IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-6, Issue-7, July 2018 E-ISSN: 2347-2693

Characterizing Human Opinion in Social Network Using Machine Learning Algorithms

Lavanya V S[1*], Savita K Shetty[2]

[1]Dept. of ISE, Ramaiah Institute of Technology, Bangalore-54, Karnataka, India
[2] Dept. of ISE, Ramaiah Institute of Technology, Bangalore-54, Karnataka, India
[*]*Corresponding Author: lavanya.vs09@gmail.com, Tel.: +91-9164056336*

Available online at: www.ijcseonline.org

Accepted: 23/Jul/2018, Published: 31/July/2018

***Abstract*—** Social media emergence has gained significant impact on how people communicate and socialize. Twitter provides the social media platform from where opinions of the people can be heard. Sentimental analysis can be applied to obtain the useful information by analyzing these tweets carefully. To characterize the human opinion, this paper studies users perception regarding a controversial product, namely self-driven cars. To find people's opinion regarding this new technology, self-driven car Twitter dataset is used. Based on the people's reaction about the self driven car in the social media(Twitter), human opinions are characterized like whether the people gave positive statement, negative or neutral statement regarding the self-driven car tweets. To classify the tweets, different machine learning algorithms, such as Logistic regression, Support Vector Machine, Random forest classifier and AdaBoost classifier are used. By using these tweets, opinions are characterized as "positive", "negative" and "neutral". To evaluate the performance of four algorithms, comparisons is carried out over the metrics like accuracy, recall, precision and f1-score. From the experimental results Logistic regression outperforms Support Vector Machine, Random forest classifier and AdaBoost classifier algorithms.

***Keywords*—** Random forest, Support Vector Machine, Logistic Regression , AdaBoost classifier, Sentiment analysis

I. INTRODUCTION

In today's world information plays a very important role in every field. These information can be generated the social media such as Twitter, Facebook etc. Social media has got vast scope in between the people [1]. The way of peoples thinking and deliberating their expressions has rapidly changed. To convey the opinion-based information, people used the platform which is convenient from their viewpoints and expressions [2]. Billions and millions of people and many business organizations are using Twitter and other social media applications for posting their opinions or views. Twitter, Facebook and some other social media applications are inherently share the opinion-based information among the people. These shared opinion data can be made useful for analysing the different types of opinion. These opinions can be positive, negative or neutral [3].

In a recent day, some of the new technologies are come under in the field of Artificial intelligence. In that one of the new technologies is self-driven cars. Some of the companies like Google, Tesla are contest with each other to attract the people about their products [4]. There are lot of improvements in this new technology and first the company should understand the benefits and consequences of the self driven cars. Some researchers have studied these new technologies and they understood that, it may be helpful in the society. Some people can give their opinions about the self-driven car in the form of tweets. These tweets can be classified in different forms by using machine learning techniques. After classification of these tweets which can be used to analyze the human opinion on self-driven tweets by using sentiment analysis.

To classify the tweets, we are using different machine learning algorithms, such as Logistic regression, Support Vector Machine, Random forest classifier and AdaBoost classifier [5]. The proposed method is evaluated across different metrics. For better performance, we are comparing the accuracy with all these algorithms. For comparative analysis, three measures are used, and it outperforms in terms of recall, precision and f1-score [3]. From the experimental results it can be shown that Logistic regression outperforms Support Vector Machine, Random forest classifier and AdaBoost classifier algorithms.

For a given tweet, we will able to determine the general sentiment based on the sentiment known as opinion mining. Text is the main resource to express the sentiments, based on this we can Identify and classifying opinions [6].

This research paper is organised in the following manner. Section I contains the introduction of the proposed topic, Section II contains related work of proposed method, Section III describe the Methodology with proposed architecture and steps, Section IV describe the performance evaluation and

results achieved by the proposed method, Section V describe the Conclusion, Section VI describe the future work.

