International Journal of Applied Mathematics
Volume 38 No. 3s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

TWEETING SENTIMENTS ANALYZING AND COMPARING THE EMOTIONAL PATTERN VARIANCE OF TWO USERS IN TWITTER POSTS.

Dr. K. Sundravadivelu

Assistant Professor of Computer Science

School of Information Technology

Madurai Kamaraj University

Madurai-625 021, Tamil Nadu, India.

svadivelu2021@gmail.com

https://orcid.org/0009-0001-9969-1571

Abstract:

Twitter has developed into a potent platform for expressing one's feelings and opinions in real time in the age of digital communication. This work explores and compares the emotional patterns of two distinct Twitter users by analyzing the sentiment and emotional content embedded in their tweets. Utilizing natural language processing (NLP) techniques and sentiment analysis tools, we extracted emotional cues. We classified the tweets as positive, negative, neutral, and emotion-specific types like joy, anger, sadness, and fear. The emotional variance was quantified over time to observe patterns, trends, and shifts in mood or response to external events. By employing visualizations and statistical comparisons, the research work reveals notable differences and similarities in emotional expression, frequency, and intensity between the two users. This comparative approach not only provides insights into individual emotional behavior on social media but also contributes to broader applications such as digital psychology, influencer profiling, and social media monitoring. The implementation of the proposed model provides a good performance concerning memory and time. This research work demonstrates the importance of the sentiments of human beings; their new technologies, future sentiment analysis research should be enhanced even more.

Keywords: TF-IDF, LDA, LSA, Retweets, NLP,

1. Introduction:

Billions of people use Twitter to share their ideas and feelings on day-to-day life, news, and events [10]. These tweets are analyzed to learn about the active state of people and society about the concerns mentioned, which is useful for analyzing the frequency of words, trend of words, and sentiment of tweets [12]. This form of social media content is utilized to learn about people's opinions on a particular product or event [15]. This will be used to assess the citizens' feelings and disaster situations. Twitter has over 600 million users and sends almost

Received: July 20, 2025

International Journal of Applied Mathematics
Volume 38 No. 3s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

12,000 tweets every second [11]. Nowadays, assessing tweets is primarily used to locate patterns and uncover hidden data inferences [8].

The discovery of similar patterns that are formed among people regarding generated items is one of the Twitter-based content applications [4]. The government uses similar Bayesian analysis to discover threat-related content, whereas businesses utilize it for recruitment and target marketing [17]. Individuals use this activity to identify people who are similar to them [13].

2. Related Works:

This system [9] categorizes tweets as positive, negative, and neutral before storing them in a database. The data is then compared to various classification and regression algorithms in this work. However, this work makes no recommendations regarding which method is best suited for classification and regression. This work [1] detects comparable users to profile consumers for social and security concerns. Retweets, favorites, common hashtags, common interests, profiles, mutually followed, and followers are all utilized to profile people. Humans who utilize the service review the returned results. This work employs the MapReduce technique to handle large volumes of data, although it omits key critical information such as account type. This work [3] looks into the behavior of those who use both Facebook and Twitter. It is used to discover user specificities such as buddy selection, privacy preferences, and actions undertaken. The key limitation of this work is the small sample size and the presence of latent factors [6]. This work [16] compares Twitter content to the New York Times news medium. The [14] subjects are discovered using an LDA model, and then the tweeted text is compared to the news medium using unsupervised topic modeling. However, this approach falls short when it comes to visualizing Twitter material. This review [2] gives a quick taxonomy of several topic modelling strategies as well as an overview of the most prevalent topic modelling methodology, LDA, and some of its expansions [7]. This survey also covers a few topic modeling applications in several disciplines of work. This work [5] explains hierarchical LDA, which employs a non-parametric Bayesian technique. An algorithm that generates a tree of subjects as data becomes available defines it procedurally. Every node in the topic tree represents a random variable and is assigned a word-topic distribution. Documents may be created by traversing the tree from root to leaf and sampling subjects along the way [146].

3. Comparing Twitter Archives using LDA model:

Figure 1: depicts the proposed research work architecture, which consists of three layers. The layers are as follows: 1. Tweet extraction layer, 2. Document usage layer, and 3. Sentiment patterns analysis layer. These three layers are outlined briefly in the sections that follow.

3.1 Document Usage Layer:

This step is used to determine the frequency of word occurrences in the corpus. The corpus is organized by person, and the number of times each word is used by each individual is calculated. This gives the number of words used by each individual. Determine the frequency of each word later. The pivot wider () method is used to create a data frame in order to compute the frequency of words. This layer focuses on the use of words in tweets. It is made up of two

International Journal of Applied Mathematics
Volume 38 No. 3s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

parts. 1. Determine the usage of the word and 2. Determine the change in word usage over time. Find the Usage of Words to produce precise results, the str-detect () function is used to ignore user names from the corpus. The words that are used more than 20 times are then considered. Then the word's log-odds ratio is computed.

