Review of Factors Affecting Efficiency of Twitter Data Sentiment Analysis

Sangeeta and Nasib Singh Gill

***Abstract*—Twitter sentiment analysis has been explored in various domains including Business reviews, Political forecasting, decision support, Movie reviews and many more. The nature of data collected by Twitter imposes several challenges for sentiment analysis. There are other factors also like the selected classifier, multiclass sentiment analysis, feature selection method, number of feature selected, level of preprocessing, preprocessing techniques involved that can affect the accuracy of classification. This paper discusses various factors affecting the accuracy of Twitter sentiment analysis. Consideration of these factors can be very beneficial while designing an efficient classification model for twitter sentiment analysis. The survey also focuses on various metrics used for representation of sentiment analysis result and their relevance.**

***Index Terms*—Twitter sentiment analysis, recall, class imbalance, multiclass classification, feature selection.**

I. Introduction

Twitter is a social networking site where people freely give their opinion about various topics. Tweets posted on Twitter can have maximum length of 140 characters. Sentiment analysis also called as opinion mining, analyze people opinion about various topics [1]. It extracts the information about tweet whether it is positive or Negative. There can be more number of classes rather than just dual. In that case classification is trickier. Tweet can include URL's, Hash tags, emoji, emoticons, short forms, numbers, Punctuations, misspelled words, multi lingual words etc. There can be toggle of opinion sever time in a single tweet. Even tweet may not contain any of the opinion. So it is quite difficult to accurately classify tweet in mentioned classes.

Accuracy of classification is important in Twitter sentiment analysis. But in some of the application accuracy of result is crucial. Various metrics like Accuracy, F-Score, Recall, Precision are used to represent classification efficiency.

There are vast range of algorithms for doing Twitter sentiment analysis based on Lexicon based approach, Machine learning approach or ensemble approach. All these are explored in number of researches. Naïve Bayes, SVM outperforms than other machine learning based algorithms. Algorithms based on ensemble like Boosting and Bagging also performs better than other approaches. Choice of algorithm noticeably affects the accuracy of results.

Nature of data, Algorithm used, Attribute selected, Number of attributes, Number of classes involved in classification are few factors affecting classification result. The goal of this paper is to discuss various factors affecting result of sentiment analysis.

The paper is organized as follows. In Section I, introduction to Twitter sentiment analysis is given. Section II, mention the various metrics used for the measurement of classification result. Section III, give overview of the related work included in the survey. In Section IV, various factors affecting the Twitter classification result are discussed. Survey is summarized in Section V.

II. Performance Evaluation Metric for Sentiment Analysis

Accuracy of result is the main requirement in classification. Confusion matrix is very important in finding the accuracy of classification. This matrix tells, what the real classes were and how accurate the model is able to predict the class by comparing training and testing data instances. Generally four Metrics based on confusion matrix are used for the analysis of result of sentiment analysis.

A. Accuracy

Accuracy is calculated as the ratio of correctly classified instances over total number of instances. Accuracy as a metric is good when cost of false positive and false negative are similar means data is symmetric.

$$Accuracy=(TP+TN)/(TP+TN+FP+FN)$$

Accuracy is the main measure for representing the classification result but it is alone not sufficient in many cases. Like if the data is imbalanced then accuracy alone will not give the real picture. Other measures are required to clear the picture.

B. Recall

Recall is the ratio of truly classified positive instances over total number of positive instances. It is also called as sensitivity or true positive rate.

$$Recall = TP/(TP+FN)$$

High recall value means that class is classified correctly.

C. F-Score

This parameter considers both precision and recall to compute score. F1 score is harmonic average of precision and Recall. F1 score give equal weight to both precision and recall.

$$F1=2*(precision*recall) / (precision + recall)$$

Manuscript received October 3, 2019; revised December 20, 2019.

Sangeeta and Nasib Singh Gill are with Department of Computer Science & Application with Maharishi Dayanand University, Rohtak, India (e-mail: sangeeta.yogi@gmail.com, nasibsgill@gmail.com).

DOI: 10.7763/IJCTE.2020.V12.1263

F score is more meaningful than accuracy when uneven class distribution is in training data set.

D. Precision

Precision is the ratio of correctly predicted true instances divided by total predicted True instances.

$$\text{Precision} = TP / (TP+FP)$$

Precision is a good measure to use when cost of false positive is high. All these parameters are collectively used for performance evaluation.

