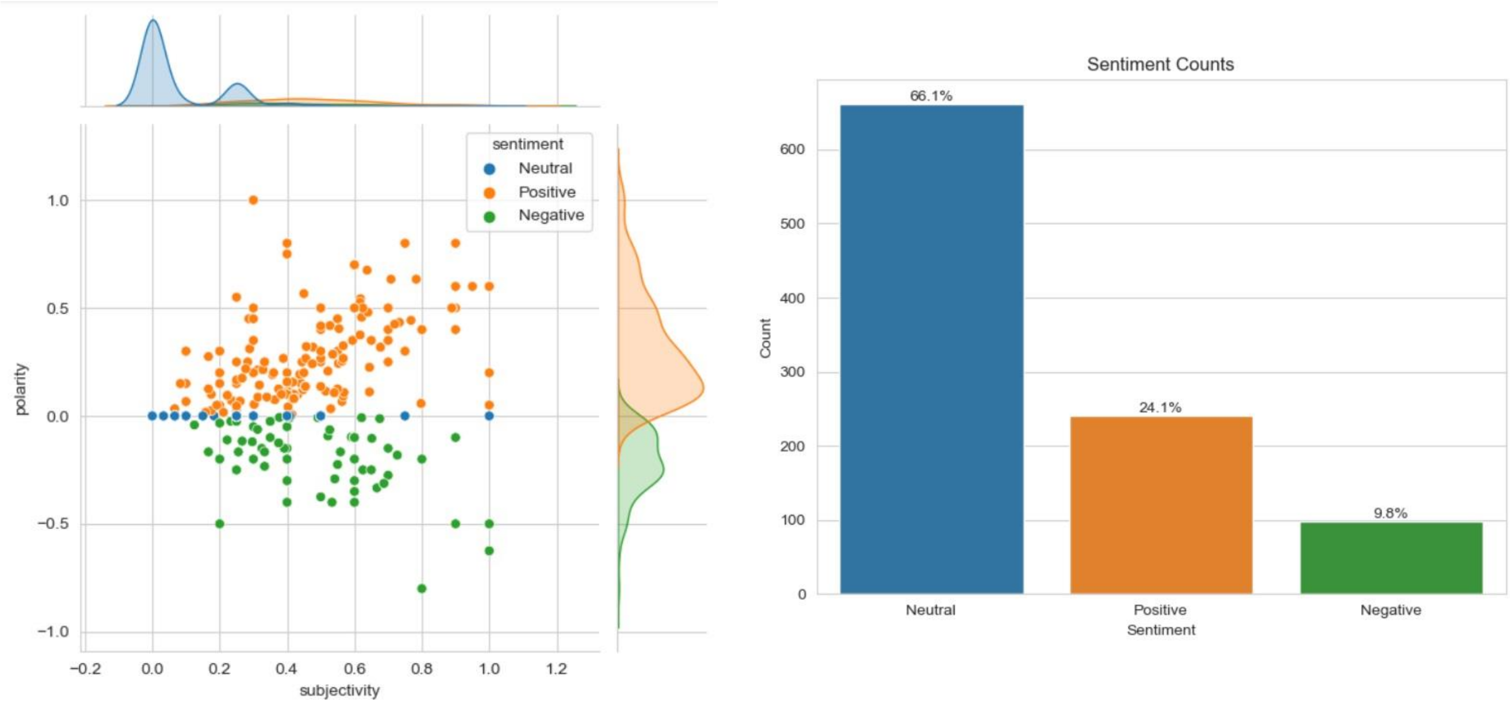


Figure 9: Tweets Sentiment Analysis

The joint plot correlates tweet polarity and subjectivity, with polarity on the Y-axis and subjectivity on the X-axis. A superimposed bar plot indicates 66.7% neutral, 24.1% positive, and 9.2% negative sentiment in Bitcoin-related tweets, offering a comprehensive sentiment analysis overview.

2.10 Visual Analysis of Sentiment Patterns in Tweets

Visualizations explored sentiment patterns in tweet data. Frequency distribution plots illustrated word distribution for each sentiment (neutral, positive, negative). Word clouds visually depicted most common words in each sentiment, with word size indicating frequency. These aids deepen understanding of language patterns and prevalent themes in tweets, enhancing overall analysis.

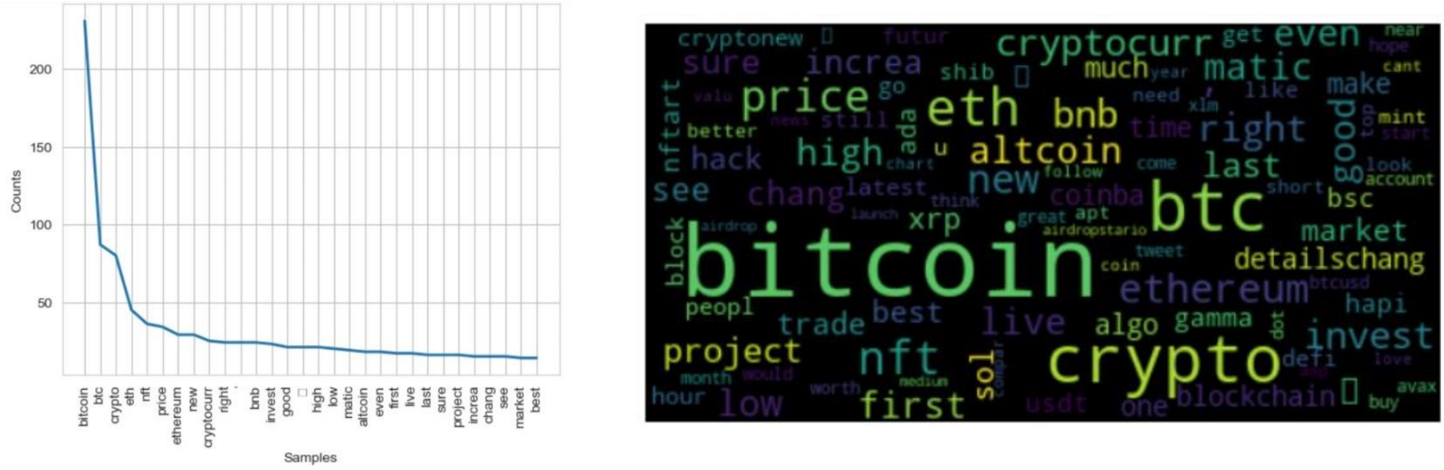


Figure 10: Word Distribution and Common Words in Positive Tweets

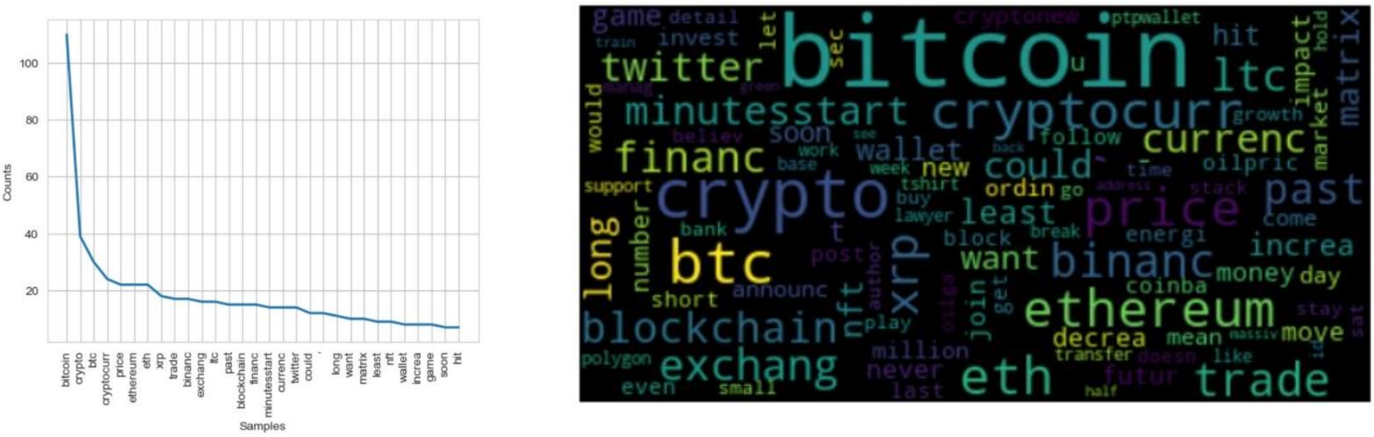


Figure 11: Word Distribution and Common Words in Negative Tweets

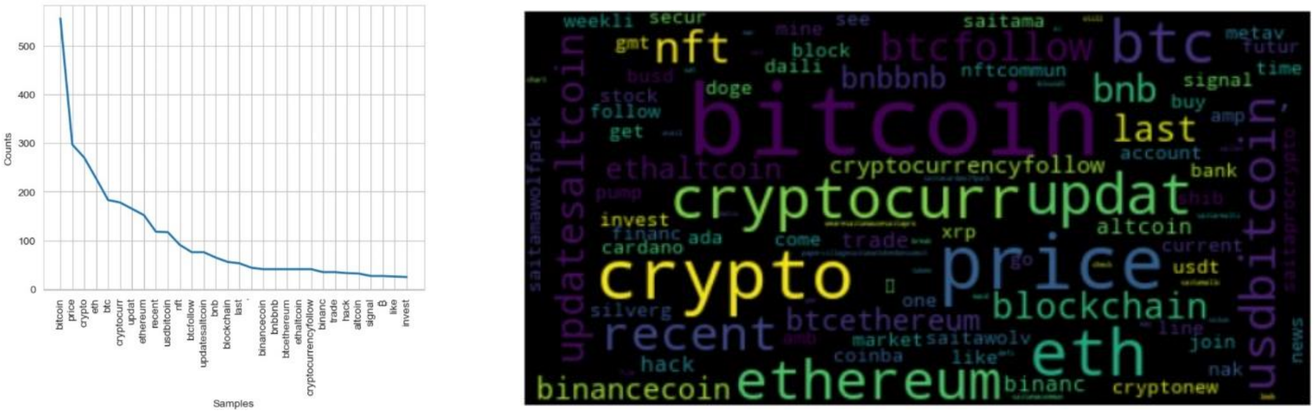


Figure 12: Word Distribution and Common Words in Neutral Tweets

3. Graph Analysis

The resurgence of social network analysis is fuelled by the widespread availability and abundance of content from social media, websites, and sensors. While this content offers valuable data for building and examining social networks, its sheer volume and lack of structure pose numerous challenges (Campbell, 2013).