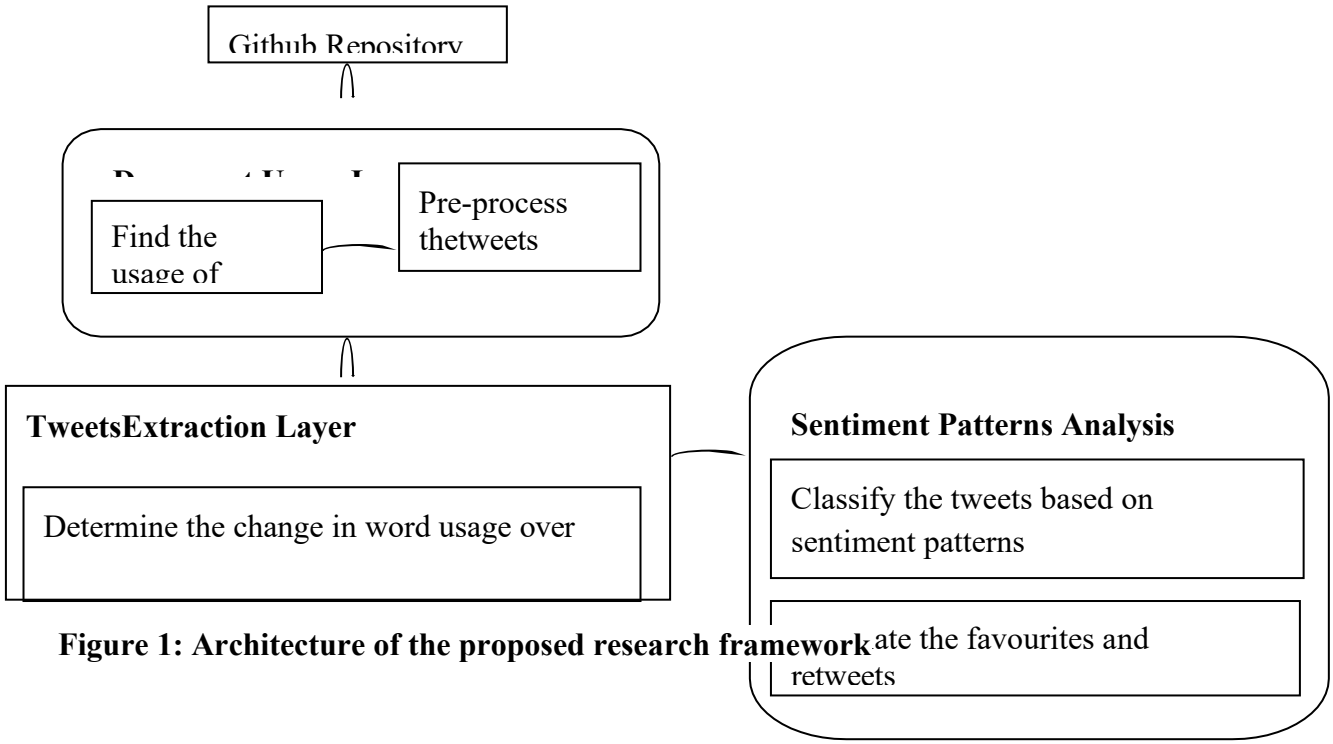


Figure 1: Architecture of the proposed research framework

Determine the terms that are more or less likely to appear from each account using the log odds ratio adaptation. Odds ratio in logarithmic form is the odds ratio. Determine the Change in Word Usage over Time. The next job is to determine which words in their individual Twitter feeds have changed rapidly over time. To begin this task, a time bin must be constructed. A time bin representing a nearly one-month time range is built for this purpose. The number of times the word appears in this timestamp bin should then be counted. Finally, look for words that appear at least 30 times. The person-word pair now has its own row in the data frame. The neatly packaged "glm" objects are then used to locate the slopes. This yields the most significant slopes, and the words' frequency in tweets changed moderately.

3.2 Tweets Extraction Layer:

The process of extracting tweets for analysis, which is the first step in the analysis process, is known as tweet extraction. To complete the job, R programming is utilized as an experiment. The Twitter dataset of two well-known celebrities, Elon Musk and Donald Trump, is taken from the Kaggle and GitHub sources, which contain more than 500,000 tweets, especially during the pandemic period from the years2019 to 2021. This layer also includes two more responsibilities. 1. Preprocess the tweets; 2. Determine the frequency of words. These two responsibilities are discussed further below. Preprocess the Tweets The collected raw tweets are not pure and very unorganized. It contains URLs, punctuation, hashtags, and a lot of redundant content that needs to be fixed in order to improve analysis quality. The tweets are

Received: July 20, 2025

International Journal of Applied Mathematics
Volume 38 No. 3s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

first organized into a corpus. The packages tm and stringr from the R language are used to do these text mining tasks. Figure 2 depicts the tweet counts for Elon Musk and Donald Trump after preprocessing.

