International Journal of Engineering & Technology, 7 (4) (2018) 4737-4742

SPC

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET
doi: 10.14419/ijet. v7i4.9275
Research paper



Machine learning based twitter data sentiment classification on real time 'clean India mission' tweets

Sangeeta [1] *, Nasib Singh Gill [1]

[1] Department of Computer Science & Applications M D University, Rohtak, Haryana, India
**Corresponding author E-mail: Sangeeta.yogi@gmail.com*

Abstract

Social networking sites are popular medium to share opinion on various topics. Twitter is one of the social networking site used for tweet posting. Sentiment classification deal with finding the polarity of these tweets as positive, negative or neutral. This analysis can be useful in decision support in different ways. The aim of this paper is to discuss various machine learning algorithm for twitter sentiment classification, compare their results on the basis of Accuracy, Precision, Recall and F-Score. A real time data set for training and testing is created after extraction and cleaning of tweets on "Clean India Mission or Swachh Bharat Abhiyan". "Clean India Mission" is a campaign by government of India to clean the country. This paper also compares machine learning algorithms with Bagging, Boosting and Random Forest ensemble approaches.

Keywords: *SVM; Maxent; RF; Naïve Bayes; Clean India Mission; Swachh Bharat Abhiyan.*

1. Introduction

Twitter is one of the social networking sites where people post their views about various topics in a precise way. This large set of information after proper analysis can be used in different areas like in Business, Recommender system, Advertisement, Politics and many more.

Twitter sentiment analysis is the process of detecting the contextual polarity of twitter posts. In other words, it determines whether a tweet of writing is positive, negative or neutral. Sentiment analysis is used to discover how people feel about a particular topic. Tweets are precise in length of maximum 140 characters and generally people do not change their opinion in singe tweet, so analysis can be very accurate and useful. Use of acronyms, emoticons, hash tags makes it little trickier. Words are frequently misspelled that effect the feature set. Sentiment analysis can be done at word level, sentence level, document level or feature level. But as the length of tweet is limited so word level is appropriate for tweet sentiment analysis.

Various approaches like Lexicon based, Machine learning and Ensemble approach [1, 3] are implemented for tweet sentiment classification. In lexicon based approach [3] predefined polarity dictionary like WordNet and SeniwordNet are used to find the polarity of features used in tweets. Average polarity score of tweet is calculated. Average polarity score above a threshold value decide positive polarity and below threshold value denotes negative polarity of tweet. This approach is very simple but due to common dictionary this is not very accurate. The second approach is machine learning approach [1], [4], [5] which use training data to train model using algorithms like SVM, Naïve Bayes, Decision tree and Maximum entropy classifier. Model is tested on testing data for accuracy check. This approach perform better than lexicon approach. The third approach is an ensemble approach [2], [3], [6] to enhance the accuracy of model. Bagging, Boosting, Random Forest are few ensemble based approaches used and give good results. Lexicon based approach can also be ensemble with machine learning approach to increase classification accuracy.

In machine learning approach three different approaches Supervised, Semi-supervise and unsupervised are followed. SVM [1, 8], Naïve Bayes [1], Maximum entropy comes under supervised approaches. There approaches highly depend upon good quality domain specific training set. If training set is not appropriate then results are not up to the mark. If trained properly these algorithms provide good results. Semi-supervise approach only require partially labeled data. Transductive and Inductive two kind of learning are applied in this approach. Self-Organizing Maps, K-Means are unsupervised approaches and does not require training set. There approaches are appropriate to apply in case unavailability or higher cost of getting training data.

Choice of feature generation technique [6] and number of feature selected highly effect the accuracy of model used. So feature generation and feature selection techniques are very important while finding high accuracy classification model. Chi Squared (CS), Gini-Index (GI), Kolmogorov-Smirnov (KS) statistic, Mutual Information (MI), Probability Ratio (PR) are few of the commonly used feature selection methods.

The present paper discuss about various machine learning approaches for sentiment classification like SVM, Naïve Bayes, Decision tree, and Maximum entropy classifiers. All these algorithm are simulated for real time "Swachh Bharat Abhiyan" data tweets. 'Clean India Mission' is campaign by government of India that aims to clean up every village and town of India. Common people are attached to the mission and have their opinion on success or failure of mission. In present paper 5000 tweets related to "swachh Bharat Abhiyan" are collected and Preprocessed for opinion mining. After preprocessing, cleaning, Removal of duplicates and Removal of retweets, two corpus having 527 and 1008 tweets are created. Also ensemble approaches Bagging, Boosting, Random Forest are tested for same set of data to analyze the classification accuracy.