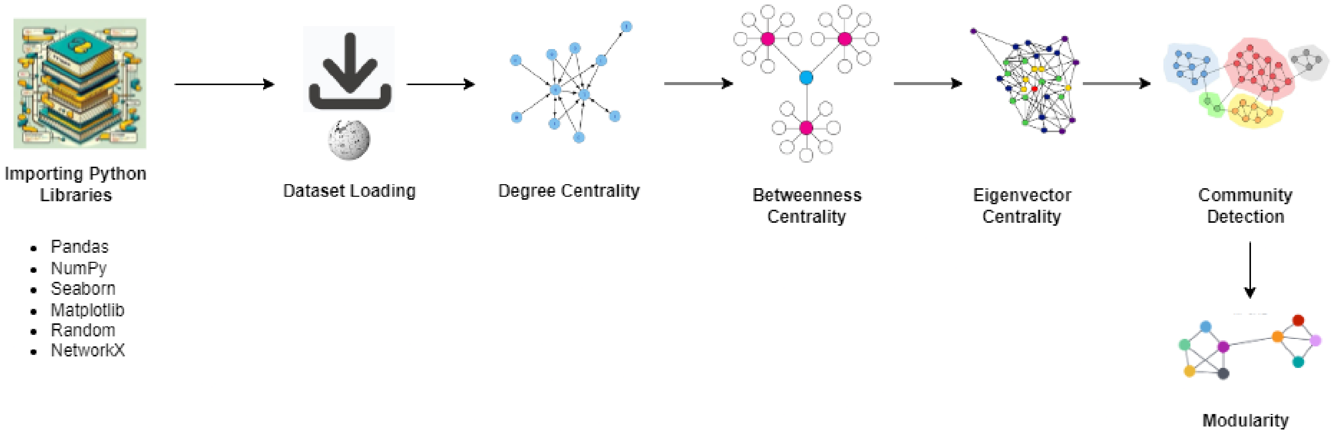


Figure 13: Flow Chart of Graph Analysis

3.1 Data Description

The dataset encompasses Wikipedia's entire voting history from its inception. In this network, nodes represent Wikipedia users, and a directed edge from node i to node j signifies that user i voted on user j. The network consists of **889 nodes and 2914 edges**.

The average node in the graph is connected to around 6.56 other nodes, suggesting moderate connectivity. These metrics gauge the graph's density and interconnectivity.

Spring layout: The spring layout of graph analysis utilizes a force-directed algorithm to position nodes, simulating spring-like forces to arrange them based on edges, aiding visualization and understanding of network structures.

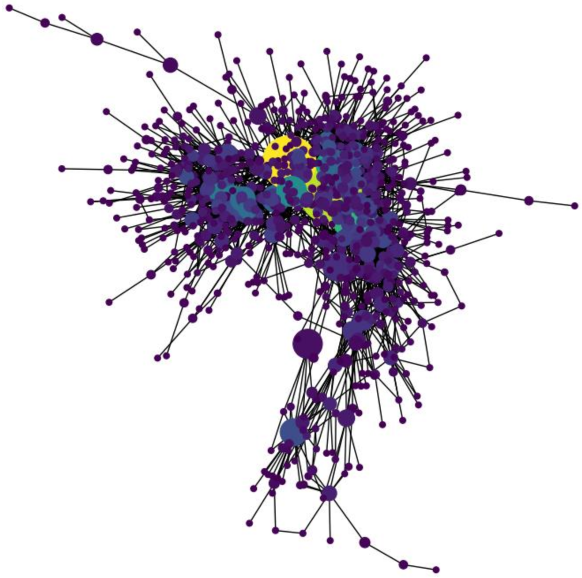


Figure 14: Spring Layout of Wiki Votes

3.2 Degree Centrality

Degree centrality measures a node's significance by its number of direct connections. It indicates the immediate connections each node has in the network, identifying influential or well-connected individuals who can disseminate information efficiently or bridge various network segments.

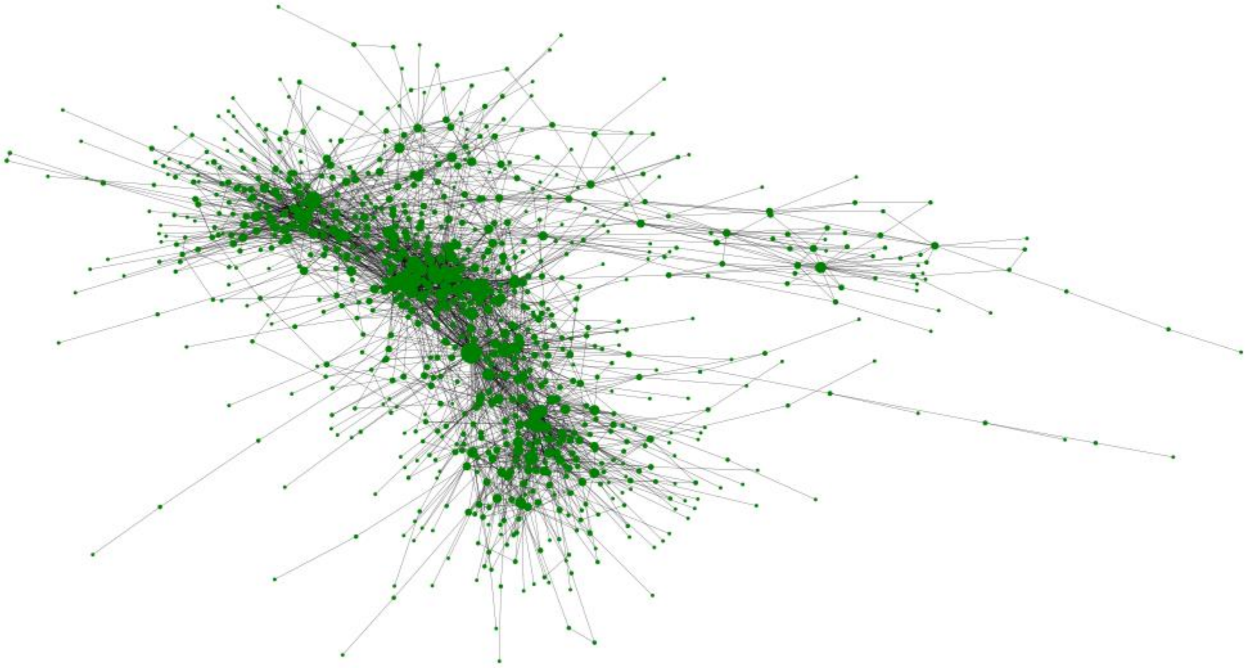


Figure 15: Degree Centrality

The output displays the top 5 nodes in the graph, ranked by their degree centrality scores. These scores represent the proportion of direct connections each node has to other nodes.

For example, node 431 has a degree centrality of approximately 0.115, indicating that it is connected to around 11.5% of the other nodes in the network

Node	Degree Centrality
431	0.11486486486486486
273	0.1036036036036036
170	0.07432432432432433
536	0.06756756756756757
399	0.06306306306306306

Table 2: Degree Centralities of Top 5 Nodes

3.3 Betweenness Centrality

Betweenness centrality quantifies how often a node serves as a bridge on the shortest path between other nodes in a network. It identifies influential nodes that control information flow within the system. The output lists the top 8 nodes in the graph by betweenness centrality scores.

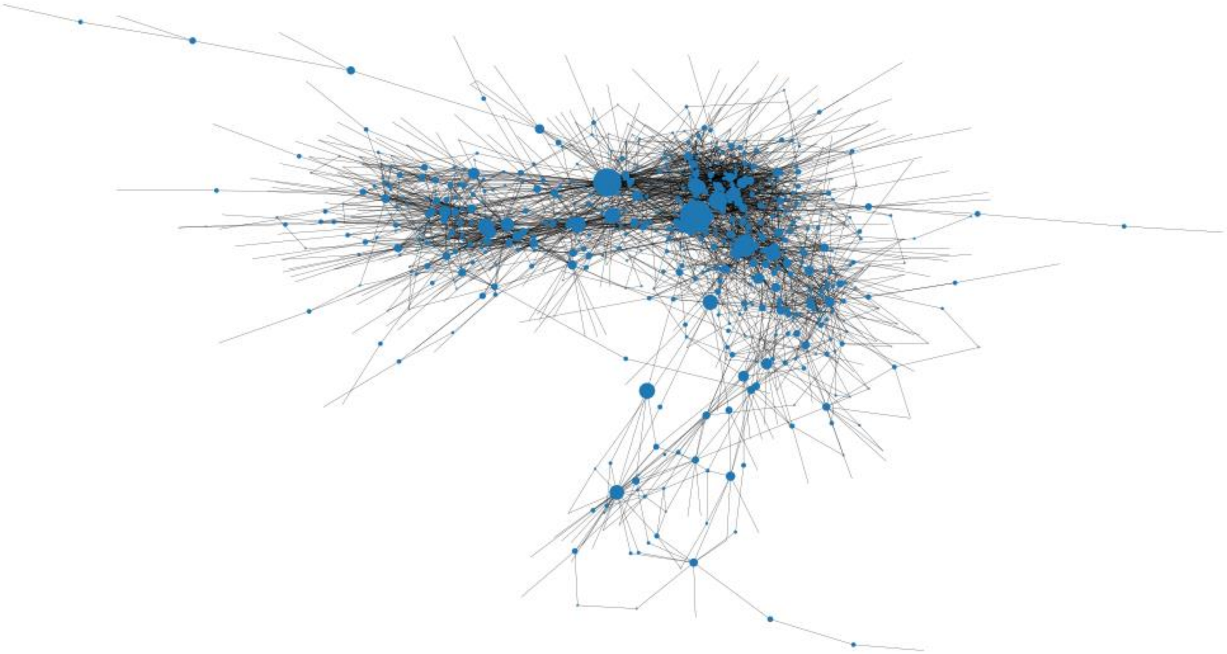


Figure 16: Betweenness Centrality

This metric quantifies how often each node acts as a bridge on the shortest paths between other nodes, indicating their influence on information flow. Higher scores signify greater importance in maintaining network connectivity and facilitating communication.