II. RELATED WORK

In this paper [1], the author has discussed the Support Vector Machine algorithm for sentiment analysis. To perform the analysis of Support Vector Machine, the author has used two different datasets, one is self-driven car and another for apple product review tweets. Here they have used WEKA tool to achieve the performance of algorithm and compare the results. Here the results can be achieved by using precision, recall and f-measure. Author has concluded that the dependency performance of SVM depends on the input dataset. In future they can be explored for large datasets.

In this paper [2], the author has discussed sentiment analysis technique for self-driven cars. The author has used topic modelling for sentiment analysis. Topic modelling is a kind of statistical model and used to extract the latent semantic structure of dataset. They have used Mallet topic modelling packages and leverages the Gibbs sampling to infer the documents. Author has used different techniques for extracting the tweets like positive and negative words. He used unigrams, meta data and linguistic features for extracting the features. Author has used random forest classifier to achieve the accuracy. In future they have focus on to expand the dataset and extract the user generated text from numerous domains like reviews and microblog.

The authors in this paper [3], discussed six types of classification algorithm for twitter sentiment analysis. He has taken twitter sanders twitter dataset features like unigrams, punctuations, emotions and opinion lexicon. Author has compared the six types of supervised classification machine learning algorithms. Here discriminative multinomial naïve bays algorithm and sequential minimal optimization algorithm gives the best results and well performed algorithm of classification. Here the simulation results consist of sentiment features like unigrams, emotions and sentiment gazzetters and it gives better sentiment analysis.

In this paper [4], the author has used different multi-class Support Vector Machine for sentiment analysis on twitter. The classifier will not perform better on the not trained topic but perform better on the specific trained topic. To overcome this problem, the adaptive training method is proposed. Author has applied non-text features for training the model and he classified the tweets as either positive, negative or neutral. Author has achieved better accuracy using multi class Support Vector Machine. Author has concluded that the proposed method performed well on different topics in terms of precision, recall and f1-score.

In this paper [5], the author has proposed different clustering and sentiment analysis technique for twitter data. Clustering is used to resemble the identical objects and form a group of clusters. On top of sentiment scores, the author has discovered the weekly positive, negative tweets and strongly positive and negative tweets. By using polarity and subjectivity of tweets, the author has applied different clustering approaches to find the relationship between tweets.

In this paper [6], the author has discussed a survey technique for Twitter sentiment analysis. The tweets opinions are unstructured, heterogeneous and either positive, negative or neutral. Author has provided a survey and comparative techniques for the existing techniques. For the existing techniques they have used machine learning and lexicon-based approaches along with evaluation metrics. Using different machine learning algorithms like Naïve baye's, Max entropy and support vector machine are used.

The authors in this paper [7], presents a sentiment analysis by using source of text, here sentiments are identified and classified using machine learning techniques. Twitter sentiment analysis is hard related to normal sentiment analysis, because it consists of slang words and misspellings. User generated data is needful to know the opinion of the crowd. Author has used knowledge based and machine learning approach to analyze the sentiments of the tweet. Here they are analyze the opinions of the electronic gadgets like mobiles, laptops etc. Author has try to infer the consequence of domain information of sentiment classification. For classifying the opinion of the products, they have used feature vector and classified as positive, negative.

In this paper [8], the author has proposed predictive and descriptive approaches for DKI Jakarta's gubernatorial election in twitter. The dataset is fetched from the candidate's search query of Twitter. To classify the dataset in analytical approaches, author has used Support vector machine and Multinomial Naïve Bayes algorithm. To classify the descriptive approaches, the author has used time series graphs and word clouds. By using these two approaches the author has come up the twitter sentiment in an election.

The authors in this paper [9], discussed a novel approach for twitter sentiment analysis. They have considered the lexicon-based approach to classifying the sentiment. The training data consists of labelled tweets. A novel spell-checking algorithm is used for data pre-processing, for disjoining compound words like "high hopes" is implemented and here emotions are replaced by using emotion words like happy or sad. After pre-processing, machine learning algorithms are used for classification. To estimate the degree of sentiment, author has proposed avant-grade sentiment scoring mechanism. They got 80% accuracy for sentiments.