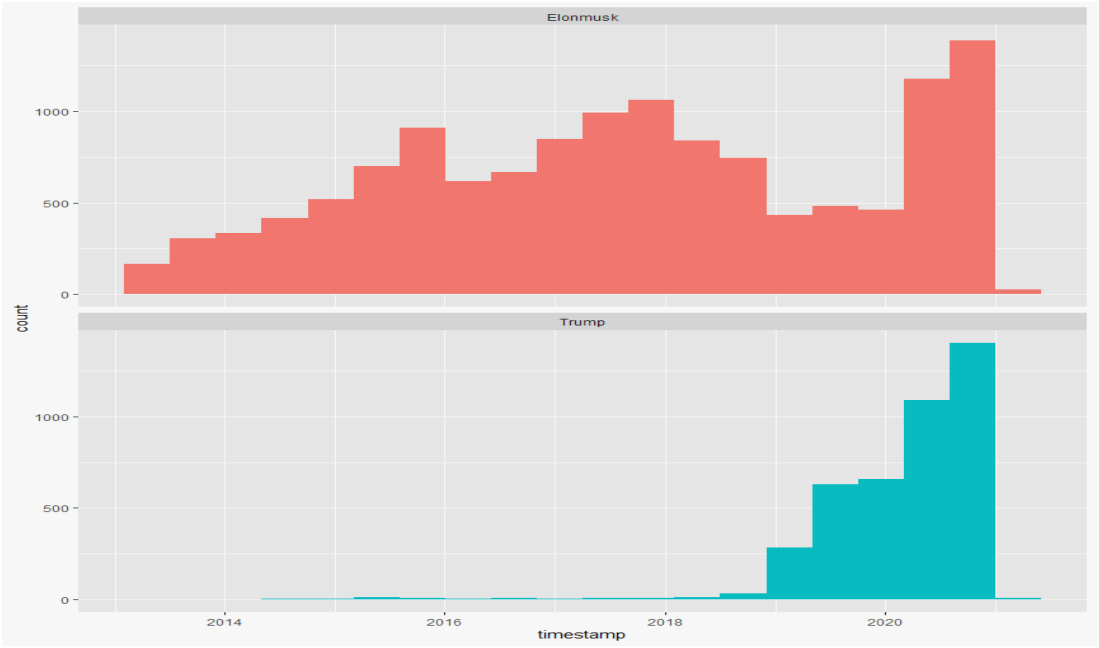


Figure 2: Twitter Counts of Elon and Trump

3.3 Sentiment Patterns Analysis Layer:

Sentiment pattern analysis is a popular natural language processing (NLP) technique for determining if a tweet is favorable, negative, or neutral. Texts that have been retweeted are retrieved from the corpus for sentiment analysis. It is made up of two jobs. 1. Find out how many retweets and favorites there are, and 2. Classify tweets depending on sentiment. These two responsibilities are discussed further.

Determine the number of favorites and retweets. In this instance, count the number of times each word was returned. Then, total the number of retweets for each person. Finally, compute the median number of retweets for each individual and word. The tidy text package's group_by() and summarize() functions, which determine the median number of retweets, can be used to accomplish this. Favorites are words that are frequently retweeted. Slice-max() returns the tweets with the highest median value. Sort the tweets according to the patterns of their sentiment. To examine popular patterns and new emotions in a text, the tidytext package provides the get_sentiments() method, which detects sentiments using the nrc, Bing, and AFINN lexicons. The repository with word-to-emotion pairs is generated by combining tidytext and the lexicons of the text data package. This work used the AFINN lexicon, which assigns a negative or positive score to each word, after which the emotion score is calculated using the get_sentiments() method. Lastly, a total score that is greater than zero indicates a positive attitude, a score that is less than zero indicates a negative attitude, and a score that is zero indicates a neutral attitude.

4. Results and Discussions

Received: July 20, 2025

International Journal of Applied Mathematics
Volume 38 No. 3s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

The proposed work was implemented in Tool R-4.2.1 in Pentium® Core i7 or above with 16 GB RAM, running on Windows 10 and minimum hardware technology. The results retrieved from the three layers of the proposed work are discussed in three aspects. 1. Based on the frequency of words, Based on word usage, and 3. Based on favorites, retweet count. Based on word frequency. The Tweets extraction layer generates the frequency of word occurrences in a plot using the ggplot2 tool, as demonstrated in Figure 3 shows that the words closest to the line return the same frequency for both people. The terms closest to the line indicate that they are used more frequently by one person than the next.

