EAI Endorsed Transactions
on Scalable Information Systems

Research Article EAI.EU

Efficient Framework for Sentiment Classification Using Apriori Based Feature Reduction

Achin Jain[1], Vanita Jain[2, *]

[1]University School of Information, Communication and Technology, GGSIPU, Sector 16 C, Dwarka, Delhi, India
[2]Bharati Vidyapeeth's College of Engineering, New Delhi, India

Abstract

This paper proposes a novel feature selection method for Sentiment Classification. UCI ML Dataset is selected having a textual review from three domains (IMDB Movie, AMAZON Product, and YELP restaurant). Text pre-processing and feature selection technique is applied to the dataset. A Novel Feature Selection approach using Association Rule Mining is presented in which Sentence is converted in binary form and Apriori Algorithm is applied to reduce the dataset. Four Machine Learning algorithms: Naïve Bayes, Support Vector Machine, Random Forest & Logistic Regression to implement experiment. The proposed approach shows an accuracy improvement of 4.2%, 4.9% & 5.9% for IMDB, Amazon & Yelp domain datasets, respectively. Compared with the Genetic Algorithm, Principal Component Analysis, Chi-Square, and Relief based feature selection, the proposed method shows an accuracy improvement of 9.8%, 0.4%, 0.6% & 1.9%, respectively.

Keywords: Sentiment Classification, Association Rule Mining, Apriori Algorithm, Feature Selection, Machine Learning

Received on 31 May 2020, accepted on 22 January 2021, published on 16 February 2021

doi: 10.4108/eai.16-2-2021.168715

*Corresponding author. Email: vanita.jain@bharatividyapeeth.edu

1. Introduction

The sentiment is considered as a viewpoint about a topic or a thing. It plays an essential role in the decision-making process of an individual. People tend to evaluate the world with their experiences, beliefs, and choices. Considering this, we can say that sentiment has a very high place in every aspect of human life. Sentiment Analysis or Opinion Mining is the technique of evaluating text documents and extracting emotions from them. Sentiments are generally classified as either positive, negative, or neutral [1,24]. Recent years have seen a surge in several research works focusing on Sentiment Analysis. Many researchers and organizations are using the concept in a vast number of fields, such as Movie Review and Product Review [2,3]. Analyzing a small number of reviews can be carried out manually, but the manual analysis is not possible when there are thousands of reviews.

Machine Learning (ML) has been extensively used to perform the task of sentiment classification. The ML model is trained on the well-labeled training dataset and then applied on a testing dataset to predict the sentiment. The authors are using various approaches in recent years. The most common of them is using Lexicon Analysis that lists words by their semantic values. Different Machine Learning Classification algorithms are also used to predict the opinion of a sentence. In recent years, researchers have started developing Hybrid models using the Lexicon technique and ML algorithms to solve exciting research problems in the field of Sentiment Analysis [4,5].

Sentiment Classification is a process of classifying a sentence into the sentence's polarity, and ML approaches prove very useful in improving the accuracy of the prediction model. There are three levels of Sentiment Classification: Document-level, Sentence-level and Feature-level. In this paper, we perform Sentence Level Sentiment Classification on the UCL ML dataset [6] obtained from three domains, namely movies, products and restaurants with equal distribution of 500 positive and 500 negative reviews.

Key Contributions of the Paper are:

i. This Paper proposes a novel feature selection technique based on the Apriori algorithm.
ii. 10-fold cross-validation was performed on all the three domain datasets from UCI ML Dataset.

iii. Four different supervised ML algorithms, namely, Naïve Bayes (NB), Logistic Regression (LR), Support Vector Machine (SVM), and Random Forest (RF) were used to compare proposed approach with existing feature selection techniques.
iv. Evaluation metrics like Accuracy, Precision, Recall, F-measure and Area Under ROC Curve (AUC) were used to compare the efficiency of the proposed feature selection approach.

The rest of the paper is organized as follows: Section 2 presents the related work in Sentiment Analysis. Section 3 describes the proposed methodology and the framework of Apriori based feature selection. Section 4 provides details about the experimental work. Section 5 discusses the results obtained. Finally, Section 6 presents future work and concludes the paper.

2. Related Work

Recently Social Media is being used to analyze data for more efficient decision making, and researchers are working on every aspect to improve the knowledge process [7]. There is much data available on Product Review Sites, Forums, Blogs, and Social Networking Websites like Facebook and Twitter. Manual analysis of this data is not possible due to its enormous size and volume. Therefore, Sentiment Classification is used to find the sentiment of the data using supervised ML algorithms. This section presents an overview of past research works based on Sentiment Classification using supervised ML and Feature Selection techniques [8,9].

Medhat et al. [10] carried out a comprehensive survey on new classification algorithms, applications, and improvements in the area of Sentiment Analysis. The author also sheds light on Emotion Detection, Transfer Learning, and Resource Building. In [11], the authors introduced a novel rule-based Sentiment classification of sentences from blog comments and reviews. SentiWordNet is used to obtain the polarity score, and their research work shows the effectiveness of the proposed method compared to ML-based methods. Our approach focuses on both Sentiment Score and ML Classifier algorithms. We are calculating the sentiment score using Vader API and applying four different supervised classifiers to predict the polarity of sentences.

In [12], Agarwal et al. use Twitter as the data source for sentiment analysis. The authors introduced the Part of Speech (POS) feature to carry out classification task. They used new features and tree kernel and showed that their approach works better than baseline techniques. In [13], the authors used three Machine Learning Algorithms, namely: Naïve Bayes Multinomial (NBM), Maximum Entropy (ME) and SVM for classification of data into respective polarity. They employed Unigram, Bigram and Hybrid N-Gram Features approach, and the results shows that NBM performed best according to their experiments. Dave et al. in [14] carried out sentiment classification work on CNET and Amazon reviews. They also used the N-Gram approach, but only the bigram and trigram feature set are considered for final evaluation. SVM and NB classifiers are used for training and testing the model.