Copyright © 2018 Sangeeta, Nasib Singh Gill. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The paper is organized as follows. In section 1, the introduction to twitter sentiment analysis and different approaches to twitter sentiment analysis is given. In section 2, related work is surveyed and summarized. Section 3 explains the machine learning approach for twitter data classification. In section 4, tweets preprocessing and cleaning of data is explained. Section 5, mention the details of real time corpus used in implementation. In Section 6, results of the implementation are presented and compared. Section [7], give conclusion and insight to further scope.

2. Related work

The increase in people interest in social networks has made sentiment analysis a popular research area. Several researchers have done lots of work in different areas of twitter sentiment classification to improve the result of classification in terms of accuracy and efficiency. Machine learning approach is an important method of sentiment classification and explored by several researchers.

Bac Le et al. [1] proposes a model based on Naive Bayes and Support Vector Machine. Information Gain, Bigram, Object-oriented extraction methods are used for feature selection. This model proved to be highly effective and accurate on the analysis of feelings.

Prusa et al. [2] worked on two popular ensemble techniques bagging and boosting. Both bagging and boosting are tested using seven different base learners. Additionally performance of ensemble techniques by using each of the base learners is compared with no ensemble technique. The resulting 21 learning algorithms are trained and tested on two datasets, a large automatically labeled lower quality dataset and a small manually labeled high quality dataset. The research shows that ensemble learners perform better for data base of every quality.

Abinash Tripathy [4] implemented four machine learning algorithms SVM, Naïve Bayes, Maximum Entropy and Stochastic Gradient Descent (SGD) on IMDb data set for sentiment classification. Precision, recall, f-measure, and accuracy are used to analyze their performance. These classification models are implemented on unigram, bigram and n-gram features and it is observed that if value of n is increased in n-gram after 2 than accuracy is decreased rather than increasing. For unigram and bigram accuracy is good but for trigram, four-gram, five-gram accuracy is decreased. Also the use of TF- IDF and Count Vectorizer techniques for converting the text into matrix of numbers improve the accuracy of classification.

Agarwal et al. [7] selected different features as unigrams, bigrams, n grams, has tags and performed comparative analysis of their results. By varying the number of features used and different feature selection techniques like information gain, chi-square they show the variation in classification accuracy.

Munir Ahmad [8] implemented Support Vector Machine (SVM) on two pre-classified data sets of tweets for textual polarity detection. Precision, Recall and F-Measure are used for comparative analysis. The result show that performance of SVM depend on dataset. So it can be an area of research that what kind of classification algorithm is good for which kind of data set and what is the reason for that.

Bholane Savita Dattu [9] implemented SVM and Naive Bayes real time downloaded tweets. They pointed out in their research that SVM gives best results than Naive Bayes algorithm in case of text categorization. Also Naive Bayes classifier is insensitive to unbalanced data which give more accurate results.

Nádia et al. [10] proposed a semi supervised approach to solve the problem of cost of getting supervised data. Similarity matrix which is a powerful knowledge-discovery tool to get information from unlabeled data is constructed from raw classless data. This unsupervised information captured from similarity matrix is used with a classifier for sentiment classification. Experimental results demonstrate that the proposed framework can improve the accuracy of tweet sentiment analysis for unlabeled data.

Patil [11] stated that SVM eliminated the need of feature selection due to the ability to generalize high dimension feature space.

Pak and Paroubek [12] proposed a model to classify the tweets as positive, negative and neutral using multinomial Naive Bayes classification model by using features N-gram and POS-tags. They collected real time tweets using twitter API and classify these tweets automatically using emoticons. The training set used was only based on tweets having emoticons.

Subbulakshmi et al. [13] implemented Bootstrap Ensemble approach using with and without equal voting on corpus Sentiment140's. BOG method was used for feature selection. Yun Wan et al. [26] implemented ensemble approach on airline data using 10 fold cross validation. In this research, an ensemble of Naïve Bayes, Decision Tree 4.5, and SVM is implemented using Majority Vote principle.

In most of the research different machine learning algorithm like Naïve Bayes, SVM, Maximum Entropy, KNN are implemented. Some are implemented on unigram feature vector other on bigram or n-gram. Different feature selection techniques are used. Some of the research are implemented on predefined corpus and other use real time classifies data for their research. Three different machine learning approach supervised, unsupervised and semi supervised are followed. Here in this paper we are following supervised approach as we are using properly labeled data.

