Journal of Engineering and Applied Sciences 13 (11): 3907-3915, 2018
ISSN: 1816-949X

Multi-Level Tweets Classification and Mining using Machine Learning Approach

[1]Abdul Ahad, [2]Suresh Babu Yalavarthi and [2]Ali Hussain
[1]Department of CSE, Acharya Nagarjuna University, Guntur, Andhra Pradesh, India
[2]Department of CSE, JKC College, Guntur, Andhra Pradesh, India

Abstract: Sentiment analysis comes under study within natural language processing. It helps in finding the sentiment or opinion hidden within a text. This research focuses on finding sentiments for twitter data as it is more challenging due to its unstructured nature, limited size, use of slangs, misspells abbreviations, etc. Most of the researchers dealt with various machine learning approaches of sentiment analysis and compare their results but using various machine learning approaches in combination have been underexplored in the literature. This research has found that various machine learning approaches in a hybrid manner gives better result as compared to using these approaches in isolation. Moreover, as the Tweets are very raw in nature, this research makes use of various preprocessing steps, so that, we get useful data for input in machine learning classifiers. This research basically focuses on two machine learning algorithms K-Nearest Neighbours (KNN) and Support Vector Machines (SVM) in a hybrid manner. The analytical observation is obtained in terms of classification accuracy and F-measure for each sentiment class and their average. The evaluation analysis shows that the proposed hybrid approach is better both in terms of accuracy and F-measure as compared to individual classifiers.

Key words: Sentiment analysis, machine learning, KNN, SVM, analytical, isolation

INTRODUCTION

Due to the presence of enormous amount of data available on web, various organizations started taking interest in this as mining this information can be very valuable to them. This gives birth to an entirely different and broad field of study known as sentiment analysis. Various names have given to this field as opinion mining, opinion extraction, etc. However, there is slight difference in meaning between these various terms. Before automatic mining of sentiments traditional survey techniques were highly biased as they were taken individually by users thus a need of an automatic system arose that can directly deal with hundreds of thousands of opinions hidden in user's posts in the form of reviews, blogs, etc. Various applications of sentiment analysis are as in product reviews, movie reviews, business, politics, recommender system, etc. Based on the opinion about a product or about different aspects of a product an organization can make changes accordingly. Similarly based on the opinion about a particular political party, government policie's changes can be made accordingly. Two main techniques used for sentiment analysis are machine learning based and lexicon based.

Supervised, unsupervised and semi-supervised comes under machine learning. Supervised approaches, e.g., SVM (Agarwal *et al*., 2011, 2015; Silva *et al*., 2014; Khan *et al*., 2017; Anjaria and Guddeti, 2014; Hu and Liu, 2004; Jianqiang and Xiaolin, 2017), KNN (Mukwazvure and Supreethi, 2015; Khan and Jeong, 2016), Naive Bayes (Pak and Paroubek, 2010; Spencer and Uchyigit, 2012; Narr *et al*., 2012; Saif *et al*., 2012), etc. Requires a good quality training set and thus are highly domain dependent but provide better results if trained properly. Unsupervised approaches, e.g., k-means, Self Organizing Maps (SOM), etc. do not make use of training set. Semi supervised approaches require partial labelling of data and are of two types transductive learning inductive learning.

Lexicon based approach makes use of dictionary consist of labelled words and with the help of these words, a text is judged whether it is subjective or objective (Appel *et al*., 2016; Saif *et al*., 2016; Hutto and Gilbert, 2014). This approach is further divided into dictionary based which does not take into account the context of word within a text and corpus based which expands the dictionary with taking associations between different words into account. A complete survey in this field is provided by Medhat *et al*. (2014).

This research analyzes sentiment of Tweets. As Tweets are much unstructured in nature this research converts them into useful information, so that, better features can be used for machine learning. Hence, in this research, we provide a good data preprocessing to

Corresponding Author: Abdul Ahad, Department of CSE, Acharya Nagarjuna University, Guntur, Andhra Pradesh, India

Tweets followed by hybrid classifier. With the help of processed Tweets or data features are generated and fed to the two machine learning algorithms KNN and SVM in a hybrid manner. Different feature have tried by researcher for improving the results.

Support Vector Machines (SVM): SVM invented by Vapnik and Chervonenkis in 1963 is a supervised machine learning algorithm used basically for classification and regression problems. It solves an optimization problem of finding the maximum margin hyperplane between the classes. This is basically required to avoid overfitting. Basically, it is a linear classifier separating the classes which can be separated with the help of linear decision surfaces called hyperplanes. Foe classes having binary features SVM draws a line between the classes and for classes having multiple features hyperplanes are drawn. However, it can be used for classifying the non linear data also by transforming the feature space into the higher dimensional space, so that, non linear data in higher dimensional can be separated easily by a hyperplane.

This transformation is made easy with the help of Kernel-trick. With the help of kernels it is not necessary to calculate all the dimensions when transform and calculation of hyperplane can be done in the same lower dimensional feature space. Kernels are not used only for this purpose but also for making the calculation easier in case of many features. Various kernels are used by machine learning approaches, e.g., RBF (Radial Basis Function), linear kernel, poly kernel, etc. Kernels make the calculation very fast and help in improving the calculation time remarkably.