One of the datasets used by us in this study is from the IMDB Movie Review, and this dataset is one of the most popular datasets used by researchers for the Sentiment Classification task. In [15], the authors also used IMDB Dataset and incorporated WordNet lexicon resource to extract opinion from review. Various ML Classifiers such as SVM, NB and Alternating Decision Tree are used to classify the dataset with more than 75% accuracy. In [16], Zhang et al. worked on Chinese Reviews of Clothing product using the word2vec approach. The authors proposed a classification approach using Semantic Approach and SVM. They used SVMpref, which is an alternative structural formulation of SVM used for Binary Classification. Semantic features of the reviews are extracted with the help of word2vec. The proposed approach shows better results in the sentiment classification task.

The authors in [17] proposed a sentiment classification approach using Fuzzy. The dataset used for the experiment is a movie review, and the results show considerable improvement in the accuracy with SVM. Before applying the supervised Machine Learning algorithm, data was preprocessed using POS Tagging, Term Frequency Inverse Document Frequency (TF-IDF), Stop Words Removal, Tokenization to obtain the final N-Gram feature set [18]. In [19], the authors used the concept of Ensemble Framework to carry out Sentiment Analysis. Three base classifiers used in work are NB, ME and SVM with various ensemble options: fixed combination, weighted combination and meta-classifier combination. The Highest accuracy of 88.65% was achieved on Kitchen Dataset. The authors in [20,21] also proposed using the Ensemble technique where SVM is chosen as a base classifier and other methods used are Boosting, Bagging, Random Subspace and Bagging Random Subspaces. The best results are obtained in ensemble techniques using random subspace and bagging subspace.

Earlier used techniques for feature selection are Chi-Square, Correlation, Information Gain, Relief F, etc. The authors in [22] used the mentioned feature selection techniques to select a subset based on the average weight approach. Sentiment Classification is then carried out using SVM and BN on the Arabic Review dataset. Feature Combination of various selection techniques is also a new trending research area in Sentiment Classification. The authors in [23] combine the feature selected from Chi-Square (CHI2), Information Gain (IG), Optimal orthogonal centroid (OCFS) and Document frequency difference (DFD) and implemented four Classifiers to carry out Sentiment Classification on English and Turkish review dataset. Agarwal et al. [4] proposed a novel Hybrid Merging method using Rough Set Theory and Information Gain. The proposed model was evaluated on different domain datasets using SVM and NB Supervised Machine Learning Classifiers.

From the above-related work, we found out that most of the work is carried out using predefined feature selection techniques such as Information Gain, Chi-square, PCA and more [26]. In this paper, we are proposing a new feature selection technique that deals with the association between words present in the sentence with the sentence's polarity.

Table 1 discusses the use of various feature selection approaches in sentiment classification. No work has been done on feature selection using Association Rule Mining (Apriori

Algorithm) in the sentiment classification field to the best of our knowledge. In this paper, a novel feature selection model using Apriori Algorithm is proposed in which the Support and Confidence value of the rule is considered to prune the word features.

Table 1. Review of Feature Selection Techniques in Sentiment Classification

S.No.	Feature Selection	Dataset	Algorithm Used	References
P1	Uni-gram, Bi-gram	Movie and E-Product Review	NB, SVM, Max Entropy	[33]
P2	IG+RSAR(Rough Set Attribute Reduction)	IMDB	SVM, NB	[4]
P3	IG, CHI, GINI	Movie Reviews	SVM, NB	[34]
P4	PCA, Relief, GA, IG	Credit Data	ANN-Bagging	[35]
P5	Unigram + overall opinion polarity (OvOp) concept	IMDB	NB	[36]
P6	Unigram+ Linearly combinable paired feature	IMDB	NB	[37]
P7	Feature Selection and Feature Weighing using CHI2 and TFIDF	IMDB	SVM	[38]

3. Proposed Feature Selection

The proposed Sentiment Classification approach is summarized as follows:

i. **Dataset Used:** In this paper, labeled datasets from the UCI ML repository are used to test our approach using five supervised ML algorithms. The dataset contains 1000 reviews with equal distribution of positive and negative reviews for three domains: IMDB Movie Review, AMAZON Product Review, and YELP Restaurant Review.

ii. **Data Preprocessing:** Following data preprocessing techniques are applied on the dataset to remove irrelevant, noisy entities.

   a. Removal of Stopwords
   b. Stemming is carried out using LovinsStemmer
   c. The sentence is broken into tokens
   d. TF-IDF for each word is calculated

iii. **Feature Extraction and Selection**: First, the feature vector is converted into a binary form to apply the proposed feature selection technique. The feature value will be *1* if the word is present in the sentence and *0* if not present. The output label (Neg or Pos) is also considered as the feature vector. Next, the proposed feature selection approach based on the Apriori algorithm is applied to select a reduced feature set.

iv. **Classification**: Finally, we train the supervised Machine Learning classifiers SVM[27], RF[28], LR[29], NB[30] with both reduced feature dataset and original feature dataset on reviews from different domains.

3.1 Methodology

In this paper, we are focusing on using the word frequency as items so that the Apriori algorithm [31] can be applied to generate Association rules between the words and sentence's polarity. This process is carried out in 4 steps: Data Collection, Data Preprocessing, Feature Selection using the proposed approach and classification using supervised ML algorithms. A generalized scheme of the proposed work is shown in Fig.1. The proposed approach works on Association Rule Mining, in which the Apriori algorithm is applied to words. For the Generation of Rules, data needs to be present in binary format. Therefore, each of the three datasets is tokenized using the relation between Word *Wi* and Review Text *T*. The association is shown in equation 1 and snapshot of the dataset in suitable form is shown in Fig 2.

$$Binary(Wi) = \begin{cases} 0 \ if \ Wi \notin T \\ 1 \ if \ Wi \in T \end{cases} \tag{1}$$

Apriori algorithm is then applied on the datasets and different support confidence values are used to generate the rules. In this work, support = 0.02 and confidence = 0.01 were used to generate the rules. These support and confidence values provide the optimum features. The formulae used to calculate support and confidence are shown in equation 2,3, and 4.

$$S(Wi) = \frac{NWi}{N} \tag{2}$$

$$Cpos(Wi \rightarrow pos) = \frac{S(Wi \ U \ Wpos)}{S(Wi)} \tag{3}$$

$$Cneg(Wi \rightarrow neg) = \frac{S(Wi \ U \ Wneg)}{S(Wi)} \tag{4}$$

where, *Wi* = tokenized words present in review; *N* is the total number of reviews and *NWi* is the total number of transactions containing word Wi.
S(Wi) = Support of each word present in the review
Wpos = Word ‘pos’ and *Wneg* = Word ‘neg’ present in the polarity of review text.
Cpos(Wi -> pos) = Confidence of Association Rule generated between Word Wi and Word pos
Cneg(Wi -> neg) = Confidence of Association Rule generated between Word Wi and Word neg.

