

RESEARCH ARTICLE

Predicting Election Results from Twitter Using Machine Learning Algorithms

Yadala Sucharitha^{1,*}, Yellasiri Vijayalata^{*,2} and Valurouthu Kamakshi Prasad^{*,3}

¹Computer Science & Engineering, CMR Institute of Technology, Hyderabad, Telangana, India; ²Computer Science & Engineering, Gokaraju Rangaraju Institute of Engineering and Technology, Hyderabad, Telangana, India; ³Computer Science & Engineering, JNTU College of Engineering, Hyderabad, Telangana, India.

Abstract: Introduction: In the present scenario, the social media network plays a significant role in sharing information between individuals. This incorporates information about news and events that are presently occurring in the worldwide. Anticipating election results is presently turning into a fascinating research topic through social media. In this article, we proposed a strategy to anticipate election results by consolidating sub-event discovery and sentimental analysis in micro-blogs to break down as well as to imagine political inclinations uncovered by those social media users.

Methodology: This approach discovers and investigates sentimental data from micro-blogs to anticipate the popularity of contestants. In general, many organizations and media houses conduct a pre-poll review and obtain expert's perspectives to anticipate the result of the election, but for our model, we use Twitter data to predict the result of an election by gathering information from Twitter and evaluate it to anticipate the result of the election by analyzing the sentiment of twitter information about the contestants.

Results: The number of seats won by the first, second, and third party in AP Assembly Election 2019 has been determined by utilizing Positive Sentiment Scores (PSS's) of the parties. The actual results of the election and our predicted values of the proposed model are compared, and the outcomes are very close to actual results. We utilized machine learning-based sentimental analysis to discover user emotions in tweets, anticipate sentiment score, and then convert this sentiment score to parties' seat score. Comprehensive experiments are conducted to check out the performance of our model based on a Twitter dataset.

Conclusions: Our outcomes state that the proposed model can precisely forecast the election results with accuracy (94.2 %) over the given baselines. The experimental outcomes are very closer to actual election results and contrasted with conventional strategies utilized by various survey agencies for exit polls and approval of results demonstrated that social media data can foresee with better exactness.

Discussion: In the future, we might want to expand this work into different areas and nations of the reality where Twitter is picking up prevalence as a political battling tool, and where politicians and individuals are turning towards micro-blogs for political communication and data. We would likewise expand this research into various fields other than general elections and from politicians to state organizations.

ARTICLE HISTORY

Received: July 09, 2019
Revised: December 08, 2019
Accepted: December 08, 2019

DOI:
10.2174/2666255813999200729164142

Keywords: Election prediction, sentimental analysis, event detection, twitter.

1. INTRODUCTION

Social Media Networks (SMN) produces a huge volume of messages, various events, information and individual's opinions over a vast range of fields. The events discussed social blogs related to topics, areas, and time frames.

The events can be in an assortment, for example, famous personalities, political issues, local news, and natural calamities or accidents [1]. Messages are posted by individuals after they have encountered or seen the events occurring in reality, and they need to share their experiences instantly. Twitter is a web-based micro-blog that empowers users to post and read 280-character messages called "Tweets". Tweets consist of prosperous information about people group inclinations, where users generally debate with one another and reveal their assessments on Twitter. In addition,

*Address correspondence to this author at the CMR Institute of Technology, Hyderabad & Research Scholar of JNTUH, Hyderabad, Telangana, India; E-mail: suchi.yadala@gmail.com

tweets are little in length, and hence they are nearly unambiguous [2]. In such matters, data analysts particularly attempt to examine possibly the most famous social media network site twitter; likewise, it is much extensive when contrasted with Facebook, where social interactions are regularly private. It is a prominent data repository for sentiment analysis utilizing micro-blog since it can give almost quick access to ongoing social sentiments [3]. More significantly, geographic area and timestamps in the tweets possibly provide data to investigate the locations, time and discussion about the subject.

Sentimental Analysis in Twitter

Opinion mining, also called sentimental analysis, is the area of research that investigates individuals' feelings, sentiment, assessment, evaluation, mentality, and opinion towards elements, such as item, institution, service, event, subject, and their characteristics. Sentimental analysis has been dealt with as a Natural Language Processing task at numerous dimensions of granularity [4]. In the political field, it is utilized to monitor political views, to distinguish consistency and inconsistency among statements and activities at the administration level. Investigating the web user's sentiments has pulled in consideration of numerous researchers in academic and business segments. Sentiment mining has been enforced in an expansive set of areas, for example, political elections, market anticipation, and emergency management [5]. Twitter is a store for a huge amount of data. Therefore, this data is incredibly valuable for predicting outcomes of political exercises, and new activities driven by the government.

Micro-blogs play a huge role in Sentiment investigation. Twitter has turned into an immense research related gateway, which gives access to users to investigate the content from the tweets contrasted with other micro-blogs. Each tweet, remark and blog post may mirror their opinions [6]. This unstructured data give significant knowledge that establishes a vast space for making new administrations for governments, organizations or people. Manipulating this data gave rise to another field called Sentimental Analysis (SA). The primary goal of assessing the data is to choose whether or not to have faith in the information. Besides, individuals trust, numerous organizations and associations endeavour to catch up on what's going on in an individual's minds. Twitter gives progressively compact and significant information to express a feeling. The benefits of utilizing tweets as an information source are as follows; first, the quantity of tweets is extremely enormous and they are accessible to users. Second, tweets consist of the feelings of individuals, including their political views [7]. A standout amongst the most generally utilized social networking site, Twitter, is the place individuals share their thoughts, feelings, and opinions as tweets, and so it turns into one of the perfect sources for sentimental investigation. This data sharing empowered it to become one of the quickest and most prevalent data sources. Social media community responses generally differ in the face of information and misinformation. False information is commonly more questionable than valid data [8]. There could be a higher negative sentiment response against misinformation. Therefore, sentimental polarity identification in twitter could

be very useful in the scope of misinformation investigation and in numerous other validity examinations.

Sentimental analysis in Twitter begins by extracting tweets against hashtags to gather every single related data. The next step is to do pre-processing of tweets such as eliminating twitter handles (@user) and punctuation, numbers, symbols, special characters, short, and stemming words. The cleaned tweets at that point could be evaluated and visualized depending on a particular reason [9]. Commonly sentiment analysis will make or discover a bundle of words related to solidly positive, negative or neutral sentiment. Numerous positive words and a few negative words demonstrate positive sentiment, while many negative words and a few positive words show negative sentiment. Political canvassing has abused this huge range of data accessible on social media to draw experiences about user feelings and, in this way, structure their campaign technique. Immense expenditure by politicians on social media campaign right before an election alongside contentions and discussions between their supporters and rivals only upgrade the allegation that perspectives and sentiments posted by users have an aspect on election results [10]. The data could be utilized to anticipate the election result by utilizing a data examination strategy such as sentimental analysis. It is an analysis used to distinguish user likes, dislikes, remarks, feelings, or feedback about a substance that will be classified into positive, negative, or neutral responses. Earlier, researchers proposed opinion mining strategies that can be utilized over numerous applications like product and movie reviews, election result anticipation, and financial sector for stock prediction [11]. These days, each election campaign has an online existence and clients utilize these micro-blogs to express their feelings and sentiments towards political groups, contestants and significant events amid election.