K-Nearest Neighbors (KNN): It is also a supervised learning algorithm based on the classes nearestto the point which is to be classified. Based on the values of the K nearest classes a test set is provided the majority voting class. However, to improve this algorithm weights are assigned to each of the K points according to their distance from the test point. The value of K depends upon the classification problem and the size of dataset.

Literature review: Various researchers have been working on Twitter (Pak and Paroubek, 2010; Spencer and Uchyigit, 2012; Kouloumpis *et al*., 2011; Silva *et al*., 2014; Khan *et al*., 2014; Appel *et al*., 2016) and from time to time they are publishing their researches. They have used various sentiment analysis techniques for improving the results of classification their research is also helpful in this research as the sentiment analysis techniques they have used, feature selection techniques, different pre-processing steps they have used is taken care of in this research. This research mainly focuses on supervised approach for sentiment analysis task and has surveyed researches both for Twitter and non-Twitter data and also for both supervised and lexicon based approaches for better clarification and understanding of the topic chosen.

Many researches defined multiple faces of sentiment analysis as opinion orientation, feature extraction, etc., machine learning classifiers need various features for learning, so, different researchers from time to time have selected different features for comparing results.

Agarwal *et al*. (2011), Pak and Paroubek (2010), Spencer and Uchyigit (2012) and Kouloumpis *et al*. (2011) selected various features as unigrams, bigrams, pos tagging, hash tags, ngrams, etc. and found mixed response in classification results. Different features and feature selection methods as semantic features and concepts, information gain, chi-square, etc. has been used by Khan *et al*. (2017) and Agarwal *et al*. (2015).

Khan *et al*. approach includes rigorous data pre-processing followed by supervised machine learning. They collected labelled datasets of different domains, so that, machine learning will not be limited to a particular domain. To learn SVM classifier they make use of different training sets each make SVM learn different feature sets. Information Gain (IG) with feature presence and feature frequency cosine similarity with feature presence and feature frequency. They found that feature presence is better than feature frequency.

Agarwal *et al*. (2015) found that for better results using machine learning approaches, finding good features is a challenging task. They gave the concept of "Semantic parser" and treated concepts as features. They used the Minimum Redundancy and Maximum Relevance (MRMR) feature selection mechanism. They used different feature sets for their classification task, e.g., unigrams, bigrams, bi-tagged and dependency parse tree along with their proposed scheme, so that, results can be compared with.

Various approaches and classifiers such as lexicon based approach, Naive Bayes (NB), Support Vector Machines (SVM), Maximum Entropy (MaxEnt), etc. have been used time to time with various parameters for evaluating the results as accuracy, precision, recall, F-measure etc. Narr *et al*. (2012) concluded 71.5% accuracy with mixed language NB classifier on unigrams. Saif *et al*. (2012) concluded that semantic features used by NB classifier increase F1-measure against unigram by 6.47% and posunigram by 4.78%. Asmi and Ishayat (2012), Hutto and Ishaya (2014) and Neviarouskaya *et al*. (2009) proposed rule based approaches for increasing the accuracy. Kawathekar and Kshirsagar (2012), Chikersal *et al*. (2015) and Prabowo and Thelwall (2009) proposed hybrid approach consisting of rule based and machine learning classifiers.

Hybrid approaches consisting of machine learning classifiers have been underexplored in the literature with very few researches in this approach as in Revathy and Sathiyabhama (2013) and Silva *et al.* (2014) proposed an ensemble based classification in which various classifiers, e.g., SVM, multinomial Naive Bayes, random forest, logistic regression are used. They proposed that if, we train the different classifiers with different training sets and then by using either average probabilities of different classifiers or maximum voting, we get better results than by using only a single classifier. Moreover, they uses two different features for learning the classifiers. Bag of Words (BOW) feature hashing.

They used four different datasets for training and testing. They found that feature hashing is not better than BOW approach in most of the datasets except one. Our research work mainly focuses on combining the machine learning classifiers and proves that combining gives better results as compared to standalone classifiers. Also, this research gives comparative results as against to the feature hashing+lexicon based features used by Silva *et al.* (2014) with only a small dataset and few features.

Bhadane *et al.* (2015) and Apple *et al.* (2016) proposed combination of sentiment lexicon with machine learning approaches and found increase in accuracy. Muhammad *et al.* (2016) handled word's polarity in terms of local and global context by giving smart SA system and found that their system is superior to baseline lexicons and systems like SVM, NB, etc. with more F1 score.

Addlight and Supreethi compared two machine learning methods KNN and SVM and found that SVM outperforms KNN. Saif *et al* (2016) gave the concept of SentiCircles for calculating the context of words. They found that it is necessary for better sentiment classification.

Jianqiang and Xialin (2017) discussed the role of rigorous preprocessing in increasing the evaluation measure and gave six different preprocessing methods for the same. Keeping this in mind our approach also uses a good preprocessing to filter the Tweets. Khan and Jeong (2016) proposed an approach for finding the sentiments about each aspect of a product and this can be a good future reserach to explore.

MATERIALS AND METHODS

The proposed hybrid model is defined as the three stage model. In first preprocessing stage of this model, the multi-aspect based filtration and impurities correction is applied. The spell correction, stemming, abbreviation expansion, stopwords removal are defined in this stage to normalize the input Tweets. In this stage, the separation of tag tokens, positive aspects and negative aspects from messages is also defined. Negation handling is also done. In second stage, the filtered text is processed to generate the statistical features. In this stage, the transformation of input training and testing set is done to corresponding feature set. These features are processed by the hybrid classifier for sentiment prediction in final stage of this model. In classification stage, the probabilistic predictive decision is applied for selection of KNN or SVM classifier for individual instance (Fig. 1). The classification is applied on the Tweets acquired from the web. The description of dataset is given in Table 1.