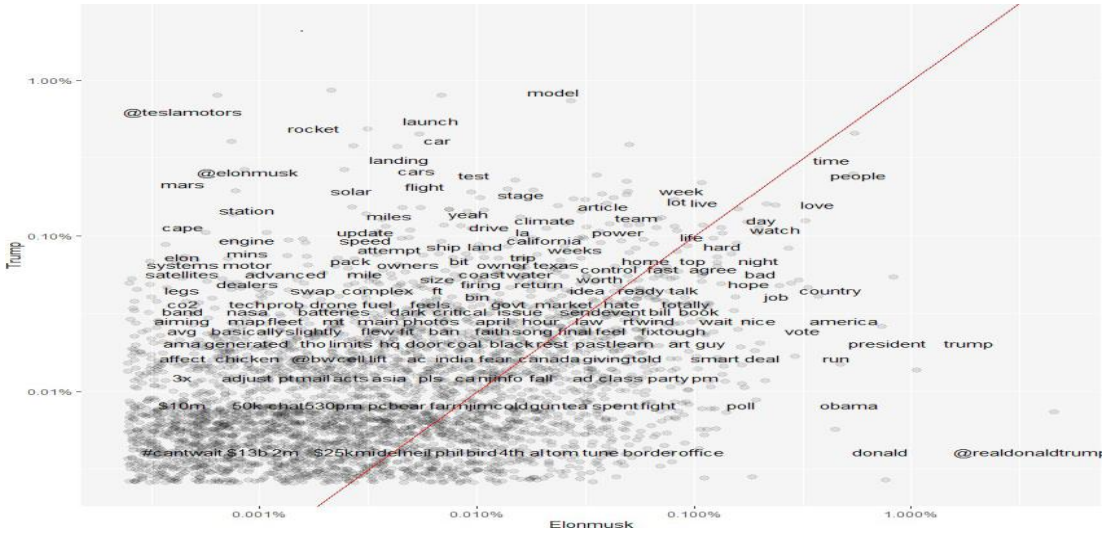


Figure3: Frequency of Word Occurrences

Based on words usage the log odds ratio approach is employed in determining word usage. Figure 4 depicts the log odds ratio for the two individuals, and the figure shows that Elon was more active in the given timestamp (in our example, 2020) than Trump.

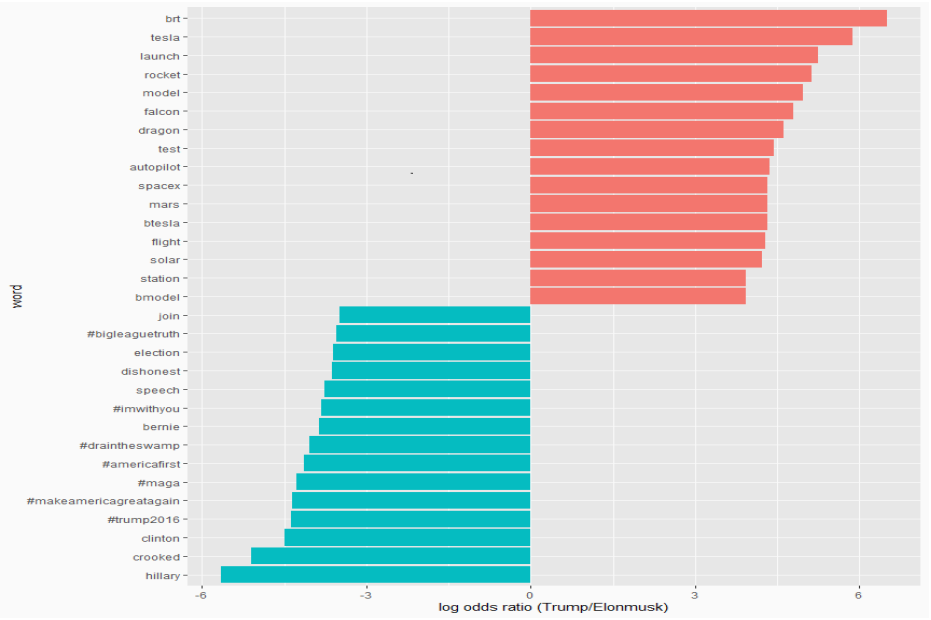


Figure4: Log odds Ratio to Calculate usage of Words

Received: July 20, 2025

International Journal of Applied Mathematics
Volume 38 No. 3s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

These two figures illustrate that the change in word usage is greater in Trump. Based on favorites, retweet count. The favorites and retweet count are included throughout the sentiment analysis jobs. Figure 6 depicts the computed median value of retweets containing each word for Elon and Trump.