For Example, Node 273, with the highest betweenness centrality score, might represent a user in a social network who frequently connects different groups or communities by sharing information or facilitating interactions across various segments of the network.

Node	Degree Centrality
273	0.25377565538555935
431	0.18841894862401887
170	0.11283653180857324
204	0.07561248876313527
736	0.05774695853512937

Table 3: Betweenness Centralities of Top 5 Nodes

3.4 Eigenvector Centrality

Eigenvector centrality assesses a node's significance by considering its links to other nodes, especially those with high centrality scores. It highlights nodes with influential connections, particularly those linked to other highly influential nodes. It's useful for identifying individuals with substantial influence in a network, considering both direct and indirect connections.

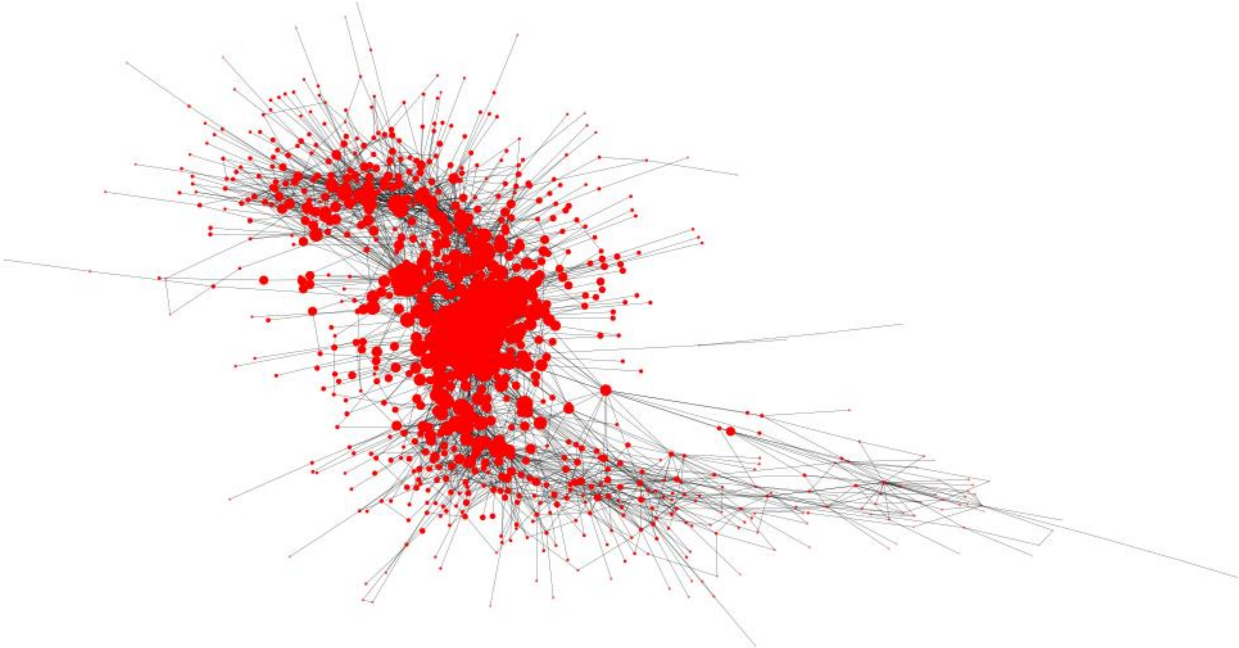


Figure 17: Eigenvector Centrality

The output lists the top 5 nodes by eigenvector centrality values. For instance, node 273 has the highest score of 0.285, indicating its strong influence due to connections with other influential nodes.

Node	Degree Centrality
273	0.2852399137391541
431	0.2791279493271889
536	0.22124652294171177
399	0.21350736615246194
416	0.21044572079769291

Table 4: Eigenvector Centralities of Top 5 Nodes

3.5 Community Detection

Communities consist of entities with closer interactions within the group than outside. Detection methods identify such groups based on network structure, emphasizing strong connections. Node attributes often define communities, linked by intra-community edges. Crucial in network analysis, especially in vast social media networks, effective algorithms are essential for accurate partitioning (Bedi, 2016).

A total of 13 communities were detected.

