Ghosh and Sanyal *J Big Data* (2018) 5:44
https://doi.org/10.1186/s40537-018-0152-5

Journal of Big Data

RESEARCH Open Access

An ensemble approach to stabilize the features for multi-domain sentiment analysis using supervised machine learning

Monalisa Ghosh* and Goutam Sanyal

*Correspondence: monalisa_05mca@yahoo.com
Dept. of Computer Science and Engineering, National Institute of Technology, Durgapur, West Bengal, India

Abstract

Sentiment classification or sentiment analysis has been acknowledged as an open research domain. In recent years, an enormous research work is being performed in these fields by applying numerous methodologies. Feature generation and selection are consequent for text mining as the high dimensional feature set can affect the performance of sentiment analysis. This paper investigates the inability of the widely used feature selection method (IG, Chi Square, Gini Index) individually as well as their combined approach on four machine learning classification algorithm. The proposed methods are evaluated on three standard datasets viz. IMDb movie review, electronics and kitchen product review dataset. Initially, select the feature subsets from three different feature selection methods. Thereafter, statistical method UNION, INTERSECTION and revised UNION method are applied to merge these different feature subsets to obtain all top ranked including common selected features. Finally, train the classifier SMO, MNB, RF, and LR (logistic regression) with this feature vector for classification of the review data set. The performance of the algorithm is measured by evaluation methods such as precision, recall, F-measure and ROC curve. Experimental results show that the combined method achieved best accuracy of 92.31 with classifier SMO, which is encouraging and comparable to the related research.

Keywords: Sentiment classification, Subjective information, Machine learning classification algorithm, Feature selection method, N-gram method

Introduction

An opinion is a viewpoint or judgment about a specific thing that acts as a key influence on an individual process of decision making. People's belief and the choices they make are always dependent on how others see and evaluate the world. So opinion holds high values in many aspect of life. Sentiment analysis is the process of determining opinions or sentiments in textual documents as positive, or negative. In recent years, this field is widely appreciated by researchers due to its dynamic range of application in various numbers of fields. There are several areas such as marketing; politics; news analytics etc. which are benefited from the result of sentiment analysis. Due to the vast range of movies these days, it has become difficult for the audience to select their preferred genre of movie. Movie reviews turn out to be very useful

reference. Despite of the willingness of people to share their thoughts and views about the movies, a problem persists due to the huge amount of total reviews. This develops a need for technology of data mining to uncover information. These solutions can be roughly categorized into machine learning approach and lexicon-based approach to solve the problem of sentiment classification. The former approach was applied to classify the sentiments based on trained as well as test data sets. The second category doesn't require any prior training data set, it performs the task by identifying a list of words, phrases that consists of a semantic value. It mainly concentrates on patterns of unseen data. There are few researchers applied hybrid approaches [1, 2] by combining both approaches machine learning and lexical to improve the sentiment classification performance.

This field becomes more challenging due to the fact that many demanding and interesting research problems still exist in this field to solve. Sentiment based analysis of a document is quite tough to perform in comparison with topic based text classification. The opinion words and sentiments are always varied with situations. Therefore, an opinion word can be considered as positive in one circumstance but may be that becomes negative in some other circumstance. The opinionated word '*unpredictable*' is used in different senses in a different domain. For example, "*an unpredictable plot in the movie*" *gives* a positive opinion about the movie, while "*an unpredictable steering wheel*" is a negative expression considering the product, car [3].

Sentiment classification process has been classified into three levels: document level, sentence level, and feature level. The entire document at the document level, based on the positive or negative opinion, is expressed by the authors. Sentiment classification at the sentence level, considers the individual sentence to identify whether the sentence is positive or negative. In feature level, we classify the sentiment with respect to the specific aspects of entities. Aspect level sentiment classification needs deeper analysis on features, mainly which are expressed implicitly and usually are hidden in a large text dataset. During this study, the focus has been made on feature level sentiment classification. We present the impact of supervised learning method on labelled data.

The main contribution of the paper can be stated in particular as:

1. We provide a novelty sentiment classification method based on feature selection and ML technique and the proposed method evaluate on three standard benchmark datasets such as: movie reviews of IMDb, Electronics and kitchen review datasets. We carried out experiments considering the 10-fold cross validation, as product review dataset consists of separate files for positive and negative reviews but training and testing data are not isolated.

   For movie review dataset, 25,000 samples are categorized as for training and another 25,000 for testing purpose. However, we noticed the distribution is sub-optimal since the training samples are not sufficient according to 25,000 testing reviews. Finally, to improve the performance of classifier we decided to use cross validation for movie as well as product review datasets.
2. We employed IG method as a single univariate method with low complexity, which ranks the features based on high information gain entropy in decreasing order. However, the IG method cannot handle redundant features. We addressed this problem

by considering CHI and Gini Index as a multivariate and mutual information based method to find and filter the redundancy among relevant features. To achieve better accuracy we combine univariate and multivariate method by applying some statistical method UNION, INTERSECTION and revised UNION.
3. We trained the above feature representation on four different classification models namely, SMO, MNB, LR and RF to classify the sentiment polarity of review datasets.
4. Finally, the performance of the proposed approach has compared based on evaluation parameters like precision, recall, F-measure and AUC with the results in an existing work obtained by different researchers.

The rest of the paper is constructed as the following: "Related work" section consists of the existing literature that can connect to our approach. Then "Proposed approach" section describes the approaches used in this paper for polarity detection. "Methodology" section and "Combination of feature selection methods" section explains methodology includes features and proposed feature selection technique. The detail regarding implementation of proposed classification algorithm discussed in "Classification". The particulars about experiments and results are expounded in "Experiments and results" section. Finally, "Conclusion" section concludes with a discussion of the proposed method and with ideas on future steps.

Related work

In current years, sentiment analysis of social media [4] content has become one of the most sought area among researchers because the number of product review sites, social networking sites, blogs, forums are developing extensively. This field mainly utilizes supervised, unsupervised and semi supervised technique for sentiment prediction and classification task. In this section we provide a brief overview of the previous studies regarding [5] supervised multiple machine learning (ML) algorithms [6].

Boiy et al. [7] employed three different ML algorithms such as SVM, NBM and ME. They considered N-gram features such as Unigram, Bigram and their combination. The performance of NBM algorithm was convincing according their analysis. Research work of Dave et al. [8] used some tools for analysis the reviews from Amazon and CNET for classification. They select bigram and trigram features using N-gram model and some scoring methods are applied finally to determine whether the review holds positive or negative opinion. SVM and NB classifier were implemented for sentence level classification with the accuracy of 87.0.

The movie reviews dataset IMDb was used to study by Annett and Kondrak [9]. They adopted lexical resource WordNet for sentiment extraction. Different classifier such as SVM, NB, alternating decision tree used for review classification and more than 75% accuracy was achieved.

Zhang et al. [10] proposed a classification approach of Chinese reviews on clothing product. They applied word2vec and SVM^{pref} technique while word2vec helps to capture the semantic features based on semantic relationship. SVM^{pref} is nothing but an alternative structural formulation of SVM optimization problem for binary classification. They achieved good outcomes of this combination for sentiment classification. Mouthami et al. [11] proposed new approach as sentiment fuzzy classification algorithm on the

movie review dataset to improve the classification accuracy. Preprocessing method tokenization, stop word removal, TF-IDF, and POS tagging are used for initial pruning [12] they researched on travel blogs and applied various machine learning algorithm NB, SVM by considering the N-gram model to obtain the feature set. In this study, SVM worked best with 85.14% accuracy.

The author [13] approached an ensemble framework to perform sentiment analysis by combining different feature subsets and classification algorithms. The feature selection techniques, they applied are POS based feature sets and the word-relation based feature sets and thereafter these features are fed to three base classifiers such as, NB, ME and SVM. The aim of this paper is to perform the sentiment classification by employing three types of ensemble methods, including the fixed combination, weighted combination and meta-classifier combination. The highest accuracy they achieved 88.65 with kitchen dataset.