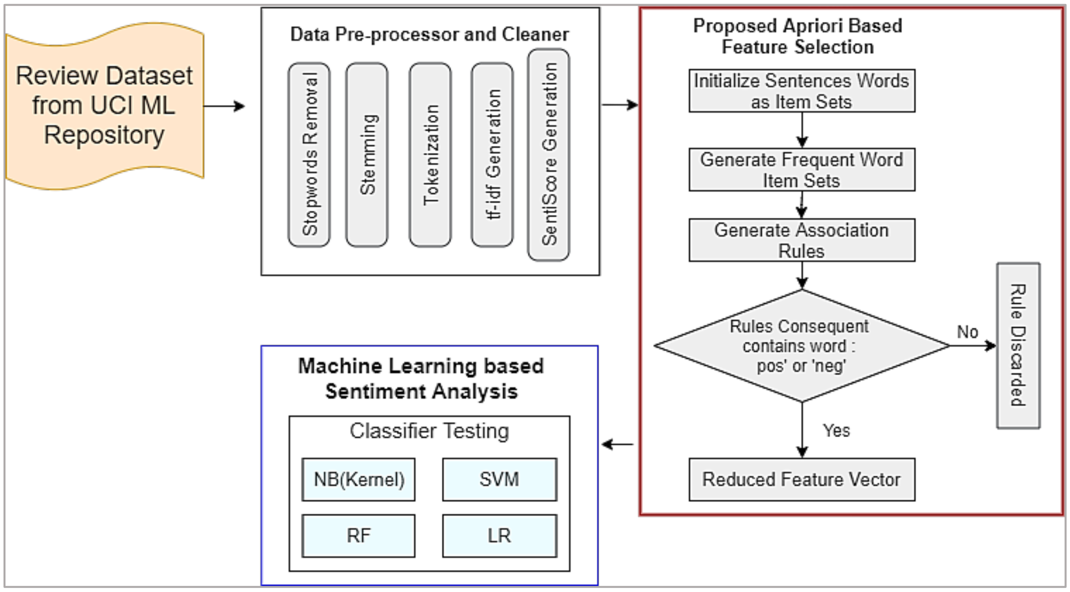


Figure 1. Proposed Feature Selection Technique using Apriori Algorithm

neat	need	neg	network	ngag	nic
0	0	1	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0
0	0	1	0	0	0
0	0	0	0	0	0
0	0	1	0	0	0
0	0	1	0	0	0
0	0	0	0	0	0
0	1	1	0	0	0
0	0	1	0	0	0

Figure 2. Representation of Sentence into Binary Tokenize form to generate rules

Criteria for consideration of rule for feature selection is simple. If the Consequent part of the rule contains either 'pos' or 'neg' sentiment class, then the rule is considered else it is discarded. Table [2,3,4] shows some of the proposed approach's rules whose algorithm is shown in Algorithm 1.

Table 2. Association rules generated on IMDB Movie Review Dataset

Rule	Consequent contains 'neg' or 'pos'	Decision
film->neg	Yes	Rule Considered
act->neg, bad	Yes	Rule Considered
movi->pos	Yes	Rule Considered
wast->tim	No	Rule Discarded

Table 3. Association rules generated on Yelp Restaurant Review Dataset

Rule	Consequent contains 'neg' or 'pos'	Decision
food-neg	Yes	Rule Considered
plac->neg	Yes	Rule Considered
eat->pos	Yes	Rule Considered
great->servic	No	Rule Discarded

Table 4. Association rules generated on Product Review Amazon Review Dataset

Rule	Consequent contains 'neg' or 'pos'	Decision
good->neg	Yes	Rule Considered
phone->neg,i	Yes	Rule Considered
work->pos	Yes	Rule Considered
lif->bat	No	Rule Discarded

Algorithm 1: Feature Selection using Apriori Algorithm

Input:
W_k = Words itemsets of k-gram
L_k = Frequent words itemsets of k-gram
L_1 = {frequent words}:
S_{min} = 0.02(Minimum Support Value)
C_{min} = 0.01(Minimum Confidence Value)
/ pos and neg are the polarity label that are considered as words itemsets*/*
Output:
F_r = Reduced Feature Set

Read Review Text **R_T**
For (k=1; L_k! = null; k=k+1) **do**
//Calculate Support for Each Frequent Word Itemset

$$S(L_k) = \frac{Transactions\ with\ L_k}{Total\ Transactions}$$

If $S(L_k) < Smin$
Discard **L_k**
Else
For each **R_T** in dataset **D** do
Increment count of Candidate in **W_{k+1}** that are present in **R_T**
L_{k+1} = Candidates in **C_{k+1}** with $\boldsymbol{Smin}$
For each (**L_{K+1}, pos**) and (**L_{K+1}, neg**) Association Rule **R**
Calculate Confidence of **R**

$$Cpos(L_{k+1} \rightarrow pos) = \frac{S(L_{k+1}\ U\ pos)}{S(L_{k+1})}$$

$$Cneg(L_{k+1} \rightarrow neg) = \frac{S(L_{k+1}\ U\ neg)}{S(L_{k+1})}$$

If $Cpos < Cmin$ or $Cneg < Cmin$
Discard **R**
Else
$\mathbf{F_r = F_r\ U\ L_{k+1}}$
End
End

4. Experiment Setup

The three datasets used in this work are processed in two different phases. In the first phase, data is preprocessed by tokenizing, stop words removal, stemming and extracting sentiment score using *Vader* API [32]. Once the data is cleaned, the TF-IDF value is calculated for each feature. In the second phase, the generation of Association Rules using the Apriori Algorithm is performed. All the rules not containing either 'neg' or 'pos' in the consequent part are discarded. The experiments are carried out on three different datasets from the UCI ML repository using tenfold (*k=10*) cross validation. The dataset is partitioned into two sets, where 9 folds (*k-1*) are used for training the model and 1-fold is used for testing.

