Journal of Theoretical and Applied Information Technology
29th February 2020. Vol.98. No 04

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

HYBRID MODEL FOR TWITTER DATA SENTIMENT ANALYSIS BASED ON ENSEMBLE OF DICTIONARY BASED CLASSIFIER AND STACKED MACHINE LEARNING CLASSIFIERS-SVM, KNN AND C5.0

[1]SANGEETA RANI, [2]NASIB SINGH GILL
[1]Assistant Professor, Department of Computer Science, M D University, India
[2]Professor, Department of Computer Science and Application, M D University, India
E-mail: [1]sangeeta.yogi@gmail.com, [2]nasibsgill@gmail.com

ABSTRACT

Social Networking sites like Twitter and Facebook has offered the possibility to users to express their opinion on various topics and events. Opinion mining is a technique to find the sentiment of people about these topics, which can be useful in decision support. Various government policies can also be monitored by doing the sentiment analysis of related tweets. The objective of this research is to enhance the accuracy of twitter sentiment classification. The paper proposes a framework for a hybrid approach with an ensemble of stacked machine learning algorithms and dictionary based classifier. Sentiment Score extracted from dictionary based classifier is added as additional feature in the feature set. Three machine learning algorithms SVM, KNN and C5.0 are stacked to build an ensemble by using two Meta learners RF and GLM. Real time manually labeled tweets based on "Clean India Mission" an Indian government policy is used for implementation of the model. Proposed model is compared with different machine learning and ensemble classifiers. Proposed hybrid model recorded higher accuracy of **0.9066377** for 5 fold cross validation and **0.9124793** for 10 fold cross validation as compared to **0.8667328** in case of stacked ensemble of SVMRadial, KNN and C5.0 by using RF as Meta classifier. RF Meta classifier performed better as compared to GLM in all stacked based ensemble. Proposed model also recorded higher accuracy as compared to machine learning classifiers-SVM, Naïve Bayes, Decision Tree, Random forest and Maximum Entropy. The contribution of the research is to enhance the accuracy of stacked based ensemble classifiers for twitter sentiment classification by using additional sentiment score provided by dictionary based classifier.
Keywords: *Clean India Mission, C5.0, KNN, Sentiment Analysis, Stack Ensemble, SVM, Swatch Bharat.*

1. INTRODUCTION

People use social networking sites like Twitter and Facebook to put their opinion about various issues. Data collected from these social media sites are frequently used in analysis and decision support. Sentiment analysis or opinion mining is used to detect opinion or mood of text as positive, Negative or Neutral. Twitter is a social networking micro blogging site frequently used to give opinion on various topics [1]–[3]. As per various researches Twitter is more suitable ground for sentiment analysis as compared to Facebook due to limited text size of 140 characters and generally single opinion. But it can be trickier also due to frequent use of Hashtags, emojis, URL's, Numbers etc. Sentiment analysis on twitter can be very important in various fields like business, tourism, Medical, Game, Politics, Science and industry [4]–[9].

Various approaches like Dictionary based, Machine learning approach, and ensemble approach are implemented in various researches for sentiment analysis of twitter data. Ensemble based classifier can be implemented either by using multiple machine learning classifier or as an ensemble of dictionary based approach and machine learning classifiers. Lexicon score of dictionary based approach can be added as one of the feature in feature set before applying machine learning ensemble. Ensemble can also be done at feature selection level by using multiple feature selection algorithms to have an optimal feature

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

subset. Bagging, Boosting, Random forest and Stacking of different machine learning algorithms are different ensemble approached implemented in various researches. Ensemble approach performs better as compared to other two approaches. Improvements in results are observed by preprocessing of data and proper feature selection.

Preprocessing of data is an important step prior to implementation of machine learning or ensemble classifier. As tweets downloaded have unwanted data like numbers, stop words, Hash tags, short forms, Punctuations, URL's, connectors, emoji that need to be removed before further processing. Stemming, tokenization, TF-IDF weighting are also required as preprocessing step to make data ready for classification. In various researches it is clearly observed that preprocessing step remove unnecessary data and help to improve classification accuracy.

Feature selection is the process of selecting optimal subset of features from large number of features. Feature selection methods and number of features selected highly effect classification efficiency but after a limit increase in size of feature set does not increase classification accuracy. It has been shown in various researches that optimal subset of features produce equally good result as compared to very large feature set.

This paper implements an enhanced framework for twitter sentiment analysis. This framework is an ensemble of dictionary based approach and stacked machine learning classifiers. Sentiment Score extracted from dictionary based approach is added in the feature set which is given as input to proposed stacked ensemble model. Three machine learning algorithms SVM, KNN and C5.0 are stacked to build an ensemble. This work compares dictionary based approach and machine learning approach with the proposed model. For the implementation of this model real time tweets are collected related to "Clean India Mission" for training and testing data set. For this more than 10000 real time tweets are downloaded. After preprocessing and removal of duplicate tweets training and testing data set of 1008 tweets are created by manual labeling.

"Clean India Mission" also called as "Swachh Bharat Abhiyan" is mission started by Government of India to clean India and make people aware about cleanliness. This mission is quite important and lots of people are coming with the mission [10], [11]. It is a common men mission and it would be very interesting and helping to know people opinion about its success. It could help to make the mission more successful as per opinion of the common men.

Section 1 gives the introduction of sentiment analysis. Literature survey of related work is given in section 2. Section 3 gives the data cleaning approaches used. Section 4 gives introduction to feature selection. Section 5 gives introduction to various ensemble classification approaches. Proposed framework is explained in Section 6. Section 7 discusses the implementation and results. Section 8 concludes the results.