Sentimental Analysis can be performed on tweets because of the below-stated reasons.

1. Tweets are short messages, abstract in format, and they have precise information.
2. Investigation in real-time can be made possible.
3. A huge collection of tweets can be taken to evaluate the performance.

Motivation

Twitter sentimental analysis has various advantages and a broad application zone. Not prohibitively but rather specifically, its excellent applications and advantages would be illustrated. Dealing with a brand or a political campaign may require monitoring an organization's popularity; the sentimental analysis gives an acceptable method to take the pulse of the tweeting people. Individuals are not typically proficient to survey naturally whether the data is true or fake. Investigation on the fakeness of the data would be so valuable for expanding the advantage from the micro-blogs data [12]. In such a manner, the sentiment analysis is exceptionally significant as a prior task of the fake investigation. With everything taken into consideration, twitter sentiment analysis has to create huge research scope in various applications.

Lately, it has been observed that conventional surveys may neglect to make precise forecasting. The scientific researchers have turned their enthusiasm for breaking down web-based data, such as blog posts or social user's act as an alternative method to anticipate election results, ideally more accurate. Moreover, conventional surveys are more expansive, while online data is free of cost and simple to get [13]. This is a fascinating research topic that joins politics and micro-blogs, both are concerned with the present society. Attempting to determine the precision and surprising expense issue, we go for using social data as the information source to foresee the result of an election.

Initial investigations regularly exhibited idealistic outcomes concerning the predictive limit of Twitter data and election outcomes. The most significant factor for better forecasting relies upon the data gathering strategy. The proposed research concentrated on getting applicable data from these micro-blogs and builds up a forecast model which helps in better comprehension of election results and significant patterns. We designed an approach for forecasting Andhra Pradesh (AP) Assembly Election 2019 results and sentimental analysis by utilizing the micro-blogs data. Indian micro-blog users have quickly emerged over the recent years to frame a total ecosystem which deals in certain regions, for example, news, legislative issues, politics, government strategies, and health. Existing works have concentrated on tallying of preferences or sentimental investigation on a contestant or party. They forget the way that the voters' preferences and feelings of individuals might be diverse, relying upon explicit political themes and in various geographic regions. Moreover, similar voters taking an interest in various debates may have distinctive political preferences. In this article, we are keen on anticipating the outcome of elections from Twitter data by consolidating sub-event recognition and sentimental analysis to identify their political inclinations and foresee the political results at a state just as a national level.

The primary contributions of this research work are as follows. (1) We present a way to predict the vote of a sample client dependent on the investigation of his/her Twitter messages and check the votes of clients to foresee the election outcomes. (2) Sub-event identification and Sentimental analysis are integrated to foresee the vote of clients as various levels of sub-events client occupied with the talks will influence the anticipated results. (3) Preparation and utilization of events and time explicit training dataset to prepare the classifier for better exactness. (4) Design of mapping capacities that map Twitter's sentiment score to the seat counts of top three parties, with great accuracy varieties. We assess our proposed methodology with real-world Twitter data posted by Andhra Pradesh, India based clients during the 2019 Assembly Election. The general outcomes are near to ground reality, which emphasizes our convictions. The rest of the article is organized as follows. The second section describes related work and the subsequent section explains the proposed methodology adopted. The fourth section outlines the experimental outcomes and discussions while, the conclusion is given in the last section.

2. RELATED WORK

Twitter is a social media network that has been pulling in growing consideration from researchers and at present, huge

research is going on micro-blogs in election result forecasting. In this section, we discuss some existing works proposed by different authors relevant to sub-event detection and sentimental analysis using predictive analytics. S Unankard, Xue Li, M Sharaf, J Zhong and Xue Li [14] proposed a model for election results forecasting, which includes sub-event identification and sentimental analysis in micro-blogs to investigate user preferences about different political parties and leaders. In this work, authors considered only positive and negative sentiments, but a neutral sentiment also affects the performance of the model. G Aceto, D Ciuonzo, A Montieri, and A Pescapé [15,16] proposed Deep Learning (DL) strategy to develop practical mobile traffic classifiers that depend on automatically mined features, capable to manage with encrypted traffic, and dazzling their difficult traffic trends. In the end, various DL models and traffic classifiers were reproduced and performance is compared with benchmarks. Ritesh S, H Kumar, M.P.S Bhatia and Shruti J [17] designed a model for Delhi Assembly election 2015 results forecasting based on twitter data and they performed an investigation of textual documents of twitter data. The proposed methods that were used to predict the number of seats of the top three parties that will win are by using mapping functions. The model outcomes are very close to actual results. In this study, the authors worked on limited data due to lack of social media users, but now users of social media have increased and the accuracy of the model is raised due to a large number of tweets. Yadala Sucharitha, Y Vijayalata and V Kamakshi Prasad [18] proposed early discovery of rising hotspot events in micro-blogging sites with area sensitivity. In this framework, we identify considerable correlations among client areas and event areas in discovering the rising-events; we evaluated our model in view of Twitter API data. Our investigational results illustrate that the recommended method can accurately identify rising events with respect to the client locations that have various granularities. Vinay K. Jain and Shishir Kumar [19] designed an election results prediction framework utilizing micro-blogs based on keyword search methods. They described a novel data pre-processing technique followed by analyzing the popularity of contestants of various parties. The results of the proposed method are compared with survey agencies' outcomes and results show that micro-blogs data produced accurate results. In this study, the authors consider only the English language, but multi-language data pre-processing can improve the results. However, they do not explain any forecasting model.

Lei Wang and John Q Gan [20] described a model for forecasting French Presidential election 2017 results dependent on social media. First, they identified keywords or search terms related to main contestants and gathered all tweets related to search terms and performed sentiment analysis to determine popularity. In this investigation the authors considered only candidate popularity as a constraint and did not look in to other factors like the party, contestants, star campaigners etc., and it is limited to two contestant's election only. Nevertheless, the authors required to interpret political messages by hand to calculate slant weight and just the primary message of each client is considered. Moreover, the strategy depends on surveying data to address for forecasting election results from social media.