4.1 Evaluation Parameters

In this paper, the performance of supervised ML algorithms is evaluated using the Confusion Matrix. There are four entities in the Confusion Matrix: True Positive (TP), False Positive (FP), True Negative (TN) and False Negative (FP). The Formula for each entity is shown in Table 5. The results are compared with the following evaluation metrics that are based on values of Confusion Matrix entities.

Table 5. Evaluation Parameter

Evaluation Metric	Formula
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$
Precision (*P*)	$\frac{TP}{TP+FP}$
Recall(*R*)	$\frac{TP}{TP+FN}$
F-measure	$2\frac{P.R}{P+R}$

4.2 ROC curve and AUC

ROC curve is a plot that depicts the performance of a classification model on different thresholds. The x-axis of the ROC Curve represents False Positive values and the y-axis represents True Positive values. Our classification problem is an example of binary classification, where higher Area under the Curve (AUC) value means better classification.

5. Results and Discussions

The experiment conducted in this study helps select features with a novel technique based on the Apriori algorithm. Experiments were conducted using three different datasets: IMDB Movie Review, Restaurant and Product Reviews, taken from the UCI ML repository. This section gives an in-depth analysis of results obtained by the proposed feature selection approach using four supervised classifiers: SVM, RF, NB and LR. First, the data is preprocessed to make it ready for classifiers. To preprocess the data, various operations are applied, such as Tokenizing the sentence into unigram, removing stop words, and stemming using *Lovins Stemmer* and representing feature vector as *TF-IDF*. The proposed feature selection approach is then applied to select the best features based on support and confidence values of the association rule. Finally, the classifiers SVM, RF, NB and LR are used on the original feature set (without feature selection) and on the reduced feature set (using proposed feature selection).

Table [6,7,8] shows the performance of ML classifiers with original features and reduced features (with the proposed scheme) on different domain datasets. Results show that using the proposed feature selection approach improves all the evaluation parameters for Sentiment Classification. To observe the impact of the proposed feature selection technique, ROC Curves are plotted and shown in Figure 3,4,5 for IMDB, AMAZON, and YELP Dataset respectively. To understand each supervised classifier's effect, each graph contains 4 ROC plots of the respective ML algorithm: SVM, NB, RF and LR.

Figure 6,7 and 8 show the accuracy comparison of ML Techniques on IMDB, Amazon and Yelp Reviews respectively. From the graphs we can observe that accuracy of Apriori based reduced feature set results in improved accuracy in all classifiers. For IMDB Movie Review and Amazon Product Review datasets, NB classifier shows maximum accuracy with a value of 78.4% and 81.08%. For Yelp Movie Review dataset, RF performs best with an Accuracy value of 77.6%. Apart from Accuracy, a detailed comparison of four ML classifiers was performed using Apriori based reduced feature set. Figure 9 shows Precision, Recall, and F-Measure scores for various classifiers.

Table 6. Results of the proposed model for IMDB movie review data set

ML Classifier	Unigram + SentiScore					Unigram + SentiScore + Proposed FS				
	Accuracy (%)	Precision	Recall	F-Measure	AUC	Accuracy (%)	Precision	Recall	F-Measure	AUC

LR	71.70	0.7304	0.6880	0.7085	0.729	78.20	0.8026	0.7480	0.7743	0.861
SVM	75.90	0.8278	0.6540	0.7307	0.856	76.50	0.8193	0.6800	0.7432	0.859
RF	70.90	0.7029	0.7240	0.7133	0.814	76.20	0.7120	0.8800	0.7871	0.863
NB	78.30	0.8137	0.7340	0.7718	0.847	78.40	0.8060	0.7480	0.7759	0.866

Table 7. Results of the proposed model for Amazon Product review data set

ML Classifier	Unigram + SentiScore					Unigram + SentiScore + Proposed FS				
	Accuracy (%)	Precision	Recall	F-Measure	AUC	Accuracy (%)	Precision	Recall	F-Measure	AUC
LR	65.87	0.6501	0.7098	0.6743	0.636	80.88	0.8220	0.7873	0.8024	0.889
SVM	80.48	0.4935	0.8228	0.8065	0.887	78.78	0.8791	0.6628	0.7537	0.888
RF	80.57	0.8217	0.7904	0.8013	0.890	80.98	0.8117	0.8097	0.8085	0.894
NB	78.98	0.7878	0.7851	0.7851	0.856	81.08	0.8291	0.7818	0.8025	0.901

Table 8. Results of the proposed model for Yelp Restaurant review data set

ML Classifier	Unigram + SentiScore					Unigram + SentiScore + Proposed FS				
	Accuracy (%)	Precision	Recall	F-Measure	AUC	Accuracy (%)	Precision	Recall	F-Measure	AUC
LR	60.90	0.6096	0.6225	0.6136	0.569	77.30	0.7547	0.8136	0.7805	0.861
SVM	76.10	0.7650	0.7619	0.7600	0.845	74.80	0.6908	0.9043	0.7810	0.852
RF	76.70	0.7581	0.7989	0.7728	0.865	77.60	0.7557	0.8231	0.7847	0.869
NB	74.60	0.7480	0.7502	0.7457	0.793	77.10	0.7523	0.8124	0.7788	0.850

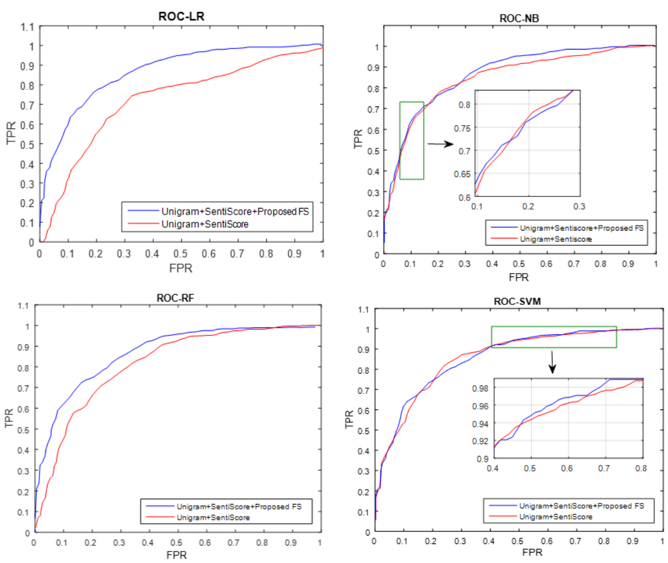
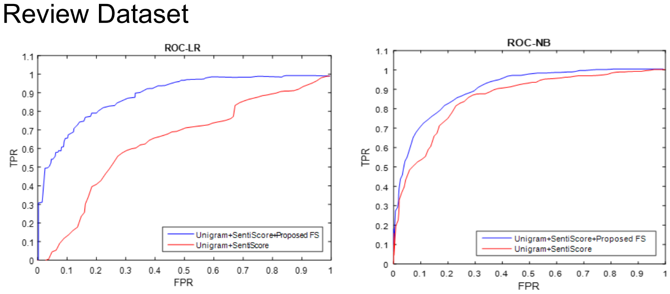