Table 1: The description of dataset

Features	Values
Dataset name	Twitter-sentiment-analysis-finalizedFull
Dataset URL	https://github.com/TharinduMunasinge/Twitter-Sentiment-Analysis
Number of Tweets	997
Classes	Positive Tweets , negative Tweets and neutral Tweets
File type	CSV

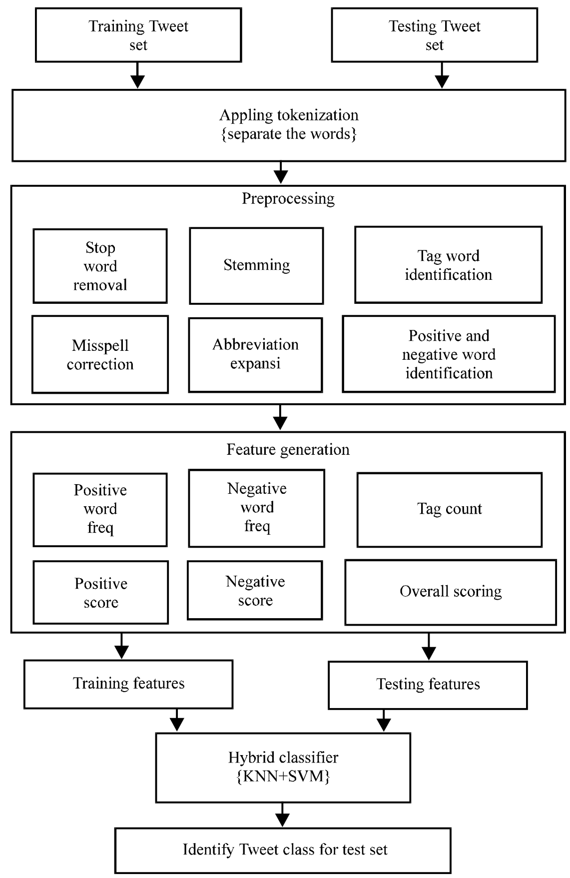
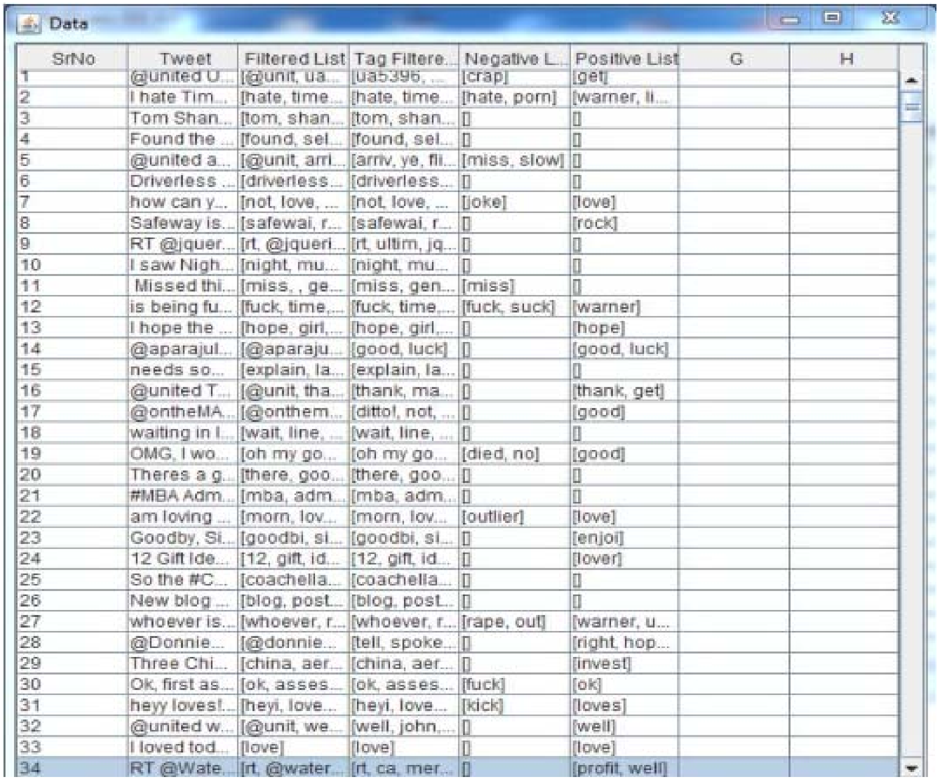