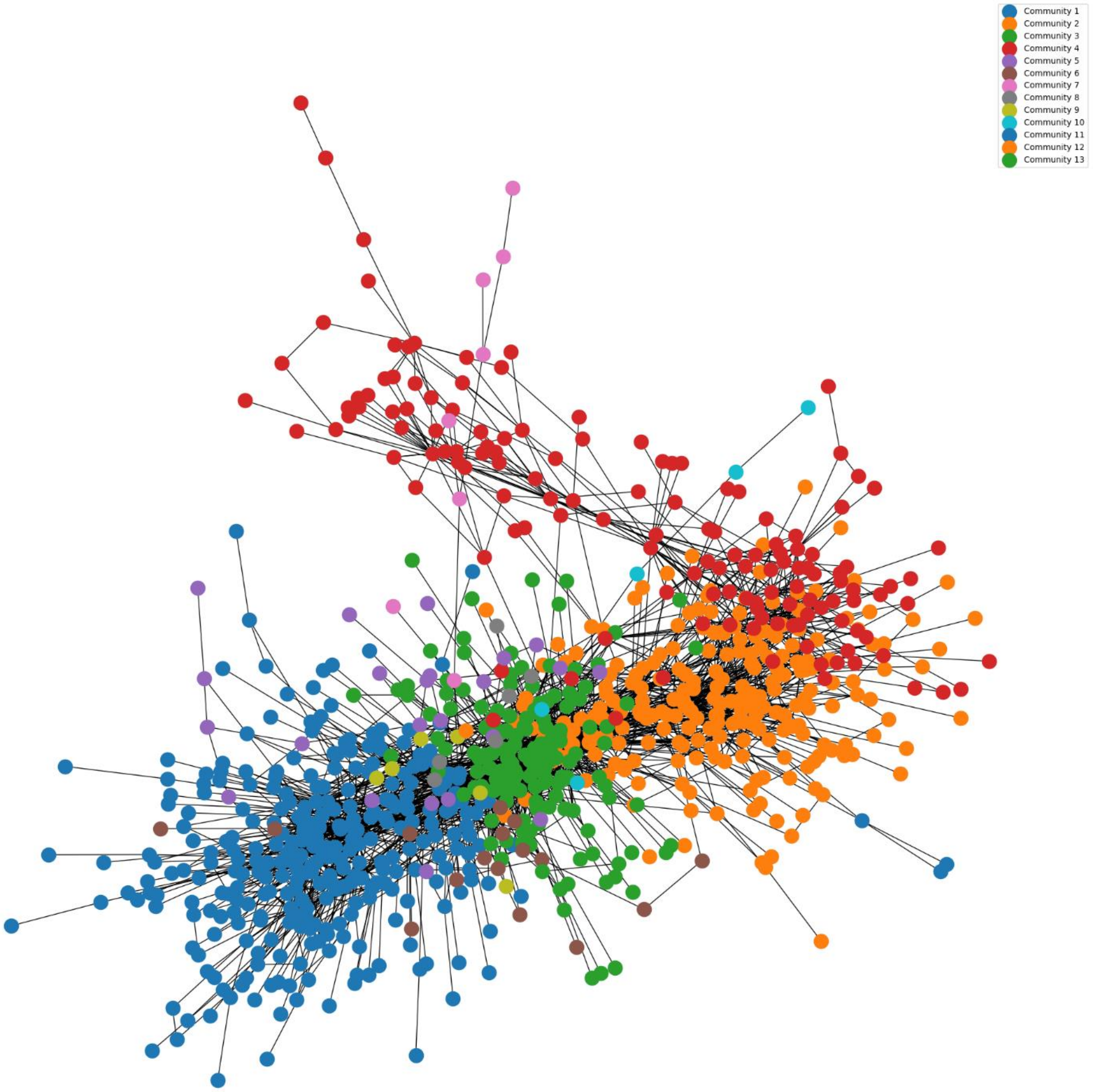


Figure 18: Community Detection

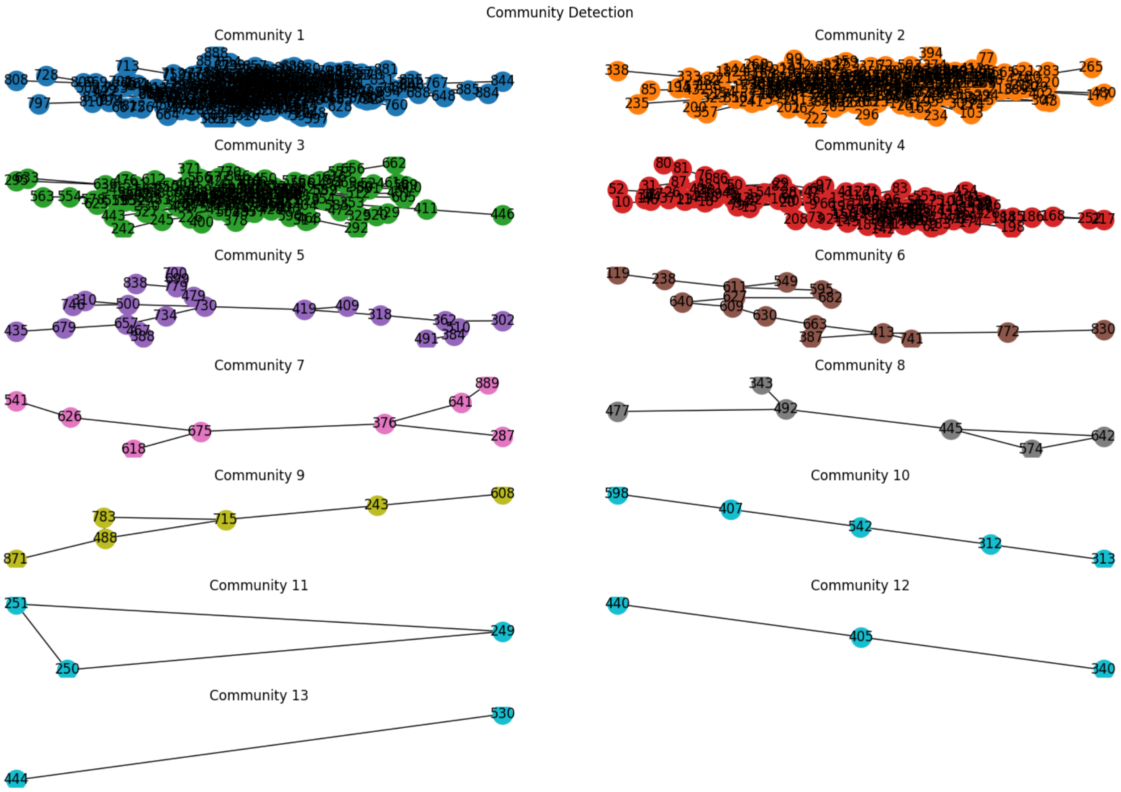


Figure 19: 13 Communities Detected

3.6 Modularity

Modularity, either positive or negative, suggests community presence. Seeking network divisions with positive, ideally large, modularity values help precisely identify community structure within the network.

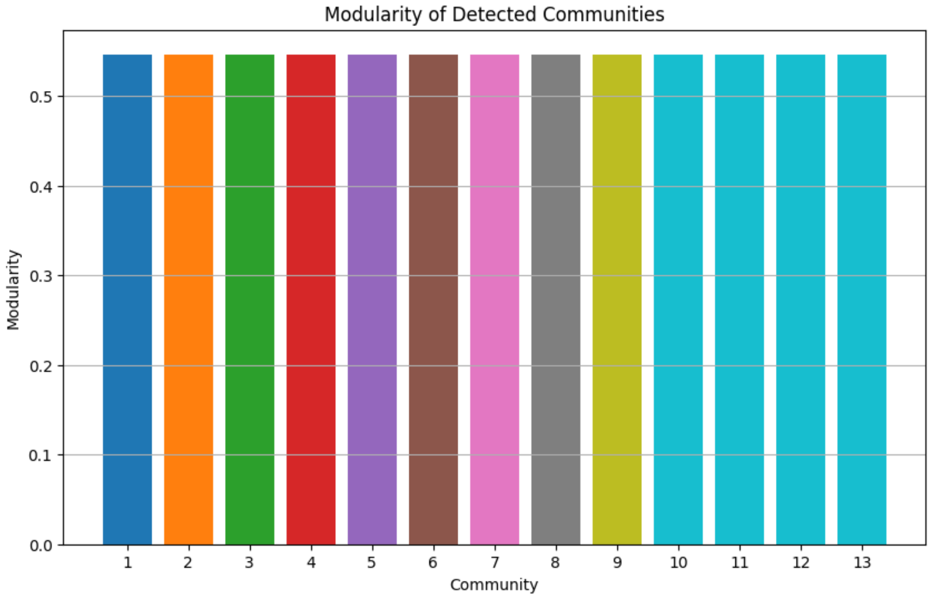


Figure 20: Modularity of Detected Communities

4. Event Twitter Sentiment Analysis

The event chosen is the 2015 Nepal Earthquake. Through tweet analysis, I aim to explore public discourse surrounding this significant event, uncovering insights into its impact and sentiment.

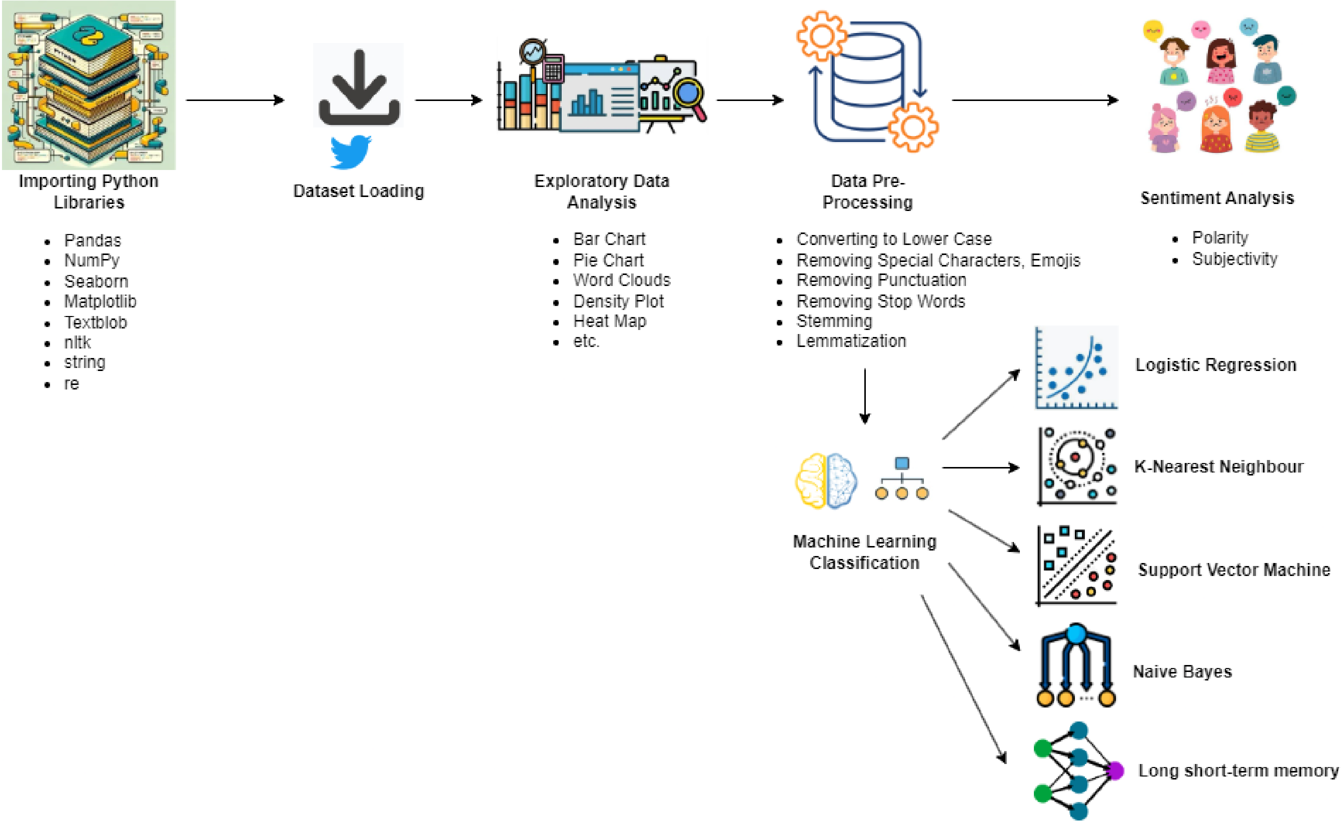


Figure 21: Flow Chart of Event Twitter Analysis

4.1 Data Description

Data extracted form Kaggle contains 18,233 rows and 3 columns: Tweet Class, Tweet ID, and Tweet Text. The Twitter ID represents the user, the Tweet Text includes the content of the tweet, and the Tweet Class categorizes the tweet as follows:

- 0 = General Tweets
- 1 = Tweets indicating a need for resources (Need Tweets)
- 2 = Tweets indicating the availability of resources (Availability Tweets)

Feature Name	Data Type	Description
TweetClass	Integer	Tweets Categorization
TweetID	String	User ID
TweetText	String	Tweet Content

Table 5: Nepal Earthquake Tweets Data Description

4.2 Distribution of Tweets

General Tweets dominate, comprising 95.27% of the dataset, indicating widespread discussion, information sharing, and sentiment expression directly related to the earthquake event.

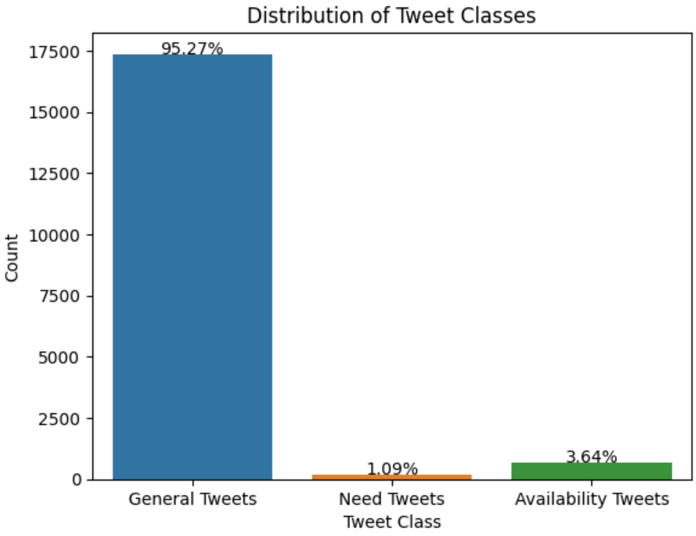


Figure 22: Nepal Earthquake Tweets Distribution

4.3 General Tweets

The word cloud of General Tweets highlights terms like 'earthquake,' 'Nepal,' 'help,' 'people,' and 'victim relief,' reflecting discussions on disaster, assistance, impacted individuals, and relief operations.