The evaluation of historical shifts in word usage is the next task assigned to the word usage layer. The changes in how words are used by Elon and Trump are depicted in Figures 5 and 6.

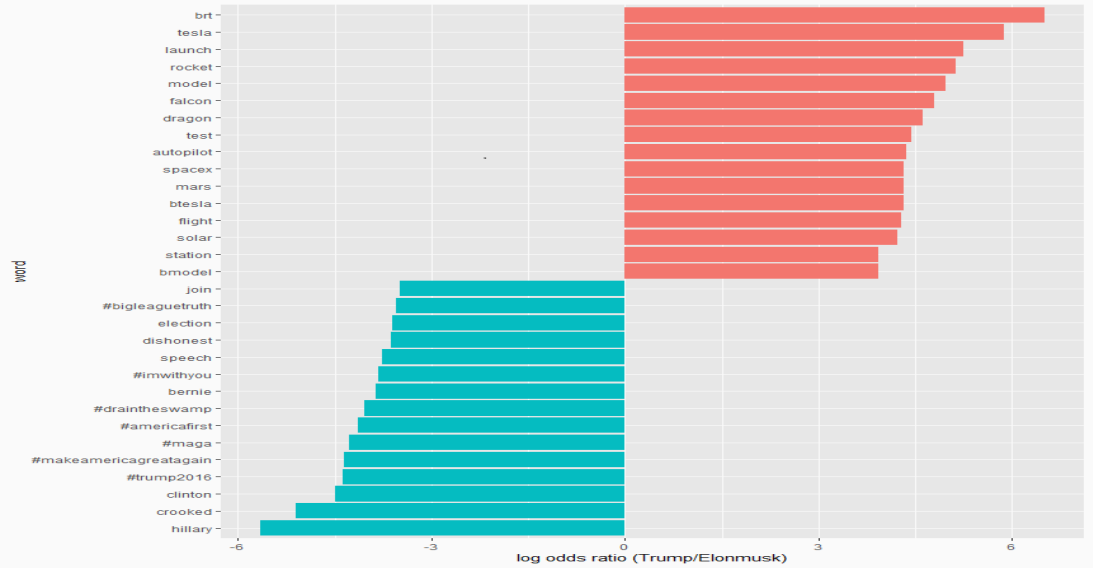


Figure 5: Change in usage of words in the timeframe Jan 2020 to December 2020 for Elonmusk.

