Using Support Vector Machine to Predict Political Affiliations on Twitter: Machine Learning approach

Muhammad Javed[1, *], Kiran Hanif[1] , Arslan Ali Raza[2], Syeda Maryum Batool[1] , Syed Muhammad Ali Haider[1]

[1]Institute of Computing and Information Technology, Gomal University, Dera Ismail Khan, K.P.K, Pakistan.
[2]Department of Computer Science, COMSATS University Islamabad, Vehari Campus, Vehari 45550, Pakistan

Abstract
The current study aimed to evaluate the effectiveness of using Support Vector Machine (SVM) for political affiliation classification. The system was designed to analyze the political tweets collected from Twitter and classify them as positive, negative, and neutral. The performance analysis of the SVM classifier was based on the calculation of metrics such as accuracy, precision, recall, and f1-score. The results showed that the classifier had high accuracy and f1-score, indicating its effectiveness in classifying the political tweets. The implementation of SVM in this study is based on the principle of Structural Risk Minimization (SRM), which endeavors to identify the maximum margin hyperplane between two classes of data. The results indicate that SVM can be a reliable classification approach for the analysis of political affiliations, possessing the capability to accurately categorize both linear and non-linear information using linear, polynomial or radial basis kernels. This paper provides a comprehensive overview of using SVM for political affiliation analysis and highlights the importance of using accurate classification methods in the field of political analysis.

Keywords:
Classification method, Support Vector Machine, Political affiliation, Twitter.

1. Introduction

Web 2.0 is another name for second generation internet connectivity which helps Internet users to collaborate or help, communicate and share information online. In addition, social media means websites and applications that help different Internet users to create and share content from miles away. Wikipedia, Facebook, Twitter, and various blogs are also included in the web 2.0. All of these are designed to convey information and ideas from one user to another.

Web 2.0 has revolutionized the world of research into content-based analysis and research designed for Internet users. The expression of messages reflects your thinking. Nowadays, their ideas are presented through various means, in which the world of internet is the most important, which can also be called digital medium or digital media. Observing consumers' opinions, debates and discussions can be very helpful in getting information about people's attitudes towards their attitudes, political, social, business, religious tendencies & also many related ideas. Here mining is called the technique or process which may help to understand & sorting of the consumer's opinion.

For the sake of public interest various techniques are used to calculate the opinion mining. It uses different fields of routine life, marketing in business which may help any institute or company to evaluate its company's effectiveness & helps to know the popularity of its products.

These methods allow & help companies to gain a clear and transparent perspective than old-fashioned methods such as interviews and other older analyses. More recently, the Twitter social media platforms are being used as the most popular & best platform for campaigning for any public figure or any politician & also for any political events to gain more popularity.

Numerous techniques & methods are used to analyze & predict regarding the concerned elections results we may use these techniques of mining of opinion, these techniques may be as supervised, unsupervised techniques. In recently Javed et al. [1] experimented that, social media platforms such as Twitter has become one of the most popular platforms used by public for socio monitoring, Political contents and political parties to express sentiments regarding political events, campaigns and elections.

Previous Studies Boutet et al.[2], Conover et al. [3] , Ellen & Parameswaran [4] to analyze & predict the political situations & affiliation of any user. They have used the limited datasets & number of machine classification. For political arrangement of users we may use the supervised machine learning classifier. This classifier with having limited number of structures.

The first step in effective research is to categorize tweets containing political affiliations. In it, we will further classify on the basis of the political training and expression datasets of different users. The political

Manuscript received May 5, 2024

Manuscript revised May 20, 2024

https://doi.org/10.22937/IJCSNS.2024.24.5.25

leanings of those who use such tweets can help to predict future election results. We can come to a definite conclusion by examining the identities of tweets containing such political content and related elements.

We have obtained the dataset linked to political affiliations & then preprocess it. Using the supervised machine learning technique namely Support Vector Machine, we have categorized user feedback regarding with their political associations& views.