The authors in this paper [10], has taken online transportation like GoJek twitter post, they have proposed public sentiments for online transportation services, the proposed system collect the tweets form the twitter. They have eliminated the noise for the text and it is the pre-processing step. Author has used n-gram and TF-IDF for feature extraction. Here they have used Support Vector Machine algorithm to classify the

statement as positive, negative. Result of the paper can be predicted as positive, negative to the GoJek.

With the help of [1][2][4] and [5], we have adopted various machine learning techniques, to characterize the human opinion using sentiment machine learning algorithms. We have adopted Support vector machine, random forest algorithm for this project. Supervised machine learning algorithms can achieve tasks based on the predefined rules given by the user. The existing literature survey shows that results gained by applying these predictive techniques on social media twitter dataset will run visions on the basic attributes and parameters of the same.

III. METHODOLOGY

Figure 1 shows the architecture of proposed model. The proposed system architecture aims in the characterization of human opinion in social media using machine learning algorithms.

In proposed architecture, the data collection is an important step. First step is to collect the self-driven car Twitter data and the data will be collected by twitter API. Extracted Twitter tweets are in the unclean form and require to do clean the data using Pre-processing techniques. In the proposed system, we have used python word cloud and stop words for data pre-processing. After pre-processed data, require to do sentiment analysis. Based on the Sentiment compound polarity score, require to classify the tweet. Next step is to apply the classification algorithm based on the sentiment compound polarity score, to analyze the opinion to check whether they have given positive, negative or neutral opinions on self-driven car tweet. Finally, these results can be compared in terms of accuracy, precision, recall and f1-score.

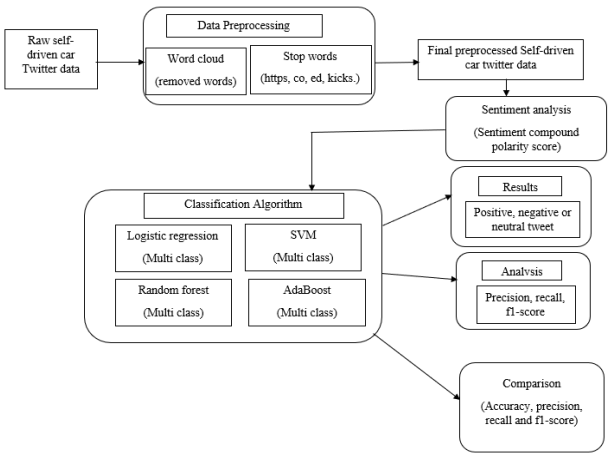

Figure 1. System architecture of proposed model

1. *Data collection from twitter*

Tweet collection consists of collecting related tweets about the particular area of significance. The tweets are collected using Twitter stream API. In this paper, self-driven car dataset is used. The pattern of these tweets can be converted into JSON/CSV files. In the proposed system we have used CSV files, which contains around 4306 twitter posts. Twitter posts. One of the example is shown like *"Driverless cars could reduce accidents, pollution and congestion"*. In these tweet statement, need to clean some word like "could" using data pre-processing. The data pre-processing is carried out in next step.

2. *Data Pre-Processing*

The data pre-processing is a main step for cleaning the data, it removes the unwanted words and it includes syntactical correction of tweets. Tweets are in the form of unstructured text data and it usually requires a transformation of text into the representation of machine processable format. In order to reduce the uncertainty in extraction of features data, the data should be in machine readable format. By applying python word cloud technique on the tweet dataset, data can be pre-processed, and it can be represented in the required format. In word cloud technique, the size of each word represents its frequency. Important textual data can be focused using word cloud. Word clouds are extensively handled for analyzing data from social network sites. The few steps are included in the data pre-processing are as follows.

- Removal of Stop words: It is the best familiar words which is avoided by the most search engines. During crawling or indexing this stop words saves the time and space.
- Twitter Feature Removal: Feature processing is ineffective and from the perspective of future the usernames and URLs are not important. To save the space in the databases search engine helps efficiently.
- Removal of re-tweets.

3. *Sentiment Analysis*

After pre-processed data, we need to do sentiment analysis. Through Natural Language Processing(NLP) some of the sources like database, tweets emit the opinions, views and emotions automated by the process can be defined by sentiment analysis [8] [14].

