International Journal of Applied Metaheuristic Computing
Volume 13 • Issue 1

Product Review-Based Customer Sentiment Analysis Using an Ensemble of mRMR and Forest Optimization Algorithm (FOA)

Parag Verma, Uttaranchal University, Dehradun, India*
https://orcid.org/0000-0002-3201-4285

Ankur Dumka, Computer Science and Engineering, Women Institute of Technology, Dehradun, India
Anuj Bhardwaj, Computer Science and Engineering, Chandigarh University, Punjab, India
Alaknanda Ashok, College of Technology, G. B. Pant University of Agriculture and Technology, India

ABSTRACT

This research presents a feature selection problem for classification of sentiments that uses ensemble-based classifier. This includes a hybrid approach of minimum redundancy and maximum relevance (mRMR) technique and forest optimization algorithm (FOA) (i.e., mRMR-FOA)-based feature selection. Before applying the FOA on sentiment analysis, it has been used as feature selection technique applied on 10 different classification datasets publicly available on UCI machine learning repository. The classifiers for example k-nearest neighbor (k-NN), support vector machine (SVM), and naïve Bayes used the ensemble based algorithm for available datasets. The mRMR-FOA uses the Blitzer's dataset (customer reviews on electronic products survey) to select the significant features. The classification of sentiments has improved by 12-18%. The evaluated results are further enhanced by the ensemble of k-NN, NB, and SVM with an accuracy of 88.47% for the classification of sentiment analysis task.

KEYWORDS

Classification Techniques, Feature Selection, Forest Optimization Algorithm, K-Nearest Neighbor, Minimum Redundancy and Maximum Relevance (mRMR) Technique, Sentiment Analysis, Supervised Learning

1. INTRODUCTION

Over the past few years, the dataset dimensionality has been increased in various domains like text-based sentiment analysis or bioinformatics.(Zhai et al., 2014) This reality has brought an intriguing challenge to the research field as much Artificial Intelligence (AI) or Machine Learning (ML) methods unable to manage high dimensional input data that involve products. Indeed, on the occasion that we examine the dimensionality of data posted in the well-known UCI repository and libSVM database,(Chang, 2001) we can see that the largest dimensionality of the dataset has expanded to over 30 million (approximately). Therefore, a part of these calculations is additionally when they face larger instance sizes. In this new situation, it is usual to manage information collection that is much larger than both the number of highlights and the number of tests, so current learning techniques must be adjusted.