Whitehead et al. [14, 15] proposed to apply SVM as a base classifier with four different ensemble techniques such as boosting, bagging, random subspace, and bagging random subspaces. They achieved best performance through random subspace and bagging subspace method.

This research work [16] investigated on the behaviour of five feature selection method such as: Chi square, Correlation, GSS Coefficient Information Gain and Relief F. The final feature subset has been selected based on the average weight of the features assigned by combined feature selection method. SVM and NB classifier are employed to classify the sentiment of Arabic review corpus. The authors claimed that combined feature selection method outperformed the individual method with SVM classification algorithm.

The researchers [17] obtained the highest accuracy 86.9 after combining the feature selection method CHI, DFD and OCFS. They implemented a maximum entropy modelling (MEM) classifier to accomplish sentiment classification and the performance of classifier evaluated on movie review dataset with fivefold cross validation.

Agarwal et al. [1] have proposed a hybrid method merging rough set theory and Information Gain for sentiment classification. These methods are evaluated on four standard datasets such as: Movie review (IMDb) and product (book, DVD, and electronics) review dataset. SVM and NB classifier is used with tenfold cross validation for classifying sentiment polarity of review documents. F1-measure value is considered as a performance measure with maximum 87.7 and 80.9 for SVM and NB classifier.

Kolog et al. [18] implemented machine learning techniques to perceive sentiments in text form, regarding social influences on student's life story. They applied k-means algorithm for clustering purpose and after that the main influences are identified and those are considered as class level for classification task. The supervised machine learning classifier MNB, SMO and J48 are employed with tenfold cross validation to detect the sentiment either as positive or negative.

The feature selection stage primarily helps in refining features, which are considered as input for classification task. Feature selection is definitely a beneficial task considered by Narayanan et al. [19] based on the experimental result. They have applied only Mutual Information feature selection method with Naïve Bayes (NB) classifier in the domain of movie review.

Amolik et al. [20] proposed a model for sentiment prediction of more than 21,000 tweets by applying the machine learning classifier SVM and NB. Feature vectors were also created to handle the problem of repeating characters in Twitter. They achieved the higher accuracy with SVM was 75% in comparison with NB (65%) by using evaluation matrices precision and recall. A huge number of research papers with different ML classifiers namely Naive Bayes (NB) [11, 21] Support Vector Machine (SVM) [10, 22–24], maximum entropy [17, 25, 26], decision trees [9, 21, 27] have been used mostly to build classification model in different domain (Table 1).

Proposed approach

The proposed classification methods are summarized into several steps as described below:

1. *Data collection* In this work, movie review database (IMDB) and product review (Electronics, Kitchen) database are considered to solve the problem regarding sentiment classification.
2. *Pre-processing* This technique is required to remove noisy, inconsistent and incomplete information by considering tokenization, stop words removal, stemming method.

Table 1 Research work related to machine learning classifiers for sentiment analysis

Author/year	Technical approach	Accuracy in %	Dataset domain
Pang et al. (2002) [25]	Applied N-gram model with NB, SVM, ME	77.4–82.9	Internet Movie Database (IMDb)
Dave et al. (2003) [8]	Used N-gram model for feature extraction with SVM, NB classifier	87.0	Product review from Amazon and CNET
Annett and Kondrak (2008) [9]	Considered WordNet as Lexical resource with SVM, NB, Decision Tree classifier	75.0	Movie reviews (IMDb)- 1000 (+) and 1000 (−) reviews
Ye et al. (2009) [12]	NB, SVM classifier used for classification	85.14	Travel blogs
Mouthami et al. (2013) [11]	TF-IDF and POS tagging with fuzzy classification algorithm	87.4	Movie review dataset
Zha et al. (2014) [17]	SVM, NB, ME classifier adopted with evaluation matrices F1-Measure	83.0–88.43	Customer reviews (feedback)
Habernal et al. (2014) [26]	N-gram and POS related features and emoticons are selected using MI, CHI, OR, RS method. Classifier ME and SVM used for classification	78.50	Dataset from social media
Zhang et al. (2015) [10]	Use word2vec for features with SVM classifier for classification	89.95–90.30	Chinese review dataset
Luo et al. (2016) [21]	First transform the text into low dimensional emotional space (ESM), next implement SVM, NB, DT classifier	63.28–79.21	Stock message text data

3. *Feature extraction and selection* Initially, to create a feature vector with numeric value we consider binary presence or absence of a feature in a document. The features score will be 1 if it presents in a document otherwise score will be 0. Next, feature selection method IG, CHI, and Gini Index applied to select different feature subsets. Then to generate a top ranked feature sub list we combined all three feature subsets.
4. *Classification* Finally, train the supervised machine learning classifiers SMO, MNB, RF and LR with the different feature sub list for different domain.

Methodology

Text classification as a research field was introduced long time ago [28], however, sentiment based categorization was initiated more recently [18, 25, 29]. The main purpose of this research work is to investigate the performance of various machine learning classifier (MLC) with three combined feature set. The whole process can be completed in four step including Data acquisition, pre-processing, feature selection and classification. A general overview of the proposed framework is introduced with Fig. 1, and the following subsections consist of a detailed description about each preliminary function.

Dataset preparation

We conducted experiments on movie review data set [30], which were prepared by Pang and Lee 2004 [29]. This study uses movie review and product review dataset (Electronics and kitchen) to perform sentiment classification task. The movie review dataset is one of the popular benchmark dataset, which has been exploited by several researchers in order to analyze the experimental outcomes. The standard movie review dataset consists of overall 2000 reviews where 1000 reviews are tagged as positive and 1000 s are negative. The amazon products review dataset [31] provided by Blitzer et al. [32] are considered for investigation and we adopted the data set of Electronics and Kitchen domain

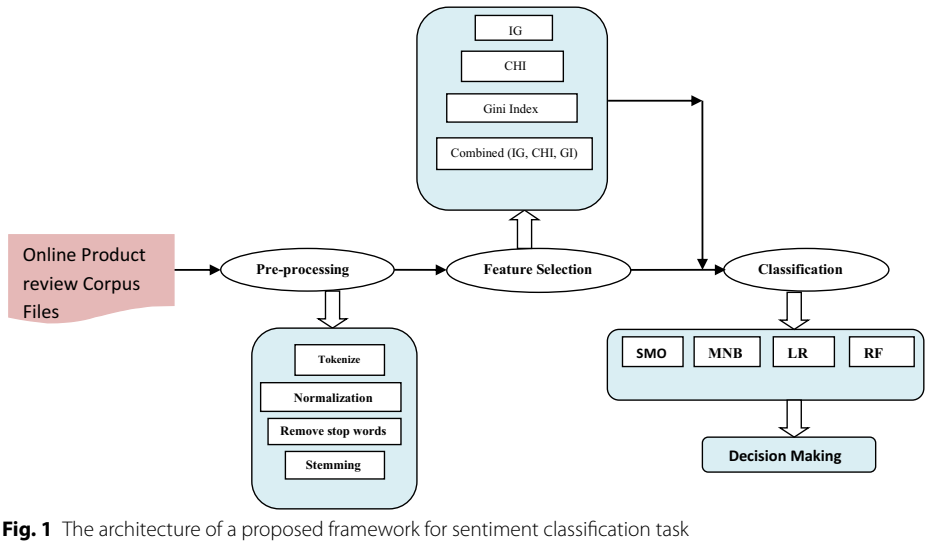

Fig. 1 The architecture of a proposed framework for sentiment classification task

from the corpus produced by Blitzer et al. Each domain of this corpus has 1000 pos+ and 1000 neg− labeled reviews. The pre-processing is approved to make this three dataset prepare for experiment. The statistics of this data set is given in Table 2.

Pre-processing

- *Tokenization or segmentation* It can be accomplished by splitting documents (crawled reviews) into a list of tokens such as word, numbers, special characters etc. and make the document ready to be used for further processing.
- *Normalization* This process converts all the word token of a document into either lower case or upper case because most of the reviews consist of both case i.e. lowercase and uppercase characters. The purpose of shifting all tokens into a single format can easily be used for prediction.
- *Removal of stop words* Stop words are very common and high-frequency words. This process carried out by removing frequently used stop words (prepositions, irrelevant words, special character, ASCII code), new line, extra white spaces etc. to enhance the performance of feature selection technique.
- *Stemming* It is the process of transforming all the tokens into their stem, or root form. Stemming is a swift and easy approach that makes the feature extraction process more effortless.