A Hasan, S Moin , A Karimand Shahaboddin S [21] designed a hybrid prediction model that incorporates the machine learning strategies with subjectivity and sentimental analysis for opinion text mining. A combination of SVM and naïve byes algorithm is utilized to investigate political views. In this work authors proposed hash-tags based model, if they considered other search keywords, model accuracy might have improved. Moreover, it does not foresee the election results. To solve the above-mentioned problems, we proposed an ML-based Algorithm called Support Vector Machine classification with Sentimental analysis for the prediction of AP election 2019 results that are discussed in Section 3.

3. PROPOSED METHODOLOGY

To understand whether the action on Twitter can provide as a forecaster of the political decision results, we propose a way to deal with sentimental analysis for each sub-occasion for anticipating client's political inclination. Data extracted from tweets is input for the proposed model, and we computerize the procedure of tweet extraction and classify it into two classes i.e., for positive or negative. This examination gives a few novel discoveries that have direct implications in the design of micro-blogs and information analysis. The main objective of this research is to investigate the AP Assembly Election 2019 results utilizing Twitter's data. Fig. (2) illustrates the flow diagram of the proposed methodology and it consists of five steps. Our methodology depends on the data accumulated from the online networking site (i.e., Twitter) and performs data examination to understand the idea of discussions and individual behaviour on social media.

The proposed system includes five main phases which are:

- 1) Extracting tweets from Twitter API based on hash-tags and search keywords
- 2) Data pre-processing
- 3) Polarity Classification and performing Sentiment analysis
- 4) Determining the sentiment score of each political party
- 5) Converting sentiment score to seat share to predict the election results.

We used R-Tool to collect Twitter data utilizing Twitter API, and it also utilized for information cleaning and extraction of each tweet's related metadata.

Data Collection and Pre-processing

In this section, we come up with a unique technique for data gathering, such as the significant variable for data accumulation from micro-blogging sites that are the search keys that help in getting applicable tweets. During the election period (1 to 3 months) lots of events/topics relevant to the election are discussed on social media, which may be significant for anticipating the results. Taking this hypothesis that utilizes more keywords associated with the election can gather progressively significant information and thus, improve the forecast accuracy. Proposed data gathering proce-

dure displayed in Fig. (1), is unique from the other existing methods because we considered trending and dynamic keywords for this. These are collected by utilizing three types of data sources such as investigation top newspapers, RSS channels and discover top topics, filter out most happening topics and key terms utilized in data gathering [22, 23]. This process gives dynamic keywords that are trending amid the election and relevant to public opinions. Also, the utilization of top trending hash-tag taken from Twitter and thirdly, gathered from tweets. Ongoing investigations have demonstrated that with Twitter [24, 25], it is conceivable to get individuals' insight from their profiles rather than conventional methods for obtaining data about observations. Moreover, researchers proposed a model for exploiting the feelings from tweets while thinking about a huge size of data for SA. To distinguish social networks with compelling effect, a novel technique was proposed and executed by appointing metric incentive to every one of the client's enthusiastic posts. Next, the contribution of this article incorporates the investigation of political election preferences assembled from Twitter profiles, with different opinion analyzers. Also, this article introduces the validation of results obtained from every analyzer with machine learning classifiers. Our assessment depends on the comparison of various opinion analyzers and approves the outcomes with the various classifiers. The investigation on Twitter data will demonstrate which procedure has a superior ability to estimate sentiment anticipation accuracy.

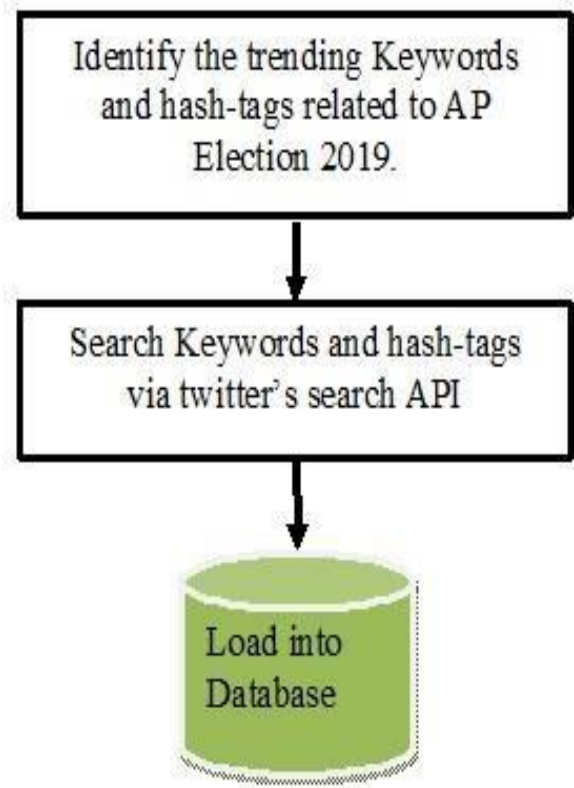


Fig. (1). Data collection process using twitter API.

On the ground of the result of the past AP Assembly Election 2014, we confined our data collection to only three major contesting parties, namely: TDP, YSRCP, and JSP. An inquiry terms list is manually crafted for each party that in-

Table 1. List of search terms and hash-tags related to the main three parties of AP assembly election 2019

S. No.	Party Name	Search keywords and hash-tags
1	TDP	TDP, JaiNTR, ChandraBabu, Lokesh ,BalaKrishna , DivyaVani,...
2	YSRCP	YSRCP, ysjagan,YSR, Bharathi, Rojaselvamani, VijayasaiReddy, ByeByebabu, yssharmila...
3	JSP	pk, pavankalyan, powerstar, janasena, nagababu, jdlaxminarayana, mega...