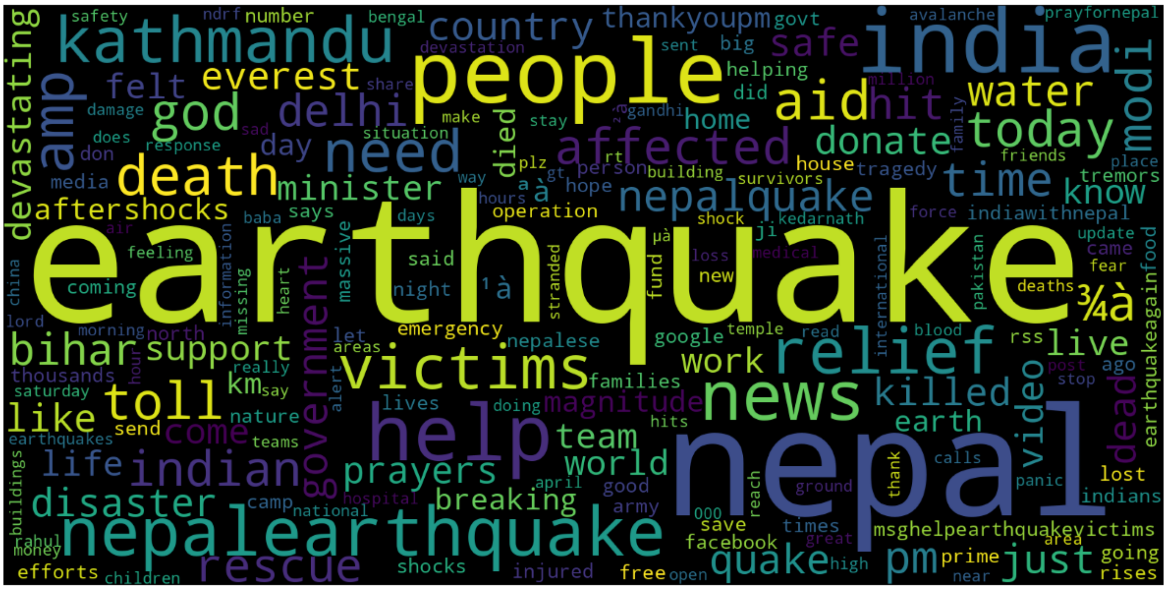


Figure 23: Prevalent General Tweets

4.4 Need Tweets

The Need Tweets' word cloud emphasizes urgent needs in Nepal, notably 'food,' 'water,' and 'blood.' It succinctly communicates critical necessities like sustenance and medical supplies for disaster relief efforts.

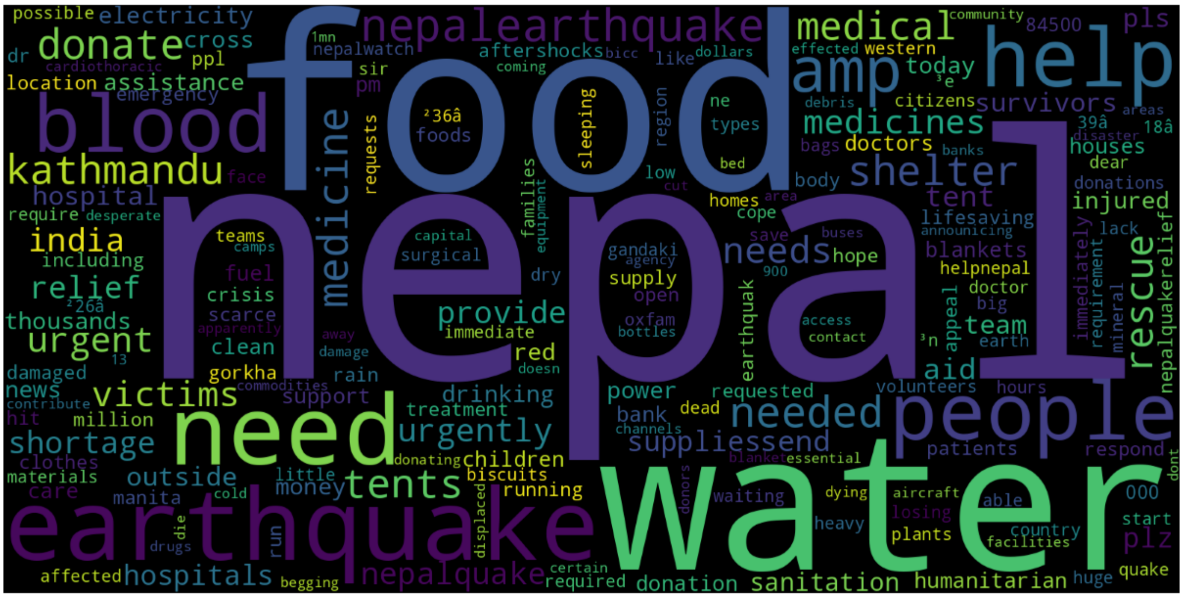


Figure 24: Prevalent Need Tweets

4.5 Availability Tweets

Availability Tweets' word cloud highlights accessible resources in Nepal, notably 'water,' 'blood,' 'medical,' and 'relief.' It encapsulates offered aid, indicating support networks and relief efforts mobilized during the crisis.

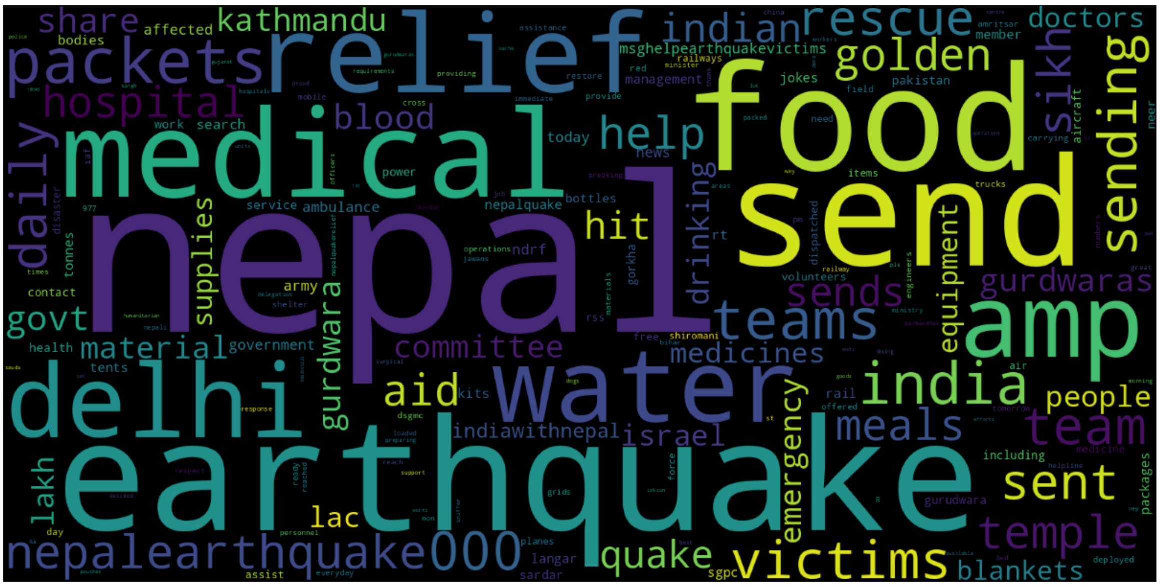


Figure 25: Prevalent Availability Tweets

4.6 Sentiment Analysis

The joint plot illustrates the correlation between tweet polarity and subjectivity. A bar plot overlay reveals that 66.7% of Bitcoin-related tweets are neutral, 24.1% positive, and 9.2% negative, offering comprehensive sentiment analysis (Shahzad, 2021).

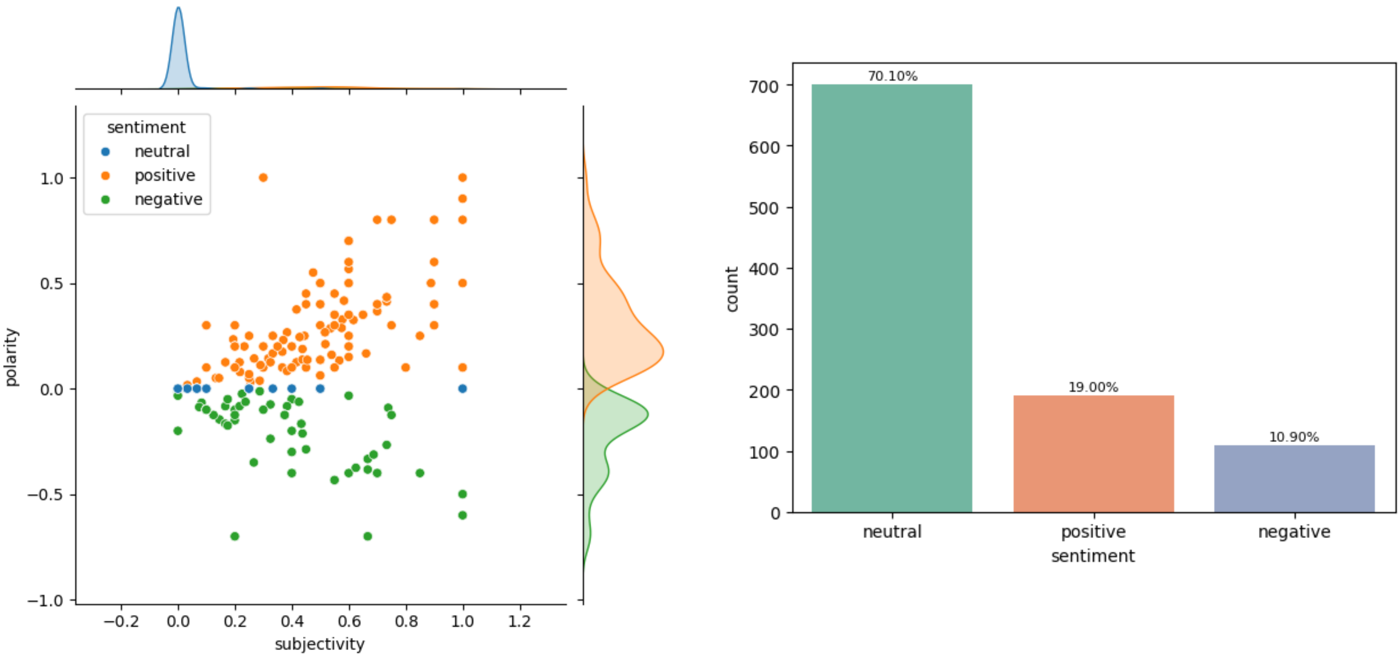


Figure 26: Tweets Sentiment Analysis

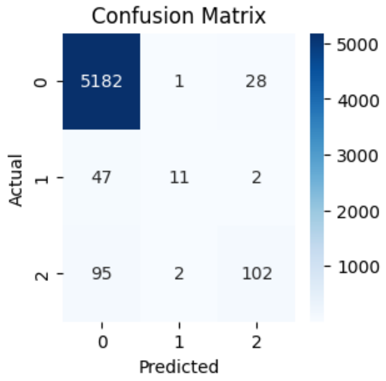
4.7 Machine Learning Models

Several machine learning models have been employed for classification tasks, such as Logistic Regression, KNN, SVM, Naive Bayes, and LSTM, to determine the class to which tweets belong.