3. Machine learning approach

In supervised machine learning approach [1, 3, 5] the main task is to train the classifier. Classifier need training examples which can be labeled manually or obtained from online sources. Support Vector Machines (SVM), Naive Bayes, Decision Trees, Maximum Entropy Algorithm, AdaBoost, Regression method (Logistic Regression LR), J48, Simple CART, Random Tree are some commonly used machine learning based classifiers. To implement machine learning approach first classifier algorithm is selected. Number of features and feature selection technique is decided. Next the classifier is trained with labeled training data. Then the model is tested on labeled testing data set for accuracy. It is observed that machine learning approach performs better than lexicon based approach.

After getting data it is very important to preprocess and clean the data to get appropriate feature set. Unprocessed data always provide undermined results as compared to clean data due to lots of useless features. Feature selection highly effect the result of classification accuracy. Number of feature selected also effect the result. Increasing number of features after a limit does not increase the classification accuracy.

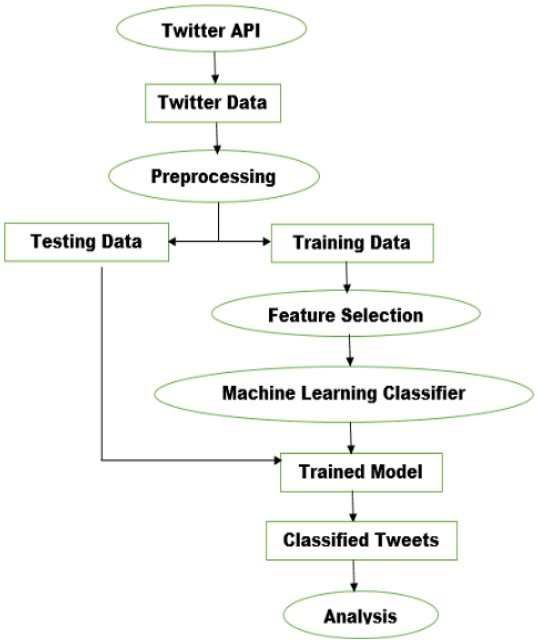


Fig. 1: Machine Learning Based Twitter Data Sentiment Classification.

F-Score, Accuracy, Recall and Precision are used as for analysis of classification.

4. Tweet preprocessing

Downloaded real time tweets contains lots of noise like URL's, Hashtags, Stop words, punctuation and irrelevant content that need to be removed before further processing to get better results [7,8] . Lots of other information is also downloaded with tweets like account holder name, longitude, latitude, Retweet or not, tweet id, date etc. Preprocessing of content is required to make the tweets suitable for sentiment analysis.

Some basic steps in preprocessing are mentioned as:

- Removal of re-tweets: It is very important as same tweets are re-tweeted multiple time. It is useful to include tweet just once.
- Tokenization: Tokenization is a process of creating a BOG from tweets. Unigram, Bigram or N-gram can be used as feature for sentiment classification. Here in this research unigram approach is followed.
- Twitter feature removal: User names and URLs, date, Longitude, Latitude are not important from the perspective of sentiment analysis, hence their presence is futile. All these features can be removed while preprocessing data.
- Special character and digit removal: Digits and special characters do not convey any sentiment. Sometimes they are mixed with words; hence, their removal can help in associating two words that were otherwise considered different.
- Stemming: Stemming is a procedure of replacing words with their roots like replacing "going" with "go". This helps in reducing the dimensionality of the feature set.
- Stop-words removal: Stop words like "the", "is", "at", "which", "on", "and", "or", "still" that do not affect the meaning of the tweet are removed tom increase efficiency of classification.
- Negations handling: Negation revert the sentiment of the tweet by using words like "no", "not", "don't" etc.. The simplest approach to handle negation is to revert the polarity of all words that are found between the negation and the first punctuation mark following it.
- But-clauses: The phrases like "but", "except that", "except for" generally change the polarity of the part of the sentence following them. In order to handle these clauses the polarity of the text before and after these phrases should be set opposite to each other.
- Converting upper case to lower case: It is better to convert all tweets in same case. This will further help to reduce the dimensionality of feature set.

5. Data set

Real time tweets related to keywords "Swachh Bharat Abhiyan or "Clean India Mission" are downloaded. Twitter allows 1% percent of data to be downloaded for research purpose. After preprocessing, cleaning, removal of duplicates and retweets, Corpus-1 having training and testing data set of 592 tweets is created. Another Corpus-2 having training and testing data set of 1008 tweets is created. Table 1 show the details of data set used for analysis. Classification is done on both the data set for comparative analysis.