Figure 3. ROC Comparison Chart for IMDB Movie Review Dataset



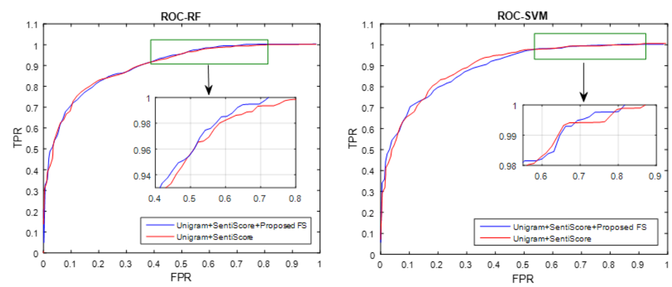


Figure 4. ROC curves for Amazon Product Review Dataset

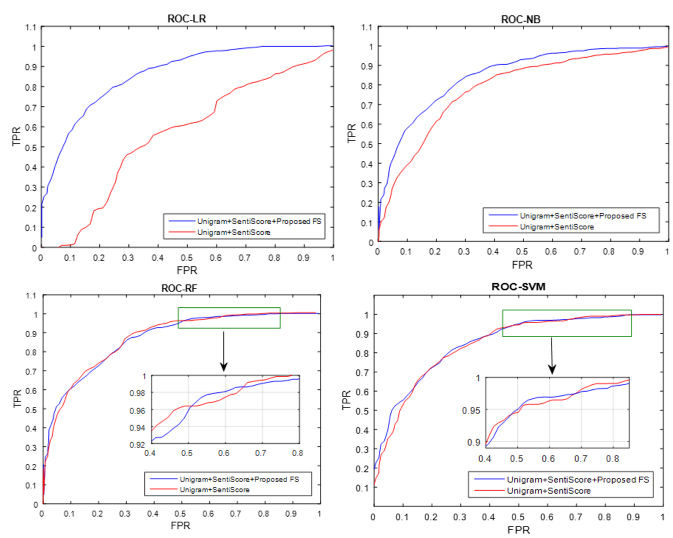


Figure 5. ROC curves for Yelp Restaurant Review Dataset

For IMDB Movie Review dataset, it was observed from Table 6 and Fig. 9, that SVM has the maximum Precision score of 0.8193 followed by NB having a score of 0.806. From the Accuracy curve, it was found that NB shows maximum Accuracy with 78.4% which is aligned with the results obtained for Precision. For Recall, it was observed that RF outperform all other classifiers with a score of 0.88 followed by NB and LR with a score of 0.748. For F-measure, RF scored the maximum with a value of 0.7871 which is closely followed by all other classifiers.

For the Amazon dataset, it was observed from Table 7 and Fig 9, that SVM has the highest Precision value of 0.8791, closely followed by other three classifiers. For Recall and F-measure, RF shows the best results with values 0.8097 and 0.8085 respectively. Among the other three classifiers SVM shows lowest value of 0.6628 and 0.7537 for Recall and F-measure respectively.

Lastly, for the Yelp dataset, it was observed from Table 8 and Fig 9, that LR, RF and NB show closely related performance for Precision value in the range of 0.75 and SVM shows the lowest value of 0.6908. For Recall, SVM achieved the highest value of 0.9043 followed by other three ML classifiers. There is substantial difference in the Recall scores achieved by SVM and other three classifiers. For F-measure all four classifiers show similar results in the range of 0.78.

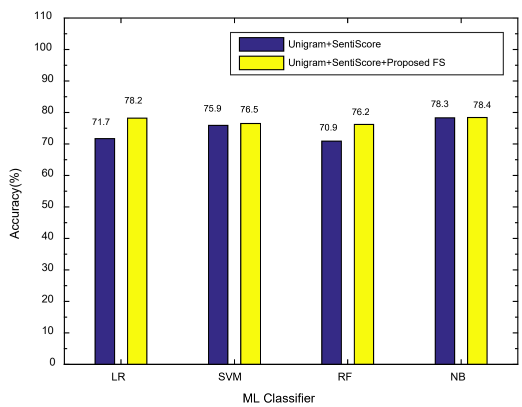


Figure 6. Accuracy Comparison of feature selection techniques on IMDB Movie Reviews

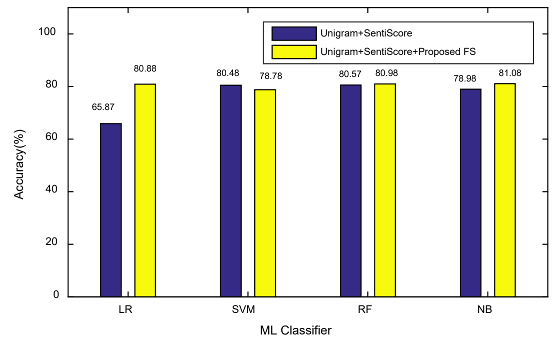


Figure 7. Accuracy Comparison of feature selection techniques on Amazon Product Reviews

Confusion Matrix of the proposed feature selection approach is shown in Table 9,10,11 for IMDB Movie Review dataset, Amazon Product Review dataset and Yelp Restaurant Review dataset respectively. It was observed that maximum reduction in error rate for all datasets considered in this paper is found for LR with a value of 6.5%, 15.01% and 16.4% for IMDB, Amazon and Yelp datasets respectively.