4.7.1 Logistic Regression

Logistic Regression, a statistical method, and ML algorithm excels in classification, like tweet class detection. It predicts probabilities using the sigmoid function. Widely used for categorical target variables (Brzezinski, 1999).

The logistic regression model achieved 97% overall accuracy in classifying tweets. Category 0 had strong performance, while categories 1 and 2 showed lower precision, recall, and F1-scores, indicating room for improvement.



	precision	recall	f1-score	support
0	0.97	1.00	0.98	5211
1	0.70	0.12	0.20	60
2	0.79	0.48	0.60	199
accuracy			0.97	5470
macro avg	0.82	0.53	0.59	5470
weighted avg	0.96	0.97	0.96	5470

Figure 27: Confusion Matrix of LR

Table 6: Classification Report of LR

4.7.2 K-Nearest Neighbour (KNN)

KNN is a versatile supervised learning method, determining classification or prediction based on proximity to neighboring points. "K" signifies the number of nearest neighbors considered (Shamrat, 2021).

The KNN model achieved 97% overall accuracy in classifying tweets. While category 0 performed well, categories 1 and 2 showed lower precision, recall, and F1-scores, indicating room for improvement.

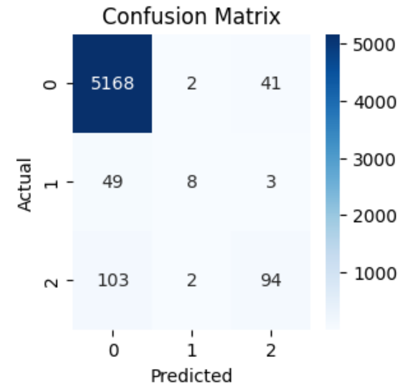


Figure 28: Confusion Matrix of KNN

	precision	recall	f1-score	support
0	0.98	0.99	0.98	5211
1	0.76	0.27	0.40	60
2	0.66	0.58	0.62	199
accuracy			0.97	5470
macro avg	0.80	0.61	0.67	5470
weighted avg	0.96	0.97	0.96	5470

Table 7: Classification Report of KNN

4.7.3 Naive Bayes

Naïve Bayes, a generative learning algorithm, models input distribution within classes without prioritizing key features. Unlike discriminative classifiers like logistic regression, it's common in text classification, spam filtering, and recommendation systems (Abbas, 2019).

The classifier achieved 95% accuracy. While class 0 had high precision and recall, categories 1 and 2 struggled, with zero precision and recall for category 1. Improvement strategies are needed.

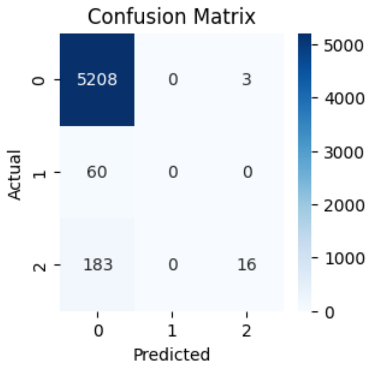


Figure 29: Confusion Matrix of Naïve Bayes

	precision	recall	f1-score	support
0	0.95	1.00	0.98	5211
1	0.00	0.00	0.00	60
2	0.70	0.04	0.07	199
accuracy			0.95	5470
macro avg	0.55	0.34	0.35	5470
weighted avg	0.93	0.95	0.93	5470

Table 8: Classification Report of Naive Bayes

4.7.4 Support Vector Machine (SVM)

Support Vector Machine (SVM) is a robust ML algorithm for linear or nonlinear classification, regression, and outlier detection. It's versatile, handling tasks like text and image classification, spam detection, and anomaly detection efficiently (Puri, 2019).

The model achieved 97% accuracy, excelling in categories 0 and 2 but facing challenges in category 1. Overall balanced performance suggests effectiveness with room for improvement in category 1 classification.

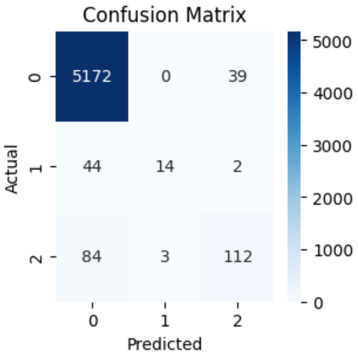


Figure 30: Confusion Matrix of SVM

	precision	recall	f1-score	support
0	0.98	0.99	0.98	5211
1	0.84	0.27	0.41	60
2	0.76	0.58	0.66	199
accuracy			0.97	5470
macro avg	0.86	0.61	0.68	5470
weighted avg	0.97	0.97	0.97	5470

Table 9: Classification Report of SVM

4.7.5 Long short-term memory (LSTM)

LSTM, a recurrent neural network, is pivotal in Deep Learning, adept at capturing long-term dependencies for sequence prediction. Unlike conventional neural networks, LSTM's feedback connections enable processing entire data sequences, enhancing pattern understanding and prediction (Garlapati, 2022).

The classifier achieved 94% accuracy. While excelling in category 0, it struggles with categories 1 and 2, indicating the need for significant improvement in their classification.

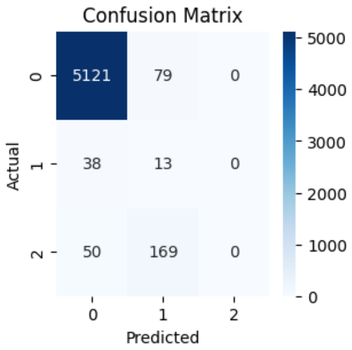


Figure 31: Confusion Matrix of LSTM

	precision	recall	f1-score	support
0	0.99	0.98	0.99	5200
1	0.06	0.33	0.11	51
2	0.00	0.00	0.00	219
accuracy			0.94	5470
macro avg	0.35	0.44	0.36	5470
weighted avg	0.94	0.94	0.94	5470

Table 10: Classification Report of LSTM

4.8 Comparison among Machine Learning Models

SVM led the Tweets Classification Project with 97% accuracy, followed closely by Logistic Regression at 96.8%. KNN achieved 96.3%, while LSTM trailed at 93.9%. Logistic regression had the shortest training and prediction times, whereas LSTM had the longest.