Fig. 1: Flowchart of the proposed system



Data

SrNo	Tweet	Filtered List	Tag Filtere...	Negative L...	Positive List	G	H
1	@united U...	[@unit, ua...	[ua5396, ...	[crap]	[get]		
2	I hate Tim...	[hate, time...	[hate, time...	[hate, porn]	[warner, li...		
3	Tom Shan...	[tom, shan...	[tom, shan...	[]	[]		
4	Found the ...	[found, sel...	[found, sel...	[]	[]		
5	@united a...	[@unit, arri...	[arriv, ye, fli...	[miss, slow]	[]		
6	Driverless ...	[driverless...	[driverless...	[]	[]		
7	how can y...	[not, love, ...	[not, love, ...	[joke]	[love]		
8	Safeway is...	[safewai, r...	[safewai, r...	[]	[rock]		
9	RT @jquer...	[rt, @jqueri...	[rt, ultim, jq...	[]	[]		
10	I saw Nigh...	[night, mu...	[night, mu...	[]	[]		
11	Missed thi...	[miss, , ge...	[miss, gen...	[miss]	[]		
12	is being fu...	[fuck, time,...	[fuck, time,...	[fuck, suck]	[warner]		
13	I hope the ...	[hope, girl,...	[hope, girl,...	[]	[hope]		
14	@aparajul...	[@aparaju...	[good, luck]	[]	[good, luck]		
15	needs so...	[explain, la...	[explain, la...	[]	[]		
16	@united T...	[@unit, tha...	[thank, ma...	[]	[thank, get]		
17	@ontheMA...	[@onthem...	[ditto!, not, ...	[]	[good]		
18	waiting in l...	[wait, line, ...	[wait, line, ...	[]	[]		
19	OMG, I wo...	[oh my go...	[oh my go...	[died, no]	[good]		
20	Theres a g...	[there, goo...	[there, goo...	[]	[]		
21	#MBA Adm...	[mba, adm...	[mba, adm...	[]	[]		
22	am loving ...	[morn, lov...	[morn, lov...	[outlier]	[love]		
23	Goodby, Si...	[goodbi, si...	[goodbi, si...	[]	[enjoi]		
24	12 Gift Ide...	[12, gift, id...	[12, gift, id...	[]	[lover]		
25	So the #C...	[coachella...	[coachella...	[]	[]		
26	New blog ...	[blog, post...	[blog, post...	[]	[]		
27	whoever is...	[whoever, r...	[whoever, r...	[rape, out]	[warner, u...		
28	@Donnie...	[@donnie...	[tell, spoke...	[]	[right, hop...		
29	Three Chi...	[china, aer...	[china, aer...	[]	[invest]		
30	Ok, first as...	[ok, asses...	[ok, asses...	[fuck]	[ok]		
31	heyy loves!...	[heyi, love...	[heyi, love...	[kick]	[loves]		
32	@united w...	[@unit, we...	[well, john,...	[]	[well]		
33	I loved tod...	[love]	[love]	[]	[love]		
34	RT @Wate...	[rt, @water...	[rt, ca, mer...	[]	[profit, well]		

Fig. 2: Tweets after preprocessing, for stop words, abbreviation expansion[1] and misspell correction[2] database is created. Filtered list: -contains Tweets after tokenization and applying the above written filters. Tag filtered list:-contains the filtered list with @ tags removed. The @ tags are used in feature generation as tag count in each Tweet. Negative list: -contains negative adjectives in each Tweet. Positive list: -contains positive adjectives in each Tweet

Table 2: Describing various attributes of an adjective

Attributes	Description
ID	Numeric unique id to all adjectives
Adjective	Stores the textual information to represent the actual adjective
Pscore	Positive score to represent the positive acceptability of an adjective Lies between 0 and 1
Fscore	Negative score, lies between 0 and 1
Score	Overall score of adjective lies between -1 and 1 +ve values for +ve adjective -ve value for –ve adjective

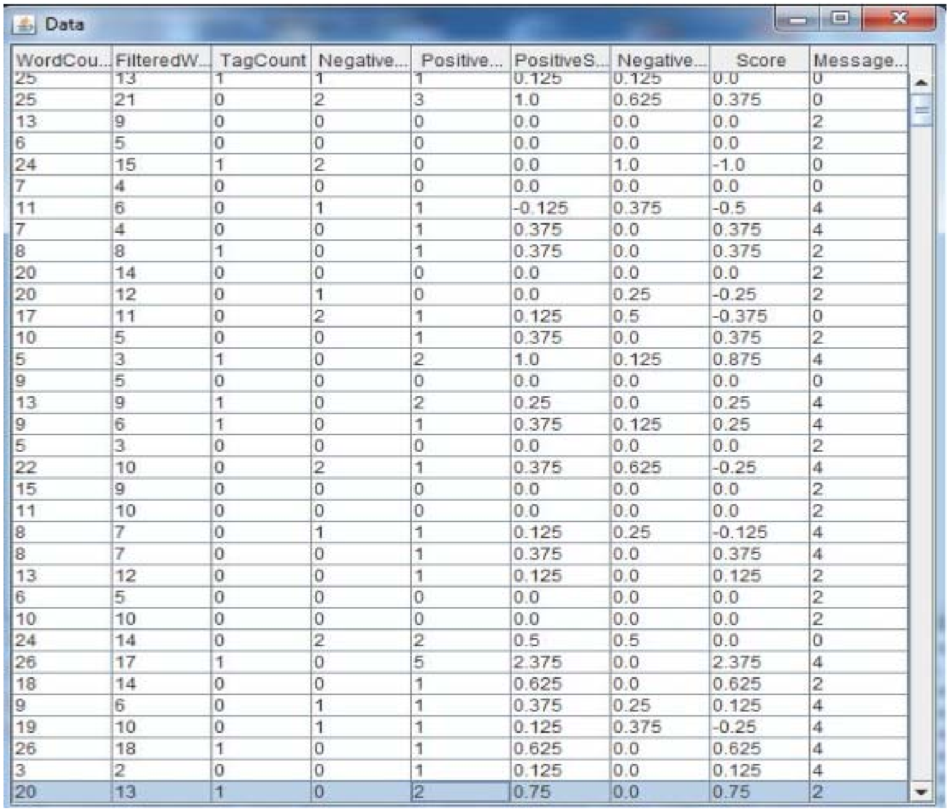
Preprocessing: During preprocessing various steps are taken as stopword removal, handling of negation, abbreviation expansion, misspell correction, stemming, positive word lists of each Tweet, negative word lists of each Tweet. Porters algorithm is used for stemming.