Feature selection

Feature selection method (FSM) is an essential task to enhance the accuracy of sentiment classification process. Generally, FSMs are statistically represented by the relationship between feature and class category. The performance of the classifier mostly depends on the feature set, if feature selection method [33] performs well then the simplest classifier may also give a good accuracy through training. These FSMs are often defined by some probabilities to realize the theoretical analysis of these probabilistic methods. We use a list of notation which is depicted in Table 3.

Analytical information from the training data is required to determine these probabilities and notations about the training data are listed in Table 2 given as follows:

We denote $C_{i=1}^{m} = \{c_1, c_2, \ldots, c_m\}$ is the set of classes (Table 4).

Information Gain (IG)

This statistical property used as an effective solution for feature selection. IG method is used to select important features based on the class attribute rules of features classification. The IG value of each term can measures the number of bits of information

Table 2 The detailed statistics of above mentioned datasets

Dataset		Dataset size	Dataset size	Objective	Class
Movie review	IMDB	50,000	Sentiment	2/binary	279
MDSD	Electronics	2000	Sentiment	2/binary	115
	Kitchen	2000	Sentiment	2/binary	112

Table 3 Notation use for feature selection

Symbol	Description
$P(c_i)$	Probability that a document d in class c_i
P(f)	Probability that document d contains feature f
$P(\bar{f})$	Probability that a document d does not contains feature f
$P(c_i/f)$	Probability that document d contains feature f in class c_i
$P(c_i/\bar{f})$	Probability that document d does not contains feature f in class c_i

Table 4 Notation use for feature selection

Symbol	Description
N_{all}	The total no. of documents in training dataset
N_i	No. of documents in class c_i
W_i	No. of documents in class c_i contain feature f
X_i	No. of documents not in class c_i but contain feature f
$Y_i = N_i - W_i$	No. of documents in class c_i don't contain feature f
$Z_i = N_{all} - N_i - X_i$	No. of documents neither in class c_i nor contain the feature f

acquired for class prediction by knowing the presence or absence of that term in the document [34]. The IG value of a certain term or feature is calculated by the following equation:

$$\begin{aligned} \mathrm{IG(f)} =& \left\{ -\sum_{i=1}^{m} P(c_i) \log P(c_i) \right\} + \left\{ P(f) \left[\sum_{i=1}^{m} P(c_i|f) log P(c_i|f) \right] \right\} \\ &+ \left\{ P(f) \left[\sum_{i=1}^{m} P(c_i|f) \log P(c_i|f) \right] \right\} \end{aligned} \tag{1}$$

And it is defined as

$$\begin{aligned} \mathrm{IG(f)} =& \left\{ -\sum_{i=1}^{m} \frac{N_i}{N_{all}} \log \frac{N_i}{N_{all}} \right\} + \left(\sum_{i=1}^{m} W_i/N_{all} \right) \left[\sum_{i=1}^{m} \frac{W_i}{W_i + X_i} log \frac{W_i}{W_i + X_i} \right] \\ &+ \left(\sum_{i=1}^{m} Y_i/N_{all} \right) \left[\sum_{i=1}^{m} \frac{Y_i}{Y_i + Z_i} log \frac{Y_i}{Y_i + Z_i} \right] \end{aligned} \tag{2}$$

IG offers a ranking of the features depending on their IG score, thus a certain number of features can be selected easily.

Chi square (χ^2)

Chi square (χ^2) is a very commonly applied statistical test, can quantify the association between the feature or term f and its related class c_i. It tests a null-hypothesis that, the two variables feature and class is completely independent of each other. The

CHI value of feature *f* for class C_i is higher, the closer relationship exists between the variables feature *f* and class C_i. The features with the highest χ^2 values for a category should perform best for classifying the documents. The formulation of this method as follows:

$$\chi^2(\mathrm{f}, \mathrm{c_i}) = \frac{N_{all..}(W_i Z_i - Y_i X_i)^2}{(W_i + Y_i).(X_i + Z_i).(W_i + X_i).(Y_i + Z_i)} \tag{3}$$

It can also be defined by considering $\mathrm{Y_i}$ as $(\mathrm{N_i} - \mathrm{W_i})$ and $\mathrm{Z_i}$ as $(\mathrm{N_{all}} - \mathrm{N_i} - \mathrm{X_i})$ and the above formula is rewritten as follows

$$\chi^2(\mathrm{f}, \mathrm{c_i}) = \frac{N_{all}.[W_i(N_{all} - N_i - X_i) - (N_i - W_i)X_i]^2}{N_i.(N_{all} - N_i).(W_i + X_i).[N_{all} - (W_i + X_i)]} \tag{4}$$

Gini Index (GI)

Gini Index measures the features ability to discriminate between classes. This method was mainly proposed to use for decision tree algorithm based on an impurity split method. The main principle of Gini Index is to consider S as a dataset of the sample having m number of different classes $C_{i=1}^{m} = \{\mathrm{c_1}, \mathrm{c_2}, \ldots, \mathrm{c_m}\}$. According to the class level, the sample set can be splitted into n subset (*Si*, i = 1, 2,…,n). The Gini index of the set S is

$$Gini\ Index(S) = 1 - \sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{P_i^2} \tag{5}$$

where probability $\mathrm{P_i}$ of any sample belongs to class $\mathrm{C_i}$, can be computed by S_i/S [35]. Gini Index for a feature can be estimated independently for binary classification. We adopted Gini index Text (GIT) method for calculating the feature score, which was introduced by Park et al. [36]. This algorithm enhanced to overcome the limitations of Gini Index method.

According to previous notation defined in Table 3, we can compute the Gini Index for a feature f of document d belongs to class $\mathrm{C_i}$.

$$GIT_{wi}(f, C_i) = P(C_i|f)^2 \tag{6}$$

$$GIT_{Xi}(f, C_i) = \left| \frac{P(Ci|f)^2}{\log_2 P(f)} \right| \tag{7}$$

Combination of feature selection methods

As each feature selection method applied with different rules to extract a feature subset, it outcomes various feature subsets for same dataset. We merged these different feature sub lists by adopting either statistical method UNION to select all features or INTERSECTION to select only common features. In our paper, modified UNION method has been considered to obtain all top ranked including common selected features.

Let F {f1, f2,…,fn}be the primary feature set selected by preprocessing the review dataset D review dataset D. The feature subsets $\mathrm{F_{SUB}}1$ $\{\mathrm{f_{11}}, \mathrm{f_{12}}, \ldots, \mathrm{f_{IG}}\}$, $\mathrm{F_{SUB}}2$ $\{\mathrm{f_{21}}, \mathrm{f_{22}}, \ldots, \mathrm{f_{CHI}}\}$ and

$F_{SUB}3$ $\{f_{31}, f_{32},\ldots,f_{CHI}\}$, are selected with Info Gain, CHI and Gini Index, respectively. All the features exists in these subsets are must be sorted according to their score or weight.

To generate a feature sub list with UNION, we just combine all the features from above three feature selection method.

$$F_{SUB}4 = F_{SUB}1 \cup F_{SUB}2 \cup F_{SUB}3 \tag{8}$$

To obtain a feature sub list with INTERSECTION, we only select common features in all three feature subsets.

$$F_{SUB}5 = F_{SUB}1 \cap F_{SUB}2 \cap F_{SUB}3 \tag{9}$$

Next we applied revised UNION approach to catch all top ranked features along with common features. As the features are already sorted, the highest scored features and lowest scored features are got there accurate positions. Therefore, again we tested UNION and INTERSECTION method over top ranked (T1) and lowest rank (L1) subsequently.