2. Literature Review

Boutet et al. [1] proposed a system for the prediction of UK general Elections. The main purpose of this study was to compile a list of political tweets that may related to help, analyses & to predict the UK general election in 2010. This data set is used by more than 220,000 Internet users for approximately 1,150,000 political tweets after ratings.

For better results, they divided the dataset into three categories,

Volume rating 2- retweet rating 3- SVM rating.

Conover et al. [2] suggest a number of effective ways to gauge the political orientation, political inclination of Twitter users during 2010 midterm elections of US. The following twitter data based on content and structure of political contacts. Through the help of linear support vector machines (SVMs), we may differentiate users between in the 'left' and 'right' categories. Which led to political classification in the dataset. Latent Semantic Analysis (LSA) of Used Hashtags. SVM systematically trained the full details of the tweets of Twitter users, resulting in 91% accurate predictions of political affiliation, 373 Twitter users were identified by human observers as an expression of 'left' political alignment, and 506 users as 'right'. Gone, and the remaining 77 were classified as 'ambiguous'.

Though the author's group affiliation, the work done by Ellen & Permishorn [3] ranks posts through the website medium. On the basis of political information, they classify the data in two datasets. They are actually based on real world information, which is the real issues related to the Israeli / Palestinian dialogue and their current issues, their solutions, in their favor or against a supervised ranking algorithm, k-Nearest Neighbor (KNN) algorithm and Support Vector Machines (SVM), are used to organize their information into datasets.

This work provides a detailed comparative overview of REPTree, Simple Cart and Random Tree to classify &categorize the Indian views, news contexts using the wiki platform Kalmegh [5] about 649 news items were collected. The term frequency matrix was used to further test the news.

The NLP process may also use to highlight the output news like text, audio, and video to relevant e-learning data. This leads to the conclusion that the best in all other news ranking performance is Random Tree algorithm. The overall performance of the REPTree and Simple Cart algorithms is inefficient and unacceptable, as it has been observed that these two algorithms only accurately categorize political news

The main importance here was on better ranking of tweets keeping in mind "liberal opinions". Tokenization, stop word removal etc. were used through various pre-processing steps. Which helped to achieve effective satisfactory results.

The main objective of the research conducted by Sharma & Moh [6] was to collect more than 42,000 tweets and then categorize their emotional analysis on Twitter in different ways to predict Indian elections. By using dictionary based & SVM techniques, namely Bayes &Supervised Vector Machine they classify the data through the basic steps of preprocessing. This process and rating reflect the performance test results. The rating yielded 68% accurate results, while NB and SVM provided results of 62% and 78% respectively. It requires other classifications for better results, such as regression and taking advantage of even better results from random forest research.

The work done by Khatua et al. [7] which aimed to achieve results by predicting a political trend based on Twitter before the election results. The volume of tweets as well as the emotions of the users were also analyzed. During the election campaign from March 2014 to May 2014, they analyzed& predicted the election results on the bases of people interests. It has been observed that tweet coefficients (%) are statistically significant during all assessment. It shows & prove all the concerned results are remained same through the analyzing and also after featuring the national & regional factors. The Research & analysis of this study relates to the loss of small example size.

Wicaksono [8] proposed a system based on two classifiers called Binary Multi, which sustained emotional buildup. In term of correct approaches, the results were more hopeful. Effective research on the application of other ratings is needed to achieve better performance.

Sahu and Nanda [9] performed work for the prediction of elections results for Indian political parties from Twitter data. They check the positive and negative Tweets of twitter users. They used python tools to predict the positive and negative tweets. They process the twitter data with the help of lexicons and dictionaries. After the positive and negative classes evaluation of each political parties that clear most of the voters have a positive attitude for Bharatiya Janata Party and Predict that Bharatiya Janata Party wins log saba General election 2019.

Nayak and Natarjan [10] presented that, by using Twitter data we can categorize any user's political

affiliation and views by using supervised machine learning technique & support vector machine. The basic concept of Linear Support Vector Machine in favor of sentiment classification is to decide the hyper plane that divides the corpus.