Features generation: A list of adjectives (Fig. 2) is used for features generation. This list contains positive score, negative score, overall rating of an adjective among other attributes (Table 2 and Fig. 3).

- http://www.illumasolutions.com/omg-plz-lol-idk-idc-btw-brb-jk.htm
- https://noisy-text.github.io/norm-shared-task.html
- http://www.sentix.de/index.php/en/item/sentix-website.html

Various features are generated after filtering of Tweets for learning the classifiers. Various features used for learning the classifiers are:

- Word count: total words in each Tweet after filtration
- Tag count: total @ tags used in each Tweet
- Negative word count: total negative words in each Tweet
- Positive word count: total positive words in each Tweet
- Positive score: total positive score obtained by adding the positive scores of each positive adjective
- Negative score: total negative score obtained by adding the negative scores of each negative adjective



WordCou...	FilteredW...	TagCount	Negative...	Positive...	PositiveS...	Negative...	Score	Message...
25	13	1	1	1	0.125	0.125	0.0	0
25	21	0	2	3	1.0	0.625	0.375	0
13	9	0	0	0	0.0	0.0	0.0	2
6	5	0	0	0	0.0	0.0	0.0	2
24	15	1	2	0	0.0	1.0	-1.0	0
7	4	0	0	0	0.0	0.0	0.0	0
11	6	0	1	1	-0.125	0.375	-0.5	4
7	4	0	0	1	0.375	0.0	0.375	4
8	8	1	0	1	0.375	0.0	0.375	2
20	14	0	0	0	0.0	0.0	0.0	2
20	12	0	1	0	0.0	0.25	-0.25	2
17	11	0	2	1	0.125	0.5	-0.375	0
10	5	0	0	1	0.375	0.0	0.375	2
5	3	1	0	2	1.0	0.125	0.875	4
9	5	0	0	0	0.0	0.0	0.0	0
13	9	1	0	2	0.25	0.0	0.25	4
9	6	1	0	1	0.375	0.125	0.25	4
5	3	0	0	0	0.0	0.0	0.0	2
22	10	0	2	1	0.375	0.625	-0.25	4
15	9	0	0	0	0.0	0.0	0.0	2
11	10	0	0	0	0.0	0.0	0.0	2
8	7	0	1	1	0.125	0.25	-0.125	4
8	7	0	0	1	0.375	0.0	0.375	4
13	12	0	0	1	0.125	0.0	0.125	2
6	5	0	0	0	0.0	0.0	0.0	2
10	10	0	0	0	0.0	0.0	0.0	2
24	14	0	2	2	0.5	0.5	0.0	0
26	17	1	0	5	2.375	0.0	2.375	4
18	14	0	0	1	0.625	0.0	0.625	2
9	6	0	1	1	0.375	0.25	0.125	4
19	10	0	1	1	0.125	0.375	-0.25	4
26	18	1	0	1	0.625	0.0	0.625	4
3	2	0	0	1	0.125	0.0	0.125	4
20	13	1	0	2	0.75	0.0	0.75	2

Fig. 3: The results of features generation

- Score: positive score-negative score for each Tweet
- Message class 0: for negative Tweets
- 1: for neutral Tweets
- 2: for positive Tweets

Classification: After features generation classification is done with our hybrid approach in which prediction probability of both the classifiers is used which is shown in the algorithm.

Algorithm 1: Classification (TrainingSet, TestingSet):
/*TrainingSet is the training Tweet Set and TestingSet is the Testing Tweet Set on which features are generated */
{
1. Train FeaturesSet = Feature Generation(TrainingSet) /*Generate Features for Training Set*/
2. Test Features Set = FeatureGeneration(TestSet)/*Generate Features for Testing Set*/
3. SWeight = GenerateWeight(TrainFeaturesSet,S VM)
/*Process the Classifier, train with the Training Feature set and Generate Feature weights for SVM*, KNN is trained directly during testing/
4. For I = 1 to TestFeatureSet.Length /*Process the Testing Instances*/
{
5. K1 = Predict(TestFeatureSet(i),TrainFeaturesSet)
/*Apply Prediction on Test Instance respective to KNN Classifier Weight*/
6. S1 = Predict(TestFeatureSet(i),SWeight)
/*Apply Prediction on Test Instance respective to SVM Classifier Weight*/
7. If (K1>Th1 And S1>Th1)
/*Apply Hybrid Classifier for Test Class Identification, Th1 is the threshold used for prediction probability, Th1 = 0.5 is used here*/
{
8. TestFeatureSet(i).Class = IdentifyClass(greater(K1,S1))
}
9. Else If (K1>Th1)
/*Apply KNN Classifier for Test Class Identification*/
{
10. TestFeatureSet(i).Class = IdentifyClass(K1)
}
11. Else
12. /*Apply SVM Classifier for Test Class Identification*/
{
13. TestFeatureSet(i).Class = IdentifyClass(S1)
}
}
Return TestFeatureSet.Class
}

RESULTS AND DISCUSSION

In this present research, the SVM and KNN based hybrid classification model is presented to process the Tweet features and to identify the hidden sentiments from these Tweets. Implementation is done in Netbeans 8.0 with Weka (3.8) integrated into it. Weka is widely used in data mining for preprocessing, clustering, classification, etc. and gives results in terms of accuracy, precision, recall, F-measure, etc. MySql is used for storing the various datasets as list of adjectives, abbreviations, misspell corrections, training dataset and testing dataset.