According to revised UNION approach, we applied union on top T1% of features and intersection on remaining L2%.

$$\begin{aligned} F_{SUB}6 = {} & \{T1\%\{ F_{SUB}1\}UT1\%\{F_{SUB}2\}UT1\%\{F_{SUB}3\}\} \\ & U\{L1\%\{F_{SUB}1\} \cap \{L1\%\{F_{SUB}2\} \cap \{L1\%\{F_{SUB}3\}\}\} \end{aligned} \tag{10}$$

These merged feature sub list will be employed to learn to the supervised classifiers to compare the performance of classifiers with feature subsets obtained from individual feature selection method.

Classification

Machine learning techniques are widely used in artificial intelligence and document classification. Extracted feature sets are used to train the classifier to classify the review of the data set as positive or negative. We applied generative classification model (MNB) and discriminative model (SMO, LR and RF) as a prominent classification approach. Generative model captures p (*d, c*) and p (*c*), and then directly computes *p* (*c, d*) with conditionally independent assumption between features, on smaller dataset. The reasons to select these classifiers are attributed as follows:

- For this research work, we applied first multivariate Bernoulli naïve bayes (BNB) classifier which considers features vector only with binary or boolean values. The vector can focus on the presence or absence of the feature and not worried about how many times that feature occurs in the document.
- We considered multinomial naïve bayes (MNB) to be applied to overcome the aforementioned limitations of BNB. MNB classifier utilized to classify the document based on frequency counts of multiple features.
- MNB classifier is not comfortable with imbalanced (one class having more data samples than others) training dataset, while SMO can equally deal with imbalanced as well as balanced dataset.
- Based on the priority of the data status, we assign one more classifier suitable for binary classification that is linear regression (LR). We utilized the main advantage

[16] of LR classifier i.e., to use the non linear function with a linear combination of features.

- According to previous studies, we designed Random forest classifier (RF) designed as consistently accurate to predict the sentiment's class labels for large scale data set. This classifier always a good choice as it can tackle different types of features without scaling including, numerical, categorical, binary features.

Naive Bayes (NB)

Naive Bayes classification method is used for both purpose; classification as well as training. The fundamental theory of NB classifier [37] is based on the independence assumption; where the joint probabilities of features and categories are used to roughly calculate the probability score of categories of a given document. It is a simple probabilistic classifier, that helps in classifying a document d_r, out of classes $c_i \in C$ ($C_{i=1}^{m} = c_1, c_2, \ldots . c_m$). The best class returns in NB classification is the most probably or maximum posterior (MAP) class C_{map}.

$$C_{\underset{c_i \in C}{map}} = argmax P(c_i) P(d_r|c_i) \tag{11}$$

where the class $P(c_i)$ can be estimated by dividing the number of documents of class c_i by the total number of documents. $P(d_r|c_i)$ indicated the number of occurrence of the feature in document d_r belongs to class c_i. The probability value $P(c_i|\ d_r)$ will be computed for each possible class, but $P(d_r)$ doesn't change for each class. Thus we can drop the denominator.

We thus select the highest probable classes' c_{map} of given document d by calculating the posterior probability of each class.

There are several Naive Bayes variations. In this paper, we consider the Multi-nominal Naive Bayes classifier.

Multi-nominal Naïve Bayes (MNB)

The multi-nominal Naive Bayes model [38] is distinctly used for discrete counts. We consider MNB classifier for text classification task, where a document d is represented by a feature vector ($f_1, f_2 \ldots$, fn) with the integer value of word frequency in the given document. For multinomial NB model, The conditional distribution P ($d|c_i$) of document d given the class c is as follows:

$$\text{Multi-nominal } P(|c_i) = P((f_1, f_2, \ldots, f_n)|c_i) = \prod_{1 \le j \le n} P(f_j|c_i) \tag{12}$$

The final equation with Bayes' rules the highest probable classes by a Naive Bayes classifier as follows:

$$C_{map} = \underset{c_i \in C}{argmax}\, \hat{P}(c_i) \prod_{1 \le j \le n} \hat{P}(f_j|c_i) \tag{13}$$

Now, to estimate the probability $\hat{P}(f_j|c_i)$ we consider the feature as a word appears in the document's bag of words. Thus we'll compute $\hat{P}(wj|c_i)$ by considering N_{jr} as the number

of occurrence of word wj in documents d_r from class c_i among all words in all documents of class c_i. Then the estimated probability of a document given its class is given as follows.

$$P(d_r|c_i) = \left(\sum_{j}^{|v|} N_{jr}\right)! \prod_{j=1}^{|v|} \frac{\hat{P}(w_j|c_i)^{N_{jr}}}{N_{jr}!} \tag{14}$$

where, v is the union of all the word types in all classes.

The probability of w_j in c_i is estimated from training dataset and it is defined as follows.

$$\hat{P}(wj|c_i) = \frac{\text{count}(w_j c_i)}{\sum_{w \in V} count(w, c_i)} \tag{15}$$

Support Vector Machine (SVM)

Support Vector Machines (SVMs) are supervised learning model introduced [15] for binary classification in both linear and nonlinear versions. Generally, datasets are nonlinearly inseparable, so the primary aim of the SVM classifier is to catch the best accessible surface to make a separation between positive and negative training samples based on empirical risk (training set and test set error) minimization principal. SVM method can try to define a decision boundary with the hyper-planes in a high dimensional feature space. This hyper-plane separates the vectorized document into two classes as well as determines a result to make a decision based on this support vector [5]. The optimization problem of SVM can be minimized as follows.

Given N linearly separable training set with feature vector x of d dimension. For dual optimization where $\alpha \in R^N$ and $y \in \{1, -1\}$. Then the solution of SVMs (dual) can be minimized as follows:

$$\vec{\alpha}^{\,*} = argmin\left\{ -\sum_{i=1}^{n} \alpha_i + \sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j \langle \vec{x_i}, \vec{x_j} \rangle \right\} \tag{16}$$

where,

$$\sum_{i=1}^{n} \alpha_i y_i = 0; \quad 0 \le \alpha_i \le C \tag{17}$$

The classical SVM seems to be able to separate the linear dataset with a single hyperplane, which can separate two classes. For nonlinear dataset where more than two classes to be handled, kernel functions are used in that situation to lay out the data to a higher dimensional space in which it is linearly separable.

Sequential Minimal Optimization (SMO)

The algorithm Sequential Minimal Optimization (SMO) is employed as a learning algorithm to train Support Vector Machine with a linear kernel. It is highly efficient for solving the issue regarding quadratic programming (QP) problem which appears during the

training of SVM. In this paper, we implemented SMO [18] by using data mining tools in python.

SMO is preferred because of its better scaling abilities for large and complicated SVM problems with less computational time than standard SVM. SMO solves the QP problem by decomposing the large problem into a sequence of small sub problems. Then SMO follows analytic QP steps to deal with smallest possible optimization problem.

The above QP problem in Eq. (17) will be solved by SMO as a smallest optimization problem, which consists of Lagrange multipliers α_i. To get a definite solution of this QP problem, all the Lagrange multipliers should satisfy the Karush–Kuhn–Tucker (KKT) conditions.

Random forest (RF)

The Random forests are one of the most popular and widely used methods or framework for classification and regression problems. It has evolved as an ensemble learning approach based on multiple numbers of decision trees. According to the description of Wikipedia, Random Forest classifier operates by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. This classifier can solve the problem of over fitting by reducing the correlation between randomly selected trees and it helps in increasing the prediction power. The prediction for unseen samples can be done by averaging the predictions of the p number of trees. Random forest model performs well even when the feature size quite larger than number of samples. In case of high dimensional feature space RF classifier give poor accuracy, therefore to generate more accurate trees we applied some efficient feature selection method for dimensionality reduction. The basic procedure to build RF model with training dataset D is as follows.