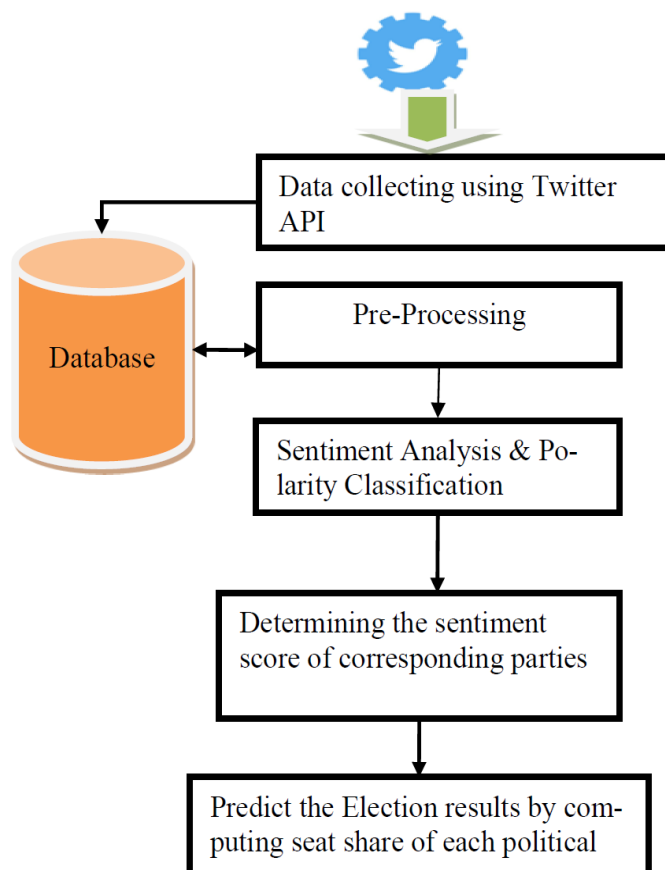


Fig. (2). Framework for a forecast of AP assembly election 2019 utilizing Twitter.

cludes the name of the parties, names of leaders, names of star campaigner of each party, election-related hash-tags, campaign relevant hash-tags and the derivative terminologies belonging to each party (e.g., “YSR” for Jagan Mohan Reddy) are listed in Table 1.

After selecting keywords and hash-tags, the data gathering procedure begins by fetching tweets utilizing Twitter API. It offers two APIs: Streaming API and REST API for fetching tweets. The data gathering process begins from the declaration of election schedule by the Election Commission of India [26, 27]. The tweets related to the general election to Legislative Assemble 2019 in AP state were collected using Twitter’s search API. The data was gathered for a period of 31 days (from 10th March 2019 to 11th April 2019), by expecting this period as a peak period for election pertinent activities like an election campaign. The gathered dataset contains 1012411 tweets, of which 62589 are re-tweets or duplicates, leaving a sum of 949822 tweets such as party- wise displayed in Fig. (3).

To collect individuals sentiment dependent on the gathered hash-tags and search keywords relevant to be identified with perspectives about political parties, including twitter top patterns, we utilized Twitter API. We have registered on twitter account to extract the tweets, which linked to Twitter API and its authentication procedure is completed utilizing the OAuth package of R language. Extracted tweets, from Twitter accounts, were stored in the database under the accompanying fields:

Twitter_id, Hashtag, Tweet_created, User_id, Tweet_text, Retweet_count, Follower_count, Favourite_count of each tweet.

We select hash-tags that were inclining on Twitter, speaking to the political perspectives on individuals. Furthermore, we scanned for some more hash-tags of political behaviour. Each word in a tweet is significant in decision management, so pre-processing of these tweets is a significant assignment on the grounds that these messages are

brimming with slang, incorrect spellings, and words from other languages [28]. So as to handle the issues with the noise in writings, standardization of tweets is performed by applying text pre-processing steps like tokenization, numbers and stop words elimination, stemming, lower case transformation lemmatization, removing white spaces and dimensionality reduction.

For data pre-processing, we perform the following steps:

- 1) We commonly eliminate all redundant tweets, if retweeted by the same user ID. But not of that which has another user ID.
- 2) All non-English and multilingual tweets were discarded from the dataset.

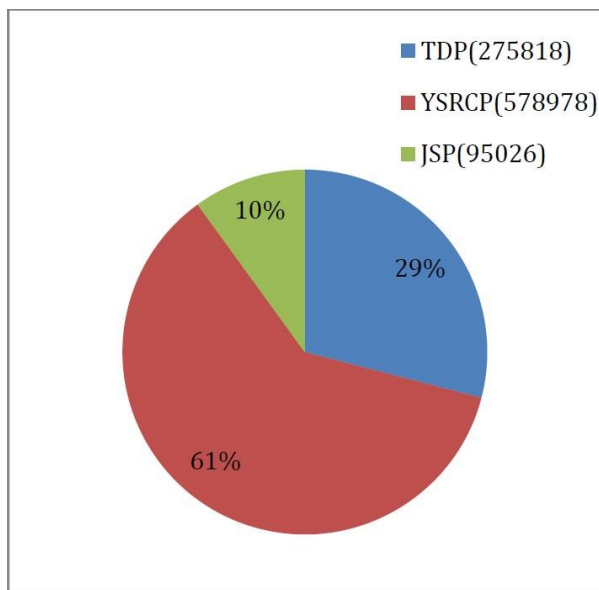


Fig. (3). Distribution of tweets among 3-parties

Polarity Classification and Sentiment Analysis

Sentiment analysis can be a supervised or an unsupervised methodology or a blend of both. In the supervised methodology, the way of labeling training datasets requires extensive time and effort and gathering datasets for all application areas is very tedious and troublesome. Sentiment analysis can give significant bits of knowledge from microblogs by recognizing feelings or sentiments from a huge volume of data present in the unstructured pattern. We first gather Twitter data documents, then the procedure of sentiment investigation is to compute the synchronization of the keywords of the tweets regarding positive, neutral and negative word lists. Sentiment analysis incorporates three polarity classes, which are positive, neutral and negative. Each tweet polarity is calculated by assigning a rank -1 to 1 depending on the words utilized, where a negative rank implies a negative sentiment and a positive rank implies a positive sentiment while the zero esteem is viewed as a neutral sentiment. Each tweet subjectivity rank assigned depends on whether it is speaking to subjective or objective meaning. The scope of subjectivity rank is also from 0 to 1, where a rank close to 0 speaks to objective and close to 1 subjective. We used a fundamental approach to infer votes from tweets

that are exhibited in numerous kinds of literature, that are shown in the below hypothesis.

Hypothesis H: H is the proposal to the real vote share of party X

Where

$$H \rightarrow \left(\frac{\text{positive tweets belongs to the party X}}{\text{Total positive tweets all parties}} \right) \text{ (Vote share of party X)}$$

In this paper, the sentiment analysis approach contains two stages: (i) subjectivity test, i.e., the discovery of those tweets that are subjective and possess some sentiment about any party irrespective of their polarity. For this, we utilized the opinion word list from SentiWordNet and a list of positive and negative emotions. The Pie diagram of Fig. (4) demonstrates the division of total tweets into subjective and objective and objective tweets is hence eliminated from the data as they do not have any opinion about any party. Next, we isolated tweets by associating each tweet on the basis of the search keyword list, as mentioned above in three sets (i.e., TDP, YSRCP, and JSP). Fig. 5 portrays this division of subjective tweets into three sets.