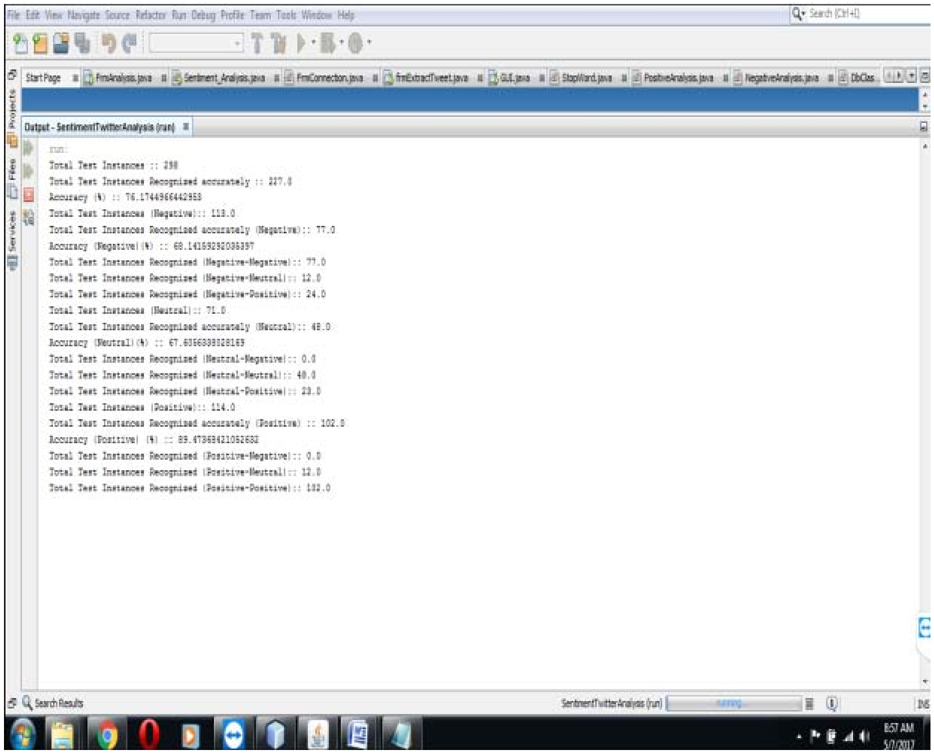


Fig. 4: The analysis results

For running classifiers KNN and SVM in isolation, weka is used directly but for their combination Weka is integrated into netbeans and confusion matrix and results are manually calculated with the help of results. The comparative analysis is provided against KNN and SVM based methods separately. K = 15 is used for comparison between KNN and hybrid approach. The description of processing training and testing set is shown in Table 3.

The classification algorithm combining KNN and SVM is given in this study 2. Confusion matrix for KNN and SVM is taken from weka by opening the train features set and test features set there directly and is provided as:

Confusion matrix for hybrid approach is calculated manually by the analysis results and with the help of confusion matrix precision, recall and F-measure is calculated (Table 4 and 5).

Analysis results shows True Positives (TP), False positives (FP), True Negatives (TN) and False Negatives (FN) for each sentiment class. Thus, confusion matrix is derived (Table 6).

With the help of confusion matrices accuracy, Precision, recall and F-measure for positive, negative and neutral classes are calculated and is also compared for the 3 approaches used above (Fig. 4-8).

Table 3: The description of processing training and testing set

Features	Values
Size of training set	699 (267-positive, 264-negative, 168-neutral)
Size of testing set	298 (114-positive, 113-negative, 71-neutral)
Tweet classes	Positive, negative, neutral
Existing methods	KNN and SVM
Proposed	Hybrid KNN+SVM

Table 4: Confusion matrix for KNN

	Predicted		
Actual class	Negative	Neutral	Positive
Negative	77	19	17
Neutral	9	46	16
Positive	21	14	79

Table 5: Confusion matrix for SVM

	Predicted		
Actual class	Negative	Neutral	Positive
Negative	77	14	22
Neutral	11	47	13
Positive	16	20	78

Table 6: Confusion matrix for KNN+SVM

	Predicted		
Actual class	Negative	Neutral	Positive
Negative	77	12	24
Neutral	0	48	23
Positive	0	12	102