III. Literature Survey

Twitter data sentiment analysis has wide variety of applications. Classification accuracy is very important while sentiment analysis. Data collected from Twitter is preprocessed and cleaned before actual classification. After that feature selection is done and classifier is applied. Various factors affecting the performance of classification are addressed in various researches.

Doaa Mohey El-Din Mohamed Hussein [2] analyzed various challenges in sentiment analysis by a comparative study of 47 research papers.

By his research he show the improvement in accuracy in the area of huge lexicon, Negation handling, Extracting features, Domain dependency, spam and fake data, NLP overheads with the advancement in research area.

If one of the classes is having remarkably lesser number of instances then this issue is termed as class imbalance problem. Class imbalance highly affects the result of classification. Instances in under sampled classes are more prone to misclassification. Various researches have referred class imbalance issue and provide different techniques to solve the problem [3]-[10].

Various researches have shown the variation in accuracy in case of multiclass sentiment classification [11], [12]. Different approaches as dictionary based, Machine learning Based and ensemble are explored in various researches and their effect on efficiency of classification is compared [13]-[17]. Classification accuracy highly depends on data quality and various researches [18]-[21] show the effect of preprocessing technique on efficiency of classification. Effects of feature selection methods and number of features on classification accuracy are explored in various researches [22]-[28]. All these factors collectively effect overall efficiency and need to be addressed in sentiment analysis. Table I show the analysis and outcome various researches in this area.

TABLE I: Researches Showing the Factors Affecting Accuracy of Twitter Data Sentiment Classification

Author	**Factor Affecting Efficiency (Challenges)**	**Approach / Technique**	**Classifier**	**Outcome**
Doaa Mohey El-Din Mohamed Hussein [2], 2018	Sentiment Analysis Challenges: Big Lexicon, Bi Polar, Extracting Features, NLP overheads, Negations, Domain Dependency, Spam and fake data	Comparison of BOG, POS, Semantic Analysis, Lexicon Technique, Maximum Entropy, N-gram	Theoretical and technical survey of 47 researches to find most prominent factors affecting sentiment analysis.	Percentage in average accuracy enhancement 79.9% in case of huge lexicons, 85.83% in case of feature extraction, 77.72 in case of negation handling, 87.85 in case of fake and spam handling, 76.62 in domain dependency.
Adnan Amin *et al*. [5], Oct 2016	Imbalance Data (Skew Data)	Oversampling Techniques- MTDF, SMOTE, ADASYN, MWMOTE, ICTE, TRkNN	Exhaustive, Genetic, Covering and RSES LEM2 Algorithm	MWMOTE outperforms with accuracy- .969 , Precision - .976, F-measure - .969 , MI- .803.
Prabhjot Kaur *et al*. [8], 2018	Imbalance Data	Oversampling and Under sampling technique	SMOTE, RUS	SMOTE outperforms than RUS for 70% synthetic noisy data
M. Bouazizi *et al*. [11] 2019	Multiclass Sentiment Analysis	1-7 sentiment class sentiment analysis	Random Forest classifier	Classification accuracy decreases from 86.0% to 60.2% as number of classes increases from 1 to 7.
Shuo Wang *et al*. [12] 2012	Multiclass and imbalanced Sentiment Analysis	Ensemble Approach for Multiclass imbalance data set	AdaBoost.NC, SMOTEBoost with and without OAA Method	Performance decreases as the number of imbalance class increases
Yun Wan *et al*. [18] 2015	Classifier selected	Lexicon Based, Machine learning and Ensemble based	Naïve Bayes, SVM, Bayesian Network, C4.5 Decision Tree, Random Forest, AdaBoost,	Comparison in term of precision show maximum in ensemble approach – Lexicon Based – 60.5% C4.5 Decision Tree – 83.7% Ensemble – 84.2%
Zhao Jianqiang *et al*. [19] 2017	Preprocessing Techniques	Six different preprocessing methods on 5 different Twitter data sets.	NB, SVM, LR, RF	Increased accuracy and F-measure after applying preprocessing techniques and Removal of Numbers, Stop words, URL's hardly effect efficiency.
Emma Haddia *et al*. [21] 2013	Text Preprocessing in Sentiment Analysis	Pre-Processing Vs No pre-Processing, Chi- square Feature selection method.	SVM Classifier	**Accuracy increases** from 78.33 to 81.5 in TF-IDF, 72.7 to 83 in FF and 82.7 to 83% in FP after applying pre-processing. **Accuracy increases** from 81.5 to 92.3 in TF-IDF, 83 to 90 in FF and 83.1 to 93 after using Chi-Square method.
Joseph D. Prusa *et al*. [23]	Feature Selection Technique	Feature selection using CS, GI, KS, MI, PR, PRC, ROC, S2N, SAM, WRS **Vs None feature Selection**	5NN, C4.5, LR, MLP,	**(With / Without feature selection)** **5NN –** 0.70387 / 0.65191 **C4.5 –** 0.70404 / 0.66392 **LR –** 0.75226 / 0.59623 **MLP -** 0.72117 / 0.53133 **Efficiency improves with feature selection.**