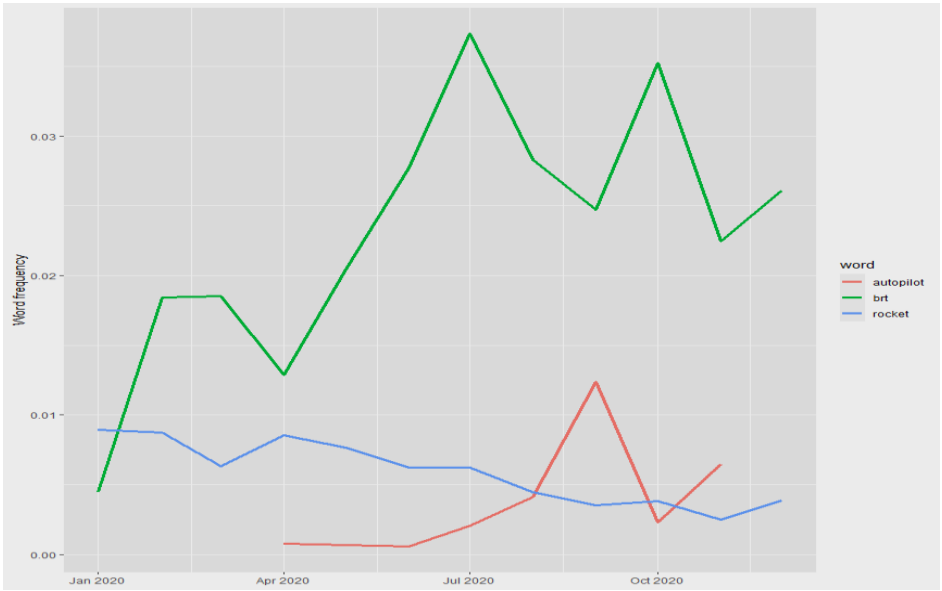


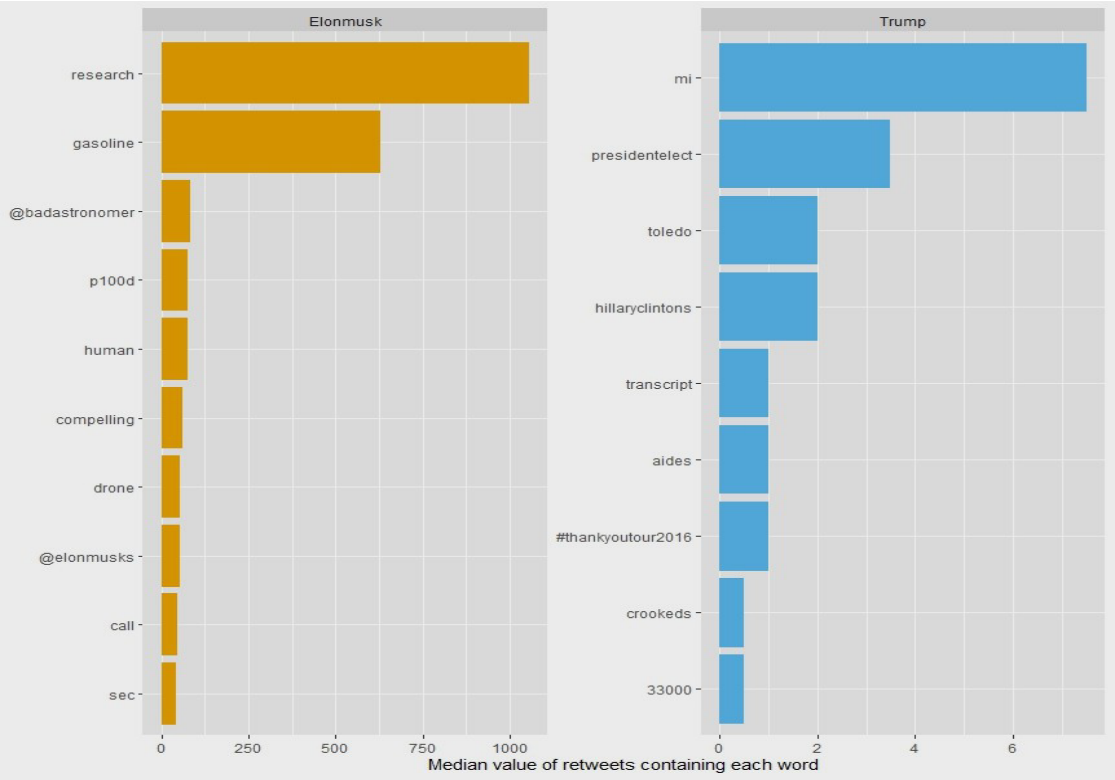
Figure 6: Change in usage of words in the timeframe Jan 2020 to December 2020 for Trump.

Received: July 20, 2025

International Journal of Applied Mathematics
Volume 38 No. 3s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

The final goal is to assess Elon's and Trump's tweet sentiment trends. Figure 7 depicts the sentiment classification results.

Figure 7: Median Values of Retweets containing each word for Elon and Trump



The term Trump is depicted as positive in the above graphic, indicating the previous president of the United States, and thus it receives a positive score. The suggested work will then address the TF-IDF score. The frequency with which a word appears in the dataset is represented by the term frequency (TF).

TF is calculated as follows:

TF=Frequency (word) /Total number of

The Inverse Document Frequency (IDF) is an additional metric that assigns less weight to terms that are frequently used and more weight to words that are rarely used.

Received: July 20, 2025

International Journal of Applied Mathematics
Volume 38 No. 3s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