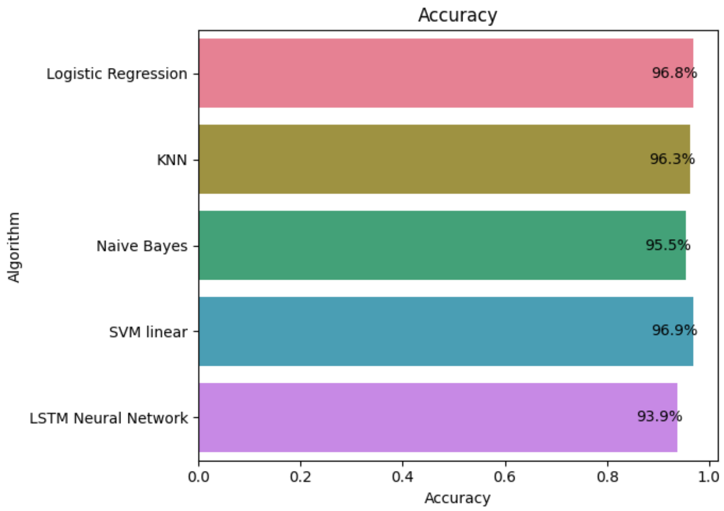


Figure 32: Comparison of Machine Learning Algorithms based on Accuracy

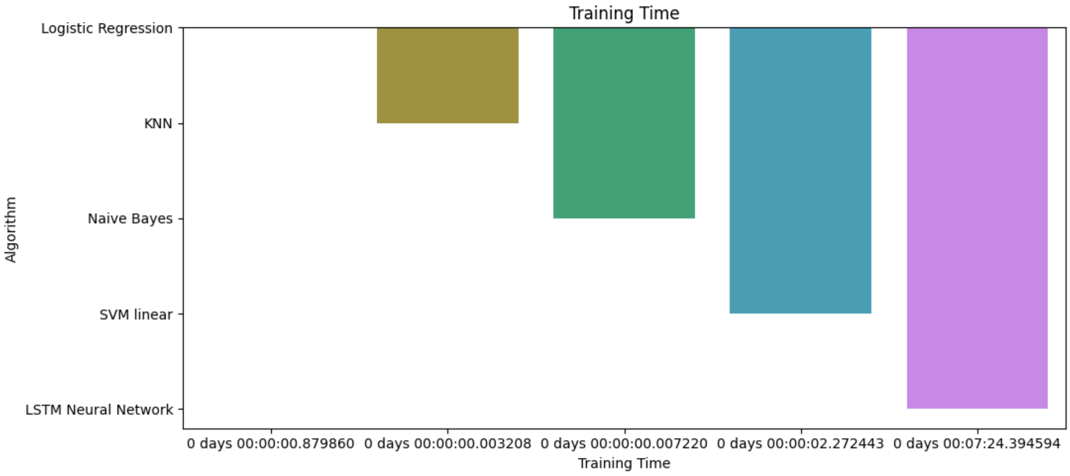


Figure 33: Comparison of Machine Learning Algorithms Training Time

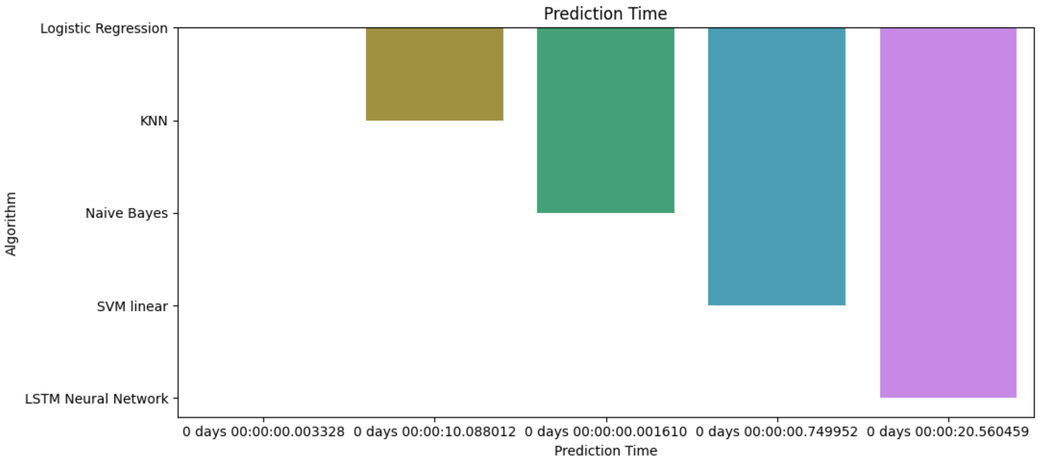


Figure 34: Comparison of Machine Learning Algorithms Prediction Time

5. News Article Analysis

The News API is a straightforward REST API that provides access to news articles across the web in JSON format. It allows users to retrieve top stories from news websites or search for news on specific topics (Visvam Devadoss, 2019).

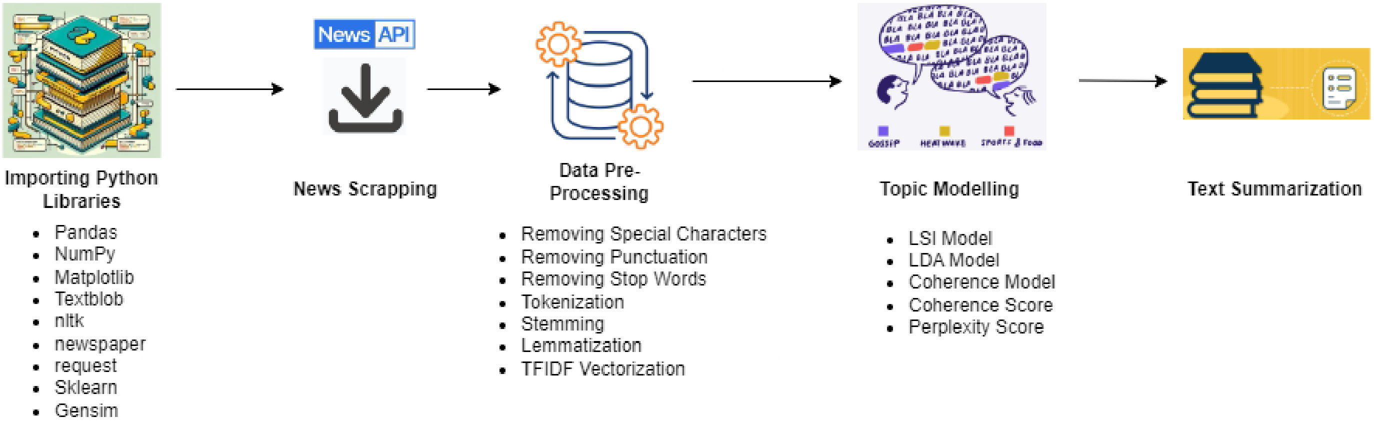


Figure 35: Flow Chart of News Analysis

5.1 News Article Descriptive Analytics

Five news articles have been selected, and tokenization, lemmatization, stemming, and stop words processes have been applied to them. After that, the count of words and sentences has been calculated.

S.No.	Query	News Article Title	Word Count	Sentences Count
1	Israel Palestine War	USC Faces Backlash Over Alleged ‘Censorship’ of Pro-Palestinian Valedictorian’s Speech	872	28
2	2024 Elections in India	Congress didn’t want Ram temple in Ayodhya; Modi honoured people’s sentiments	1034	41
3	Taiwan Earthquake 2024	Taiwan Rocked by Massive Earthquake	789	23
4	Ukraine Russia War	Russia's firing new, long-range Kh-69 cruise missiles, war experts say, piling on the misery for Ukraine's dwindling air defence	946	33
5	Artificial Intelligence	How AI is paving the way to smoother streets using autonomous robots	1011	39

Table 11: Descriptive Analytics of 5 News Articles

5.2 Topic Modelling

Topic modeling involves identifying the words associated with topics found within a document or dataset. This is beneficial because it simplifies the task compared to extracting words directly from the document, which is more time-consuming and intricate (Tong, 2016).

Coherence: Coherence gauges how effectively the words within a topic connect, indicating how well it align with human understanding of a quality topic. A high coherence score indicates clarity, consistency, and relevance, while a low score suggests vagueness, noise, or irrelevance.

Perplexity: Perplexity serves as a gauge of a topic model's capability to predict unfamiliar or unseen data, illustrating its ability to generalize. A low perplexity score indicates the model's confidence and accuracy in predictions, whereas a high score suggests uncertainty and inaccuracy in predictions.

For example, four-word cloud stations representing Topic 0, Topic 1, Topic 2, and Topic 3, suggest that the topic modeling algorithm Latent Dirichlet Allocation (LDA) identified four main themes or subjects within the news articles.

By examining the words in each word cloud, gained insights into the prominent terms and concepts associated with each topic, helping to understand the content and structure of the article better.

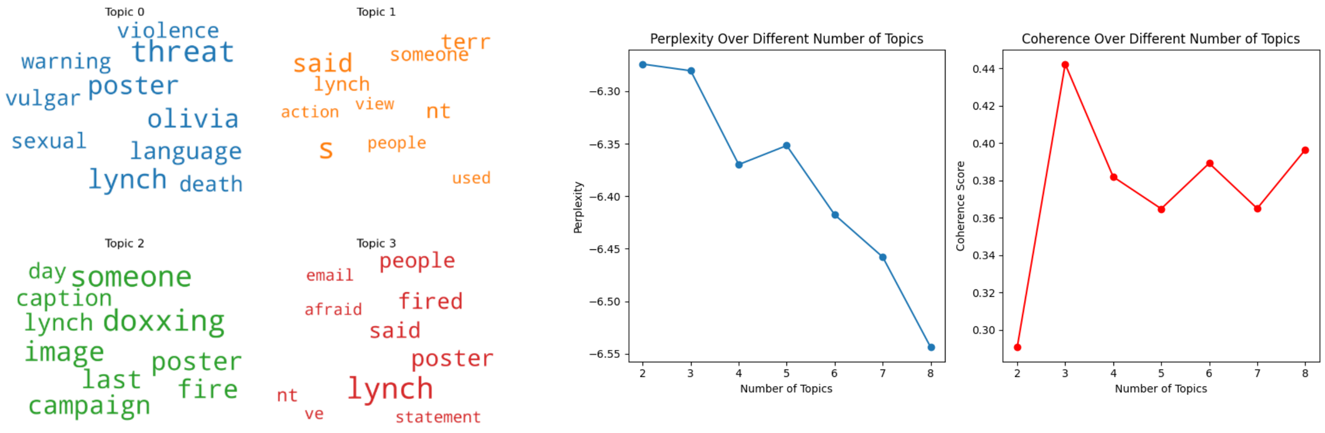


Figure 36: Word Cloud, Perplexity and Coherence for Israel-Palestine War News Article

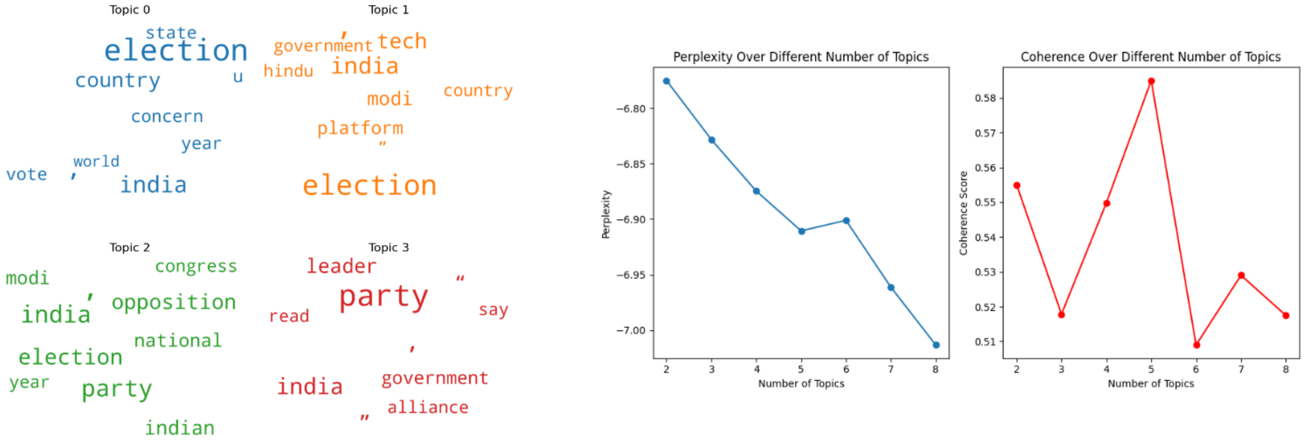


Figure 37: Word Cloud, Perplexity and Coherence for India Election 2024 News Article