2. RELATED WORK

Sentiment analysis or opinion mining is the task of classifying tweets in target classes (Positive, Negative or Neutral). This sentiment analysis can be further used in decision support. Research work has been done related to Dictionary based classifier [12], [13], Machine Learning classifier [14]–[20], Ensemble Based classifier, Neural and Deep learning based classifier. Feature selection is very important area of research and different feature selection methods are used their effect on accuracy of classifier is monitored. Data pre-processing is also very important for preparing data prior to classification[21]. Either BOW- Bag of Words (Uni-gram, Bi-gram, N-gram), POS-Tagging, TF-IDF, can be used for data set representation. Recently lots of work has been done on ensemble approach based on feature selection ensemble or classifier ensemble [22]–[25]. Also neural network and deep learning are hot topic for twitter data sentiment analysis. Here in the present research we are focusing on ensemble based classification and feature ensemble.

M. Naz at al. [26] in their research implemented an ensemble of two classifier K-Nearest neighbor and Naïve Bayes by using two feature selection techniques Forest Optimization algorithm(FOA) for feature selection and minimum redundancy and maximum relevance(mRMR) for removal of irrelevant features. . Ensemble with feature selection technique performed better as compared to individual machine learning algorithms. Results are further improved by using an ensemble of KNN, NB and SVM with an accuracy of 95%. It is also shown in the research that hybrid of FOA-KNN and FOA-NB has outperformed single KNN and NB classifiers. Accuracy is increased when FOA and nRMR feature selection techniques are applied.

M. M. Fouad et al. [27] implemented ensemble classifier based on majority voting ensemble of SVM, LR, NB by using IG as feature selection

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

technique. The result show that IG feature selection boosted the accuracy of classification. The ensemble classifier try to improve the accuracy, But if one of the ensemble participant algorithm does not suit the data set than accuracy is decremented. On the other hand emoticon based features did not have that much of addition to the efficiency.

M. Ghiassi et al. [28] developed a Twitter Specific lexicon for twitter sentiment analysis by using a supervised learning technique using n-gram statistical analysis. Their model was tested on 3440 manually collated tweets on Justin Bieber twitter account. The results show the improvement in accuracy for proposed model over SVM with an accuracy of 95.1%.

J. J. Bird et al. [29] in this research propose an approach to ensemble sentiment classification to score in the range 1-5 of negative to positive scoring. Different single classifiers named OneR, MLP, NB, NBM, RT, J48, SMO SVM and ensemble classifiers named RF, Vote (NBM,RT,MLP), Vote (RF, NBM, MLP), AdaBoost (RT), AdaBoost (RF) are implemented. The result of research show that the best performing classifier was an ensemble of RF, NBM, MLP by majority voting with an accuracy as 91.02%. In individual classifiers best model was RT (Random Tree) with an accuracy of 78.6%. All ensemble methods outperforms than single classifier.

In the above mentioned researches variation in implementation is done either by stacking different combinations of machine learning classifiers or using different feature selection algorithms but sentiment score from dictionary based classifier is not used with stacked classifier for enhancing accuracy. This information can be very valuable for performance improvement. Also stacking more number of machine learning algorithms for enhancing the accuracy of classification adds performance overhead. Proposed methodology use sentiment score as one of the feature in feature set to improve classification accuracy with little overhead.

3. RESEARCH QUESTION

What is the effect of adding lexicon score extracted from dictionary based classifier in feature set for the implementation of stacked based ensemble sentiment classifier?

Dictionary based classifier is the simplest approach for sentiment classification but it's performance does not match with machine learning or ensemble classifiers. The information provided by dictionary classifier in terms of lexicon score can be very valuable for enhancing the accuracy of machine learning and ensemble classifiers with minimum overhead.

The objective of this paper is to enhance the accuracy of sentiment classifier based on stacked ensemble of SVM, KNN and C5.0 by adding sentiment score of dictionary based classifier as one of the feature in feature vector. This research helps to improve performance of classification with minimum overhead.

4. DATA PREPROCESSING

Data preprocessing remove unnecessary information and clean the data. Like most of the social media sites, Twitter also has lots of noise like URL's, punctuation, numbers, has tags, stop words and irrelevant content in tweets. Raw tweets cannot be used for classification and preprocessing is required to get better results. Some basic steps in preprocessing are:

Tokenization: Tokenization is a process of splitting longer sentences into small phrases or single words to create BOW (Bag of words). Generally for tweets unigram tokens are considered due to single opinion and small size of text. Tokenization of tweets is considerably more difficult than tokenization of the general text since it contains numerous emoticons, URL links, abbreviations that cannot be easily separated from other content of tweet.

Stemming: In stemming words are replaced by their roots that lead to reduce the dimensionality of the BOW. For example replacing "stemming" by "stem". But it can increase the bias if not used carefully.

Stop-words removal: Prepositions, articles like "the", "is", "at", "which", "on" are only for connecting and does not affect the sentiment of tweet. It is better to remove these words to reduce dimensionality of BOW.

Handling negations: The words like "no", "never", "not", "don't" are negations which reverse the sentiment of text. To handle negations the polarity of all the words between negation and first punctuation mark after negation are reverted.

Part of speech tagging: In this preprocessing step each word of the text is tagged with the part of speech it belongs to like noun, verb, pronoun, adverb, adjective etc. In some researches it is explored, but in twitter sentiment classification unigram is preferred due to small size of tweet text.

Grammatical tagging: In grammatical tagging words of the text are tagged with the type of word category it belongs to like verb, noun, adjective etc. In some researches grammatical tagging is

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

explored, but in twitter sentiment classification unigram without grammatical tagging is preferred due to small size of tweet text.

Removal of Numbers: Numbers rarely participate in sentiment detection, so can be removed without effecting the accuracy of classification.

Removal of URL's: Most of the tweets contain URL's with hardly effect the tweet sentiment, So can be removed from tweet text without effecting the accuracy of classifier.