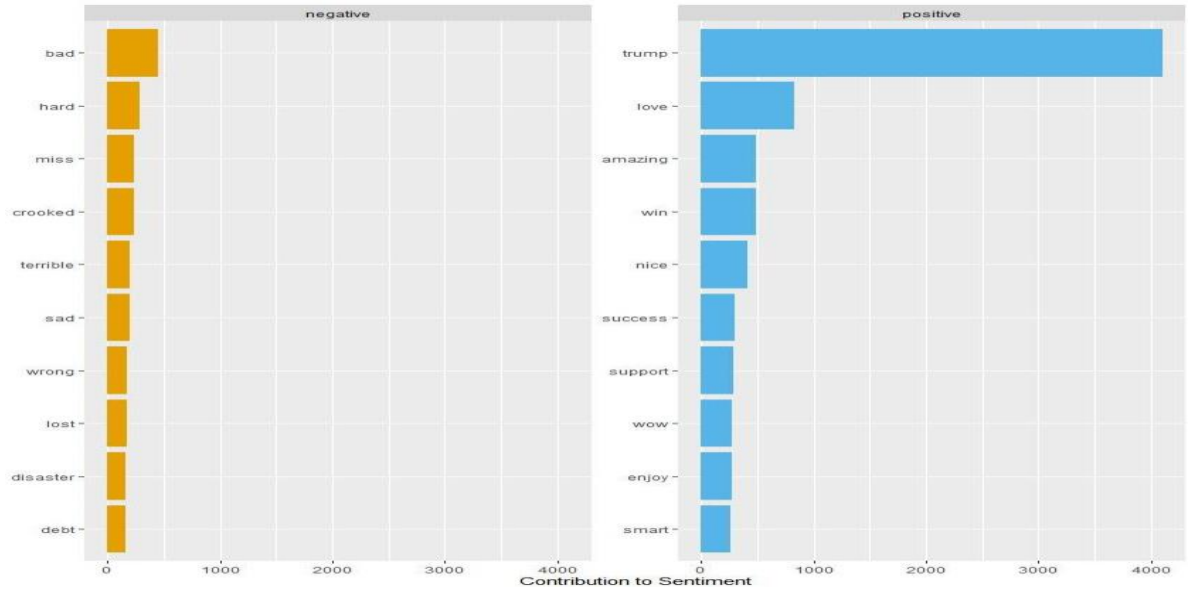


Figure 8: Results of Sentiment Classification

The TF and IDF metrics are combined to produce the TF-IDF score, which indicates the value of a word. Figure 8 depicts the TF-IDF results for the Elon Musk and Donald Trump tweets dataset.

The TF-IDF score is used to determine the value of a word, which returns the value of a word as more accurate, nearly 95% of the time, than other methods. The proposed model classifies the sentiments from the respective Twitter archives with 90% more precision than LSA-based sentiment classification. In addition to that, the proposed model also calculates favorite and retweet counts accurately.

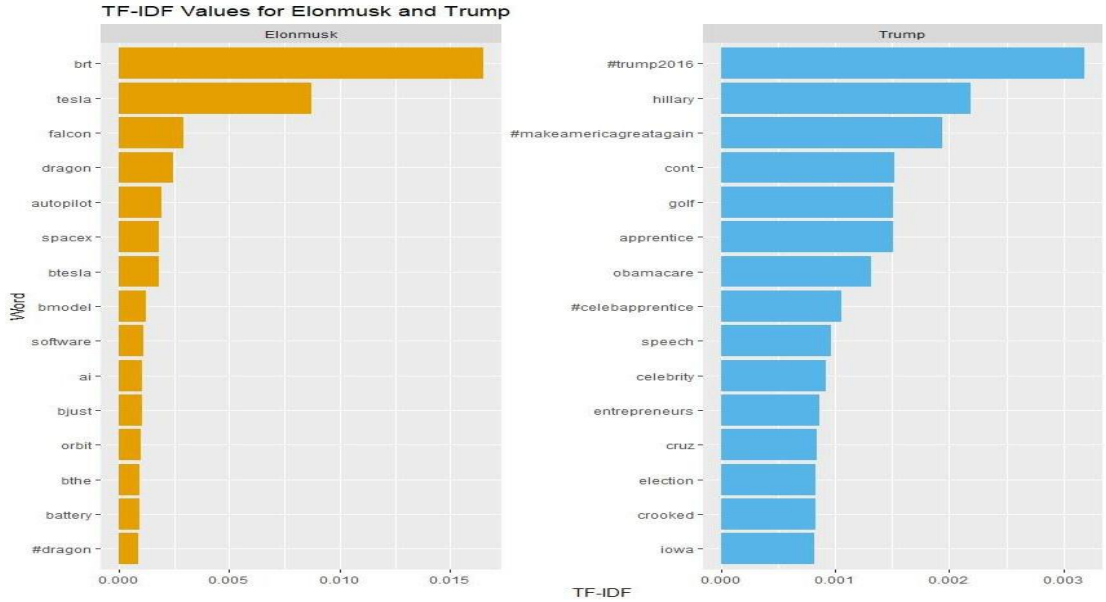


Figure9: TF-IDF Results Related to Elon Musk andTrumpTweets

Received: July 20, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

It was discovered that the proposed model performed well for all metrics when compared to the existing model for a variety of datasets and applications.

5. Conclusion:

The proposed model yields the most significant slope of terms, which represent the words that are frequently changed in tweets. It also calculates the frequency of word occurrences in tweets concerning the given timestamp. The log odds ratio approaches are used to determine the word usage and also return the usage of words in a given timestamp concerning two people's tweets. It returns who uses which data of words and when it changes in the given timestamp (in our example, 2020). By the LDA algorithm, the proposed model classifies the sentiments from the respective Twitter archives. In addition, the proposed model also calculates words that are favorited and retweeted. The proposed system returns the TF-IDF score, which is a popular approach for weighing terms for NLP tasks. A value assigned to a term based on how important it is in a document, scaled by how important it is across all of the documents in this corpus. This is done by mathematically removing all of the naturally occurring English words and choosing only the valuable words. The implementation of the proposed model provides a good performance with respect to memory and time. The generated TF-IDF score will be utilized in the future to expand the research towards neural network-based categorization. Furthermore, additional research is to be conducted to deal with diverse sentiment patterns. LDA has been criticized for its inability to scale due to the linearity of the technology upon which it is built. Hence, in the future, the other variations, such as PLSI, the probabilistic form of LSI, are going to be used.