Table 1: Data Set Description

Swachh Bharat Abhiyan Corpus-1 - 527			
Training Data - 271		Testing Data - 256	
Positive	151	Positive	136
Negative	120	Negative	120
Swachh Bharat Abhiyan Corpus-2 - 1008			
Training Data - 608		Testing Data - 400	
Positive	300	Positive	200
Negative	308	Negative	200

6. Implementation and result

Machine learning based classification is implemented using different models as SVM, Naïve Bayes, Maximum entropy and Decision Tree. Ensemble techniques as Bagging, Boosting and Random Forest are also implemented on same set of training and testing data. Table 2 below show precision, Recall and F-Score of different classification model in corpus 1. Results show that in machine learning classifiers SVM performs best from others. Naïve Bayes does not perform well due to class imbalance problem and common set of features for different classes. Ensemble classifier perform better as compared to other machine learning classifier other than SVM.

Table 2: Precision Recall and F-Score on Corpus-1

Classification Method	Precision	Recall	F-Score
Machine Learning Classifiers			
SVM	.81	.85	.83
Naïve Bayes	.49	.38	.49
MAXENT	.70	.85	.77
TREE	.70	.72	.71
Ensemble Classifiers			
RF	.73	.89	.80
Boosting	.68	.87	.76
Bagging	.70	.86	.68

Fig-2, Fig-3 and Fig-4 shows comparative analysis for classification models based on Precision, Recall and F-Score for Corpus-1.

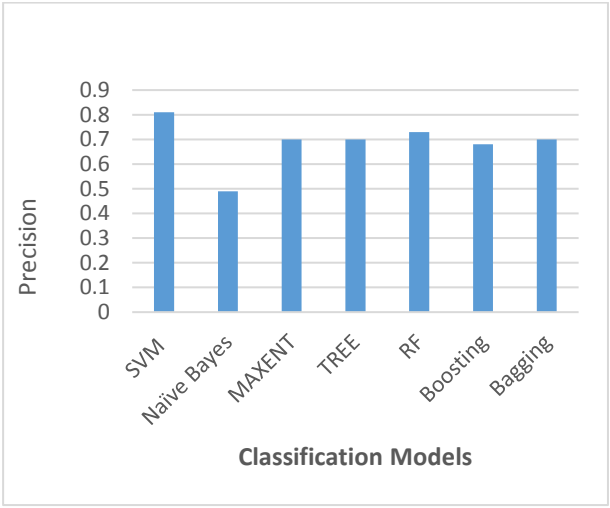


Fig. 2: Comparative Analysis for Classification Models based on Precision for Corpus-1.

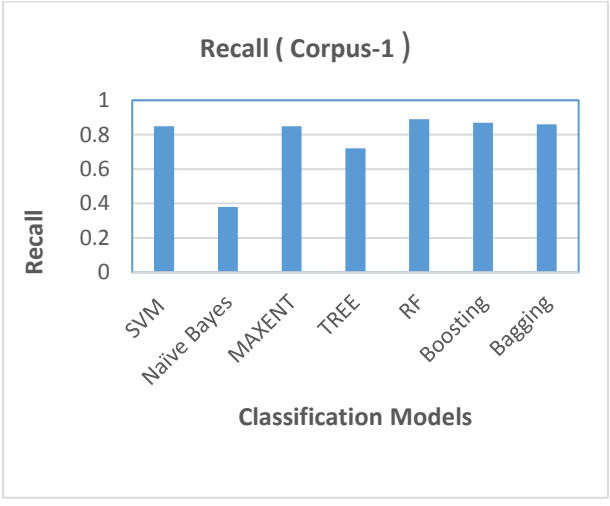


Fig. 3: Comparative Analysis for Classification Models Based on Recall for Corpus-1.

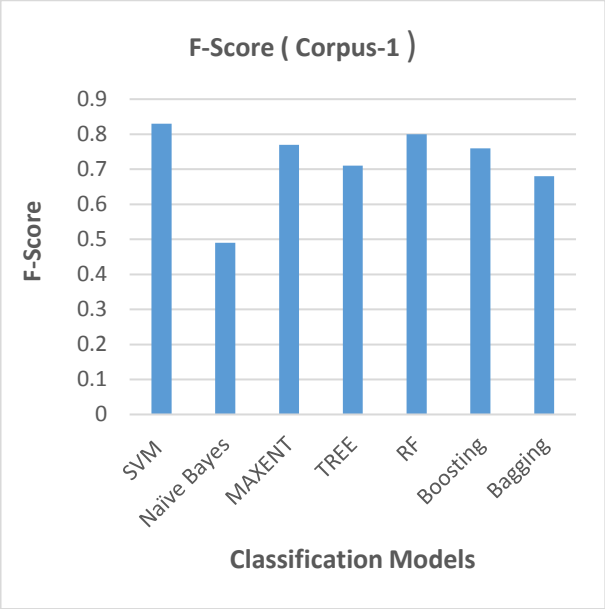


Fig. 4: Comparative Analysis for Classification Models based on F-Score for Corpus-1.