As the second (ii) step, we applied sentiment analysis to decide the polarity of each tweet corresponding to a party by associating the keywords list with each tweet. We performed sentiment analysis of tweets using Support Vector Machine (SVM), and we applied it for training set to the design of the model, and the test set was used to assess the model.

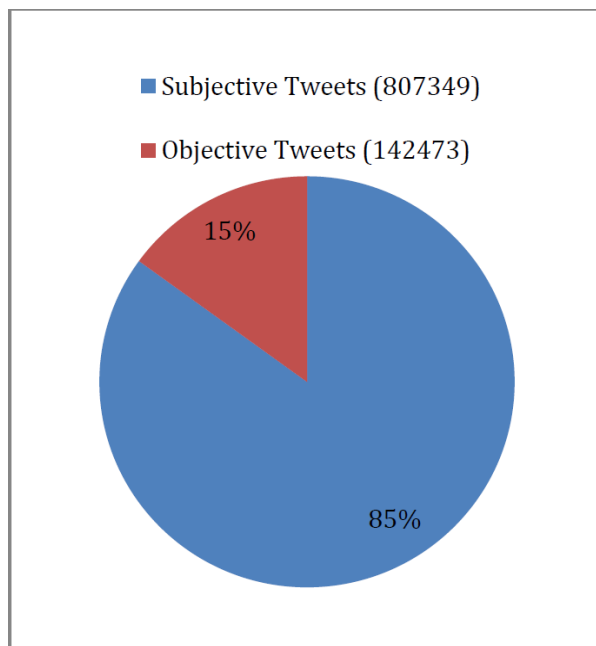


Fig. (4). Division of Subjective and Objective tweets.

Tweets sentiment score or polarity of the tweet is may be positive or negative or neutral and it computed as follows

The Polarity of tweet or Sentiment Score = No. of positive words – No. of negative words

If Score > 0, this implies the sentence has a generally speaking 'positive sentiment'

If Score < 0, this implies the sentence has a by and large 'negative sentiment'

If Score = 0, at that point the sentence is viewed as a 'neutral sentiment'

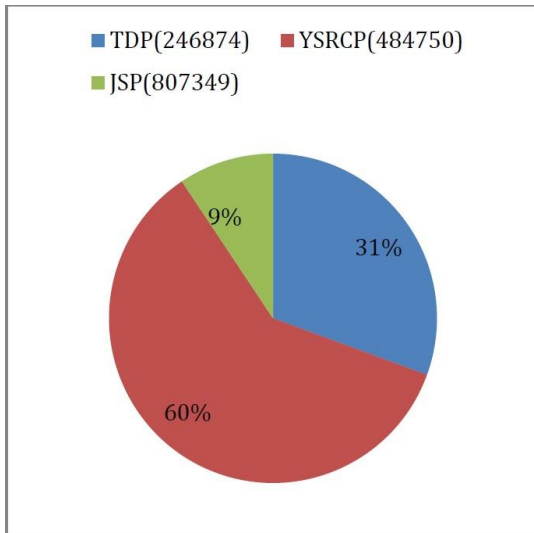


Fig. (5). Division of subjective tweets among 3- parties

The sentiment score for both positive and negative of all the three parties are described with the help of bar graphs in Fig. (6) shown below, and the Positive Sentiment score (PSS) of a party X is denoted as PSS(X) and can be computed using Equation 1. By utilizing Equation (1), the estimation of PSS (YSRCP), PSS (TDP), and PSS (JSP) is equivalent to 0.7865, 0.1805 and 0.0328 for YSRCP, TDP and JSP individually, is presented in Fig. 7.

$$PSS(X) = \frac{\text{Total number of positive tweets belongs to the party X}}{\text{sum}(\text{Total positive tweets of three parties})} \quad \text{Eq. (1)}$$

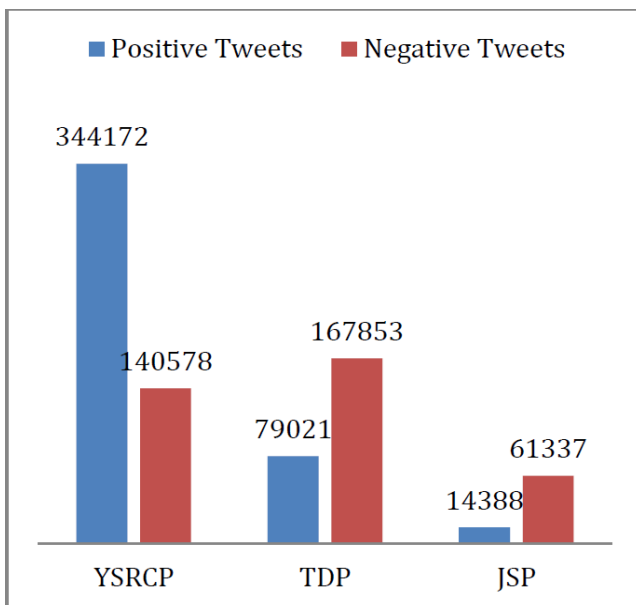


Fig. (6). Sentiment score of 3-parties

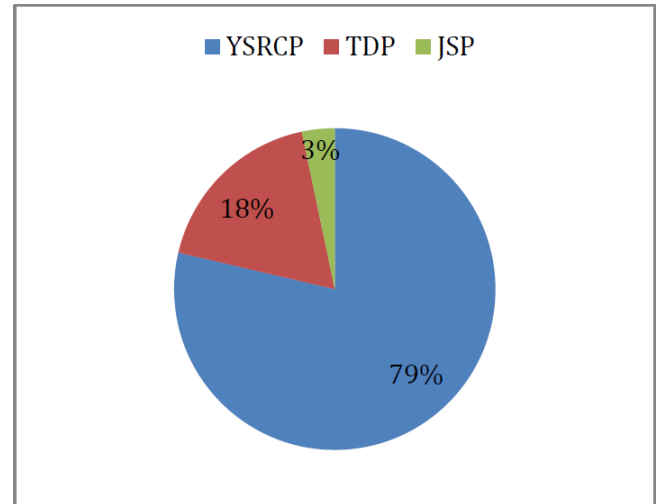


Fig. (7). PSS (%) of 3-parties.

Converting Sentiment Score to Seat Score

Most investigations that are done for anticipating elections result utilizing Twitter's data are commonly depicted in the election situation where the election is bipolar: for example, elections were challenged among two parties, which are having two noted leaders. Because of which it is very simple to decide the winning of a party basically by favouring the higher emotion score. To the extent Indian elections are considered, the situation that decides the winning of any party is very convoluted as there are five national parties with many enlisted and perceived state parties.