To express the sentiments in social media here we are using the "VADER" (Valence Aware Dictionary and Sentiment Reasoner), it is a lexicon and rule-based sentimental analysis tool and inherently lexicons is related to the sentiment-related texts. Based on the sentiment compound polarity, we are going to classify the tweets in the next step.

4. Classification techniques

For twitter sentiment classification, we used different kinds of classifiers and it is mainly used for twitter text classification. For classification we used Random forest, AdaBoost, Logistic regression and Support Vector Machine classifiers.

4.1 Random forest Classifier

Random forest is an ensemble learning algorithm. The main assumption of this algorithm is to build the small decision tree along with some new features and computing the competitive features. If we frame several small and weedy decision trees is in parallel, then we can club the trees and to form a single, robust learner can take the majority vote [9] [11].

The beginning of the random forest algorithm starts by selecting "f" features out of "k" features. Using the best split approach, here the execution of algorithm shows the sub nodes. The execution of the random forest classifier algorithm explained in the below steps.

Step 1: Loading of Self-driven car dataset
Step 2: Split the dataset into X and Y. //For Training and Testing the data
Step 3: Extract the Features from the dataset // To Train the Data
Step 4: Apply TfidfVectorizer to the dataset //To convert the feature into Vector Form
Step 5: Apply Random forest classifier to dataset //To Classify the different statements
Step 6: Apply Kfold for cross validation //To validate the dataset
Step 7: Calculate the Metrics like Accuracy, Precision and recall //Evaluate the Performance Calculation.
Step 8: Result

4.2 Support Vector Machine

Support Vector Machine is one which analyze the data used for regression analysis and classification. Support Vector Machine are the supervised models with associated learning algorithms [10] [14]. In this project we are using the linear SVM for classification because it's a fast machine learning algorithm which solves the problems on large datasets. The execution of the random forest classifier algorithm explained in the below steps.

Step 1: Loading of Self-driven car dataset
Step 2: Split the dataset into X and Y. //For Training and Testing the data
Step 3: Extract the Features from the dataset // To Train the Data
Step 4: Apply TfidfVectorizer to the dataset //To convert the feature into Vector Form
Step 5: Apply linear SVC classifier to dataset //To Classify the different statements
Step 6: Apply Kfold for cross validation //To validate the dataset
Step 7: Calculate the Metrics like Accuracy, Precision and recall //Evaluate the Performance Calculation.
Step 8: Result

4.3 AdaBoost classifier

AdaBoost is a kind of ensemble learning algorithm, where various learners are worked to figure out the stronger learning algorithm. AdaBoost is one of the most commonly used approach for the performance of the decision trees. It is mainly used for the classification problems rather than regression problems. It is used to achieve the boost performance of machine learning algorithm.

For weak learners its well suitable. The accuracy of the models is well suitable when choosing random classification problem. AdaBoost are decision trees with one level and it's the most common algorithm. The execution of the random forest classifier algorithm explained in the below steps.

Step 1: Loading of Self-driven car dataset
Step 2: Split the dataset into X and Y. //For Training and Testing the data
Step 3: Extract the Features from the dataset // To Train the Data
Step 4: Apply TfidfVectorizer to the dataset //To convert the feature into Vector Form
Step 5: Apply AdaBoost classifier to dataset //To Classify the different statements
Step 6: Apply Kfold for cross validation //To validate the dataset
Step 7: Calculate the Metrics like Accuracy, Precision and recall //Evaluate the Performance Calculation.

Step 8: Result

4.4 Logistic Regression

Logistic regression is one in which estimates variable categorically and its shows the relationship between dependent and independent. For sigmoid and soft max logistic regression helps to estimates the probabilities. To estimate the final target class based on the high probability [12]. The execution of the random forest classifier algorithm explained in the below steps.