A. Krouska et al. [30] in their research show the effect of preprocessing on the classification accuracy. Results also show that significant improvement in result is shown when attribute selection is based on information gain.

Z. Jianqiang et al. [31] monitored the effect of six preprocessing techniques by using four classification algorithms and two feature selection methods. The result shows that accuracy is improved after using preprocessing techniques on dataset. But removal of URL's, Numbers and stop words hardly effect the accuracy so can be removed. Random deletion of word reduces the accuracy as deleted word might be important in sentiment detection. It is also shown that same preprocessing technique used on different classifier have similar effect of accuracy of result.

E. Haddia et al. [32] show the role of preprocessing on sentiment analysis. 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.

5. FEATURE SELECTION

After preprocessing and cleaning of data next step is feature selection. Feature generation techniques result in a large numbers of features being generated to represent tweets. Many of these features may degrade classifier performance and increasing computational cost. Feature selection techniques such as Information gain, Gain Ratio, Chi-square, Correlation based feature selection (CFS) can be used to select an optimal feature set. Feature selection helps in reduction of dimensionality and reducing computational cost of classifier without significantly affecting the accuracy of classifier. Feature selection on the basis of ranking of feature is important for efficient classification of tweets. As with the data set used in present research without preprocessing the total no of features are 2654 and with pre-processing are 2578. This is a very large set of features and an efficient feature subset selection technique is required.

Emma Haddia et al. [32] in their research show that chi- square method used for feature selection helps to reduce the dimensionality and noise in data and increase the performance of classifier. Accuracy is increased from 81.5 to 92.3 in TF-IDF, 83 to 90 in FF and 83.1 to 93 in FP by using Chi-square method.

R. Mansour et al. [33] in their research used multiple set of features for sentiment classification by using 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.

6. ENSEMBLE BASED CLASSIFICATION METHOD

Ensemble or hybrid base classification use combination of multiple classifiers or feature selection methods to improve classification accuracy. Ensemble is tested on different algorithms but not all the times accuracy is improved. If one of the classifier selected is not performing well in that case participation of that classifier in ensemble can degrade the efficiency. Ensemble can be performed at feature selection level also. More than one feature selection method can be used to select relevant features. Some classifier based ensemble techniques are given below:

6.1 Bagging

Bagging or bootstrap aggregating is an ensemble of multiple similar classifiers designed to improve the accuracy of classification. Some models are accurate only for the data sets they were trained on. This problem is called as over fitting. Bagging tries to minimize the variance from the models. L. Bbeiman [34] initially mentioned Bagging Predictors in his research. In Bagging sample from observation is selected randomly with replacement. Along with subset of instances, subsets of features are also selected to create the model. Features from dataset are selected such that it gives the best split on training data. This is repeated multiple times and all the weak learners run in parallel and learn independently from each other. These weak learners are combined by deterministic averaging method. Prediction is given on the basis of all the models.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

6.2 Boosting

This technique uses a group of classification algorithms that use weighted average to make a strong learner from weak learners. Boosting require bootstrapping. The dataset with more number of misclassifications are given higher weight because they are more complex. Then the next base learning algorithm is applied. Next base model depend on the previous one and combine them by following a deterministic strategy. Each model that runs instructs what features the next model will focus on. This is repeated till highest accuracy is achieved. Boosting try to produce a strong model with less bias than it's components. Adaptive Boosting, gradient Tree Boosting and XGBoost are the various techniques used as Boosting technique. Boosting technique builds ensemble model to decrease the bias.

6.3 RF

RF (Random Forest) was initially given by Leo Breiman in 2001. Random forest is an ensemble of bagging and random trees using voting technique. It is an enhancement of bagged decision tree. Bagging process is used to train the instances and multiple decision trees are generated. Data is randomly selected and trees are generated to fit the data set. After training process is completed all the generated tree will vote and class is predicted on the basis of majority voting [35]. Random Forest enhances the accuracy of classification and is very good for large datasets.

6.4 Stacking

Stacking is an ensemble technique that gives the way of combining multiple models of different types by using a meta classifier. The base learners are trained on complete data set. Meta classifier is then trained on the output of base classifiers as feature in the feature set. Base level algorithms are generally heterogeneous and stacking these algorithms combines the strength of all these base models. In most of the cases stacked model perform better than the best classifier in the base layer but training the stacking based classifier is computationally more expensive.

6.5 Voting

Voting is an ensemble base learning technique in which multiple machine learning classifiers classify instances by combined voting. In voting process, the first step is to create multiple base models using training dataset. Multiple base models can be created either by training a single classifier with multiple split on same training data set or by using multiple classifier with single training dataset. Each model will be allowed to vote on each class and the final selection can be done by various methods given below:

Majority Vote - Every model vote on the class it predicts for each test instance and the class having maximum numbers of votes is final class of instance. Say for an instance x, Ci is the class predicted by i_{th} classifier.
Than final class of the instance x is given by y as:

$$y = \text{mode}\ \{ C_1(x), C_2(x), C_3(x), \dots C_n(x) \} \quad (1)$$

Weighted Majority Vote: We can compute a weighted majority vote by associating a weight w_j with classifier C_j. The final predicted class of the instance x is given by y as:

$$y = \arg\max_i \sum_{j=1}^{m} \boldsymbol{wj}\ \mathrm{XA}\ (\mathrm{Cj(x) = i}) \quad (2)$$

where XA is characteristic function $[\mathrm{Cj(x) = i} \in A]$ and A is the set of class labels.

Simple Averaging: In this method, the final prediction for every instance in test dataset is done by taking average of all the predictions given by the base classifiers. This help to reduce the over fit.

Weighted Averaging: Weighted averaging is modified over simple averaging. In this weight is given to all classifier involved in ensemble as base classifier. Prediction of each model is multiplied by the weight assign to classifier and then their average is calculated.