An election prediction system for users opinions, posted on an election prediction website http://www.electionprediction.org/, was proposed by Kim & Hovy [11]. They used SVM classifier to predict opinions, outperforming all the baselines methods.. Previously, researchers in opinion analysis, mostly focused on judgment opinions, which expressed positive or negative sentiment about a topic.

Mejova et al. [12] proposed a supervised machine learning classification system by applying Tuned SVM light classifier for the prediction of sentiments expressed by political candidates. The main emphasis was on classifying tweets in terms liberal sentiments.

Arslan et al. [13] developed a framework to predict political affiliation of online users during the political protest of PTI which was held in 2014. They used lexicon based approach at sentence level, they achieved efficient outcomes in comparison of baseline studies.

Macrohon et al. [14] performed sentiment analysis to predict the political moods of Philippines for presidential election by using Semi-supervised technique, they gained promising results up to 84.83% of accuracy as compare to the existing research.

Chauhan, P et al. [15] explored and predicted the power of social media networks by conducting a survey to predict the political affiliation of on users by utilizing various techniques, they also highlighted different issues and challenges to predict the moods of pubic associated with social media contends,

It is obvious for accurate and efficient sentiment analysis the contents of social media should have normalized. Javed et al. [16] proposed mechanism for the normalization of unstructured text of politics. They employed unsupervised approach at sentence level for the evaluation of political contents of microblogging website. They achieved 82.357% average accuracy in comparison of existing studies.

2. Methodology

The methodology employed in the system design involves the use of Support Vector Machine (SVM) as the classifier to analyze the political affiliation of tweets. The utilization of Support Vector Machine (SVM) in this study was rooted in the principle of the Structural Risk Minimization (SRM), which seeks to identify the maximum margin hyperplane among two sets of data. The methodology involves dividing the corpus into two classes of positive and negative words, and then classifying them into either of the classes. To assess the efficacy of the SVM classifier, various performance metrics, such as precision, accuracy, recall and F1-Score are calculated utilizing the sklearn library. The results are analyzed to determine the effectiveness of the classifier in predicting political affiliation in tweets.

3.1 System Design Overview

The system design focuses on using support vector machines (SVM) to classify political affiliation based on tweets data collected from Twitter. The system consists of two main parts, training and testing.

In the training phase, the system trains a classifier model using a dataset of tweets containing information on users' political views. The classifier model is trained to categorize tweets into different political affiliations, such as liberal, conservative, socialist, etc.

In the testing phase, the system crawls real-time tweets from Twitter using the Twitter Streaming API. The tweets are preprocessed by performing text cleaning and normalization techniques such as transforming all text to lowercase, removing URLs, hashtags, and mentions, and performing lemmatization. The preprocessed data is then used as input features for the SVM classifier model.

The classifier model outputs the political affiliation of each tweet and all results are stored for further analysis. Diagnostic metrics such as accuracy, recall, and F-measure are used to evaluate the performance of the classifier model.

The system design provides a systematic approach for classifying political affiliation from tweets using SVM. The use of preprocessing and normalization techniques, along with the implementation of SVM classifier, helps to improve the accuracy and reliability of the results.

3.2 Data Collection

The data collection process in this system involves using the Twitter API to gather tweets related to political and non-political affiliations. The API is used to search for tweets containing specific keywords related to politics and non-politics, and the relevant tweets are collected and stored in a database. The collected data is then preprocessed to remove irrelevant information, handle missing values, and correct any errors in the data. This preprocessed data is used to train the Support Vector Machine (SVM) classifier to recognize the political and non-political affiliations in the tweets. Finally, the system uses the trained SVM classifier and the T.F-I.D.F approach for feature extraction to determine the relevance

and weight of each word in the tweet. In the proposed system, a crucial aspect of the design is the selection of features that are deemed important for the classification of tweets into political and non-political affiliations. This process is carried out by utilizing a Knowledge Base to identify the relevant features that are crucial to the classification task. Subsequently, the selected features are utilized in the classification phase, which determines the political and non-political affiliations of the tweets.