The models are tested for 3 fold, 5 fold and 10 fold cross validation. The table 3 show the comparison of all algorithms on basis of their accuracy for 3 fold, 5 fold and 10 fold cross validation. In most of the cases as value of n is increased in n-fold, results are improved. 10 fold cross validation show better results than others.

Table 3: Accuracy for different Cross Validation on 'Swachh Bharat Abhiyan' on Corpus-1

Classification Method	Accuracy		
	3 fold cross validation	5 fold cross validation	10 fold cross validation
Machine Learning Classifiers			
SVM	0.817679	0.858695	0.92
Naïve Bayes	0.460937	0.543478	0..5632
MAXENT	0.824120	0.892156	0.9
TREE	0.744186	0.776595	0.745454
Ensemble Classifiers			
RF	0.844444	0.788461	0.884615
Boosting	0.897590	0.892857	0.965517
Bagging	0.753246	0.792452	0.866666

The above results show that SVM, Maximum entropy and Boosting show the best results as compared to other algorithms. Bagging and Boosting show remarkable increase inaccuracy if number of folds are increased. These results are for real time data set and may vary from standard corpus.

Fig. 5 show the comparative analysis of Machine learning and ensemble classifier based on Accuracy for corpus-1. [10] fold cross validation show better results as compared to [3] fold and [5] fold cross validation.

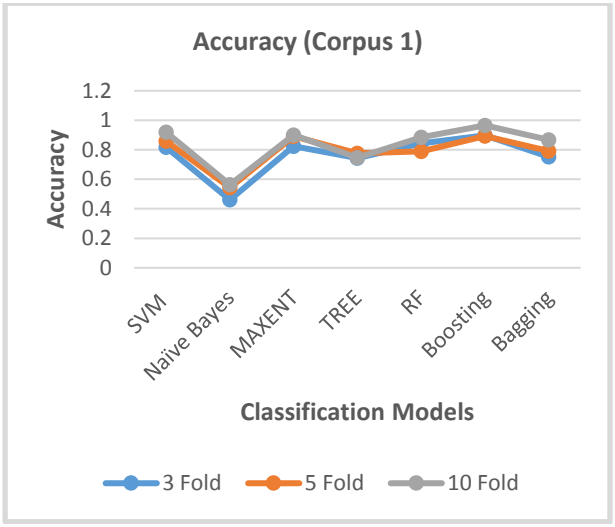


Fig. 5: Comparative Analysis for Classification Models based on Accuracy for Corpus-1.

The same set of algorithms is implemented on corpus-2 and results are compared if any change in performance with increase in number of training data. Results are improved with large set of training data.

Table 4: Precision Recall and F-Score on Corpus-2

Classification Method	Precision	Recall	F-Score
Machine Learning Classifiers			
SVM	0.885	0.885	0.885
Naïve Bayes	0.46	0.44	0.4497
MAXENT	0.865	0.860	0.860
TREE	0.760	0.760	0.755
Ensemble Classifiers			
RF	0.880	0.875	0.875
Boosting	0.845	0.840	0.840
Bagging	0.830	0.830	0.830

Table 4 show the results of comparative analysis based on Precision, Recall and F-Score for corpus-2.

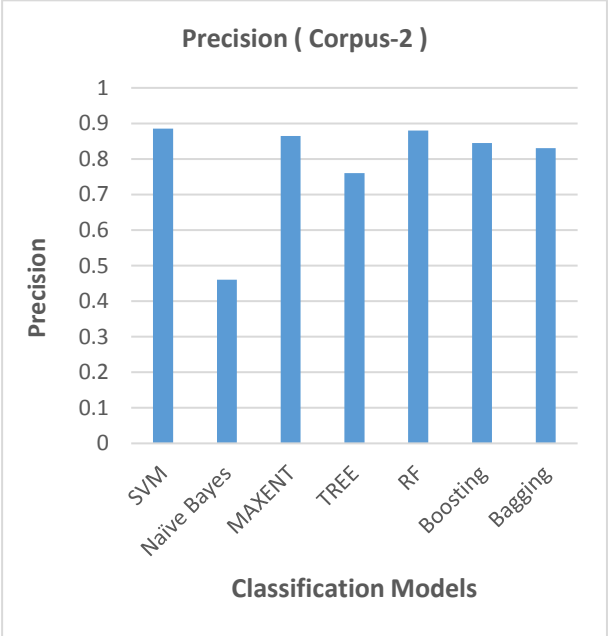


Fig. 6: Comparative Analysis for Classification Models Based on Precision for Corpus-2.