7. PROPOSED FRAMEWORK

Proposed classification model is hybrid model based on ensemble of dictionary based classifier and stacked machine learning classifier. Three machine learning algorithms SVM, KNN and C5.0 are stacked to build an ensemble by using two Meta learners RF (Random forest) and GLM (Generalized linear model). The model in Figure 1 is implemented on manually labeled real time tweets based on "Clean India Mission" to get the sentiment analysis. After preprocessing and cleaning of data, duplicate tweets are removed. Training and testing data sets are labeled as positive and negative. After applying the dictionary based classifier, extracted sentiment score is added as one of the feature in feature set. Then proposed stacked based model is applied to get the classified tweets. The trained model is tested for 5 fold and 10 fold cross validation. Data is collected from Twitter as

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Twitter allows 1% of its data repository for research purpose. 10,000 tweets are downloaded with search keyword "Clean India Mission", and "Swatch Bharat Abhiyan" at different span of time. After preprocessing, cleaning and removing duplicate tweets, a data set of 1008 tweets is manually labeled for testing and training purpose.

The model is implemented using Caret library of R tool. Uni-gram BOW (Bag of Words) is used for the representation of tweets. Preprocessing technique is applied to remove Hashtags, URL's , Stop words, Conjunctions and Numbers. Stemming is performed to get the root words and reduce the dimensionality of feature set. Tweets are frequently re-tweeted, so duplicate tweets are removed to get reduced data set. Data set is represented in BOW using 2579 number of features. Sentiment score of dictionary based classifier calculated and is added in feature set. IG (Information Gain) is used as feature selection technique. Stacked ensemble model is built using C5.0, SVM and KNN machine learning algorithms. Two meta classifier RF and GLM provided in R library are used as base classifier for building stack model. The model is tested using 5 fold and 10 fold cross validation and repeated 10 numbers of times to validate the results. Accuracy, F-Score, Precision and Recall are used to represent and compare the results of classification. Results of proposed model are compared with different machine learning and ensemble classifiers.

Proposed hybrid Model is diagrammatically explained in Figure 1.

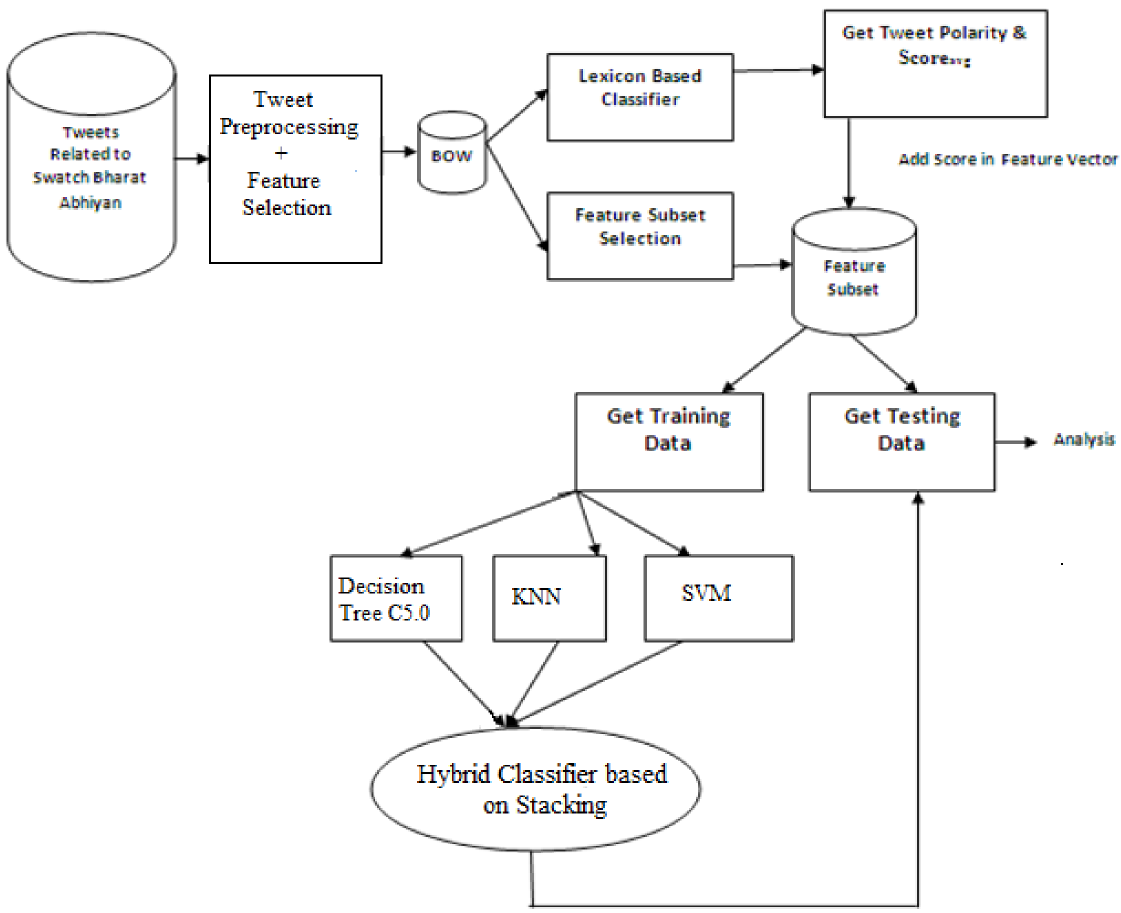

Figure 1: Proposed Hybrid Model Based on Stacked Machine classifiers (C5.0, KNN, SVM) and Dictionary Based Classifier for Twitter Sentiment Classification

Journal of Theoretical and Applied Information Technology
29th February 2020. Vol.98. No 04

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

8. IMPLEMENTATION AND RESULTS

In proposed model for tweet sentiment analysis hybrid approach based on stacking of machine learning algorithms is implemented with a little twist of adding outcome sentiment score of dictionary based approach as one of the feature in feature vector to facilitate more accurate sentiment prediction. Three machine learning algorithms SVM, C5.0 and KNN are stacked by using two Meta learners GLM and RF.