3.3 Preprocessing

Next, the stop words were removed from the data, stop words are words which do not contain important meaning and are usually removed from texts. These words can include words such as "a", "an", "the", etc. After removing stop words, words were then stemmed to their root words using the Porter Stemming algorithm. Stemming helps reduce the words to their core structure and helps in reducing the data size. After stemming, words were then lemmatized using the WordNetLemmatizer in Python's NLTK library. Lemmatization helps in reducing words to their core structure while also taking into consideration their context, thus preserving the context of words. The final preprocessed data was then ready for feature extraction and selection to be used in training the SVM classifier for political and non-political affiliation classification on twitter.

3.4 Classifying and ordering of political views & affiliations

In this study, the political views and affiliations of Twitter users were categorized through the use of a supervised machine learning technique, specifically Support Vector Machines (SVM). The classification was performed by identifying the hyperplane that separates the corpus using linear SVM. A comprehensive dataset containing information on political views of users was used for training and testing the classifier. The results were evaluated using various diagnostic metrics such as accuracy, precision, recall, and F1-score. This study provides insight into the use of SVM for classifying and ordering political views and affiliations from Twitter data.

3.5 Support Vector Machines (SVMs)

The use of Support Vector Machines (SVM) in classification is a well-established technique that follows the Structural Risk Minimization (SRM) approach. The objective of Structural Risk Minimization (SRM) is to find the hyperplane with the greatest separation between two sets of data. This classification method can handle both linear and nonlinear data, and the kernel used can be linear, polynomial, or radial basis.

One of the core components of using SVMs for prediction is the calculation of a hyperplane. The hyperplane separates the data into different classes, with the objective of maximizing the margin between the classes. The equation for a hyperplane can be expressed as:

$$W^T x + b = 0 \quad (1)$$

Where w is the weight vector, x is a data point, and b is the bias.

Another important aspect of SVMs is the calculation of the support vectors. These are the data points closest to the hyperplane and have the greatest impact on the position of the hyperplane. The equation for the decision function, which determines the class membership of a data point, is:

$$F(x) = \text{sign}(w^T x + b) \quad (2)$$

Where sign(x) returns the sign of x, either -1 or 1.

One practical example of using SVMs to predict political affiliations on Twitter could be to classify tweets into either “left-leaning” or “right-leaning” based on their content. The process would involve collecting and preprocessing a large number of tweets, selecting relevant features, and using an SVM classifier to train on the data. The classifier would then be used to predict the political affiliation of new tweets based on their content.

Another example could be to classify users based on their past tweets and the tweets of other users they follow. This would involve collecting and preprocessing a large number of tweets, calculating the relationships between users, and using an SVM classifier to train on the data. The purpose of the classifier is to assign political affiliations to new users on Twitter based on their previous behavior and connections, using prediction techniques.

In both of these examples, the accuracy of the predictions would be evaluated using metrics such as precision, recall, and F1-score. The results could then be used to better understand the role of Twitter in shaping political opinions and affiliations.

There are other formulas commonly used in the analysis of SVMs for prediction. One such formula is the optimization problem for finding the optimal hyperplane, also known as the maximum margin classifier. The objective is to find the hyperplane that maximizes the margin between the classes while correctly classifying the training data. This can be expressed as the following optimization problem:

$$\text{Min } \tfrac{1}{2} \|w\|^2$$

Subject to: $y_i (w^T x_i + b) \geq 1$, for $i = 1, 2, \ldots, n$

Where w is the weight vector, x_i is a training data point, y_i is the class label for x_i (-1 or 1), b is the bias, and n is the number of training data points.

Another formula that is commonly used in SVM analysis is the kernel function. The kernel function is used to map

the data into a higher dimensional space where a linear boundary can be more easily defined. Commonly used kernel functions include linear, polynomial, and radial basis function (RBF) kernels. The equation for an RBF kernel is:

$$K(x_i, x_j) = \exp(-||x_i - x_j||^2 / 2\sigma^2) \quad (3)$$

Where x_i and x_j are data points, $||x_i - x_j||$ is the Euclidean distance between x_i and x_j, and σ is the bandwidth parameter.