Suppose, Training dataset $\mathrm{D} = \left\{ (\mathrm{f_i},\ \mathrm{C_i})_{\mathrm{i}=1}^{\mathrm{N}} |\ \mathrm{f_i} \in \mathrm{R^F},\ \mathrm{C} \in \{1, 2, \ldots, \mathrm{c}\} \right\}$ is given, where $\mathrm{f_i}$ are features, $\mathrm{C_i}$ is the set of classes and N denotes the number of training samples. Sample the training set D with replacement to create bagged samples $\mathrm{D_1}$, $\mathrm{D_2}$,...,$\mathrm{D_p}$ and each decision tree is grown from these bagged sample set. In each decision tree, for every node we consider a random and separate subset of predictive features as candidate feature for splitting the node. The class prediction of RF model with p number of trees can define as follows. Let assume $\hat{C}p$ be the prediction of tree T_p given input f.

$$\hat{C} = \text{majority voting } \{\hat{C}^{\mathrm{p}}\}_1^{\mathrm{p}} \tag{18}$$

In this research work, random forest (RF) classifier provides a striking precision in contrast to other classification model namely MNB, SVM, logistic regression, and GBM for medium length dataset. The F-score of RF classifier differ with size of different dataset in classification. The hyper parameters of RF classifier such as number of trees, number of features, and depth of trees plays a crucial role in maintaining the higher accuracy. The major characteristics of this algorithm are given below:

- This algorithm is easy to build and interpret.
- Classification model considered as robust and accurate.
- If some parameters are there they may be inserted easily, in such a way eliminating the requirement for pruning the trees.

Logistic regression

Logistic regression model consists of a set of classification rules extensively used for binary classification problem, to solve multiclass problem the model must be extended. This logistic function of this classifier utilizes to extract a set of weighted features from the input and estimates the correlation between the occurrence class, and extracted features. The researcher Allison [39] states that logistic regression becomes a suitable fit to the data by maximizing the log-likelihood function. Containing of all predictor into single model generally results poor predictions. The proper variable selection makes the model more accurate and generalized.

The probability of a feature vector i existing with positive class can be measured by logistic regression as given:

$$P(\mathrm{c}=1|\mathrm{f}) = l(f) = \frac{1}{1+e^{wT_f}} \tag{19}$$

where $P(\mathrm{c}=1|\mathrm{f})$ refers to the probability of document 'f' of class 'c'. 'w' indicates the feature-weight parameters to be estimated.

Experiments and results

Experimental settings

We applied Java SE (version 6) with Netbeans IDE (version 6.9) and Python with simple as well as efficient scikit-learn library to perform the experiments. We implemented standford POS tagger in java Netbeans IDE for POS tagging and remaining parts including tokenization, normalization, stop words removing are also performed by Java tools. Therefore, we can have a large feature space, from product review datasets [5], which are incorporated in python package. For feature selection and classification purpose, we used python tools. Specifically, we have utilized scikit-learn module with NumPy 1.8.1, SciPy 0.14.0 python library to enhance and extend the core python capabilities. We carried out experiments considering the tenfold ($\mathrm{k}=10$) cross validation, as the dataset we have considered for our experiment are comparatively smaller than other existing data sets. We separate the dataset into two portions, where $(k-1)$ 9-folds are used for training and 1-fold is used for testing.

Evaluation parameters

The performance of supervised ML algorithm can be evaluated based on the term or elements of confusion matrix on a set of test data. The confusion matrix consists of four terms are True Positive (TP), False Positive (FP), True Negative (TN) and False Negative (FP). According to the value of these elements, the evaluation matrices precision, recall, F-score and ROC are determined to estimate the performance score of any classifier.

$$\text{Precision } (\pi) : \frac{TP}{TP+FP} \tag{20}$$

$$\text{Recall } (\rho) : \frac{TP}{TP + FN} \tag{21}$$

$$\text{F-Score} : \frac{2 * Precision * Recall}{Precision + Recall} \tag{22}$$

ROC curve

In machine learning, AUC *or* 'Area under the ROC Curve' is most popular measurement metric to determine which of the model can predict the classes best. **ROC curve** (**receiver operating characteristic curve**) is a graphical plot showing the performance of a classification model at all thresholds by considering the parameters True positives (TF) on X-axis and True negatives on Y-axis. The AUC values are lies in between 0.5 and 1.0. For binary classification, the higher accuracy indicates the model performs best whereas the classifier with AUC value 1.0 is an excellent performer and AUC value 0.5 is consider as a bad performer.

Results and discussion

The following experimental results help in the study the effects of an individual as well as a combination of the different feature selection methods on the performance of the classifier. This result clearly exhibits how each the classifier behaves with different feature selection method. In this section, an in-depth investigation was carried out to measure the effectiveness of the proposed approach i.e., to compare the performance of four supervised classifiers SMO, MNB, RF and logistic regression based on the combination of the different feature selection method.

Firstly, we proposed the dataset with standard pre-processing method such as tokenizing, stop words removal, normalizing, stemming. Then we applied the feature selection method Information Gain (IG), Gini Index (GI), Chi square (CHI) to assign a score to each feature and three different feature subsets are generated based on the score. Next, we combined three different feature subsets by adopting statistical method UNION to select all features, INTERSECTION to select only common features and revised UNION to collect all top ranked features along with common features of three subsets. Thus a prominent feature vector by merging IG, CHI, GI feature subsets can be generated easily for classification. Finally, the classifiers SMO, MNB, RF and logistic regression machine learning classifier used individual feature subset as well as prominent feature vector for classifying the review document into either positive or negative.

Tables 5, 6, 7 displays the performance of machine learning methods SMO, MNB, RF and logistic regression with respect to different feature selection method and their combination. The method IG performed well in comparison with other FSMs.

The following table indicates that the combination of (IG, CHI, GI) produce better results than applying those method individually.

According to Table 5, the SMO classifier performs best for linear SVM (F-score 92.31) with combined (IG, CHI, GI) method. The result shows that feature selection method IG and GI also performed well with SMO classifier. The F-Score of SMO becomes 89.77 and 88.62 for IG and Gini Index, respectively. SMO achieves higher accuracy for large

Table 5 Results of the proposed model for movie review data set

Method	Classifier											
	SMO			MNB			RF			Logistic regression		
	Prec	Recall	F-Score	Prec	Recall	F-Score	Prec	Recall	F-Score	Prec	Recall	F-Score
IG	91.5	86.2	89.77	87.1	84.5	85.78	86.2	82.4	84.25	84.2	87.1	85.62
CHI	88.7	86.2	87.43	85.8	83.6	84.75	79.7	76.4	78.01	86.4	83.8	85.08
Gini Index	90.1	87.3	88.62	86.2	84.5	86.35	86.5	84.8	*85.64*	87.6	84.5	86.02
Combined (IG, CHI, GI)	92.5	86.1	*92.31*	88.9	86.0	*88.12*	87.4	83.6	83.45	85.6	88.2	*86.88*

Table 6 Results of the proposed model for electronics review data set

Method	Classifier											
	SMO			MNB			RF			Logistic regression		
	Prec	Recall	F-Score	Prec	Recall	F-Score	Prec	Recall	F-Score	Prec	Recall	F-Score
IG	86.5	83.9	85.18	86.3	82.9	84.56	84.6	83.2	83.89	83.3	84.5	83.89
CHI	84.8	82.5	84.54	83.9	85.4	86.63	85.9	82.5	84.16	86.0	82.6	84.26
Gini Index	87.3	87.1	*87.19*	86.2	84.7	85.44	88.6	83.4	85.92	86.9	84.1	85.75
Combined (IG, CHI, GI)	87.3	84.7	85.98	91.2	89.6	*90.53*	87.9	85.6	*86.73*	85.1	87.3	*86.02*

Table 7 Results of the proposed model for Kitchen ware review data set

Method	Classifier											
	SMO			MNB			RF			Logistic regression		
	Prec	Recall	F-Score	Prec	Recall	F-Score	Prec	Recall	F-Score	Prec	Recall	F-Score
IG	86.8	84.9	85.83	88.8	85.3	*87.49*	86.9	83.2	85.00	82.7	85.6	84.12
CHI	83.2	81.1	82.13	86.7	84.6	85.63	84.4	87.7	86.01	84.3	87.1	85.67
Gini Index	86.4	82.2	84.24	82.1	80.7	81.39	83.9	81.4	82.63	84.0	81.2	82.57
Combined (IG, CHI, GI)	87.9	85.2	*86.52*	85.0	83.5	84.24	88.7	86.8	*87.73*	89.8	87.6	*88.47*

problems. With the classifier evaluation, the SMO performed better than MNB (86.38), RF (83.45) and logistic regression (86.88) classifiers for movie review dataset.