Initially dictionary based approach is applied on testing and training data set to get the sentiment score of tweets and accuracy of approach is recorded.

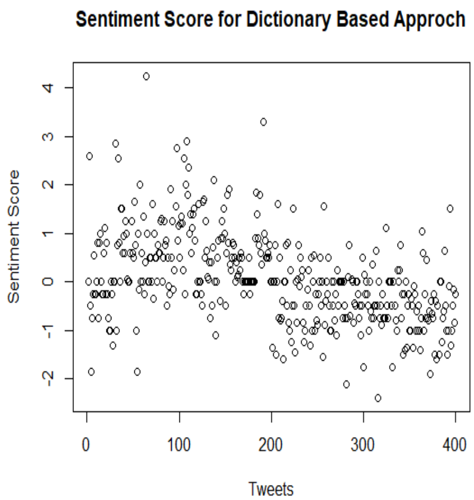

Figure 2: Sentiment Score for Dictionary Based Classifier for Testing Data Set (400 Tweets)

Figure 2 show the sentiment score of 400 real time tweets, which will be used as testing data set in machine learning algorithms. Dictionary based approach classify tweets with an accuracy of 74%. Figure 3 represents the sentiment score of 1008 tweets classified with an accuracy of 75.1% using dictionary based approach.

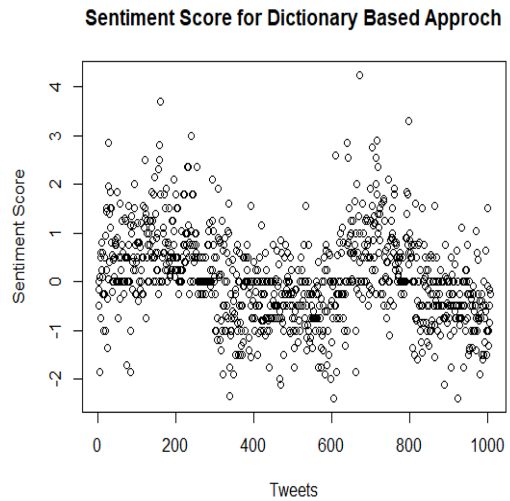

Figure 3: Sentiment Score for Dictionary Based Classifier for Testing + Training Data Set (1008 Tweets)

Table 1 shows the results of dictionary based classification. The sentiment predicted by dictionary based approach is compared against labeled data. The efficiency of dictionary based approach is shown in terms of Accuracy, Precision, Recall and F-Score.

Table 1: Performance of Dictionary Based Classifier

Classifier	**Accuracy**	**Precision**	**Recall**	**F-score**
Dictionary Based Approach (Testing Data) – 400 tweets	.74	.81	.71	.7567
Dictionary Based Approach (Complete Data Set) – 1008 tweets	0.7509921	.85	.706	.771

Table 2 show the efficiency of various classifiers in terms of accuracy, Precision, Recall, F-Score for unprocessed dataset. Figure 4 diagrammatically show the comparison of these classifiers for unprocessed data.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 2: Classifier Performance for Unprocessed Data

Classifier	Accuracy	Precision	Recall	F-score
SVM	0.87	0.889	0.845	0.8664
Naïve Bayes	0.4375	0.4643	0.815	0.5915
Tree	0.7475	0.72	0.81	0.762
Maximum Entropy	**0.875**	0.864	0.89	0.8768
Random Forest	0.855	0.838	0.88	0.8584
Boosting	0.8425	0.8044	0.905	0.8517
Bagging	0.84	0.806	0.895	0.8481

Table 3: Classifier Performance for Preprocessed Data

Classifier	Accuracy	Precision	Recall	F-score
SVM	**0.8825**	0.918	0.84	0.8772
Naïve Bayes	0.44	0.4664	0.835	0.5985
Tree	0.7575	0.7309	0.815	0.7706
Maximum Entropy	**0.8625**	0.8756	0.845	0.8600
Random Forest	**0.8675**	0.855	0.885	0.8697
Boosting	0.8375	0.8	0.9	0.847
Bagging	0.8275	0.799	0.875	0.835