Different factors or challenges are mentioned and implemented in various researches and their effects are monitored. But this survey collectively analyzes the effect of all these factors. Consideration of these factors is helpful in designing an efficient model for Twitter sentiment classification.

IV. Factors Affecting Result of Sentiment Analysis

Accuracy of results matters a lots in some applications. So some work has already done on factors affecting result of Twitter data sentiment analysis. Here are various factors affecting classification accuracy and their relevance.

A. Class Imbalance Data

When all the classes in sample data set have remarkably different number of samples then the data set is imbalanced. Imbalanced data highly affect the result of classification accuracy. Adnan Amin *et al*. [5] addressed the effect of imbalanced data on classification accuracy and used various techniques to solve the problem.

Three different level of techniques as Data level approach, Algorithm level approach and ensemble level approach are used to address class imbalance problem. Data level approach include under sampling and oversampling. Under sampling randomly removes samples from the majority class and in oversampling number of instances in under represented class are increased. AdaCost, CSB1, CSB2, Z-SVM, GSVM-RU, AdaC1, AdaC2, AdaC3 are few algorithm level approaches. Ensemble approach is the ensemble of the above two approaches to efficiently handle class imbalance problem. SMOTEBoost, RUSBoost, MSMOTEBoost, RBBagging DataBoost-IM are different ensemble approaches used.

Fig. 1 shows the various techniques to handle class imbalance data.

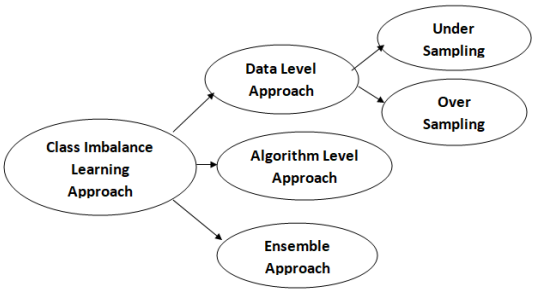

Fig. 1. Class imbalance approach for skewed data.

Yang Lu [6] proposed new measures IBI^3 and BI^3 to estimate the impact that is individually caused by imbalance data set. IBI^3 measures how sample in the minority class is influenced by the imbalance data. BI^3 is used to measure the degree of degradation of imbalanced dataset, so that one can determine whether or not to apply imbalance recovery methods.

Show-Jane Yen *et al*. [7] implemented cluster based under sampling techniques to improve the classification accuracy of minority class and investigate the effect of under sampling methods for skewed data set. The outcome shows that cluster based under sampling techniques outperforms other under sampling techniques.

P. Kaur *et al*. [8] implemented two techniques SMOTE and RUS for skewed data. SMOTE performed better than RUS.

Ines Domingues *et al*. [9] implemented over sampling data balancing techniques. Results show that data balance techniques improve classification results on ordinal imbalanced datasets.

Shaza M *et al*. [10] in their research mentioned Under sampling, Over sampling, Cost sensitive method, Recognition based method and Ensemble method for handling class imbalance issue.

B. Multiclass Sentiment Classification

Higher the number of classers involved in classification, more it is difficult to classify the data and accuracy of result is diminished. If only two classes positive and negative are involved, then classification comparatively easy and accurate.

M. Bouazizi [11] *et al*. implements sentiment analysis on various number of classes and compared the result of binary class with N-class sentiment analysis. Table II shows the comparative results of accuracy with different no of classes.