The NB performs surprisingly well for sentiment analysis in many previous studies. NB method is a simple and popular classification technique, although the conditional independence assumption is harsh. In our investigation, MNB is next best to SVM in performance. In all three datasets, MNB classifier equally provides a good result. In this paper, MNB performed great (90.53) especially with electronics product review data set.

As reported in Tables 5, 6, 7, the F score value obtained using combined methods (IG, CHI, GI) is comparatively better than that obtained using IG, CHI and GI method separately. If we consider all three domains, the combined method exhibits the better result in terms of F-score with more or less every classifier such as SVM, RF, LR and MNB. They chose the features based on their importance to the class level attribute.

The best performance of the logistic regression classifier is achieved with review data set of kitchen domain (88.47) when the combined (IG, CHI, GI) method is being used. Naïve Bayes classifier is quite convenient for small datasets as it is effortless to implement and very swift to train, but the classifier LR and SVM produce overall better performance for large dataset. In particular, unlike the NB classifier LR is capable of handling dependent features, while NB classifier is based on the independent assumption. In this case, LR underperformed SVM and MNB classifier for the datasets, such as: Movie, Electronics and outperformed RF classifier for most of the review datasets with combined feature selection method.

RF classifier got the maximum F score of 87.73 with Kitchen domain, when we consider the domain Movie and Electronics the F score for RF classifier reduced to 85.64 and 86.73, respectively.

In order to investigate the following figures, if we compare the classifier performance, SVM outperforms other classification methods MNB, RF and logistic regression for movie domain. MNB produce best result with Electronics dataset and the result obtained based on Kitchenware is the most favourable with RF classifier. According to the highest value of accuracy, recall and F-score value we estimate the results of three algorithms on testing dataset.

To see the impacts of different feature selection methods, we plot the ROC curve with highest AUC to represent the results of four different classification models for the aforementioned bench mark datasets of movie review and multi domain product review in Figs. 2, 3, 4. For readability, each graph presented four ROC curves for classifiers namely, SMO, MNB, LR and RF. According to the graph, we noticed all ROC curves incline to the upper left space in the graph. It specified that all the classifiers we have selected for this research work could reach both the maximum TPR as well as minimum FPR, as the point (0, 1) in the upper left corner of the roc space is also called a perfect classification which representing the true positive rate is 100% and the false positive rate is 0%.

We also calculated AUC and FI measure score in Figs. 5, 6, 7 where white column denoted to AUC and black for F1 measure. The highest AUC was 0.94 with SMO classifier for Movie review dataset. MNB and LR classifier obtained highest 0.93 with Electronics dataset and 0.92 with Kitchen review dataset, respectively. According to the

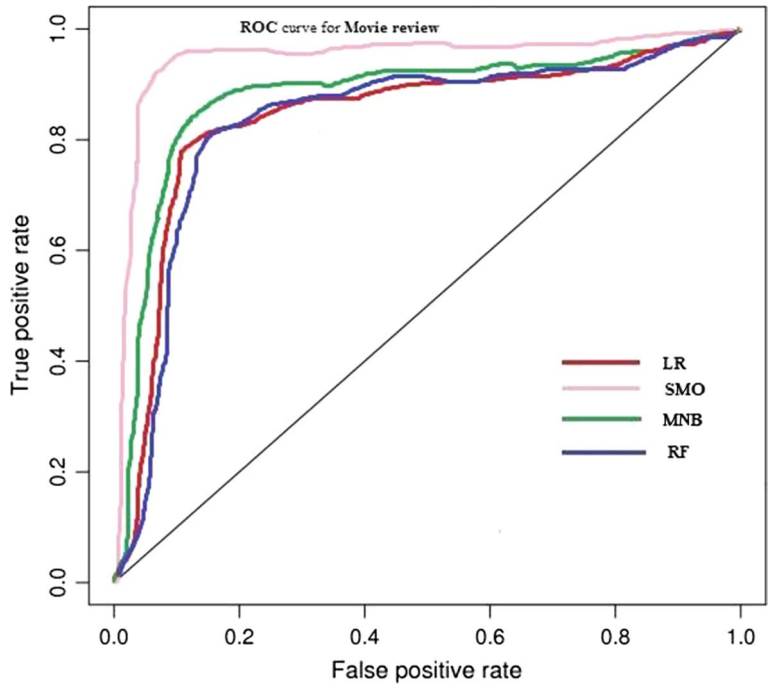

Fig. 2 ROC curves obtained by classifiers for *Movie review* dataset

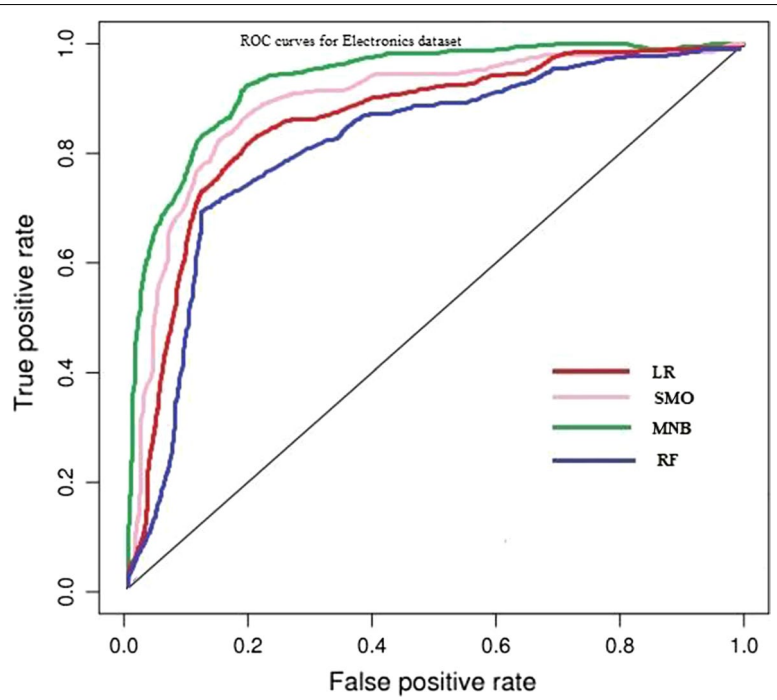

Fig. 3 ROC curves obtained by classifiers for *Electronics review* dataset

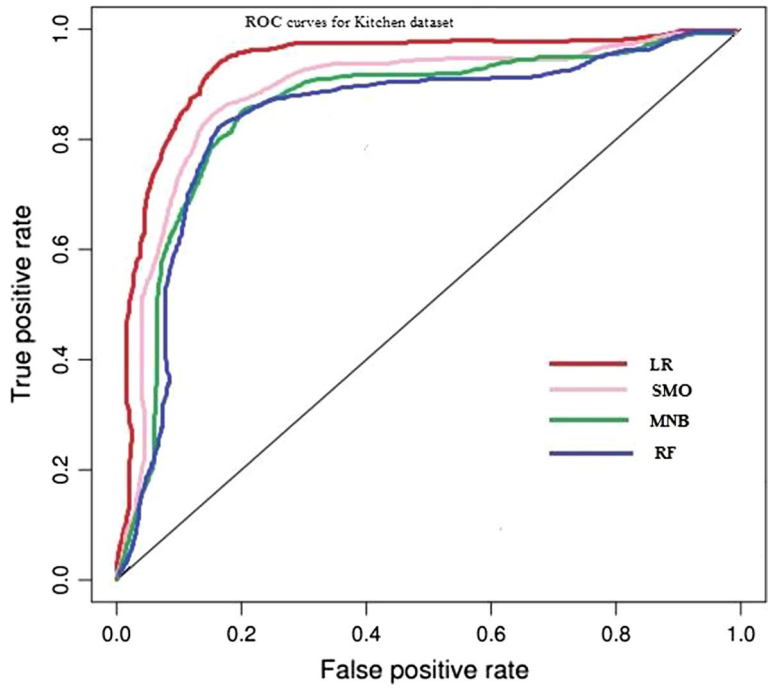

Fig. 4 ROC curves obtained by classifiers for *Kitchen review* dataset

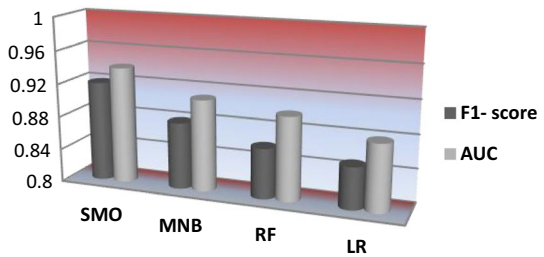

Fig. 5 Comparing the resulting value of F1-score and AUC of classifiers for *Movie review* dataset