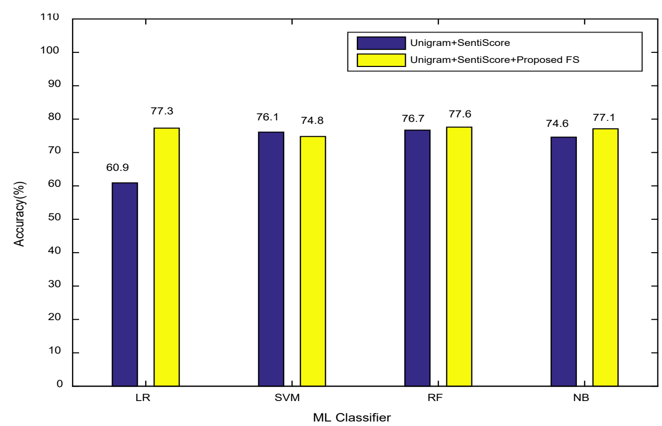


Figure 8. Accuracy Comparison of feature selection techniques on Yelp Restaurant Reviews

5.1 Comparison with Existing Approach

This section demonstrates that the proposed Apriori-based feature selection approach is more efficient than the Genetic Algorithm(GA) based approach used in [25]. In [25] the authors used the UCI ML datasets belonging to the same domains as used in this paper. The authors proposed GA based feature reduction (GA-FS) technique. Figure 10,11 and 12 shows the Accuracy comparison between the FS approach proposed in this paper and GA-FS.

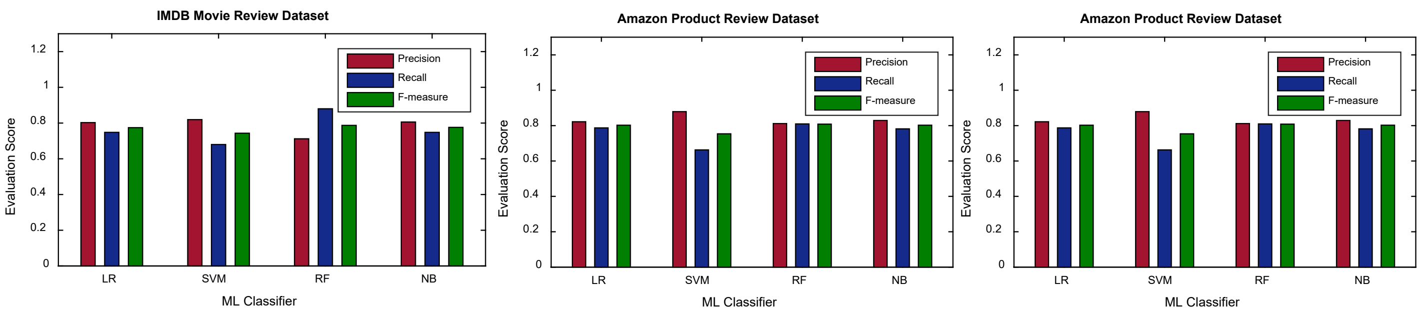


Figure 9. Precision, Recall and F-measure Comparison for Proposed FS

Table 9. Confusion Matrix of IMDB Movie Review

Class/method	LR		SVM		RF		NB	
	True pos.	True neg	True pos.	True neg	True pos.	True neg	True pos.	True neg
Unigram + SentiScore								
Predicted pos.	344	127	327	68	362	153	367	84
Predicted neg.	156	373	173	432	138	347	133	416
Error rate	28.3%		24.1%		29.1%		**21.7%**	
Unigram + SentiScore + Proposed FS								
Predicted pos.	374	92	340	75	437	178	374	90
Predicted neg.	126	408	160	425	63	322	126	410
Error rate	21.8%		23.5%		23.8%		**21.6%**	

Table 10. Confusion Matrix of Amazon Product Review

Class/method	LR		SVM		RF		NB	
	True pos.	True neg	True pos.	True neg	True pos.	True neg	True pos.	True neg
Unigram + SentiScore								
Predicted pos.	353	195	412	107	394	89	393	104
Predicted neg.	147	305	88	393	106	411	107	396
Error rate	34.13%		19.52%		**19.43%**		21.02%	
Unigram + SentiScore + Proposed FS								
Predicted pos.	395	86	333	45	404	95	391	81
Predicted neg.	105	414	167	455	96	405	109	419
Error rate	19.12%		21.22%		19.02%		**18.92%**	

Table 11. Confusion Matrix of Yelp Restaurant Review

Class/method	LR		SVM		RF		NB	
	True pos	True neg.	True pos	True neg.	True pos	True neg.	True pos	True neg.
Unigram + SentiScore								
Predicted pos.	297	188	382	121	370	103	373	127
Predicted neg.	203	312	118	379	130	397	127	373
Error rate	39.1%		23.9%		**23.3%**		25.4%	
Unigram + SentiScore + Proposed FS								
Predicted pos.	367	94	297	49	367	91	366	95
Predicted neg.	133	406	203	451	133	409	134	405
Error rate	22.7%		25.2%		**22.4%**		22.9%	