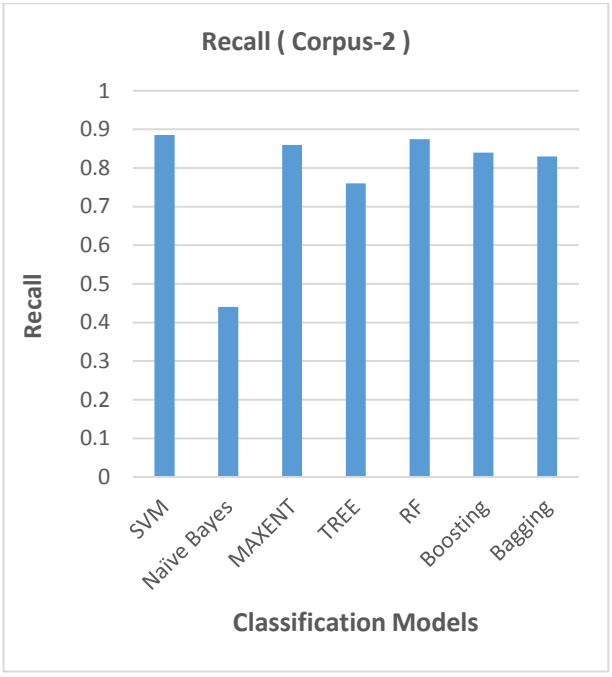


Fig. 7: Comparative Analysis for Classification Models Based on Recall for Corpus-2.

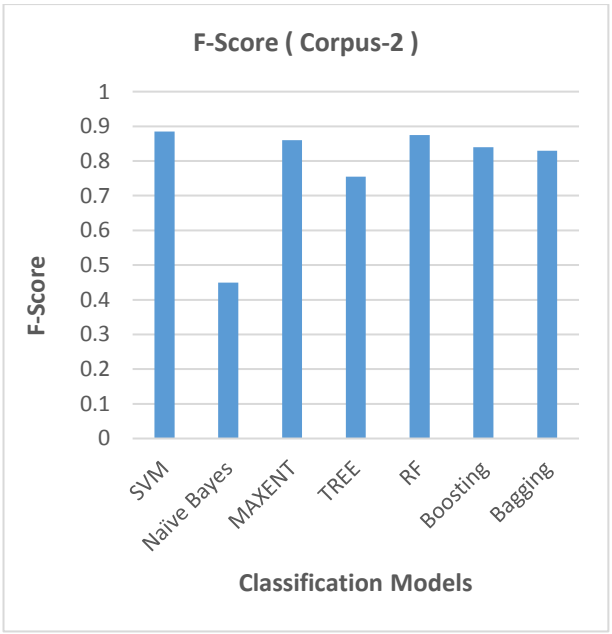


Fig. 8: Comparative Analysis for Classification Models Based on F-Score for Corpus-2.

Fig-6, Fig-7 and Fig-8 shows comparative analysis for classification models based on Precision, Recall and F-Score for Corpus-2.

Table 5: Accuracy for different Cross Validation on 'Swachh Bharat Abhiyan' on Corpus-2

Classification Method	Accuracy		
	3 fold cross validation	5 fold cross validation	10 fold cross validation
Machine Learning Classifiers			
SVM	0.896656	0.898395	0.871287
Naïve Bayes	0.44	0.4783	0.4652
MAXENT	0.912676	0.9162304	0.93
TREE	0.722713	0.740566	0.726495
Ensemble Classifiers			
RF	0.847761	0.831683	0.873786
Boosting	0.914373	0.972973	0.952381
Bagging	0.815047	0.821256	0.828571

The models are again tested for corpus-2 for 3 fold, 5 fold and 10 fold cross validation. Table 5 Show the comparison of all algorithms for corpus-2. These results show in increase in accuracy of results.

Fig. 9 show the comparative analysis of Machine learning and ensemble classifier based on Accuracy for corpus-2.

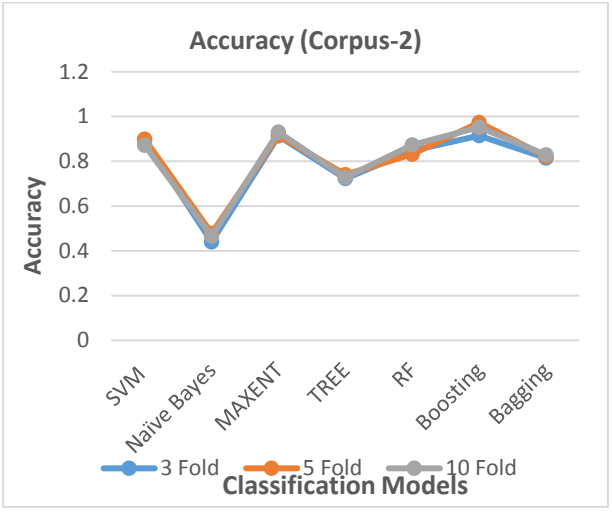


Fig. 9: Comparative Analysis for Classification Models Based on Accuracy for Corpus-2.