3.6 Transformation

In this study, the transformation step of the preprocessing phase is addressed. During the transformation step, the aim is to clean the raw tweets obtained from Twitter to transform them into a structured arrangement appropriate for further assessment.
The transformation phase of the system involves several steps to clean and structure the raw dataset of tweets. The tweets are converted from uppercase to lowercase and URLs are replaced with the generic term 'URL'. Additionally, @username is replaced with 'AT_USER' and hashtags are replaced with their exact wording without the hash symbol. To further improve the structure of the data, any punctuation present at the start or end of the tweets is removed, and multiple whitespaces are consolidated into a single space. The outcome of these transformations is a cleaned and organized dataset, which can be further processed for analysis. This step is crucial for improving the accuracy of the results obtained from the subsequent phases.

3.7 Negation Handling

In the present study, the negation handling step is an important aspect of the data preprocessing phase. The aim of the negation handling step is to precisely recognize the scope of negation and reverse the polarities of opinion-based words that are affected by negation words. The occurrence of negation words, such as "not", "no", "never", "cannot", "should not", "would not", and others, in the tweets can alter the sentiment direction of other words in the sentence. To ensure that the sentiment analysis results accurately reflect the sentiment expressed in the tweets, it is crucial to perform negation handling.

3.8 Tokenization

In the context of sentiment analysis on political affiliation and views on Twitter, tokenization is an important step in the preprocessing phase. The process of tokenization is a crucial step in natural language processing that helps to divide a sentence into meaningful segments. This is achieved by breaking down the sentence into smaller components, such as words, keywords, phrases, symbols, and other elements, that can be easily processed and analyzed. Tokenization plays a significant role in enabling the system to effectively understand the structure and context of the text, making it an important step in text processing. In this process, the uniformed sentences obtained from the negation handling step are transformed into smaller components, such as unigrams. These resultant words serve as the input for the next preprocessing step, providing a more structured and organized representation of the text data. The implementation of tokenization helps to improve the overall efficiency and accuracy of the sentiment analysis system.

3.9 Filtering

The step of filtering in the preprocessing stage involves the removal of stop words from the tokenized words obtained from the Tokenization step. Stop words are the most common words found in any language and are considered to have no significant impact on the sentiment analysis task. Therefore, they are filtered out to reduce the dimensionality of the data and increase the efficiency of the model. The list of stop words can be obtained from pre-defined lists or can be obtained from the data itself. The elimination of stop words results in a set of filtered words that are then utilized as inputs for the subsequent preprocessing stage.

3.10 lemmatization

The normalization step in the preprocessing of text data is an important step as it helps in reducing the complexity of the text data by transforming words into their base or root form. This is achieved through lemmatization, which is the process of grouping together the inflected forms of a word into a single item. The lemma or dictionary form of a word is then used to identify the word. By normalizing the text data, the system reduces the number of unique words, making it easier to extract features and improve the accuracy of the analysis.

3.11 Feature Selection

The feature selection step in the system involves selecting the most relevant features from the Knowledge Base to improve the accuracy of the classifier. The process begins by comparing the features obtained from the preprocessing step with the Knowledge Base. The

Knowledge Base contains a collection of relevant words in various domains such as education, business, and crime. The system then selects the most relevant features from the Knowledge Base that would be used as input features for the classification step. This process ensures that the system only uses essential features for the best accuracy.

3.12 Feature Extraction

In the context of political and non-political affiliation on twitter, the system utilizes Support Vector Machine (SVM) and performs feature extraction using Term Frequency-Inverse Document Frequency (TF-IDF) method. TF-IDF is a measure of the importance of a word in a document, where the importance is calculated as the product of term frequency (TF) and inverse document frequency (IDF). This method is used to extract meaningful features from the data, which can be used as input for the SVM classifier. The system calculates the term frequency of each word in a document and adjusts the weight of the word based on its rarity in the collection of documents. The resulting TF-IDF scores are used to select features and train the SVM classifier. The ultimate goal is to use the trained SVM classifier to classify tweets into political or non-political affiliations.