TABLE II: Variance in Accuracy with Increase in Number of Classes

No of Classes	Accuracy
Binary Sentiment Classes	86.0
Three Sentiment Classes	72.5
Four Sentiment Classes	69.9
Five Sentiment Classes	61.8
Six Sentiment Classes	60.4
Seven Sentiment Classes	60.2

The results in Table II clearly show the decrease in classification accuracy with increase in number of sentiment classes. Although decrease in accuracy with increase in number of classes may not be uniform but general behavior remain same.

The paper also presents various challenges that makes multi class sentiment analysis difficult. Negation handling, context dependency, Polysemy, Presence of multiple sentiments, Closeness between different sentiments, and Absence of sentiment indicators are various challenges faces in multiclass classification.

Shuo Wang *et al*. [12] addressed multiclass imbalance problems and effect of multi minority and multi majority on the performance of re-sampling techniques is analyzed. Proposed approach AdaBoost.NC and other ensemble algorithms are implemented and compared with multiclass imbalance data. The result of analysis show that proposed AdaBoostNC performed better in recognizing minority class instances.

C. Classifier Selected

Selection of Classifier is very important when talking about classification efficiency. Various approaches are used for sentiment classification like Lexicon approach, Machine learning approach and ensemble approach.

In lexicon based approach predefined polarity dictionary are used to find the polarity of uni-gram or n-gram features present in the tweets. Average polarity score of every tweet is calculated and a given threshold decide the polarity of tweet as positive or negative. This technique is quite simple but performance is low [13]-[15]. Machine learning approaches

for example Naïve Bayes, Support Vector Machine (SVM), Decision tree, Maximum entropy classifier, K-Nearest Neighbor (KNN) are few commonly used classifiers. SVM and Naïve bayes outperforms than other classifier. Bagging and Boosting are ensemble approaches that use multiple classifier to improve performance of classification [16].

Olga Kolchyna *et al*. [17] compared machine learning algorithms with lexicon based approach and results in Table III show that machine learning approach outperforms better than lexicon based approach. And SVM performs better than other machine learning approaches.

TABLE III: COMPARISON OF MACHINE LEARNING APPROACH AND LEXICON BASED APPROACH

Method	Accuracy
Naïve Bayes	81.5%
Decision Tree	80.57%
SVM	86.62%
Lexicon Based Approach	61.74%

Ensemble approach provides better result in many cases depending upon the ensemble algorithms used. Yun Wan *et al*. [18] implemented machine learning and ensemble classifiers and research show that ensemble classifier get better performance in terms of precision, Recall and F-Measure. Ensemble approach shows a high recall of 84.2 % with a minimum error rate of 15.8 %.

Some researches implemented Neural Network or Neuro-Fuzzy method based classifiers to improve classification efficiency.

D. Preprocessing Techniques

Preprocessing of data means removal of irrelevant information like URL's, Numbers, Emoticons, punctuations, stemming, stop word removal, expanding acronyms, negation handling etc. from tweets before actual classification. After preprocessing of data, Feature selection is done and than selected classifier are applied to classify tweets.

Zhao Jianqiang *et al*. [19] analyzed the effect of six different preprocessing techniques by using four classifiers and two feature selection methods. The result of the research clearly show that accuracy and F-measure are improved by using preprocessing techniques as expanding acronyms and replacing negation. But there is little difference on accuracy after removing of URL,s , Numbers and stop words. Performance of SVM is improved after expanding acronyms and replacing negations. So it is better to remove Numbers , URL,s and stop words to remove noise and it hardly effect the accuracy of result. Random deletions of words decreases the performance remarkably as the word deleted may be a keyword for polarity detection. The result show that same preprocessing techniques applied on different classifier have similar effect on classifiers accuracy, But there are more variations in result in case of Naïve Bayes and Random Forest after application of preprocessing techniques.

Akrivi *et al*. [20] analyzed the effect of preprocessing techniques of Twitter data sentiment analysis. The results show that appropriate feature selection method improve the accuracy of classification. Although the behavior represented is not much uniform but comparatively 1-to-3 gram performed better as compared to unigram and N-gram. Also results shows that feature selection improve the results over all feature selection. Results also show that significant improvement in result is shown when attribute selection is based on information gain.