References:

[1] AlMahmoud, Hind, and Shurug AlKhalifa, "TSim: a system for discovering similar users on Twitter", Journal of Big Data 5, pp. 1-20, 2018.

[2] Buccafurri, Francesco, Gianluca Lax, Serena Nicolazzo, and AntoninoNocera", Comparing Twitter and Facebook user behavior: Privacy and other aspects." Computers in Human Behavior 52 (2015): 87-95.

[3] Buenano-Fernandez, Diego, Mario Gonzalez, David Gil, and Sergio Luján Mora. "Text mining of open-ended questions in self-assessment of university teachers: An LDA topic modeling approach." IEEE Access 8, pp. 35318-35330, 2020.

[4] Goodman, Noah, and Daniel Light. "Coding Twitter, lessons from a content analysis of informal science‖, In 2016 Annual Meeting of the American Educational Research Association, 2016.

[5] Griffiths, Thomas, Michael Jordan, Joshua Tenenbaum, and David Blei", Hierarchical topic models and the nested Chinese restaurant process", Advances in neural information processing systems 16, 2003.

[6] K. Sundravadivelu, “Effective Information and Extraction an Improved Semantic Pattern Based Approach in LDA Model using Bigdata”, Communications on Applied Nonlinear

Received: July 20, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Analysis, https://internationalpubls.com,ISSN: 1074-133X, Vol 32 No. 7s (2025), pp13-24, 2025.

[7] K. Sundravadivelu, "Improved Semantic Information and Extraction based Effective Pattern Discovery Mining in Bigdata Using Latent Semantic Indexing Model", Journal of Electrical Systems, Volume 20, Issue 11s, pp: 1757-1764, ISSN: 1112-5209, 2024.

[8] Krouska, Akrivi, Christos Troussas, and Maria Virvou, "The effect of preprocessing techniques on Twitter sentiment analysis.", In 2016 7th international conference on information, intelligence, systems & applications (IISA), pp. 1-5, 2016.

[9] Kulkarni, Aditya, and Shubham Mhaske, "Tweet Sentiment Analysis and Study and Comparison of Various Approaches and Classification Algorithms Used", IRJET, April 2020.

[10] Kwak, Haewoon, Changhyun Lee, Hosung Park, and Sue Moon, "What is Twitter, a social network or a news media?", In Proceedings of the 19th international conference on World Wide Web, pp. 591-600, 2010.

[11] Razis, Gerasimos, and Ioannis Anagnostopoulos, "Discovering similar Twitter accounts using semantics", Engineering Applications of Artificial Intelligence 51, pp.37-49, 2016.

[12] Srinivasan, M. S., SrinathSrinivasa, and Sunil Thulasidasan, "A comparative study of two models for celebrity identification on Twitter", In Proceedings of the 20th International Conference on Management of Data, pp. 57-65, 2014.

[13] Sundravadivelu. K, M. Thangaraj, "Analyzing Educational Tweets using LDA Model", International Journal of Intelligent Systems and Applications in Engineering (IJISAE), Volume 10, Issue 4, ISSN:2147-6799, PP. 100- 104, Dec. 2022.

[14] Sundravadivelu. K, M. Thangaraj, ―A Novel Approach for Discovering the Patterns by using PDBD Model in Big Data", Journal of Computer Science, (Science Publications), Volume 18 Issues 5, DOI: 10.3844 / jcssp.2022.382.395, ISSN: 1549-3636, pp.382-395, May 2022.

[15] Zhang, L and Liu, B. ―Sentiment Analysis and Opinion Mining‖ , Springer US, Boston, MA, pp 1152–1161, 2017.

[16] Zhao, Wayne Xin, Jing Jiang, JianshuWeng, Jing He, Ee-Peng Lim, Hongfei Yan, and Xiaoming Li, "Comparing twitter and traditional media using topic models", In Advances in Information Retrieval: 33rd European Conference on IR Research, ECIR 2011, Dublin, Ireland, April 18-21, 2011. Proceedings 33, Springer Berlin Heidelberg, pp. 338-349, 2011.

[17] Zimmer, Michael, and Nicholas John Proferes, "A topology of Twitter research: disciplines, methods, and ethics", Aslib Journal of Information Management 66, no. 3 , pp. 250-261,2014.

Received: July 20, 2025