3.13 Text Classification

Classification of text involves the categorization of text into organized groups based on its content. The classification of sentiment orientation in text is accomplished through the utilization of classifiers, which are machine learning models capable of assigning tags to text automatically based on its contents. There are two main methodologies for sentiment classification: lexicon-based and machine learning-based. In the lexicon-based approach, individual words are divided into tokens and compared with a sentiment lexicon, which contains a polarity value for each word. The overall sentiment of the sentence is then determined by summing up these polarities. Conversely, the machine-learning-based approach employs a labeled training set to train a classifier that predicts sentiment orientation in new texts. This approach uses a supervised learning algorithm. This approach involves generating a function that maps inputs to desired outputs based on a set of predictors and the target or outcome variable. The training process continues until the model reaches a desired level of accuracy. In this system, the Support Vector Machine (SVM) and K-Nearest Neighbour (K-NN) classifiers are utilized.

3.14 Support Vector Machine (One-Versus-One)

In the context of the system developed to analyze political and non-political affiliations on Twitter, the Support Vector Machine (SVM) classifier is used in a One-versus-One (O.V.O) approach. The O.V.O strategy trains K(K-1)/2 binary classifiers for a K-way multi-class problem, where each classifier is trained on the samples of a pair of classes from the original training set. The goal of this approach is to learn to distinguish between these two classes. At prediction time, all K(K-1)/2 classifiers are applied to an unseen sample and the class that received the highest number of "+1" predictions is predicted by the combined classifier. The system uses this approach to handle multi-class classification problems, as it is not limited to binary classification like the traditional SVM. It is worth mentioning that despite the advantages of the One-Versus-One (O.V.O) strategy, it may still face challenges in certain areas of the input domain where multiple classes receive the same number of votes, resulting in ambiguity.

3. Performance Analysis

The system evaluation concentrates on analyzing the accuracy of the Support Vector Machine (SVM) classifier. This analysis involves computing various metrics, including accuracy, precision, recall, and f1-score, to determine the classifier's effectiveness. The accuracy of the classifier is evaluated by computing the ratio of the number of words that have been accurately classified as positive, negative, or neutral, over the total number of words in the corpus. The precision of the classifier provides insight into its exactness, while the recall metric reflects its sensitivity and thoroughness. The precision and recall metrics are calculated using sklearn libraries based on the SVM classifier. The F1-Score is the harmonic average of precision and recall and ranges from 0 to 1, with a score of 1 meaning perfect precision and recall.

4. Results and Discussions

The study investigated the application of Support Vector Machine (SVM) in predicting political and non-political affiliations on Twitter. The study collected tweets related to political and non-political topics and preprocessed the data through steps such as transformation, negation handling, tokenization, filtering, and normalization. The preprocessed data was then used to extract meaningful features for training the SVM model. The SVM model was trained on the extracted features and was tested on a separate dataset of tweets to evaluate its performance in predicting political and non-political affiliations. The results showed that SVM can effectively classify tweets into political and non-political affiliations. This indicates that SVM has the potential to be a useful tool for analyzing the political leanings of individuals on social media platforms like Twitter.

5. Conclusion

In conclusion, the use of Support Vector Machine (SVM) in classifying political and non-political affiliations on Twitter has proven to be an effective approach. SVM is a supervised machine learning technique that uses a linear hyperplane to divide the corpus and categorize different datasets. By analyzing the tweets of Twitter users, the system is able to accurately classify their political and non-political views. The use of diagnostic metrics such as accuracy, memorization, and F-measurement further strengthens the validity of the results. This study highlights the potential of using SVM in social media sentiment analysis, particularly in understanding the political affiliations of citizens on social media platforms. The results of this research can provide valuable insights for organizations, governments, and individuals to better understand the impact of social media usage on political affiliations. Moreover, This research will be helpful to catagarize the social application Twitter information into +ive and –ive poltical views of users to asses the poltical affilation through imposing the supervised machine learning technique, namely Support Vector Machine to differentiate the performance, views & other results in elections.