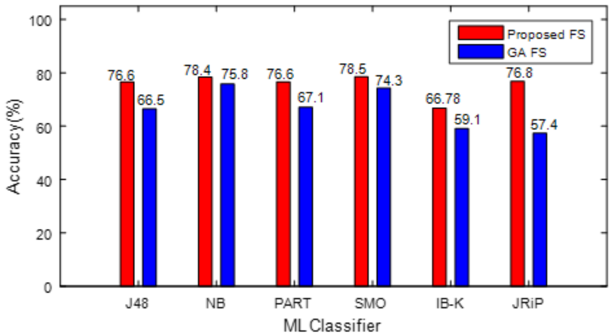


Figure 10. IMDB Dataset Accuracy Comparison

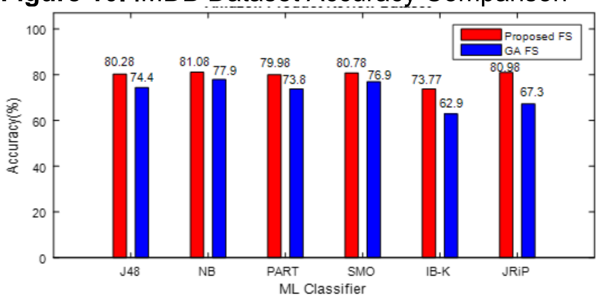


Figure 11. Amazon Dataset Accuracy Comparison

5.2 Comparison of Proposed Feature Selection with Other Feature Selection techniques

In the final part of our evaluation, we demonstrate that our Apriori based feature selection techniques perform better than other feature selection techniques such as PCA, Chi-Square and Relief. Figure 14 shows the accuracy graph of all four feature selection techniques on three datasets i.e. IMDB Movie Review, Amazon Product Review and Yelp Restaurant Review. Naïve Bayes classifier is used to compare our proposed feature selection with other feature selection techniques. As we can see that ARM based feature reduction technique has better accuracy in all three datasets. For IMDB Movie review dataset we achieve same accuracy score of 78.4% for both ARM based and Chi Square feature selection. For Amazon and Yelp Dataset our proposed approach is giving maximum accuracy score of 81.08% and 77.1% respectively.

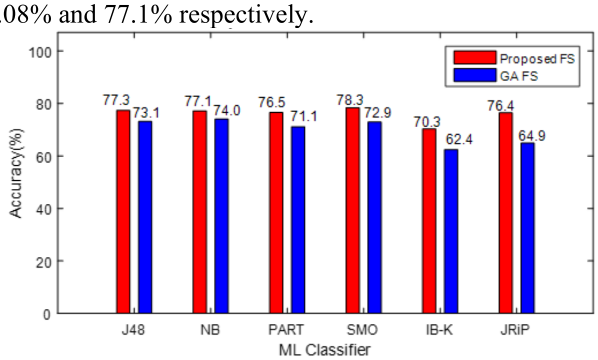


Figure 12. Yelp Dataset Accuracy Comparison

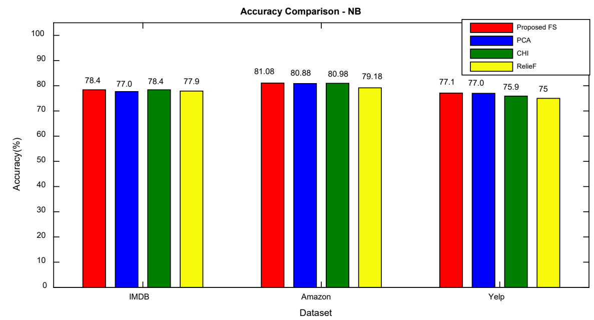


Figure 14. Accuracy comparison of ARM based feature selection with PCA, Chi Square and Relief on IMDB, Amazon and Yelp Review Dataset

6. Conclusion and Future Scope

This paper proposed a novel feature selection approach based on the Apriori algorithm for performing Sentiment Classification. We employed four Different Classifiers i.e., SVM, NB, LR and RF to compare our proposed approach on dataset with proposed feature selection and without feature selection. Detailed analysis of results shows that Naïve Bayes classifier achieved maximum accuracy in 78.4% and 81.08% for IMDB and Amazon datasets respectively. While in case of Yelp dataset, Random Forest classifier outperforms other classifiers, achieving an Accuracy score of 77.6%. Further, the proposed approach's results were compared with [25] using six classifiers (J48, NB, PART, SMO, IB-K, and JRiP). The proposed approach manages to outperform the GA based approach by an average of 9.78% increase in Accuracy for all the three datasets. Detailed analysis of results shows that JRiP classifier achieves maximum Accuracy increase of 25.26%, 16.89%, and 15.05% for IMDB, Amazon and Yelp dataset respectively. The proposed approach is also compared with existing feature selection techniques: PCA, CHI2 and RF. The results show 0.41%, 0.557% and 1.87% average accuracy increase than PCA, CHI2 and RF feature selection respectively. The results achieved during this study strengthen the claim that the proposed feature selection technique gives better accuracy than existing feature selection techniques and also helps in reducing dataset by considerable size.

This work provides several interesting future directions. The proposed approach can be used to generate n-gram (Bigram, Trigram….) feature set using unigram dataset. This will help in reducing feature preprocessing time by considerable amount. Then we will incorporate other features also such as Part of Speech(POS), Negation handling to build fused dataset.

References

[1] Pang, B., & Lee, L. (2008). Foundations and Trends® in Information Retrieval. Foundations and Trends® in Information Retrieval, 2(1-2), 1-135.

[2] Davies, A., & Ghahramani, Z. (2011). Language-independent Bayesian sentiment mining of Twitter. In The 5th SNA-KDD Workshop'11 (SNA-KDD'11).

[3] Prabowo, R., & Thelwall, M. (2009). Sentiment analysis: A combined approach. Journal of Informetrics, 3(2), 143-157.

[4] Agarwal, B., & Mittal, N. (2013, June). Sentiment classification using rough set based hybrid feature selection. In Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis (pp. 115-119).

[5] Govindarajan, M. (2014). Sentiment classification of movie reviews using hybrid method. International Journal of Advances in Science Engineering and Technology, 1(3), 73-77.

[6] (2015). UCI ML Repository_Sentiment Analysis Dataset. Accessed: Jan. 8, 2020. [Online]. Available: http://archive.ics.uci.edu/ml/datasets/Sentiment+Labelled+Sentences

[7] Dhaoui, C., Webster, C. M., & Tan, L. P. (2017). Social media sentiment analysis: lexicon versus machine learning. Journal of Consumer Marketing.

[8] Samal, B., Behera, A. K., & Panda, M. (2017, May). Performance analysis of supervised machine learning techniques for sentiment analysis. In 2017 Third International Conference on Sensing, Signal Processing and Security (ICSSS) (pp. 128-133). IEEE..

[9] Catal, C., & Nangir, M. (2017). A sentiment classification model based on multiple classifiers. Applied Soft Computing, 50, 135-141.

[10] Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. Ain Shams engineering journal, 5(4), 1093-1113.

[11] Khan, A., Baharudin, B., & Khan, K. (2011). Sentiment Classification Using Sentence-level Lexical Based. Trends in Applied Sciences Research, 6(10), 1141-1157.

[12] Agarwal, A., Xie, B., Vovsha, I., Rambow, O., & Passonneau, R. J. (2011, June). Sentiment analysis of twitter data. In Proceedings of the workshop on language in social media (LSM 2011) (pp. 30-38).