The results show that ensemble approaches perform better than machine learning approaches. Performance is increased with increase in number of folds and size of training data. Naïve Bayes performance is worst as compared to other algorithms. In machine learning SVM and Maximum entropy performs very well.

7. Conclusion

This paper presents sentiment analysis on real time twitter data. Sentiment analysis of twitter is very important as can be helpful in decision support in various areas. Real time tweets are downloaded, preprocessed and two data set of size 527 and 1008 are created. Machine learning algorithms SVM, Naïve Bayes, Decision Tree and Maximum Entropy are implemented on both data sets. In Machine learning approaches SVM Performs better as compared to others. Naïve Bayes does not perform well due to lack of independence of the explanatory feature variables or due to large set of common features in testing and training data. Also decision tree performance was poor in all cases as compared to other algorithms. Ensemble techniques Boosting, Bagging and Random Forest performed better than machine learning approaches. Random Forest give best results. In general results are improves as value of n is increased in n-fold cross validation. Also with increase in size of training data results are improved. In this paper unigram features are used. In further approach n-gram features can be used to increase result accuracy. It is observe that lots of tweets related to 'Clean India Mission or Swachh Bharat Abhiyan' does not hold any contextual polarity, so better to include neutral class. The paper proposes an ensemble technique based on lexicon based approach and machine learning approaches. Sentiment score of features retrieved from lexicon approach can be treated as a feature in machine learning approaches.

References

[1] BAC Le, and Huy Nguyen, "Twitter Sentiment Analysis Using Machine Learning Techniques", Advanced Computational Methods for Knowledge Engineering, Advances in Intelligent Systems and Computing 358, DOI: 10.1007/978-3-319-17996-4_25, Springer International Publishing Switzerland 2015, pp. 279-289.

[2] Joseph Prusa, Tahhi M. Khoshgoftaar, David J. Dittman, "Using Ensemble Learners to Improve Classifier Performance on Tweet Senti-

ment Data", Information Reuse and Integration (IRI), IEEE International Conference, 13-15 Aug. 2015, pp 252 – 257, INSPEC Accession Number: 15556647.

[3] Olga Kolchyna, Th´arsis T. P. Souza, Philip C. Treleaven and Tomaso Aste, "Twitter Sentiment Analysis: Lexicon Method, Machine Learning Method and Their Combination", Cornell University Library, 18 Sep 2015, arXiv: 1507.00955.

[4] Abinash Tripathy, Ankit Agrawal, Santanu Kumar Rath, "Classification of sentiment reviews using n-gram machine learning approach", in "Expert System With Applications", ELSEVIER, 2016, pp. 117-126.

[5] M.S.Neethu, R. Rajasree, "Sentiment analysis in twitter using machine learning techniques", in "Computing, Communications and Networking Technologies (ICCCNT)", 4th International Conference, ISBN: 978-1-4799-3926-8,2013. https://doi.org/10.1109/ICCCNT.2013.6726818.

[6] 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)", Volume 415, pp 1-13, Nov 2015.(ISBN: 978-3-31927211-5 (Print) 978-3-319-27212-2.

[7] A. Agarwal, B. Xie, I. Vovsha, O. Rambow, R. Passonneau, "Sentiment Analysis of Twitter Data", LSM '11 Proceedings of the workshop on Languages in Social Media, Columbia University New York, ISBN: 978-1-932432-96-1, pp. 30-38, 2011.

[8] Munir Ahmad, Shabib Aftab, Iftikhar Ali, "Sentiment Analysis of Tweets using SVM", International Journal of Computer Applications (0975 – 8887) Volume 177 –No.5, pp 25-29, Nov 2017.

[9] Bholane Savita Dattu, Prof.Deipali V. Gore, "A Survey on Sentiment Analysis on Twitter Data Using Different Techniques", (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 6 (6) 2015, 5358-5362, ISSN 0975 9646, pp 5358 -5362.

[10] Nádia Félix Felipe da Silva, Luiz F.S. Coletta, Eduardo R. Hruschka, Estevam R. Hruschka, "Using unsupervised information to improve semi supervised tweet sentiment classification", Elsevier journal of Information Sciences, Volumes 355–356, 10 August 2016, Pages 348–365.