Step 1: Loading of Self-driven car dataset
Step 2: Split the dataset into X and Y. //For Training and Testing the data
Step 3: Extract the Features from the dataset // To Train the Data
Step 4: Apply TfidfVectorizer to the dataset //To convert the feature into Vector Form
Step 5: Apply logistic regression classifier to dataset //To Classify the different statements
Step 6: Apply Kfold for cross validation //To validate the dataset

Step 7: Calculate the Metrics like Accuracy, Precision and recall //Evaluate the Performance Calculation.
Step 8: Result

IV. PERFORMANCE EVALUATION AND RESULTS

Self driven cars tweets are extracted by using the Twitter Application Programming Interface. A total of 4306 twitter posts has been selected to create the dataset. 3445 tweets are selected for training set and 861 tweets are selected for testing.
Since we have taken self-driven car Twitter dataset and to characterize the human opinion of tweets. We have used sentiment analysis technique and different machine learning techniques to characterize the human opinion. By using Anaconda navigator, the classifiers can be tested. Random Forest, Support Vector Machine, AdaBoost and Logistic Regression are used for classification and ensemble classifier for sentiment classification.
The performance evaluation can be done in terms of Accuracy, Precision, Recall and F1-score.

Accuracy
The accuracy is an evaluation of number of properly classified documents over the sum of the number of correctly classified documents and incorrectly classified documents [8].
The evaluation accuracy of the classifier relies on four features. They are as follows

- True positives(TP) are the correctly classified which are relevant documents as per the proper label.
- True negatives(TN) are the correctly classified which are irrelevant documents as an improper label.
- False positives(FP) are the incorrectly classified which are irrelevant documents as proper label. Hence it is type I error.
- False negatives(FN) are the incorrectly classified which are relevant documents as improper label. Hence it is type II error.

Accuracy is calculated by using the below eq (1)

$$\text{Accuracy}=\frac{TP+TN}{(TP+TN+FP+FN)} \quad \text{.....eq (1)}$$

Precision: Precision is the one of the parameter which specifies as proper label how many number of documents are classified [4].

$$\text{Precision} = \frac{TP}{TP+FP} \text{.....eq (2)}$$

Recall: Recall is another parameter which specifies how many number of relevant documents which are classified [4].

$$\text{Recall} = \frac{TP}{TP+FN} \text{..... .....eq (3)}$$

F1-score: The calculation F1-score is based on these two parameters precision and recall is calculated by using the below formula [4].

$$F1 = 2 \times \frac{Precision*Recall}{Precision+Recall} \text{.....eq (4)}$$

5.1 Comparison of Performance evaluation

	Accuracy	Precision	Recall	F1-score
Random forest	80.31	84.42	73.79	77.22
Logistic Regression	84.98	88.96	79.17	82.60
SVM	84.25	85.04	80.64	82.46
AdaBoost	74.22	74	68.14	69.61

Table 1: Comparison of Performance evaluation

The above Table 1 represents the comparison of performance evaluation in terms of accuracy, precision, recall and f1-score. The result can be compared by using Random forest, Logistic regression, Support Vector machine and AdaBoost classifier. Logistic regression got the highest accuracy with 84.98 and AdaBoost got the least accuracy with 74.22 due to the low samples of dataset.

5.2 Confusion matrix
The performance of the implemented model can be tabulated in the confusion matrix. Confusion matrix is used to define the performance in terms of true label and predicted label.

5.2.1 Confusion matrix of Random Forest classifier

		Predicted Label		
True Label	Class	Negative	Neutral	Positive
	Negative	**36**	33	4
	Neutral	9	**266**	10
	Positive	0	71	**125**

Table 2 : Confusion matrix of Random forest classifier

The above Table 2 shows the confusion matrix of random forest classifier. The above table depicts the value 36 are really negative, 266 represents really neutral, 125 represents really positive and other values are misclassified values.

5.2.2 Confusion matrix of Support Vector Machine

		Predicted Label		
True Label	Class	Negative	Neutral	Positive
	Negative	**45**	19	9
	Neutral	7	**259**	19
	Positive	0	40	**156**

Table 3: Confusion matrix of Random forest classifier

The above Table 2 shows the confusion matrix of random forest classifier. The above table depicts the value 36 are really negative, 266 represents really neutral, 125 represents really positive and other values are misclassified values.