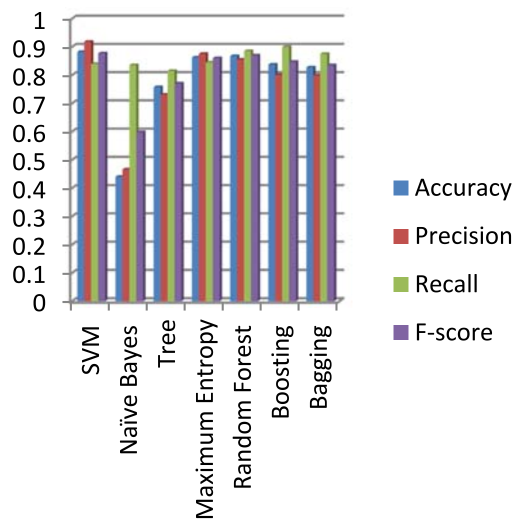

Figure 4: Comparative Analysis of Various Classifiers for Unprocessed data

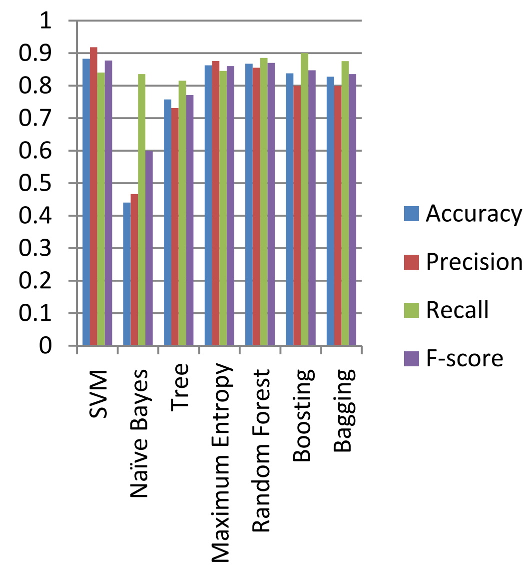

Figure 5: Comparative Analysis of Various Classifiers for Processed data set

In various researches it is observed that data preprocessing and cleaning of data reduce the dimensionality of data set without degrading the performance of most of the classifiers. Table 3 show the performance of various classifiers in terms of accuracy, Precision, Recall, F-Score for preprocessed data set. Figure 5 shows the comparative analysis of various classifiers for processed data set. Accuracy is improved with little margin after preprocessing and feature selection. So accuracy of classification is not degraded but dimensionality of data set is reduced.

SVM, Random Forest and Maximum Entropy outperformed among all the above classifiers. In general even after being a good classifier Naïve Bayes under performed in all the cases. The reason is the binary predictors are defined as numeric variables with values zero and one. Naïve Bayes implemented number of times with different parameters and different cross validation but the results were uniformly low.

By comparing the data in Table 1, Table 2 and Table 3 it is clear that dictionary based approach alone could not match the performance of machine learning based classifiers. But dictionary based approach classified the data with a good accuracy of 75%. This class labeling score done by dictionary based classification can be used in

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

machine learning approach to further increase the accuracy of machine learning classifiers. This idea is used in proposed model with stacking based ensemble approach.

In stacking based classification multiple algiorithms are used to design a more efficient classifer. Table 4 show the accuracy of ensemble approach based on stacking of machine learning classifiers- SVM, KNN, C5.0 and by using 5 fold and 10 fold cross validation. For stacking ensemble two meta learner algorithms GLM and RF are used. Table 4 also shoe the finding of proposed hybrid classifier. In proposed approach, prior to stacking, sentiment score of tweets are calculated by using dictionary based approach and this score is added as one of the feature in feature vector. After that ensemble based on stacking of SVM, KNN and C5.0 is applied for classification of data.

It is clear from the results in Table 4 that proposed approach performed better as compared to individual machine learning approaches and stacked based ensemble approaches with accuracy of **0.9066377** for 5 fold cross validation **and** **0.9124793** for 10 fold cross validation. In proposed approach among of two Meta learner RF and GLM, classification with meta learner RF performed better as compared to GLM with an accuracy of **0.9124793** in 10 fold cross validation .

In individual machine learning classifiers SVM performed best as compared to other machine learning algorithms with accuracy 0.8825, but not as good as proposed approach.

Three classifier SVM, KNN, C5.0 are stacked to build ensemble classifier. If NB (Naïve Bayes) is also added in stacked base classifier than results are not improved. So adding NB as extra base classifier in stacked ensemble only increases the overhead but not the accuracy.

The results in Table 4 show that adding of additional information from dictionary classifier in proposed model improves the accuracy in all the cases of stacked ensemble for both Meta learners RF, GLM and for all the cross validations. Highest accuracy achieved by proposed model is **0.9124793** in case of RF Meta classifier and 10 fold cross validation.

Figure 6 Show the comparative diagrammatic analysis of various best performing classification techniques based on stacked based ensemble and the proposed hybrid model.

Table 4: Accuracy of Stack Based Ensemble Classification Models and Proposed Hybrid Model

Ensemble Technique	**Meta Learner**	**5 Fold Cross Validation Accuracy**	**10 Fold Cross Validation Accuracy**
Stack Based Ensemble using SVMRadial, KNN, **NB**, C5.0	GLM	0.8334436	0.8443563
Stack Based Ensemble using SVMRadial, KNN, **NB**, C5.0	RF	0.8544974	**0.8664021**
Stack Based Ensemble using SVMRadial, KNN, C5.0	GLM	0.8334436	0.8443563
Stack Based Ensemble using SVMRadial, KNN, C5.0	RF	0.8526235	**0.8667328**
Proposed Hybrid Approach			
Dictionary Based Classifier + Stack Based Ensemble using KNN, SVMRadial, C5.0	GLM	**0.8973784**	**0.9081809**
Dictionary Based Classifier + Stack Based Ensemble using KNN, SVMRadial, C5.0	RF	**0.9066377**	**0.9124793**