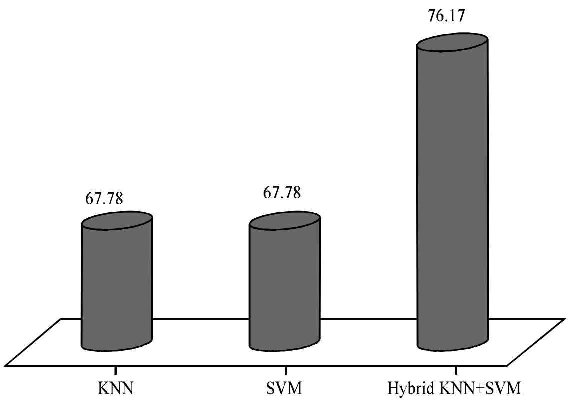


Fig. 5: Overall accuracy for 3 approaches

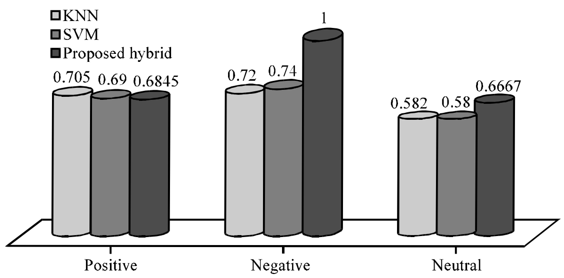


Fig. 6: Precision analysis

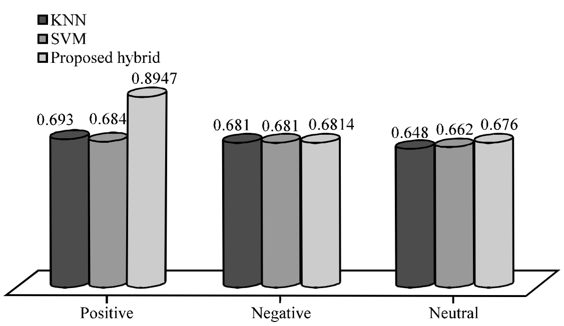


Fig. 7: Recall analysis

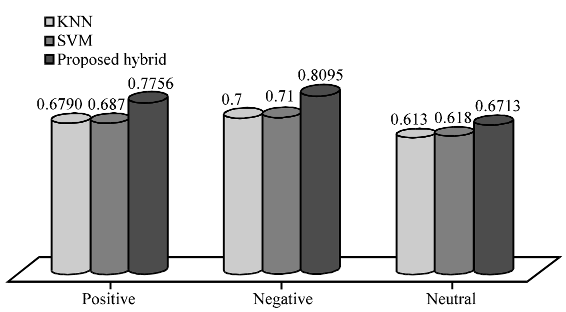


Fig. 8: F-measure analysis

Thus, the above results show that our hybrid approach reserach better both in terms of accuracy and F-measure. Comparison of our approach with that of (Silva *et al*., 2014).

Table 7: Comparison of values

Datasets	Avg. F-measure (%) (pos neg)	Avg. F-measure(%) (including neutral)	Accuracy (%) (pos, neg)	Accuracy (%) (including, neutral)
OMD (Silva et al., 2014)	65.35	-	70.62	-
Strict OMD (Silva et al., 2014)	71.80	-	74.56	-
Sanders (Silva et al., 2014)	76.25	-	76.63	-
Stanford (Silva et al., 2014)	78.25	-	79.11	-
HCR (Silva et al., 2014)	62.20	-	78.35	-
Our dataset (Tharindu Munasinge)	79.25	75.21	78.80	76.17

This research with only a small dataset and few features gives comparative results and even better results as compared to combination of Logistic Regression (LR), Random Forest (RF) and Multinomial Naive Bayes (MNB) along with feature hashing and lexicon features used in (Silva *et al*., 2014). The comparison is shown in Table 7, Avg. is used in place of average, positive in place of positive, negative in place of negative.

CONCLUSION

In this study, a SVM and KNN based hybrid model is presented to improve the classification accuracy. The proposed method classified the Tweets in positive, negative and neutral sentiments whereas much of the literature in this field is associated with 2-way classification. The reserach of proposed model has gone through preprocessing stage, features generation stage and classifiers learning stage. The analytical evaluation of proposed model is done in terms of accuracy and F-measure. The comparative observations are taken against the SVM and KNN methods. The comparative results show that the proposed model has improved the accuracy and F-measure of Tweet class prediction.

RECOMMENDATIONS

As number of features for learning the classifiers are limited in our approach, we will be using more features and better feature selection methods like information gain, Chi-square, etc. in our future reserach. Our comparison with literature shows that increasing our dataset with more Tweets and features can also help in increasing reasonable accuracy and F-measure. Other machine learning methods in combined way can also be explored in the future.

REFERENCES

Agarwal, A., B. Xie, I. Vovsha, O. Rambow and R. Passonneau, 2011. Sentiment analysis of twitter data. Proceedings of the Workshop on Languages in Social Media, June 23, 2011, Association for Computational Linguistics Stroudsburg, PA, USA, pp: 30-38.

Agarwal, B., S. Poria, N. Mittal, A. Gelbukh and A. Hussain, 2015. Concept-level sentiment analysis with dependency-based semantic parsing: A novel approach. Cognit. Comput., 7: 487-499.

Anjaria, M. and R.M.R. Guddeti, 2014. A novel sentiment analysis of social networks using supervised learning. Soc. Netw. Anal. Min., 4: 181-195.

Appel, O., F. Chiclana, J. Carter and H. Fujita, 2016. A hybrid approach to the sentiment analysis problem at the sentence level. Knowl. Based Syst., 108: 110-124.