5.2.2 Confusion matrix of Logistic regression

		Predicted Label		
True Label	Class	Negative	Neutral	Positive
	Negative	**70**	31	6
	Neutral	2	**373**	7
	Positive	0	60	**177**

Table 4: Confusion matrix of Random forest classifier

The above Table 4 shows the confusion matrix of Logistic Regression. The above table depicts the value 70 is really negative, 373 represents really neutral, 177 represents really positive and other values are misclassified values.

5.2.2 Confusion matrix of AdaBoost classifier

		Predicted Label		
True Label	Class	Negative	Neutral	Positive
	Negative	**35**	35	3
	Neutral	14	**260**	11
	Positive	5	86	**105**

Table 5: Confusion matrix of Random forest classifier

The above Table 5 shows the confusion matrix of AdaBoost classifier. The above table depicts the value 35 is really negative, 260 represents really neutral, 105 represents really positive and other values are misclassified values.

5.3 Comparison of results

The execution of four classification algorithm are evaluated on Self-driven car tweets**.** Confusion matrix table are represented in Tables 2-5.

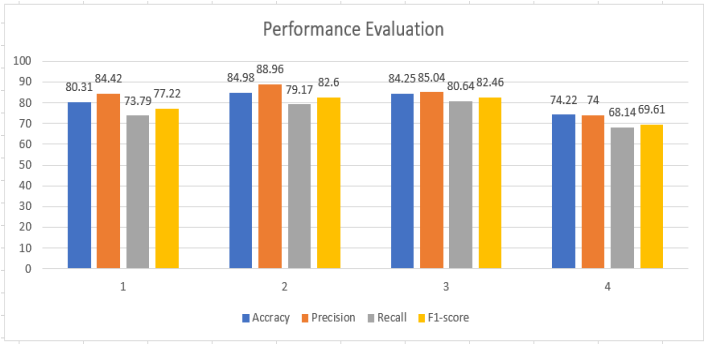

Figure 2: Performance Evaluation

Where 1= Random forest, 2= logistic regression, 3= SVM, 4= AdaBoost

The performance evaluation can be compared in terms of accuracy, precision, recall and f1-score using different machine learning algorithms such as Random forest, logistic regression, SVM and AdaBoost classifier as shown in the above figure 2. The Accuracy of all these algorithms are calculated based on True positive, True negative, False positive and False negative values as shown in eq (1). The precision of all algorithms can be calculated based on True positive and false positive values as shown in eq (2). Recall of all algorithms specifies how many number of relevant values which are classified based on True positive and false negative values as shown in the eq (3). F1-score is the weighted average of precision and recall, as shown in the eq (4). In this paper recall and f1-score is low due to the short number of samples in the tweets. The performance of AdaBoost classifier algorithm is low in contrast to other algorithms are considered in this paper. From the above performance evaluation in figure 2, it can be shown that logistic regression performs better than Random forest, Support vector machine and AdaBoost classifier to characterize the human opinion in self-driven car.

V. CONCLUSION

To characterize the human opinion in the social network, we considered Twitter dataset for this project. Based on the people's or user's opinion about self-driven car twitter dataset, we are tried to characterize the opinion of human-like whether they have given the positive, negative or neutral statement about the self-driven car. In a current trend, the self-driven car is a controversial topic in all over the globe. The work of the proposed model carried out through the pre-processing stage and classifier stage. Pre-processing helped to get the good accuracy for the proposed model. Based on the sentiment compound polarity score, we are going to classify the tweet as positive, negative or neutral. We have used four types of classifiers such as Random forest, AdaBoost, Support Vector Machine, and Logistic Regression to classify the tweets. The performance evaluation of tweets can be carried out in terms of accuracy, precision, recall, and

f1-score and compare the results with these four algorithms. The comparative results show that the proposed model got the good accuracy with Logistic Regression.