Moreover, the winning of a party cannot be solely represented by vote share, despite the fact that it is extremely fundamental to convert the vote score to the seat score. Be that as it may, in any multiparty election framework like India and states of India, psychology is exceptionally challenging as well as hazardous too. In a way to change over the opinion score to seat score, we structured a model that decides the seat score utilizing sentiment score. This model is absolutely founded on the statistical information of the last state elections (AP General Elections 2014) in different conditions of India. The model includes three disparate mapping functions that can anticipate the seat score of the best three parties.

Each function must have been determined all together, i.e., from the first-party function to third party function. The first mapping function is utilized to decide the seat score of the first party. It is basically obtained by operating curve-fitting on vote score versus seat score information of those parties that got the primary position in the last elections. We have performed moving average smoothening pursued by polynomial curve fitting between vote score and seat score. This curve is henceforth to be utilized for acquiring the level of seat score of the first party; if PSS of the party is known (i.e., H is true). The curve inferred here, holds with coefficients C1 and C2 and, S is the level of seat offer of the first party.

$$\text{Number of seats of party (YSRCP)} = \frac{S \times \text{Total no. of seats}}{100} = A$$

Table 2. Forecasting number of seats using the proposed method.

S. No.	Party Name	Vote Share (%)	PSS (%)	Actual Seats	Predicted Seats
1	YSRCP	79.21	78.65	150	150.48
2	TDP	16.58	18.05	24	22.74
3	JSP	2.12	3.28	1	1.79

Table 3. Confusion Matrix.

Actual/predicted	Positive	Negative	Accuracy
Positive	325(TP)	15(FN)	
Negative	14(FP)	146(TN)	
Total	339	161	94.2

*TP= True positives, FN= False negatives, FP = False positives, TN= True negatives.

Where $S = \frac{\text{Actual} - \text{Predicted}}{\text{Actual}}$ Eq. 2(a)

Let us assume that, the numbers of seats assigned to second and third are depicted by B and C respectively, and furthermore, let Y, represents all the number of seats in a state, at that point B and C can be acquired by utilizing mapping functions (3) and (4) separately, where (3) relies on (2),

while (4) is reliant on both (2) and (3). In all mapping functions, we have attempted to be the most secure fit however, much as could be expected.

Number of seats of party (TDP) = $\frac{(100 - \text{YSRCP}) \times \text{PSS}(\text{TDP})}{100} = B$ Eq. (3)

Number of seats of party (JSP) = $\{[Y - (A + B)] \times [\text{PSS}(\text{YSRCP}) + \text{PSS}(\text{TDP})]\} = C$ Eq. (4)

4. EXPERIMENTAL RESULTS AND DISCUSSIONS

This segment gives an analysis of the outcomes we get while doing this work. This whole area is separated into two subsections: (i) Predicting number of seats utilizing our model, (ii) Accuracy of assessment investigation and its effect to seat score.

The number of seats won by the first, second and third parties in AP Assembly Election 2019 has been determined by utilizing PSS’s of these parties by means of Equation (2), (3), and (4), respectively. In Table 2, the actual results of the election and our model prediction results are shown, and these outcomes are very close to the actual results. We utilized SVM with 15-fold cross-validation, for sentiment polarity classification utilizing our training set, which gives us the precision of 94.2%. There are 7500 tuples in our training dataset, with 3750 positive tweets and 3750 negative tweets. With 15-fold cross-validation, we get 500 tuples in our test dataset. Table 3 demonstrates the confusion matrix to compute the accuracy, the precision of the prediction model.

Based on empirical results (test data), we selected 15 samples for constructing the ROC curve shown in Fig. (8) for election prediction values. For that, TPR (True Positive Rate) and FPR (False Positive Rate) values calculated by following equations.

$TPR \text{ (Sensitivity)} = \frac{TP}{TP + FN}$ Eq. (6)

$TNR \text{ (Specificity)} = \frac{FP}{FP + TN}$ Eq. (7)

$FPR = 1 - TNR$ Eq. (8)

Election prediction is one of the most challenging tasks in India because each state so many local parties and several national parties are going for contesting. We consider three parties for our research based on previous election results (2014), and Table 4 displays exit poll results (number of seats won) conducted by various agencies. In Fig. (9), we compared our model outcomes with various survey agencies results, and we state that our model predicted results are very close to actual results and far better from exit polls results done agencies.

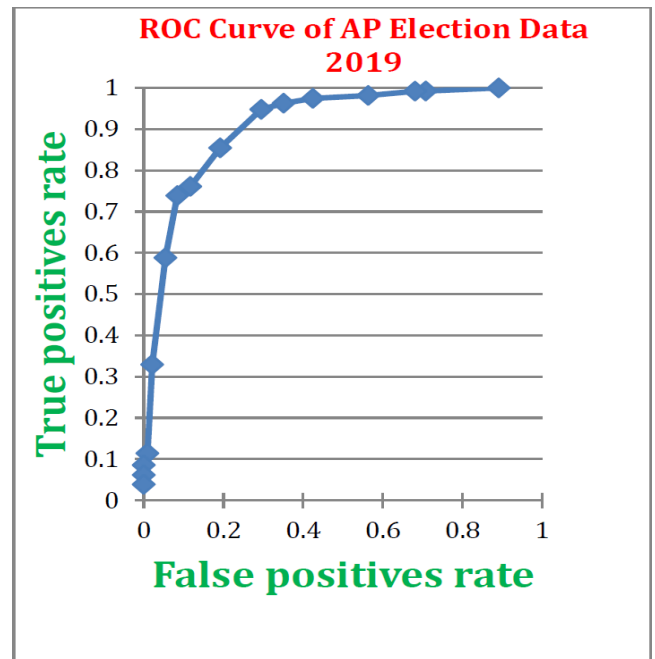


Fig. (8). ROC curve of SVM classifier for test data

4.1. Comparative Analysis

In this research, we proposed an AP Assembly Election 2019 results based on a machine learning approach called sentimental analysis using Twitter data. The proposed method was conducted on a vast scale, and the performance of the proposed model was assessed in terms of accuracy and the investigational results display that the proposed method achieved 94.2% accuracy in the prediction of election results. It is compared with other existing methods and the performance is displayed in Table 5. Non-Linear Sentimental Analysis of Twitter Data [15] is used to predict the Delhi Election 2015 results utilizing Textual Content of Social Network and the model is evaluated with an accuracy of 93.7 %. Novel intelligent prediction technique [16] proposed

based on Machine learning techniques for election results prediction was used and the authors conclude that SVM produces better accuracy (79.4 %) when compared with Naïve Bayes (69.25%), Random Forest (77.25 %) and Decision Tree (70.80 %). The authors proposed an election results prediction model called Naïve Bayes (NB) [18] with different Sentiment analyzers and they evaluated the performance of the model in terms of accuracy in three levels based on various sentiment analyzers namely TextBlob (76%), SentiWordNet (54.75%) and W-WSD (79 %) and finally stated that NB with W-WSD sentiment analyzer produces good results when compared to others. Our proposed method (SVM based Sentiment Analysis) produces better accuracy when compared to the existing models.