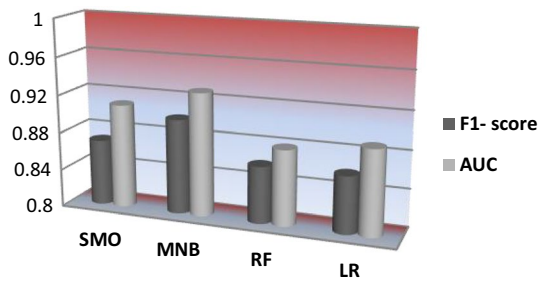

Fig. 6 Comparing the resulting value of F1-score and AUC of classifiers for *Electronics review* dataset

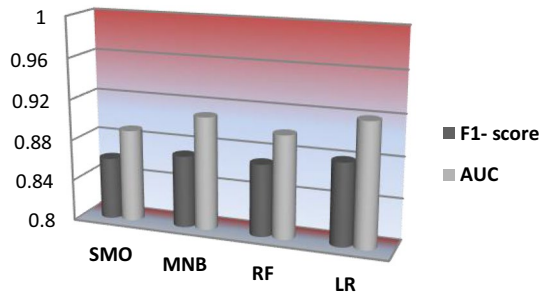

Fig. 7 Comparing the resulting value of F1-score and AUC of classifiers for *Kitchen review* dataset

Table 8 Comparisons of performance of proposed approach with different literature using different domain review Dataset

	Dataset	Feature selection method	Classifier	Performance
Pang et al. [29]	Internet Movie Database (IMDb)	N-gram features	SVM NB ME	82.9 (Accuracy) 81.5 81.0
Agarwal et al. [1]	Movie (IMDb) Product (book, DVD, electronics)	N-gram, IG, RSAR, Hybrid(IG + RSAR)	SVM NB	87.7 (F measure) 80.9
Al-Moslmi et al. [44]	Movie reviews in the Malay language	IG, CHI, Gini Index	SVM NB KNN	85.33 (F-measure) 80.88 74.68
Kolog et al. [18]	Sentiment from social network regarding student's life	N gram features	SMO MNB J48	80.0 83.0 69.0
Tripathy et al. [22, 43]	Movie (IMDb)	N-gram features	SVM ME NB SGD	88.94 88.48 86.23 85.11
Our approach	Movie (IMDb) Electronics product Kitchenware	N-gram, Combination of Unigram and bigram with IG, CHI, Gini Index	SMO MNB RF LR	90.18 (F-measure) 88.18 87.73 87.32

results, we can predict that employing the selected machine learning classification models must have effectiveness for sentiment polarity detection.

Performance evaluation

This section compares between the accuracy of proposed approach with other existing approaches considered IMDb dataset. This comparison was carried out according to the accuracy value which these methods achieved. The adopted approach i.e., the combination of different feature selection methods produces a better result in comparison [40] with the result obtained by applying individual feature selection method in previous research approaches are shown in following tables.

The author [41] proposed to use classifier ensembles and lexicons for sentiment analysis of tweets automatically. They tried to compare between bag-of-words model and feature hashing technique regarding how they represent features. The result exhibits that classifier ensembles configured by SVM, MNB, RF, and logistic regression can enhance

the classification accuracy in a huge amount. The highest accuracy they got is of 79.11 with ensemble classifier (Table 8).

Kalaivani et al. [42] examined how classifier SVM, NB and k-NN works with different feature sizes of movie review dataset. Feature selection method Information Gain (IG) applied to select top p % ranked features to train the classifier. In this work, SVM approach outperformed the Navie Bayes and k-NN approaches with highest accuracy 81.71. The experimental result reported the precision and recall value for positive and negative corpus separately.

In [43], the investigation by Tripathy et al. employed machine ML classifiers namely NB, SVM, ME, SGD to perform sentiment classification of online movie reviews [36] with N-gram techniques. The performance evaluation can be done by the parameters such as precision, recall, F-measure, and accuracy. The results in comparing with our approach show that FSMs have a great impact on the classification performance. The feature ranking techniques (Information Gain, Chi Square, Gini Index method) improve classification performance over no feature selection.

Al-Moslmi et al. [44] studied on feature selection methods effects on machine learning approaches in Malay sentiment analysis. It was demonstrated that improved feature selections resulted in better performance in Malay sentiment-based classification. The author approached three feature selection methods (IG, Gini Index, and CHI) to enhance the performance of three machine learning classifiers (SVM, NB, and k-NN). A dataset of 2000 movie reviews are crawled from several web contents in Malay language. The results showed that the combination of SVM classifier and IG-base method established as the best classification algorithm, with an accuracy of 85.33% with feature sizes of 300. Authors have also reported that use of the FSMs yields improved results compared to those from the original classifier.

Conclusion

The aim of this paper is to explore the ability of a combination of three feature selection methods such as IG, Chi Square, Gini Index to enhance and refine the performance of four machine learning classifiers namely SMO, MNB, RF and LR on the multiple domains. The effectiveness of classification algorithm is evaluated in terms of F measure, precision, and recall. As discussed in before the combined feature sub list of IG, CHI and GI produce very convincing results. As we considered the dataset from multiple domains, thus the classifiers such as SMO, MNB and RF performed best for reviews of movie, electronics and kitchenware subsequently. These empirical experiments, exhibiting the proposed method, are highly effective and encouraging. The method we proposed in this work still has some drawbacks that are mentioned as follows:

- In some reviews or comments, user expressed their sentiment through some images or emoticons, but we have not considered these kinds of expressions for analysis.
- The comment in text format contains sarcasm, linguistic problems etc. To predict the sentiment of that comment we have to understand the nature and ambience

of the comment thoroughly, as a single word can create a contradiction about the polarity of the comment. However, this aspect is also not considered in this paper.

All of above-mentioned downsides shall be considered for the future work to refine the quality of sentiment classification. We are also planning to merge the traditional machine learning method with deep learning techniques to tackle the challenge of sentiment prediction of massive amounts of unsupervised product review dataset in future.

Abbreviations
IG: Information Gain; IMDb: Internet movie database; MDSD: multi domain sentiment dataset; ML: machine learning; SMO: Sequential Minimal Optimization; MNB: multi-nominal Naïve Bayes; RF: random forest; LR: logistic regression; ME: maximum entropy; POS: part of speech; FSM: feature selection method; CHI: Chi square; GI: Gini Index; ROC: receiver operating characteristic curve; TF-IDF: term frequency-inverse document frequency.

Authors' contributions
All mentioned authors contribute in the elaboration of the article. Both authors read and approved the final manuscript.

Acknowledgements
Not applicable.

Competing interests
The authors declare that they have no competing interests

Availability of data and materials
All data is publicly available here: http://www.cs.cornell.edu/people/pabo/movie-reviewdata, http://www.cs.jhu.edu/~mdredze/datasets/sentiment.

Consent for publication
Not applicable.

Duplication
The content of the manuscript has not been published, or submitted for publication elsewhere.

Ethics approval and consent to participate
Not applicable.

Funding
No funding to report.