Emma Haddia *et al*. [21] shows the role of pre processing in sentiment analysis. The research use three feature metrics TF-IDF, FF, FP on unprocessed and processed data and result show the improvement in accuracy in processed data. Improvement is observed in the accuracies of TD-IDF matrix from 78.33 to 81.5, in Metric FF 76.33 to 83 and in FP matrix 82.33 to 83. By investigating further in the results we notice the increase in the accuracies when applying the classifier on the pre-processed data with a highest accuracy of 83%.

Preprocessing of data highly effect the accuracy of result. Unprocessed data may produce lots of irrelevant features. Even highly processed data may loss lots of relevant information from the data.

E. Feature Selection Method and Number of Features Selected

Feature Selection is the method of selecting optimal subset of features and eliminating irrelevant features form feature set. Unigram, Bigram or N-gram can be the strategy for feature selection. In unigram feature only one word is used as feature. In N-gram approach a combination of 2 to 6 words are used as single feature. Pattern Based feature selection can be done in which features are categories as regular words, or high frequency words. Chi- square method used for feature selection helps to reduce the dimensionality as well as noise in data and increase the accuracy of classifier from 81.5 to 92.3 in TF-IDF, 83 to 90 in FF and 83.1 to 93 in FP [21]. Number of feature selected highly effect the accuracy of result but after a limit increase in size of feature set does not increase classification accuracy [22]-[25].

There are different category of feature selection techniques such as Filter based, threshold based or First-Order Statistics based feature selection. Filter based feature selection technique use a single statistical measure that suits the data, and calculate the score for each feature. On the basis of that score only relevant features are selected. This technique rank feature without using any learning algorithm. Pearson Correlation, Mutual Information(MI), Kendall Correlation, Spearman Correlation, Chi Squared, Count based, Fisher score are few feature based selection techniques. F-Measure, Gini-Index (GI), Kolmogorov-Smirnov (KS) statistic, Mutual Information (MI), Probability Ratio(PR), area under the Precision-Recall Curve (PRC), and area under the Receiver Operating Characteristic(ROC) curve are threshold based feature selection technique. Signal-to-Noise (S2N) ratio, Significance Analysis of Microarrays (SAM) and Wilcoxon Rank Sum (WRS) are few First-Order Statistics based feature selection techniques used in various researches.

Riham Mansour *et al*. [26] in their research used multiple set of features for sentiment classification and used an ensemble classifier. The classification complexity comes out linear with the increase in number of features. The ensemble is implemented on two feature set one optimal set with 20000 features and other NRC data set with 4 Million features. The feature set with selected 20000 features have shown relative 9.9% and 11.9% performance gain over 4 million feature set.

Jović, K *et al*. [27] in this survey mentioned various feature

selection methods for classification, regression and clustering. Information gain, Gain ratio, Chi-square, Fast correlation-based filter (FCBF) are some of the feature selection techniques used for classification.

Jason Van Hulse *et al*. [28] implemented 11 threshold based feature selection methods and compared them with six standard filter based feature selection methods for SVM and Na ïve Bayes classifiers. The results show the improvement of results by using threshold based methods as compared to filter based methods.

F. Nature of Data

Nature of data is very important while doing sentiment classification. As addressed in the previous section skewed data can negatively affect the result of classification. Quality of sample has great effect on result.

In case of Twitter data lots of symbols, Emoticons , Informal language, Symbols, URL,s, Short forms for words are used. That makes it tricky to handle data.

G. Sample Size Used for Training

If sample size is not sufficient for all the classes involved, then results are not reliable. In Twitter sentiment analysis for any topic sufficient opinionated tweets may not be available. In that case a decision based on that insufficient data is unreliable.

H. Methods of Collecting and Labeling Samples

It is very important that the source of data collection should be reliable. In some applications highly accurate data is required like medical, scientific applications, disaster management, weather forecasting, Decision support etc. Wrong results can be more dangerous. In some other applications like social media sentiment analysis, Business analysis, Political forecasting etc. results can vary.

In social media sentiment analysis data collected may not be very reliable. Twitter allows to access one percentage of data for research analysis. Data may be downloaded statically from the past repository or live streaming data can be used for analysis. In Twitter data sentiment analysis tweets are re-tweeted many numbers of times. People use a very informal language that may not make any sense at all. Sample labeling for training data is also very important. Either manual labeling of instances in various classes is done or various algorithms are used. All these factors make the Twitter sentiment classification trickier and instances may be misclassified.