The suggested work is sufficient to effectively evaluate the views of political affiliation on the data sets obtained from the social media application Twitter using the supervised machine learning algorithm. It will also be helpful for researcher community to find more high-level applications for selective assessment based on various social media applications such facebook, YouTube, Instagram, WhatsApp.

References

[1] Javed.M, Ziaudin, S.Kamal, Nasir.J.A, Raza.A.A, Habib.A (2020). Socio monitoring framework (SMF): Efficient sentiment analysis through informal and native terms. International Journal of advanced and applied sciences, 7(12), 113-126.

[2] Boutet, A., Kim, H., &Yoneki, E. (2012). What's in Your Tweets? I Know Who You Supported in the UK 2010 General Election. *ICWSM*, *12*, (pp. 411-414).

[3] Conover, M. D., Gonçalves, B., Ratkiewicz, J., Flammini, A., &Menczer, F. (2011, October). Predicting the political alignment of twitter users. In Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third Inernational Conference on Social Computing (SocialCom), 2011 IEEE Third International Conference on (pp. 192-199).

[4] Ellen, J., &Parameswaran, S. (2011). Machine Learning for Author Affiliation within Web Forums--Using Statistical Techniques on NLP Features for Online Group Identification. In Machine Learning and Applications and Workshops (ICMLA), 2011 10th International Conference on(Vol. 1, pp. 100-105).

[5] Kalmegh, S. (2015). Analysis of WEKA data mining algorithm REPTree, Simple CART and Random Tree for classification of Indian news. Int. J. Innov. Sci. Eng. Technol, 2(2), (pp. 438-446)

[6] Sharma, P., &Moh, T. S. (2016, December). Prediction of indian election using sentiment analysis on hindi twitter. In Big Data (Big Data), 2016 International Conference on (pp. 1966-1971).

[7] Khatua, A., Khatua, A., Ghosh, K., &Chaki, N. (2015). Can# twitter_trends predict election results? Evidence from 2014 indian general election. In System Sciences (HICSS), 2015 48th Hawaii International Conference on (pp. 1676-1685).

[8] Wicaksono, A. J. (2016). A proposed method for predicting US presidential election by analyzing sentiment in social media. In Science in Information Technology (ICSITech), 2016 2nd International Conference on (pp. 276-280).

[9] Sahu, R. K. (2019). Election result prediction for indian political parties using sentiment analysis of twitter data. Journal of the Gujarat Research Society, 21(14s), 369-374.

[10] Nayak A., Nataranjan S. (2016) Comparative study of Naïve Bayes, Support Vector Machine and Random Forest Classifiers in Sentiment Analysis of Twitter feeds. International Journal of Advanced Studies in Computer Science and Engineering (IJASCSE) Issue 1, Volume 5.

[11] Kim, S. M., &Hovy, E. (2007). Crystal: Analyzing predictive opinions on the web. In Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL).

[12] Mejova, Y., Srinivasan, P., & Boynton, B. (2013). Gop primary season on twitter: popular political sentiment in social media. In Proceedings of the sixth ACM international conference on Web search and data mining (pp. 517-526). ACM.

[13] Raza, A. A., Habib, A., Ashraf, J., & Javed, M. (2019). Semantic orientation based decision making framework for big data analysis of sporadic news events. Journal of Grid Computing, 17, 367-383.

[14] Macrohon, Julio Jerison E., Charlyn Nayve Villavicencio, X. Alphonse Inbaraj, and Jyh-Horng Jeng. "A Semi-Supervised Approach to Sentiment Analysis of Tweets during the 2022 Philippine Presidential Election." Information 13, no. 10 (2022): 484.

[15] Chauhan, P., Sharma, N., & Sikka, G. (2021). The emergence of social media data and sentiment analysis in election prediction. Journal of Ambient Intelligence and Humanized Computing, 12, 2601-2627.

[16] Javed, M., & Kamal, S. (2018). Normalization of unstructured and informal text in sentiment analysis. International Journal of Advanced Computer Science and Applications, 9(10).