VI. FUTURE WORK

In future, we can expand the dataset to get the more user-generated text from the various domains such as blogs and review, by combining these domains will enable the better understanding of the citizens and users who will be the final recipient of the new products. To get recommendations about the self-driven cars can be obtained in the future. We can also expand the feature for predicting the input text polarity. Due to the lack of interpretability of the outcome, by using the deep learning algorithm can boost the accuracy in the future.

REFERENCES

[1] Munir Ahmad, Shabib Aftab, Iftikhar Ali, "Sentiment Analysis of Tweets using SVM", International Journal of Computer Applications (0975 – 8887) Volume 177 – No.5, November 2017.

[2] Rizwan Sadiq and Mohsin Khan, "Analyzing Self-Driving Cars on Twitter"

[3] Ajay Deshwal, Sudhir Kumar Sharma, "Twitter Sentiment Analysis using Various Classification Algorithms", 2016 5th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO) (Trends and Future Directions), Sep. 7-9, 2016, AIIT, Amity University Uttar Pradesh, Noida, India.

[4] K Lavanya, C Deisy, "Twitter Sentiment Analysis Using Multi-Class SVM", 2017 International Conference on Intelligent Computing and Control (I2C2'17).

[5] Shreya Ahuja, Gaurav Dubey2, "Clustering and Sentiment Analysis on Twitter Data", 2017 2nd International Conference on Telecommunication and Networks (TEL-NET 2017) .

[6] Vishal A. Kharde, S.S. Sonawane, "Sentiment Analysis of Twitter Data: A Survey of Techniques", International Journal of Computer Applications (0975 – 8887) Volume 139 – No.11, April 2016.

[7] Neethu M S, Rajasree R, "Sentiment Analysis in Twitter using Machine Learning Techniques", IEEE – 31661.

[8] Della Fitrayani Budiono, Anto Satriyo Nugroho, "Twitter Sentiment Analysis of DKI Jakarta's Gubernatorial Election 2017 with Predictive and Descriptive Approaches", 2017 International Conference on Computer, Control, Informatics and its Applications.

[9] Sanket Sahu, Suraj kumar Rout, Debasmith Mohanty, "Twitter Sentiment Analysis", 2015 IEEE International Symposium on Nanoelectronic and Information Systems.

[10] Ike Pertiwi Windasari, Fajar Nurul Uzzi, Kodrat Iman Satoto, "Sentiment Analysis on Twitter Posts: An analysis of Positive or Negative Opinion on GoJek", Proc. of 2017 4th Int. Conf. on Information Tech., Computer, and Electrical Engineering (ICITACEE), Oct 18-19, 2017, Semarang, Indonesia.

[11] Adyan Marendra Ramadhani, Hong Soon Goo, "Twitter Sentiment Analysis using Deep Learning Methods", 2017 7th International Annual Engineering Seminar (InAES), Yogyakarta, Indonesia.

[12] A53053719 Che-Lin, A53087422 Yu-Ching Hu, A53093903 Chien-Han Lin, "Twitter Sentiment Analysis".

[13] C. Nanda1, M. Dua, "A Survey on Sentiment Analysis", International Journal of Scientific Research in Computer Science and Engineering Vol.5, Issue.2, pp.67-70, April (2017).

[14] Gagandeep Kaur1, Kamaldeep Kaur, "Sentiment Detection from Punjabi Text using Support Vector Machine", International Journal of Scientific Research in Computer Science and Engineering Vol.5, Issue.6, pp.39-46, December (2017).

Authors Profile

Lavanya V S, She received B.E degree in Computer Science and Engineering from Visveswaraya Technological University, Karnataka in 2013. She is currently pursuing M.Tech in Software engineering at Ramaiah Institute of Technology, Bengaluru. Her research interest is in Software engineering, Software testing ,Cloud computing and Machine Learning.

Savita K.Shetty received her B.E. (1996) in Computer Science and Engineering from Karnatak University ,Dharwad, and M.Tech (2004) in Computer Science and Engineering from Visvesvaraya Technological University,Belagavi. She is currently working on her Ph.D. in Computer Science and Engineering. Visvesvaraya Technological University, Belagavi, Karnataka. Her research interests include Data Analytics, Data Mining and Machine learning