Table 4. Comparison of the proposed method results with various survey agencies results

S. NO.	Survey Agency Name	Name of the Political Party (no. of seats won)		
		YSRCP	TDP	JSP
1	India Today	133-137	38	2
2	CPS	131	37	1
3	VDP Associates	115	57	2
4	RG Flash	65	110	0
5	INSS Media	52	118	5
6	Peopls Pulse	112	59	4
7	TV5	68	105	2
8	I -News	100	49	0
9	L Raj Gopal	72	100	5
10	ARA Survey	130	51	2
11	Actual Results	150	24	1
12	Proposed Method results	150	23	2

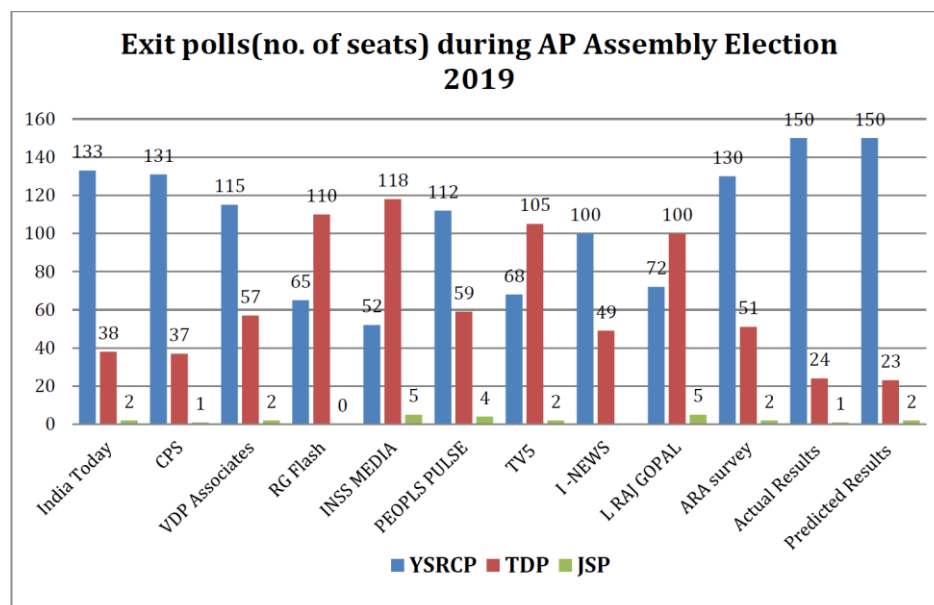


Fig. (9). Comparison of Exit polls (no. of seats) results of different agencies with the proposed method.

Table 5. Comparison of works with existing models.

S. No.	Name of the Methodology	Accuracy (%)	
1	Sentimental Analysis of Twitter Data [15]	93.7	
2	Novel intelligent prediction technique for election outcome [16]	SVM	79.4
		Naïve Bayes	69.25
		Random Forest	77.25
		Decision Tree	70.80
3	sentiment analyzers with Naïve Bayes [18]	TextBlob	76
		SentiWordNet	54.75
		W-WSD	79
4	Proposed Method	94.2	

CONCLUSION

In this research, we examined an issue of anticipating AP assembly election 2019 results based on freely accessible data on micro-blogs, such as Twitter. A precise technique for forecasting election results has been proposed. An approach to identify and gather the twitter data, pre-process and perform sentimental analysis over social media networks in order to forecast an individual's opinion is also described. In any case, finding the seat score by utilizing vote share is a standout amongst the most prevalent and least complex methodologies if the vote share of parties is known. We pursue a similar methodology for creating our mapping model for translating the sentiment score to seat score, while creating our forecasting model for foreseeing seat score with vote share from the date of the last state assembly elections in AP. Extensive examinations are directed to have assessed the performance of our methodology on the Twitter dataset. For this investigation, we have examined 949822 Twitter messages related to AP assembly elections 2019 based on names of parties, leaders and contestants. These tweets were gathered over a time of 31 days (from 10th March 2019 to 11th April 2019). The tweets were classified depending on their messages, referencing either the names of the CM candidates, names of the parties or star campaigners of the individual parties. The proposed methodology is successful in foreseeing election results against the exit polls done by different survey agencies and comes very close to the actual results with 94.2 % accuracy when compared with other existing election prediction models.

CURRENT & FUTURE DEVELOPMENTS

In the future, we might want to expand this work into different areas and nations of the reality where Twitter is picking up prevalence as a political battling tool and where politicians and individuals are turning towards micro-blogs for political communication and data. We would likewise expand this research into various fields other than general elections and from politicians to state organizations. Multi-language data pre-processing can improve forecast exactness.

FUNDING

Not Applicable

ACKNOWLEDGEMENT

The authors would like thank to Mr. P Chandra Shaker Reddy, Assistant Professor of CSE Dept. & Rosy Matilda P, and Professor of H&S in CMRCET, Hyderabad, TS, and INDIA for his great support for this research work.

CONFLICT OF INTEREST

No Conflict of interest

CONSENT FOR PUBLICATION

The authors declare no consent.

REFERENCES

- [1] S. Katragadda, R. Bentona, and V. Raghavan, "Sub-Event Detection from Tweets", *International Conference on Neural Networks*, 2017 Anchorage Available From: 10.1109/ijcnn.2017.7966112
- [2] L. Zhao, "A Dynamic Pricing Mechanism Model Based on Sentiments Analysis", *International Conference on Intelligent Transportation, Big Data & Smart City*, 2019 <http://dx.doi.org/10.1109/ICITBS.2019.00155>
- [3] M. Rathi, M. Aditya, D. Varshney, S. Rachita, and S.Mendiratta, "Sentiment Analysis of Tweets using Machine Learning Approach", *International Conference on Contemporary Computing*, 2018 <http://dx.doi.org/10.1109/IC3.2018.8530517>
- [4] U. Yaqub, S.A. Chun, A. Vijayalakshmi, and V. Jaideep, "Sentiment based Analysis of Tweets during the US Presidential Elections", *International Conference on Digital Government Research*, 2017 <http://dx.doi.org/10.1145/3085228.3085285>
- [5] W Budiharto, and M Meiliana, "Prediction and analysis of Indonesia Presidential election from Twitter using sentiment analysis", *Journal of Big Data*, 2018. Available From: 10.1186/s40537-018-0164-1
- [6] L. Oikonomou, and C. Tjortjis, "Method for Predicting the Winner of the USA Presidential Elections using Data extracted from Twitter", *International Conference on South-Eastern European Design Automation, Computer Engineering, Computer Networks and Society Media*, 2018 <http://dx.doi.org/10.23919/SEEDA-CECNSM.2018.8544919>
- [7] S Pritee, *International Research Journal of Engineering and Technology*, vol. 4, no. 10, pp. 539-544, 2017.