V. Conclusion

Accuracy of result is the most desired factor in Twitter data sentiment analysis. The paper presents various factors affecting the efficiency of sentiment analysis. Imbalanced data, Multiclass Sentiment analysis, Number of attribute selected, Attribute selection techniques, Classifiers, Nature of data are various factors that affects accuracy of classification. All these factors are addressed in various researches. Classification efficiency is improved in most of the researches after proper addressing of all these factors. Multiclass Sentiment analysis is less explored as compared to dual class.

From the review it is clear that while designing a model for sentiment classification proper classification techniques and preprocessing technique should be selected based on nature of data. Also proper subset of feature selected increases the performance of model without degrading the accuracy. This review can be helpful in addressing the relevant factors while designing a classification model with high performance in terms of accuracy and others performance parameters. There are other areas like classification of streaming and temporal data, neural networks and fuzzy based classifier, context based sentiment analysis that are newer area of research and can be further explored.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Nasib Singh Gill have given the direction of research and provided the relevant related research papers for the review. Sangeeta conducted review and research work; Nasib Singh Gill analyzed the data and work; Sangeeta wrote the paper; all authors had approved the final version.

References

[1] A. B. Xie, I. Vovsha, O. Rambow, and R. Passonneau, "Sentiment analysis of Twitter data," in *Proc. Workshop on Languages in Social Media*, Columbia University, New York, 2011, pp. 30-38.

[2] D. M. E. M. Hussein, "A survey on sentiment analysis challenges," *Journal of King Saud University – Engineering Sciences*, vol. 30, issue 4, pp. 330-338, Oct. 2018.

[3] R. Longadge, S. S. Dongre, and L. Mali, "Class Imbalance problem in data mining: Review," *International Journal of Computer Science and Network (IJCSN)*, vol. 2, issue 1, Feb. 2013.

[4] A. Ali, S. M. Shamsuddin, and A. L. Ralescu, "Classification with class imbalance problem: A review," *Int. J. Advance Soft Compu. Appl*., vol. 7, no. 3, pp. 176-204, Nov. 2015.

[5] A. Amin, S. Anwar, A. Adnan, M. Nawaz, N. Howard, J. Qadir, A.Hawalah, and A. Hussain, "Comparing oversampling techniques to handle the class imbalance problem: A customer churn prediction case study," *IEEE Access*, vol. 4, pp. 7940-7957, Oct. 2016.

[6] Y. Lu, Y. Cheung, and Y. Y. Tang, "Bayes imbalance impact index: A measure of class imbalanced dataset for classification problem," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–15, Jan. 2019.

[7] S.-J. Yen and Y.-S. Lee, "Cluster-based under-sampling approaches for imbalanced data distributions," *Expert Systems with Applications*, vol. 36, issue 3, pp. 5718–5727, April 2009.

[8] P. Kaur and A. Gosain, "Comparing the behavior of oversampling and under sampling approach of class imbalance learning by combining class imbalance problem with noise," *Advances in Intelligent Systems and Computing*, pp. 23-30, January 2018.

[9] I. Domingues, J. P. Amorim, P. H. Abreu, H. Duarte, and J. Santos, "Evaluation of oversampling data balancing techniques in the context of ordinal classification," in *Proc. International Joint Conference on Neural Networks (IJCNN)*, Oct. 2018.

[10] S. M. A. Elrahman and A. Abraham, "A review of class imbalance problem," *Journal of Network and Innovative Computing*, vol. 1, pp. 332-340, 2013.

[11] M. Bouazizi and T. Ohtsuki, "Multi-class sentiment analysis on Twitter: Classification performance and challenges," *Big Data Mining and Analytics*, vol. 2, no. 3, pp. 181–194, Sep. 2019.

[12] S. Wang and X. Yao, "Multiclass imbalance problems: Analysis and potential solutions," *IEEE Transactions on Systems, Man, and Cybernetics—Part B: Cybernetics*, vol. 42, no. 4, pp. 1119-1130, Aug. 2012.

[13] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," *Ain Shams Engineering Journal*, vol. 5, issue 4, pp. 1093-1113, Dec. 2014.