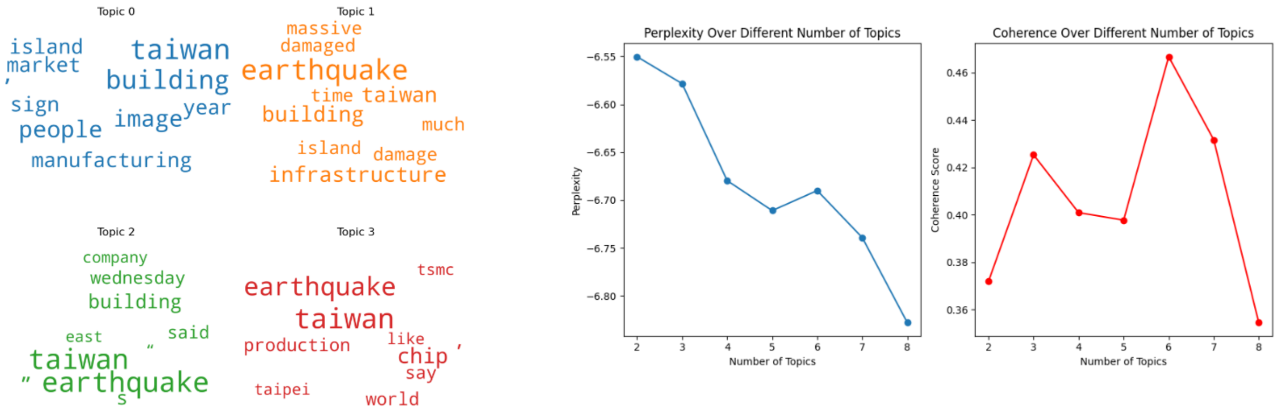


Figure 38: Word Cloud, Perplexity, and Coherence for Taiwan Earthquake 2024 News Article

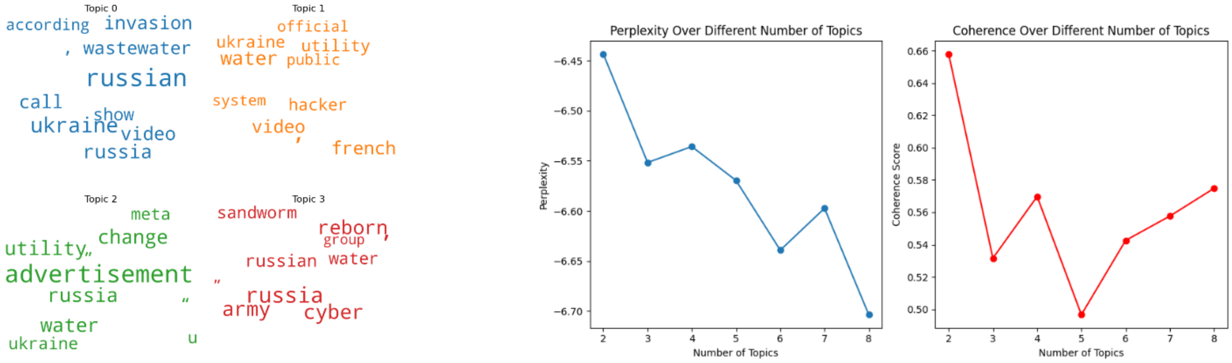


Figure 39: Word Cloud, Perplexity, and Coherence for Russia-Ukraine War News Article

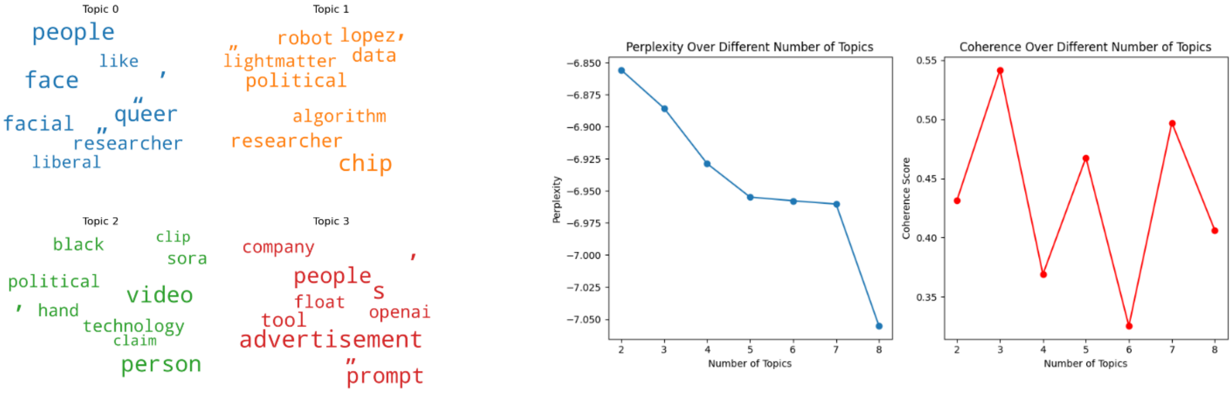


Figure 40: Word Cloud, Perplexity, and Coherence for Artificial Intelligence News Article

5.3 Text Summarization for News Article

Automatic text summarization condenses lengthy documents into concise, coherent synopses, extracting crucial information for rapid comprehension amidst the vast expanse of online textual content (Adhikari, 2020).

Using a news article about the Israel-Palestine war, word frequency was generated, followed by calculating sentence importance. The average sentence importance was then computed, and finally, the text was summarized.

News Article	Length of Article	Length of Summarized Text	Avg. Important Sentence
Israel-Palestine War	872 Words	118 Words	145.70

Table 12: Text Word Count

5.4 Summarized Text

lynch hundreds emails filled with death threats, threats sexual violence promises fired from teaching after-school nature program called wild ferns with fewer than five people staff. previous next olivia lynch hide caption warning: these images contain vulgar language, threats sexual violence death threats. previous next olivia lynch hide caption warning: these images contain vulgar language, threats sexual violence death threats. aaron terr, director public advocacy foundation individual rights expression (fire), nonprofit that works protect free speech, said that while none these actions tearing down poster, doxxing someone firing someone misconduct illegal, they hurt free speech culture u.s. americans feel like trip wires everywhere they don't know exactly what they can't say, scenario where millions americans national political conversation.

6. Summary

The report explores Statistical Analysis and Text Mining. It examines Bitcoin tweets for trends and sentiments, utilizing techniques like stemming and lemmatization. Additionally, it analyzes network structures in Wikipedia voting data. Text Mining involves sentiment analysis of Nepal Earthquake tweets using LR, KNN, NB, SVM, and LSTM models, alongside topic modeling with NewsAPI.

7. Conclusion

This project delved into social media analytics and text mining, yielding valuable insights. Exploratory data analysis revealed user behavior, while sentiment analysis of Bitcoin tweets showed a dominance of neutral sentiment. Analysis of Wikipedia voting data uncovered various centralities and communities. Sentiment analysis of Nepal Earthquake tweets favored support vector machine among classification models. News scraping facilitated topic modeling and text summarization, enhancing comprehension of online discourse. Overall, these findings underscore the significance of data-driven approaches in understanding digital content and user sentiments.

8. Challenges

8.1 Data Complexity and Noise

In both statistical analysis and text mining, working with messy data from social media and news articles. This can make it hard to clean up the data properly before analyzing it. Things like stemming, lemmatization, and tokenization, which help break down words, might not work well because of slang, shortcuts, and spelling mistakes often found in these kinds of data.

8.2 Graph Analysis Complexity

Studying complicated networks with huge datasets, like the Wiki votes dataset, needs advanced graph analysis methods such as degree centrality, betweenness centrality, and finding communities. However understanding and getting useful information from these analyses can be hard, especially for big and tightly connected networks.

9. Limitations

While valuable insights were gleaned, limitations exist. Dependence on available data may not fully capture user behavior. Model effectiveness in sentiment analysis hinges on data quality and training. Analysis scope is confined by selected datasets and methods, potentially limiting generalizability. Automated techniques for text summarization may overlook nuanced interpretations.

10. Source Code and Dataset

GitHub Link: https://tinyurl.com/5apubcfm

11. References

Abbas, M. M. K. A. J. A. A. M. S. &. A. A., 2019. Multinomial Naive Bayes classification model for sentiment analysis.. *IJCSNS Int. J. Comput. Sci. Netw. Secur,* p. 62.

Adhikari, S., 2020. Nlp based machine learning approaches for text summarization. *Fourth International Conference on Computing Methodologies and Communication (ICCMC),* pp. 535-538.

Bedi, P. &. S. C., 2016. Community detection in social networks. Wiley interdisciplinary reviews: Data mining and knowledge discovery. pp. 115-135.

Brzezinski, J. R. &. K. G. J., 1999. *Logistic regression modeling for context-based classification..* s.l., s.n., pp. 755-759.

Campbell, W. M. D. C. K. &. W. C. J., 2013. Social network analysis with content and graphs.. *Lincoln Laboratory Journal,* pp. 61-81.

Garlapati, A. M. N. &. N. G., 2022. Classification of Toxicity in Comments using NLP and LSTM.. *International Conference on Advanced Computing and Communication Systems (ICACCS),* Volume 1, pp. 16-21.

Kulcsar, L. &. v. E. F., 2023. Twitter sentiment analysis on the cryptocurrency market..

Nilekar, S. R. S. V. R. &. R. P., 2020. Twitter trend analysis.. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology.,* pp. 94-103.

Puri, S. &. S. S. P., 2019. An efficient hindi text classification model using svm. In Computing and Network Sustainability. *Proceedings of IRSCNS ,* pp. 227-237.

Shahzad, M. K. B. L. K. T. M. I. S. R. H. M. &. K. K. S., 2021. *BPTE: Bitcoin price prediction and trend examination using Twitter sentiment analysis..* s.l., s.n., pp. 119-122.

Shamrat, F. M. J. M. C. S. I. M. M. M. J. N. B. M. M. D. P. &. R. O. M., 2021. Sentiment analysis on twitter tweets about COVID-19 vaccines using NLP and supervised KNN classification algorithm.. *Indonesian Journal of Electrical Engineering and Computer Science,* pp. 463-470.

Tong, Z. &. Z. H., 2016. A text mining research based on LDA topic modelling. *International conference on computer science, engineering and information technology ,* pp. 201-210.

Visvam Devadoss, A. K. T. V. R. &. V. D. A. K., 2019. Efficient daily news platform generation using natural language processing. *International journal of information technology,* Volume 11, pp. 295-311.