- [8] K. Jaspreet, and B.K. Sidhu, "Sentiment Analysis Based on Deep Learning Approaches", *International Conference on Intelligent Computing and Control Systems*, 2018 Available From 10.1109/iccons.2018.8662899
- [9] K. Hamdullah, A.D. İbrahim, D. Murat, U. Anıl, and O. Yıldız, "Sentiment Analysis on Turkish Social Media Shares through Lexicon Based Approach", *International Conference on Computer Science and Engineering*, 2018 Sarajevo Available From 10.1109/UBMK.2018.8566481
- [10] M Firat, "'Twitter Sentiment Analysis, 3-Way Classification: Positive, Negative or Neutral?'"', *International Conf. on Big Data*, 2018. Available From 10.1109/BigData.2018.8621970
- [11] M J Adarsh, "'An Effective Method of Predicting the Polarity of Airline Tweets using sentimental Analysis'"', *International Conference on Electrical Energy Systems*, 20182018. Available From 10.1109/ICEES.2018.8443195
- [12] N. Hlongwane, Y.P. Huang, and L.J. Kao, "Using Sentiment Analysis to Determine Users' Likes on Twitter", *International Conference on Dependable, Autonomic & Secure Comp*, 2018 Available From 10.1109/dasc/picom/datacom/cyberscitech.2018.00177
- [13] B Alexandre, M Flaviano, and A M Hernán, "Validation of Twitter opinion trends with national polling aggregates: Hillary Clinton vs Donald Trump", *Journal of Scientific Reports*, 2018. Available From: 10.1038/s41598-018-26951-y
- [14] S. Unankard, X. Li, M. Sharaf, J. Zhong, and X. Li, "Predicting Elections from Social Networks Based on Sub-event Detection and Sentiment Analysis", *International Conference on Web Information Systems And Engineering*, 2014 http://dx.doi.org/10.1007/978-3-319-11746-1_1
- [15] Giuseppe Aceto, DomenicoCiuonzo, Antonio Montieri, and Antonio Pescapé, "Mobile Encrypted Traffic Classification Using Deep Learning: Experimental Evaluation, Lessons Learned, and Challenges", *IEEE Transactions on Network and Service Management*, no. Press, 2019. Available From: 10.1109/TNSM.2019.2899085
- [16] G. Giuseppe Aceto, "DomenicoCiuonzo, Antonio Montieri, and Antonio Pescapé, "Mobile Encrypted Traffic Classification Using Deep Learning", *Network Traffic Measurement and Analysis Conference (TMA)*, 2018 <http://dx.doi.org/10.23919/TMA.2018.8506558>
- [17] S. Ritesh, and M.P.S. Horesh Kumar, "Bhatia and J Shruti, "Analyzing Delhi Assembly Election 2015 Using Textual Content of Social Network", *International Conference on Computer and Communication Technology*, 2015 Available From: 10.1145/2818567.2818582
- [18] Y. Sucharitha, Y. Vijayalatha, and V. Kamakshi Prasad, "Emergent Events Identification in Micro-Blogging Networks Using Location Sensitivity", *Jour of Adv Research in Dynamical & Control Systems*, vol. 11, no. 08, pp. 596-607, 2019.
- [19] K J Vinay, and Kumar Shishir, "'Towards Prediction of Election Outcomes Using Social Media'"', *I.J. Intelligent Systems and Applications*, 2017. Available From: 10.5815/ijisa.2017.12.03
- [20] L. Wang, and J.Q. Gan, "Prediction of the 2017 French Election Based on Twitter Data Analysis", *International Conference on Computer Science and Electronic Engineering*, 2017 <http://dx.doi.org/10.1109/CEEC.2017.8101605>
- [21] A Hasan, M Sana, A Karim, and S Shahaboddin, "'Machine Learning-Based Sentiment Analysis for Twitter Accounts'"', *International Journal of Mathematical and Computational Applications*, 2018. Available From: 10.3390/mca23010011
- [22] M.W. Yaser, "M Hager Saleh and E Halim Houssein, "Twitter Sentiment Analysis using Deep Neural Network", *International Conference on Computer Engineering Conference*, 2018 International Journal of Mathematical and Computational Applications, 10.1109/ICENCO.2018.8636119
- [23] R. Qiufeng, Y. Zheng, G. Guo, and H. Yating, "Resource Recommendation Algorithm Based on Text Semantics and Sentiment Analysis", *International Conference on Robotic Computing*, 2019 International Journal of Mathematical and Computational Applications, 10.1109/IRC.2019.00065
- [24] Pundru Chandra Shaker Reddy and A Sureshbau, "An Adaptive Model for Forecasting Seasonal Rainfall Using Predictive Analytics", *International Journal of Intelligent Engineering and Systems*, 2019. Available from: 10.22266/ijies2019.1031.03
- [25] G. Aceto, D. Ciuonzo, A. Montieri, and A. Pescapé, *Multi-Classification Approaches for Classifying Mobile App Traffic*, 2018.<https://doi.org/10.1016/j.jnca.2017> <http://dx.doi.org/10.1016/j.jnca.2017.11.007>
- [26] P. Healy, G. Hunt, K. Steven, and J.P. Theo Lynn, "Morrison and V Shankar, "Evaluation of peak detection algorithms for social media event detection", *International Conference on Semantic and Social Media Adaptation and Personalization*, 2015 Available from: 10.1109/SMAP.2015.7370 090
- [27] Y. Sucharitha, Y. Vijayalatha, and V. Kamakshi Prasad, "Analysis of Early Detection of Emerging Patterns from Social Media Networks: A Data Mining Techniques Perspective", *International Conference on Soft Computing and Signal Processing*, 2019 http://dx.doi.org/10.1007/978-981-13-3600-3_2
- [28] Giuseppe Aceto, DomenicoCiuonzo, Antonio Montieri, and Antonio Pescapé, "Anonymity Services Tor, I2P, JonDonym:Classifying in the Dark (Web)", *IEEE Transactions on Dependable and Secure Computing*, no. Press, 2018. Available from: 10.1109/TDSC.2018.2804394