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

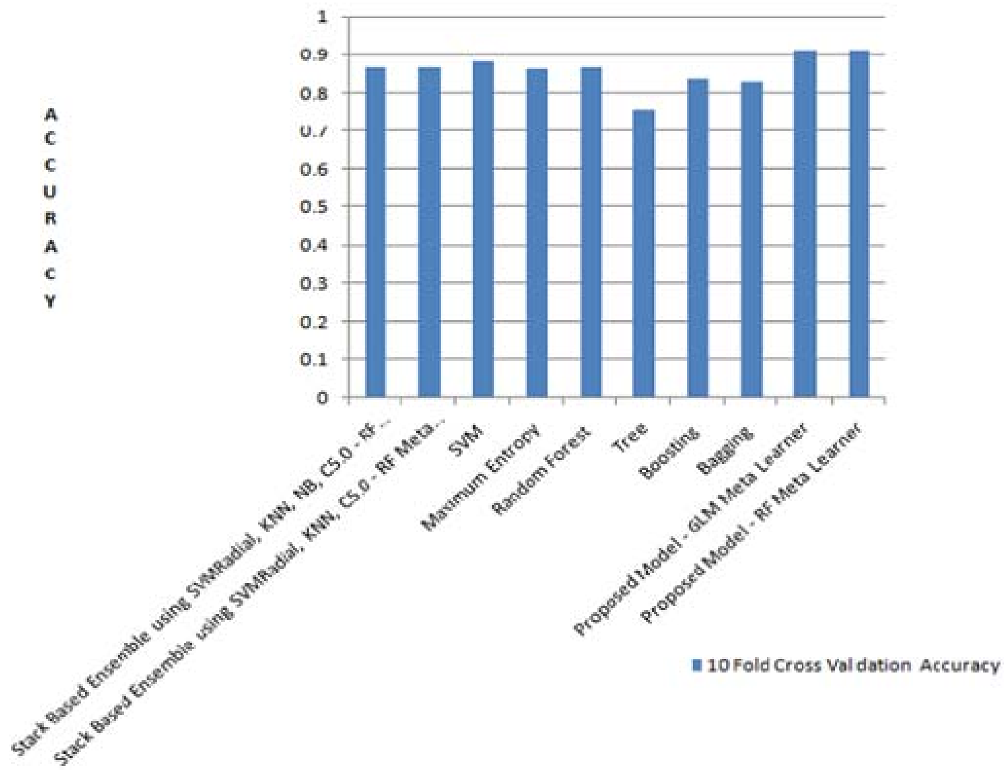

Figure 6: Comparative Analysis of Machine Learning Classifiers, Stack Based Ensemble Models and Proposed Hybrid Model

9. CONCLUSION

Twitter sentiment analysis is very helpful in decision support and policy making. In this work a model is proposed for twitter sentiment analysis to enhance the accuracy of classification. This model is an ensemble of dictionary based approach and stacked machine learning classifier by using GLM and RF as Meta classifier. Three machine learning classifiers SVM, KNN and C5.0 are used to build stacked ensemble classifier. Sentiment Score of dictionary based approach is added as one of the feature in feature set.

Out of individual machine learning classifiers SVM, Maximum Entropy, Random Forest outperformed with an accuracy of **0.8825, 0.8652 and 0.8675**. Proposed model further enhance the performance with an accuracy of **0.9066377** for 5 fold cross validation and **0.9124793** for 10 fold cross validation by using RF Meta classifier. RF Meta classifier performed better as compared to GLM Meta classifier in all the scenarios. Overall adding of sentiment score as additional feature was able to enhance accuracy of stack based ensemble for both RF and GLM Meta classifiers. The variations of the proposed model can also be implemented with stacking of different combinations of machine learning classifiers to observe any increase in accuracy and variations in results. The research contribute to enhance the accuracy of twitter data sentiment analysis by using additional information provided by dictionary based classifier with stack base ensemble of machine learning classifiers.

REFERENCES:

[1] A. B. Xie, I. Vovsha, O. Rambow, R. Passonneau, "Sentiment Analysis of Twitter Data", In Proceedings of the workshop on Languages in Social Media, Columbia University, New York, 2011, pp. 30-38.

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

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[3] S. Maheshwari, S. Shukla, D. Kumari, "Twitter Opinion Mining Using Sentiment Analysis", World Scientific News, An International Scientific Journal, 2019, pp. 78-87.

[4] A. R. Alaei, S. Becken, B. Stantic, "Sentiment Analysis in Tourism: Capitalizing on Big Data", Journal of Travel Research, vol. 58(2), Feb 2019, pp. 175-191.

[5] M. Noori, and J. Maktoubian, "Business Improvement Approach Based on Sentiment Twitter Analysis: Case Study", in EAI Endorsed Transactions on Cloud Systems, Apr 2019, pp. 1-7.

[6] M. M. Müller, M. Salathé, "Crowd breaks: Tracking Health Trends Using Public Social Media Data and Crowd sourcing.", Frontiers in public health, vol. 7:81, 12 Apr. 2019.

[7] E. Eryarsoy, D. Delen, "Predicting the Outcome of a Football Game: A Comparative Analysis of Single and Ensemble **Analytics** Methods", Proceedings of the 52nd Hawaii International Conference on System Sciences, Jan 2019, pp. 1107-1115.

[8] L. Terán, J. Mancera, "Dynamic Profiles using Sentiment Analysis and Twitter Data for Voting Advice Applications", in Government Information Quarterly, ELSEVIER, vol. 36, Issue 3, July 2019, pp. 520-535.

[9] A. Rasool, R. Tao1, K. Marjan, T. Naveed, "Twitter Sentiment Analysis: A Case Study for Apparel Brands", IOP Conf. Series: Journal of Physics, vol. 1176, 2019, pp. 1-7.

[10] V. Curtis, "Explaining the outcomes of the 'Clean India' campaign: institutional behavior and sanitation transformation in India", BMJ Global Health Journals, vol. 4 (5), 2019, pp. 1-11. Available: http://dx.doi.org/10.1136/bmjgh-2019-001892

[11] V. Rekha, R. Raksha, P. Patil, N. Swaras, G. L. Rajat, "Sentiment Analysis on Indian Government Schemes Using Twitter data", International Conference on Data Science and Communication (IconDSC), IEEE, Aug 2019, Available:10.1109/IconDSC.2019.8817036.

[12] D. Ray, "Lexicon Based Sentiment Analysis of Twitter Data" , International Journal for Research in Applied Science & Engineering Technology (IJRASET), vol. 5(X), Oct 2017.

[13] O. Kolchyna, Tharsis T. P. Souza, P. Treleaven, T. Aste, "Twitter Sentiment Analysis: Lexicon Method, Machine Learning Method and Their Combination", Cornell University Library, arXiv :1507.00955, 18 Sep 2015.