Asmi, A. and T. Ishaya, 2012. Negation identification and calculation in sentiment analysis. Proceedings of the 2nd International Conference on Advances in Information Mining and Management (IMMM), October 21-26, 2012, IARIA, Wilmington, Delaware, ISBN:978-1-61208-227-1, pp: 1-7.

Bhadane, C., H. Dalal and H. Doshi, 2015. Sentiment analysis: Measuring opinions. Procedia Comput. Sci., 45: 808-814.

Chikersal, P., S. Poria and E. Cambria, 2015. SeNTU: Sentiment analysis of Tweets by combining a rule-based classifier with supervised learning. Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval'15), June 4-5, 2015, Association for Computational Linguistics, Denver, Colorado, pp: 647-651.

Hu, M. and B. Liu, 2004. Mining and summarizing customer reviews. Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, August 22-25, 2004, ACM, Seattle, Washington, USA., ISBN:1-58113-888-1, pp: 168-177.

Hutto, C.J. and E. Gilbert, 2014. Vader: A parsimonious rule-based model for sentiment analysis of social media text. Proceedings of the 8th AAAI International Conference on Weblogs and Social Media (ICWSM'14), June 1-4, 2014, Association for the Advancement of Artificial Intelligence, Palo Alto, Californiay, USA., pp: 216-225.

Jianqiang, Z. and G. Xiaolin, 2017. Comparison research on text pre-processing methods on twitter sentiment analysis. IEEE Access, 5: 2870-2879.

Kawathekar, S.A. and M.M. Kshirsagar, 2012. Sentiments analysis using hybrid approach involving rule-based and support vector machines methods. IOSRJEN., 2: 55-58.

Khan, F.H., S. Bashir and U. Qamar, 2014. TOM: Twitter opinion mining framework using hybrid classification scheme. Decis. Support Syst., 57: 245-257.

Khan, F.H., U. Qamar and S. Bashir, 2017. A semi-supervised approach to sentiment analysis using revised sentiment strength based on SentiWordNet. Knowl. Inf. Syst., 51: 851-872.

Khan, J. and B.S. Jeong, 2016. Summarizing customer review based on product feature and opinion. Proceedings of the International Conference on Machine Learning and Cybernetics (ICMLC'16) Vol. 1, July 10-13, 2016, IEEE, Jeju, South Korea, ISBN:978-1-5090-0391-4, pp: 158-165.

Kouloumpis, E., T. Wilson and J.D. Moore, 2011. Twitter sentiment analysis: The good the bad and the OMG!. Proceedings of the AAAI 5th International Conference on Weblogs and Social Media (ICWSM), July 17-21, 2011, AAAI, Barcelona, Spain, pp: 538-541.

Medhat, W., Hassan, A. and H. Korashy, 2014. Sentiment analysis algorithms and applications: A survey. Ain Shams Eng. J., 5: 1093-1113.

Muhammad, A., N. Wiratunga and R. Lothian, 2016. Contextual sentiment analysis for social media genres. Knowl. Based Syst., 108: 92-101.

Mukwazvure, A. and K.P. Supreethi, 2015. A hybrid approach to sentiment analysis of news comments. Proceedings of the 4th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO) (Trends and Future Directions), September 2-4, 2015, IEEE, Noida, India, ISBN:978-1-4673-7230-5, pp: 1-6.

Narr, S., M. Hulfenhaus and S. Albayrak, 2012. Language-independent Twitter sentiment analysis. Proceedings of the 2012 Joint Workshop on Knowledge Discovery and Machine Learning (KDML'12) and Lernen Wissen Adaption, September 12-14, 2012, Technical University of Dortmund, Dortmund, Germany, pp: 1-7.

Neviarouskaya, A., H. Prendinger and M. Ishizuka, 2009. Semantically distinct verb classes involved in sentiment analysis. Proceedings of the IADIS International Conference on Applied Computing (ICAC'09), November 19-21, 2009, International Association for Development of the Information Society (IADIS), Portugal, Lisbon, ISBN:978-972-8924-97-3, pp: 27-35.

Pak, A. and P. Paroubek, 2010. Twitter as a corpus for sentiment analysis and opinion mining. Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC'10), May 17-23, 2010, European Language Resources Association, Valletta, Malta, ISBN:2-9517408-6-7, pp: 1320-1326.

Prabowo, R. and M. Thelwall, 2009. Sentiment analysis: A combined approach. J. Inform., 3: 143-157.

Revathy, K. and B. Sathiyabhama, 2013. A hybrid approach for supervised twitter sentiment classification. Intl. J. Comput. Sci. Bus. Inf., 7: 35-45.

Saif, H., Y. He and H. Alani, 2012. Semantic sentiment analysis of Twitter. Proceedings of the 11th International Conference on Semantic Web, November 11-15, 2012, Springer, Boston, Massachusetts, USA., ISBN:978-3-642-35175-4, pp: 508-524.

Saif, H., Y. He, M. Fernandez and H. Alani, 2016. Contextual semantics for sentiment analysis of Twitter. Inf. Process. Manage., 52: 5-19.

Silva, N.F.D., E.R. Hruschka and E.R. Hruschka, 2014. Tweet sentiment analysis with classifier ensembles. Decis. Support Syst., 66: 170-179.

Spencer, J. and G. Uchyigit, 2012. Sentimentor: Sentiment analysis of Twitter data. Proceedings of European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD'12), September 28, 2012, University of Bristol, Bristol, England, UK., pp: 56-66.