[11] Gaurangi Patil, Ms. Varsha Galande, Vedant Kekan, Kalpana Dange, "Sentiment Analysis Using Support Vector Machine", International Journal of Innovative Research in Computer and Communication Engineering, Vol. 2, Issue 1, January 2014, ISSN(Online): 2320-9801, pp. 42–49, 1999.

[12] A.Pak and P. Paroubek., "Twitter as a Corpus for Sentiment Analysis and Opinion Mining", In Proceedings of the Seventh Conference on International Language Resources and Evaluation, 2010, pp.1320-1326.

[13] T. Subbulakshmi, R. Regin Raja, "An Ensemble Approach For Sentiment Classification: Voting For Classes and Against Them", ICTACT Journal On Soft Computing, July 2016, Volume 06, Issue: 04, ISSN: 2229-6956 (ONLINE), pp 1281-1286.

[14] Narr, Sascha, Michael Hulfenhaus, and Sahin Albayrak. "Language-independent twitter sentiment analysis." Knowledge Discovery and Machine Learning (KDML), LWA (2012): 12-14.

[15] Mohammad Rezwanul Huq, Ahmad Ali, Anika Rahman , "Sentiment Analysis on Twitter Data using KNN and SVM", In (IJACSA) International Journal of Advanced Computer Science and applications, Vol. 8, No. 6, pages 19-25, 2017.

[16] A. Kumar and T. Mary Sebastian, "Sentiment Analysis on Twitter", IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 4, No 3,.July 2012, ISSN (Online): 1694-0814.

[17] G. Patil, V. Galande, V. Kekan, K. Dange, "Sentiment Analysis Using Support Vector Machine", International Journal of Innovative Research in Computer and Communication Engineering , Vol. 2, Issue 1, January 2014, pp 2607-2612.

[18] Nikhil R, Nikhil Tikoo, Sukrit Kurle, Hari Sravan Pisupati, Prasad G R, "A Survey on Text Mining and Sentiment Analysis for Unstructured Web Data", International Journal of Emerging Technologies and Innovative Research(JETIR), April 2015, Volume 2, Issue 4 , ISSN-2349-5162, pp. 1292-1296.

[19] Gaurangi Patil, Ms. Varsha Galande, Vedant Kekan, Kalpana Dange, "Sentiment Analysis Using Support Vector Machine", International Journal of Innovative Research in Computer and Communication Engineering, Vol. 2, Issue 1, January 2014, ISSN(Online): 2320-9801, pp. 42–49, 1999.

[20] Arockia Xavier Annie R, Vignesh Mohan, Sree Harissh Venu, "Sentiment Analysis Applied to Airline Feedback to Boost Customer's Endearment", second International Conference on Multidisciplinary Innovation For Sustainability And Growth (MISG 2015), ISBN: 978-969-9948-31-2, Vol. 2, pp 219-232, 2015.

[21] Christos Troussas, Maria Virvou, Kurt Junshean Espinosa, Kevin Llaguno, Jaime Caro, "Sentiment analysis of Facebook statuses using Naive Bayes classifier for language" , published in " Information, Intelligence, Systems and Applications (IISA)" , Fourth International Conference , 2013, ISBN: 978-1-4799-0771-7. https://doi.org/10.1109/IISA.2013.6623713.

[22] Akshi Kumar, Teeja Mary Sebastian, "Sentiment Analysis on Twitter", IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 4, No 3, July 2012, ISSN: 1694-0814, pp. 372-378.

[23] Walaa Medhat, Ahmed Hassan, Hoda Korashy, "Sentiment analysis algorithms and applications: A survey", Ain Shams Engineering Journal, April 2014, Vol 5, issue 4, pp. 1093-1113. https://doi.org/10.1016/j.asej.2014.04.011.

[24] Xiang Ji, Soon Ae Chun, James Geller, "Monitoring Public Health Concerns Using Twitter Sentiment Classifications", in IEEE International Conference on Healthcare Informatics, 2013, 978-0-7695-5089-3/13, pp. 235-244.

[25] Lina L. Dhande, Dr. Prof. Girish K. Patnaik, "Analyzing Sentiment of Movie Review Data using Naive Bayes Neural Classifier", in International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), Volume 3, Issue 4 July-August 2014, ISSN 2278-6856, pp. 313-320.

[26] Yun Wan, Dr. Qigang GAO, "An Ensemble Sentiment Classification System of Twitter Data for Airline Services Analysis", IEEE 15th International Conference on Data Mining Workshops, 978-1-4673-84933/15, 2015, pp 1318-1325.