[13] Boiy, E., Hens, P., Deschacht, K., & Moens, M. F. (2007, June). Automatic Sentiment Analysis in On-line Text. In ELPUB (pp. 349-360).

[14] Dave, K., Lawrence, S., & Pennock, D. M. (2003, May). Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In Proceedings of the 12th international conference on World Wide Web (pp. 519-528).

[15] Annett, M., & Kondrak, G. (2008, May). A comparison of sentiment analysis techniques: Polarizing movie blogs. In Conference of the Canadian Society for Computational Studies of Intelligence (pp. 25-35). Springer, Berlin, Heidelberg.

[16] Zhang, D., Xu, H., Su, Z., & Xu, Y. (2015). Chinese comments sentiment classification based on word2vec and SVMperf. Expert Systems with Applications, 42(4), 1857-1863.

[17] Mouthami, K., Devi, K. N., & Bhaskaran, V. M. (2013, February). Sentiment analysis and classification based on textual reviews. In 2013 international conference on Information communication and embedded systems (ICICES) (pp. 271-276). IEEE.

[18] Ye, Q., Zhang, Z., & Law, R. (2009). Sentiment classification of online reviews to travel destinations by supervised machine learning approaches. Expert systems with applications, 36(3), 6527-6535.

[19] Ghosh, M., & Sanyal, G. (2018). An ensemble approach to stabilize the features for multi-domain sentiment analysis using supervised machine learning. Journal of Big Data, 5(1), 44.

[20] Manek, A. S., Shenoy, P. D., Mohan, M. C., & Venugopal, K. R. (2017). Aspect term extraction for sentiment analysis in large movie reviews using Gini Index feature selection method and SVM classifier. World wide web, 20(2), 135-154.

[21] Whitehead, M., & Yaeger, L. (2010). Sentiment mining using ensemble classification models. In Innovations and advances in computer sciences and engineering (pp. 509-514). Springer, Dordrecht.

[22] Adel, A., Omar, N., & Al-Shabi, A. (2014). A Comparative Study Of Combined Feature Selection Methods For Arabic Text Classification. J. Comput. Sci., 10(11), 2232-2239.

[23] Parlar, T., Özel, S. A., & Song, F. (2018). QER: a new feature selection method for sentiment analysis. Human-centric Computing and Information Sciences, 8(1), 10.

[24] Hatzivassiloglou, V., & McKeown, K. (1997, July). Predicting the semantic orientation of adjectives. In 35th annual meeting of the association for computational linguistics and 8th conference of the european chapter of the association for computational linguistics (pp. 174-181).

[25] Iqbal, F., Hashmi, J. M., Fung, B. C., Batool, R., Khattak, A. M., Aleem, S., & Hung, P. C. (2019). A hybrid framework for sentiment analysis using genetic algorithm based feature reduction. IEEE Access, 7, 14637-14652.

[26] Jain, A., & Jain, V. (2019). Sentiment classification of twitter data belonging to renewable energy using machine learning. Journal of Information and Optimization Sciences, 40(2), 521-533.

[27] Xia, H., Yang, Y., Pan, X., Zhang, Z., & An, W. (2020). Sentiment analysis for online reviews using conditional random fields and support vector machines. Electronic Commerce Research, 20(2), 343-360.

[28] Jonathan, B., Sihotang, J. I., & Martin, S. (2019, December). Sentiment Analysis of Customer Reviews in Zomato Bangalore Restaurants Using Random Forest Classifier. In Abstract Proceedings International Scholars Conference (Vol. 7, No. 1, pp. 1719-1728).

[29] Al Omari, M., Al-Hajj, M., Hammami, N., & Sabra, A. (2019, April). Sentiment classifier: Logistic regression for arabic services' reviews in lebanon. In 2019 International Conference on Computer and Information Sciences (ICCIS) (pp. 1-5). IEEE..

[30] Hasanli, H., & Rustamov, S. (2019, October). Sentiment Analysis of Azerbaijani twits Using Logistic Regression, Naive Bayes and SVM. In 2019 IEEE 13th International Conference on Application of Information and Communication Technologies (AICT) (pp. 1-7). IEEE.

[31] Agarwal, R. C., Aggarwal, C. C., & Prasad, V. V. V. (2001). A tree projection algorithm for generation of frequent item sets. Journal of parallel and Distributed Computing, 61(3), 350-371.

[32] Gilbert, C. H. E., & Hutto, E. (2014, June). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In Eighth International Conference on Weblogs and Social Media (ICWSM-14). Available at (20/04/16) http://comp. social. gatech. edu/papers/icwsm14. vader. hutto. pdf (Vol. 81, p. 82).

[33] Xia, R., Zong, C., & Li, S. (2011). Ensemble of feature sets and classification algorithms for sentiment classification. Information Sciences, 181(6), 1138–1152. doi:10.1016/j.ins.2010.11.023

[34] Al-Moslmi, T., Gaber, S., Al-Shabi, A., Albared, M., & Omar, N. (2015). Feature selection methods effects on machine learning approaches in malay sentiment analysis. Proc. 1st ICRIL-Int. Conf. Inno. Sci. Technol.(IICIST) (pp. 1-2). Academic Press

[35] Koutanaei, F. N., Sajedi, H., & Khanbabaei, M. (2015). A hybrid data mining model of feature selection algorithms and ensemble learning classifiers for credit scoring. Journal of Retailing and Consumer Services, 27, 11–23. doi:10.1016/j.jretconser.2015.07.003

[36] Salvetti, F., Lewis, S., & Reichenbach, C. (2004). Automatic opinion polarity classification of movie reviews. Colorado research in linguistics, 17

[37] Beineke, P., Hastie, T., & Vaithyanathan, S. (2004, July). The sentimental factor: Improving review classification via human-provided information. In Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL-

04) (pp. 263-270).
[38] Ukhti Ikhsani Larasati, I. U., Much Aziz Muslim, I. U., Riza Arifudin, I. U., & Alamsyah, I. U. (2019). Improve the Accuracy of Support Vector Machine Using Chi Square Statistic and Term Frequency Inverse Document Frequency on Movie Review Sentiment Analysis. Scientific Journal of Informatics, 6(1), 138-149.