[14] T. S. Raghavendra, K. G. Mohan, "Web Mining and Minimization Framework Design on Sentimental Analysis for Social Tweets using Machine Learning", Procedia Computer Science, Science Direct, vol. 152, 2019, pp. 230-235.

[15] C. A. Calderón , F. O. Mohedano, M. Álvarez, M. V. Mariño, "Distributed Supervised Sentiment Analysis of Tweets: Integrating Machine Learning and Streaming Analytics for Big Data Challenges in Communication and Audience Research", EMPIRIA. Revista de Metodología de Ciencias Sociales, vol. 42, 2019, pp. 113-136.

[16] K. Lakshmi Devi, P. Subathra, P. N. Kumar, "Tweet Sentiment Classification Using an Ensemble of Machine Learning Supervised Classifiers Employing Statistical Feature Selection Methods", Proceedings of the Fifth International Conference on "Fuzzy and Neuro Computing (FANCCO - 2015)", vol. 415, Nov 2015, pp. 1-13.

[17] S. Kurnaz, M. A. Mahmood, "Sentiment Analysis in Data of Twitter using Machine Learning Algorithms", International Journal of Computer science and Mobile Computing, vol. 8(3), March 2019, pp .31 – 35.

[18] N. Rajput, S. Chauhan, "Analysis of Various Sentiment Analysis Techniques", International Journal of Computer Science and Mobile Computing, vol.8(2), Feb 2019, pp. 75-79.

[19] K. Korovkinas, P. Danenas, G. Garsva, "SVM and k-Means Hybrid Method for Textual Data Sentiment Analysis", Baltic J. Modern Computing, vol. 7 (1), 2019, pp. 47–60.

[20] D. Mohey E. M. Hussein, "A survey on Sentiment Analysis Challenges", Journal of King Saud University – Engineering Sciences, vol. 30(4), Oct 2018, pp. 330-338.

[21] 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, 2019, pp. 506-513.

[22] J. D. Prusa, T. M. Khoshgoftaar, D. J. Dittman, "Impact of Feature Selection Techniques for Tweet Sentiment Classification", Proceedings of the Twenty-Eighth International Florida Artificial

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Intelligence Research Society Conference, 2015 , pp. 299-304.

[23] R. Ahujaa, A. Chuga, S. Kohlia, S. Guptaa, and P. Ahujaa, "The Impact of Features Extraction on the Sentiment Analysis", International Conference on Pervasive Computing Advances and Applications – PerCAA 2019, Procedia Computer Science, vol. 152, 2019, pp. 341–348.

[24] M. bibi, M. S. A. Nadeem, I. H. khan, S. Shim, I. R. Khan, U. Naqvi1 and W. Aziz, "Class Association And Attribute Relevancy Based Imputation Algorithm To Reduce Twitter Data For Optimal Sentiment Analysis", vol. 7 , Oct 2019, pp. 1365535-1365544.

[25] A. Jović, K. Brkić, N. Bogunović, "A review of feature selection methods with applications", in international convention on information and communication technology electronics and microelectronics, ISBN- 978-953-233-083-0, 2015, pp. 1447-1452.

[26] M. Naz ,K. Zafar, A. Khan, "Ensemble Based Classification of Sentiments Using Forest Optimization Algorithm", vol. 4(2), May 2019, pp. 1-13.

[27] M. M. Fouad, T. F. Gharib, A. S. Mashat, "Efficient Twitter Sentiment Analysis System with Feature Selection and Classifier Ensemble", in International conference on Advanced Machine Learning and Applications, vol. 723, Jan 2018, pp. 517-527, DO - 10.1007/978-3-319-74690-6_51.

[28] M. Ghiassi, J. Skinner, D.Zimbra, "Twitter brand sentiment analysis: A hybrid system using *n*-gram analysis and dynamic artificial neural network", in Expert System with Applications , ScienceDirect, vol. 30(16), Nov 2013, pp. 6266-6282.

[29] J. J. Bird, A. Ekárt, C. D. Buckingham, D. R. Faria, "High Resolution Sentiment Analysis by Ensemble Classification", In: Arai K., Bhatia R., Kapoor S. (eds) Intelligent Computing. CompCom 2019. Advances in Intelligent Systems and Computing, Springer, Cham, vol. 997, June 2019.

[30] A. Krouska, C. Troussas, M. Virvou, "The effect of preprocessing techniques on Twitter sentiment analysis", ResearchGate Conference, July 2016, DOI: 10.1109/IISA.2016.7785373.

[31] Z. Jianqiang and G. Xiaolin, "Comparison Research on Text Pre-processing Methods on Twitter Sentiment Analysis", IEEE Access, Feb 2017.

[32] E. Haddia, X. Liua, Y. Shib, "The Role of Text Pre-processing in Sentiment Analysis" , in the International Conference on Information Technology and Quantitative Management, Elsevier, Procedia Computer Science, Dec 2013.

[33] R. Mansour, M. F. A. Hady, E. Hosam, H. Amr, and A. Ashour, "Feature Selection for Twitter Sentiment Analysis: An Experimental Study", International Conference on Intelligent Text Processing and Computational Linguistics, Springer, 2015, pp. 92-103.

[34] L. Bbeiman, "Bagging Predictors", Machine Learning, Kluwer Academic Publishers. Boston, 24, 1996, pp. 123-140.

[35] T. K. Ho, "Random decision forests," in Document analysis and recognition, In the proceedings of the third international conference on Document Analysis and Recognition, IEEE, vol. 1, 1995, pp. 278–282.

[36] M. Bouazizi and T. Ohtsuki, "Multi-Class Sentiment Analysis on Twitter: Classification Performance and Challenges", IEEE Xplore, Big Data Mining and Analytics, ISSN 2096-0654 03/05, DOI: 10.26599/BDMA.2019.9020002, vol. 2(3), Sep 2019 pp.181–194.