Publisher's Note

Received: 19 May 2018 Accepted: 27 October 2018
Published online: 14 November 2018

References

1. Agarwal B, Mittal N. 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 (Atlanta), 2013. p. 115–9.
2. Govindarajan M. Sentiment classification of movie reviews using hybrid method. Int J Adv Sci Eng Technol. 2014;3:139.
3. Liu B. Sentiment analysis and subjectivity. In: Indurkhya N, Damerau FJ, editors. Invited chapter for the handbook of natural language processing. 2nd ed. England: Taylor & Francis; 2010.
4. Dhaoui C, Webster CM, Tan LP. Social media sentiment analysis: lexicon versus machine learning. J Consumer Mark. 2017;34(6):480–8.
5. Samal BR, Behera AK, Panda M. Performance analysis of supervised machine learning techniques for sentiment analysis. In: Proceedings of the 1st ICRIL international conference on sensing, signal processing and security (ICSSS). Piscataway: IEEE; 2017. p. 128–3.
6. Catal C, Nangir M. A sentiment classification model based on multiple classifiers. Appl Soft Comput. 2017;50:135–41.
7. Boiy E, Hens P, Deschacht K, Moens MF. Automatic sentiment analysis of on-line text. In: Proceedings of the 11th international conference on electronic publishing. Vienna; 2007.
8. Dave K, Lawrence S, Pennock DM. Mining the peanut gallery: opinion extraction and semantic classification of product reviews. In: Proceedings of the 12th international WWW conference. Budapest; 2003. p. 519–8.
9. Annett M, Kondrak GA. Comparison of sentiment analysis techniques: polarizing movie blogs. In: Conference of the Canadian Society for computational studies of intelligence. Berlin: Springer; 2008; p. 25–5.

10. Zhang D, Xu H, Su Z, Xu Y. Chinese comments sentiment classification based on word2vec and SVM perf. Expert Syst Appl. 2015;42(4):1857–63.
11. Mouthami K, Devi KN, Bhaskaran VM. Sentiment analysis and classification based on textual reviews. In: Information Communication and Embedded Systems. Piscataway: IEEE; 2013; p. 271–6.
12. Ye Q, Zhang Z, Law R. Sentiment classification of online reviews to travel destinations by supervised machine learning approaches. Expert Syst Appl. 2009;36(3):6527–35.
13. Xia R, Zong C, Li S. Ensemble of feature sets and classification algorithms for sentiment classification. Inf Sci. 2011;181:1138–52.
14. Manek AS, Shenoy PD, Mohan MC, Venugopal KR. Aspect term extraction for sentiment analysis in large movie reviews using Gini Index feature selection method and SVM classifier. World Wide Web. 2016;20:135–54.
15. Whitehead M, Yaeger L. Sentiment mining using ensemble classification models. In: International conference on systems, computing sciences and software engineering (SCSS 08). Berlin: Springer; 2008.
16. Adel A, Omar N, Al-Shabi A. A comparative study of combined feature selection methods for arabic text classification. J Comput Sci. 2014;10(11):2232–9.
17. Zha ZJ, Yu J, Tang J, Wang M, Chua TS. Product aspect ranking and its applications. IEEE Trans Knowl Data Eng. 2014;26(5):1211–24.
18. Kolog EA, Montero CS, Toivonen T. Using machine learning for sentiment and social influence analysis in text In: Advances in intelligent systems and computing. Cham: Springer; 2018. p. 453–3.
19. Narayanan V, Arora I, Bhatia A. Fast and accurate sentiment classification using an enhanced naïve Bayes model, In: Intelligent data engineering and automated learning–IDEAL2013. Berlin: Springer; 2013. p. 194–1.
20. Amolik A, Jivane N, Bhandary M, Venkatesan M. Twitter sentiment analysis of movie reviews using machine learning techniques. Int J Eng Technol (IJET). 2016;7:1–7.
21. Luo B, Zeng J, Duan J. Emotion space model for classifying opinions in stock message board. Expert Syst Appl. 2016;44(2016):138–46.
22. Tripathy A, Agrawal A, Rath SK. Classification of sentimental reviews using machine learning techniques. Procedia Comput Sci. 2015;57:821–9.
23. Selvi C, Ahuja C, Sivasankar E. A comparative study of feature selection and machine learning methods for sentiment classification on movie data set. In: Mandal D, Kar R, Das S, Panigrahi BK, editors. Intelligent computing and applications. New Delhi: Springer; 2015. p. 367–79.
24. Ravi K, Ravi V. A survey on opinion mining and sentiment analysis: tasks, approaches and applications. Knowl Based Syst. 2015;89:14–46.
25. Pang B, Lee L, Vaithyanathan S. Thumbs up? Sentiment classification using machine learning techniques. In: Proceedings of the conference on empirical methods in natural language processing (ACL, 2002). 2002. p. 79–6.
26. Habernal I, Ptáček T, Steinberger J. Supervised sentiment analysis in Czech social media. Inf Process Manag. 2014;50(2014):693–707.
27. Ghosh, M, Sanyal G. Preprocessing and feature selection approach for efficient sentiment analysis on product reviews. In: Proceedings of the 5th international conference on frontiers in intelligent computing: theory and applications, Satapathy SC. et al. editors. AISC 515, Springer-India; 2016.
28. Hatzivassiloglou V, McKeown KR. Predicting the semantic orientation of adjectives. In Proceedings of the 8th conference on European chapter of the association for computational linguistics. 1997. p. 174–81.
29. Pang B, Lee L. A sentimental education: sentiment analysis using subjectivity summarization based on minimum cuts. In: Proceedings of the 42nd annual meeting of the ACL. Barcelona; 2004. p. 271–8.
30. http://www.cs.cornell.edu/people/pabo/movie-reviewdata. Accessed 21 Oct 2016.
31. http://www.cs.jhu.edu/~mdredze/datasets/sentiment. Accessed 7 Jan 2017.
32. Blitzer J, Dredze M, Pereira F. Biographies, bollywood, boom-boxes and blenders: domain adaptation for sentiment classification. In: Proc. assoc. computational linguistics. Austin: ACL Press; 2007. p. 440–7.
33. Li JJ, Yang H, Tang H. Feature mining and sentiment orientation analysis on product review. In: Management information and optoelectronic engineering. 2015. p. 79–4.
34. Yang Y, Pedersen J. A comparative study on feature selection in text categorization. In: Proceedings of ICML-97, the 14th international conference on machine learning. 1997.
35. Shang W, Huang H, Zhu H, Lin Y, Qu Y, Wang Z. A novel feature selection algorithm for text categorization Expert Syst Appl. 2007;33(1):1–5.
36. Park H, Kwon S, Kwon MF. Complete Gini-index text (git) feature-selection algorithm for text classification. In: Proceedings of software engineering and data mining (SEDM). Piscataway: IEEE; 2010. p. 366–1.
37. McCallum A, Nigam KA. Comparison of event models for naive Bayes text classification. In: AAAI-98 workshop on learning for text categorization. vol 752, 1998. p. 41–8.
38. Kibriya AM, Frank E, Pfahringer B, Holmes G. Multinomial Naïve Bayes for text categorization revisited. Adv Artif Intell. 2004;3339:488–99.
39. Gladence L, Karthi M, Anu V. A statistical comparison of logistic regression and different Bayes classification methods for machine learning. ARPN J Eng Appl Sci. 2015;10(14):5947–53.
40. Nicholls C, Song F. Comparison of feature selection methods for sentiment analysis. Berlin: Adv Artif Intell; 2010. p. 286–9.
41. Felix N, Hruschka E, Hruschka ER. Tweet sentiment analysis with classifier ensembles. Decision Support Syst. 2014;57:77–93.
42. Kalaivani P, Shunmuganathan KL. Sentiment classification of movie reviews by supervised machine learning approaches. Indian J Comput Sci Eng (IJCSE). 2013;4:286–92.
43. Tripathy A, Agrawal A, Rath SK. Classification of sentiment reviews using n-gram machine learning approach. Expert Syst Appl. 2016;57(2016):117–26.
44. Al-Moslmi T, Gaber S, Al-Shabi A, Albared M, N. Omar.: Feature selection methods effects on machine learning approaches in malay sentiment analysis. In: Proceedings of the 1st ICRIL international conference on innovation in science and technology (IICIST 2015). 2015. p. 444–7.