[14] D. Ray, "Lexicon based sentiment analysis of Twitter data," *International Journal for Research in Applied Science & Engineering Technology (IJRASET)*, vol. 5, issue X, pp. 910-915, Oct. 2017.

[15] K. L. Devi, P. Subathra, and P. N. Kumar, "Tweet sentiment classification using an ensemble of machine learning supervised

classifiers employing statistical feature selection methods," in *Proc. Fifth International Conference on Fuzzy and Neuro Computing (FANCCO - 2015)*, vol. 415, pp. 1-13, Nov 2015.
[16] S. Kurnaz and M. A. Mahmood, "Sentiment analysis in data of Twitter using machine learning algorithms," *International Journal of Computer Science and Mobile Computing*, vol. 8, issue. 3, pp. 31–35, March 2019.
[17] O. Kolchyna, T. T. P. Souza, P. Treleaven, and T. Aste, "Twitter sentiment analysis: Lexicon method, machine learning method and their combination," Cornell University Library, arXiv: 1507.00955, 18 Sep. 2015.
[18] Y. Wan and Q. Gao, "An ensemble sentiment classification system of Twitter data for airline services analysis," in *Proc. 2015 IEEE International Conference on Data Mining Workshop (ICDMW)*, pp. 1318-1325, 2015.
[19] J. Zhao and X. Gui, "Comparison research on text pre-processing methods on Twitter sentiment analysis," *IEEE Access*, pp. 2870-2879, Feb. 2017.
[20] A. Krouska, C. Troussas, and M. Virvou, "The effect of preprocessing techniques on Twitter sentiment analysis," in *Proc. Research Gate Conference*, July 2016.
[21] E. Haddi, X. Liu, and Y. Shi, "The role of text pre-processing in sentiment analysis," *Procedia Computer Science*, vol. 17, pp. 26-32, Dec. 2013.
[22] M. Khader, A. Awajan, and G. Al-Naymat, "The impact of natural language preprocessing on big data sentiment analysis," *International Arab Journal of Information Technology*, vol. 16, pp. 506-513, 2019.
[23] J. D. Prusa, T. M. Khoshgoftaar, and D. J. Dittman, "Impact of feature selection techniques for tweet sentiment classification," in *Proc. Twenty-Eighth International Florida Artificial Intelligence Research Society Conference*, pp. 299-304, 2015.
[24] J. V. Hulse, T. M. Khoshgoftaar, A. Napolitano, and R. Wald, "Threshold-based feature selection techniques for high-dimensional bioinformatics data," Springer, pp. 47-61, May 2012.
[25] I. Dilrukshi and K. Zoysa, "A feature selection method for twitter news classification," *International Journal of Machine Learning and Computing*, vol. 4, pp. 365-370, August 2014.
[26] R. Mansour, M. F. A. Hady, E. Hosam, H. Amr, and A. Ashour, "Feature selection for Twitter sentiment analysis: An experimental study," in *Proc. International Conference on Intelligent Text Processing and Computational Linguistics*, pp. 92-103, 2015.
[27] A. Jović, K. Brkić, and N. Bogunović, "A review of feature selection methods with applications," in *Proc. 2015 38th International Convention on Information and Communication Technology Electronics and Microelectronics*, pp. 1447-1452, 2015.
[28] J. V. Hulse, T. M. Khoshgoftaar, A. Napolitano, and R. Wald, "Threshold-based feature selection techniques for high-dimensional bioinformatics data," *Network Modeling Analysis in Health Informatics and Bioinformatics*, vol. 1, pp. 47-61, 2015.

Sangeeta is an assistant professor in GCG Gurugram, India and also a research scholar at M. D. University, Rohtak in the Department of Computer Science & Applications. Her area of research is social networking site sentiment classification. Her research interests include data mining, text mining, pattern recognition, character recognition, natural language processing, artificial intelligence, and big data analysis.

Nasib Singh Gill is the professor and head in Department of Computer Science & Application at Maharishi Dayanand University, Rohtak. He is the director of University Computer Centre and MDU Alumni at M. D. University, Rohtak. He has post-doctoral research (computer Sc.) from Brunel University (UK), Ph.D. (computer Sc.), master's degree in science and MBA. His research interests includes software metrics, component -based metrics, testing, reusability, data mining and data warehousing, NLP, AOSD, information and network security. He has written six books and is the author of 142 publications.