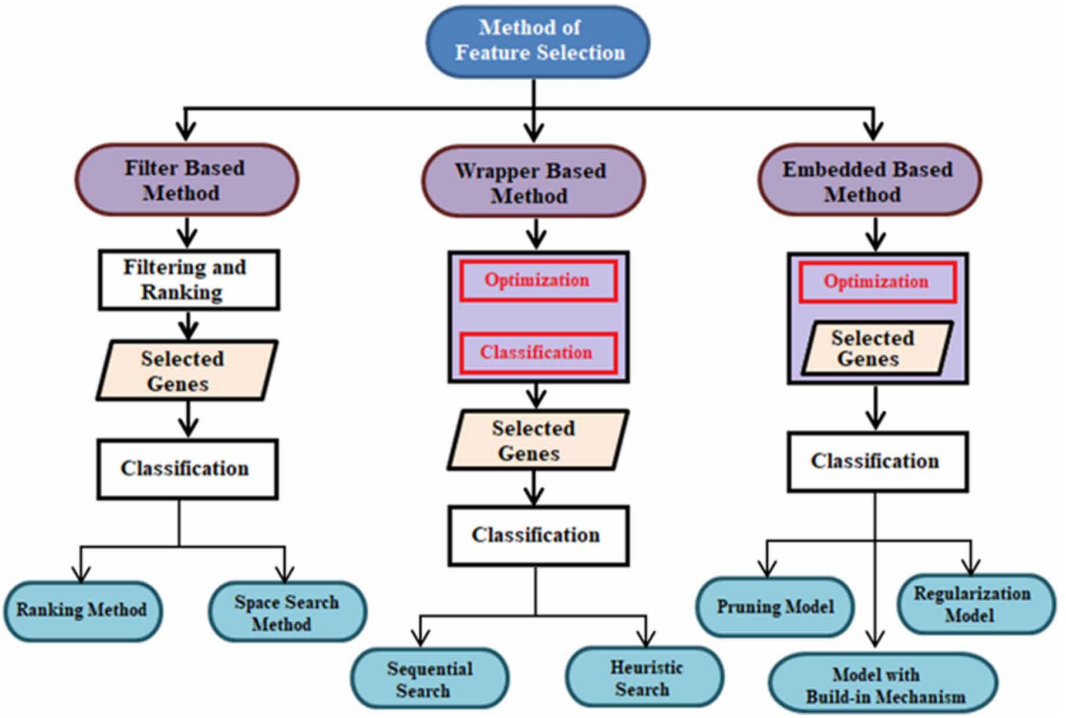
DOI: 10.4018/IJAMC.2022010107

To address this issue, dimension reduction methods can be applied to reduce the number of features and to enhance the performance of the resulting learning process. One of the most frequently used dimensionality reduction processes is the feature selection (FS), which accomplishes dimensionality reduction by emptying abstracts and additional features.(Liu & Motoda, 1998) Since FS places the highlights first, it is particularly valuable for applications where model translation and information extraction are important. In any case, existing FS techniques are not expected to scale well when managing a large-scale problem (in both various highlights and cases), in such a way that their effectiveness may be fundamentally broken or they can also be insignificant.

An analysis of sentiments is a way of identifying and classifying the emotions or opinions stated in some piece of text, sentence specifically in order to determining polarity whether the writer's disposition towards a particular topic or artefact is positive, negative, or neutral. For this purpose sentiment analysis and classification uses machine learning (ML) systems and natural language processing (NLP) together. The prevalence of rapid growth on the online social media and electronic network based societies provides all possible outcomes for customers to express their perceptions and exchange their ideas about entirety, for example, social or political issues through any article, book and films and so on through web-based networked media. These are usually in the form of survey material such as Likert type scaling data or text. Nowadays organizations are very fast, they evaluate popular perceptions about their customers or their articles of Internet-based social content.(Parvathy & Bindhu, 2016) Specific online service provider organizations are hooked in the evaluation of social media data in blogs, online forums, tweets, comments, and product feedback surveys. Publically shared reviews on sites or articles are used to recognize a customer's continued perception of any product or services to maintain a good commercialization with their decision making or the nature of its services or product quality.(Stylios et al., 2014) The critical problem that arises when collecting information from a social media networking environment is that the reviews consists mostly a large amount of unwanted data, including of HTML tags, linguistic and spelling errors, and the data is usually so bulky that removing those errors is human typical and time consuming task. An efficacious approach required to solving this problem is to select the usually relevant and significant features from the dataset and dispense repetitive or immaterial features. There are some pre-processing data cleaning techniques that rely on the choice of features selection. In the data mining process for high-dimensional dataset feature selection works as a highly effective pre-preparation strategy. Taxonomy of methods of feature selection present in Figure 1.

In the case of mining in social networks dataset, the analysis of high-dimension data is even more common. Classification of such high-dimensional dataset with a reasonable computational cost has become a vital topic of research in recent years. Most of the proposed solutions for analysis the sentiments are based on prior data processing or classification techniques to improve classification accuracy. In the same sequence Genetic algorithms (GA) and particle swarm optimization (PSO) have been used to select feature subsets from Artificial Intelligence (AI) domain. GA and PSO-based solutions have upgraded classification accuracy, but these substitute solutions are computationally high expensive due to that it affects performance of the system. GA and PSO are meta-heuristic algorithms that use a population of primary solutions. They can be used for problem optimization. This research paper presents and evaluates an ensemble based classification technique for sentiment analysis by using a newly developed evolutionary algorithm, called a Forest Optimization Algorithm (FOA).(Ghaemi & Feizi-Derakhshi, 2014) Ghaemi et al. proposed feature optimization algorithm named Forest Optimization Algorithm (FOA) in 2014,(Ghaemi & Feizi-Derakhshi, 2014) while its improved version named Feature Selection using Forest Optimization Algorithm (FSFOA) in 2016. (Ghaemi & Feizi-Derakhshi, 2016) Sowing and limiting populations based on lifetime in the tree process is simulated in these algorithm. The FOA produces the best tree (subset of features) among all other trees based on performance. The FOA supersedes GA and PSO when applied to reference functions and has the problem of optimizing weighting features using constant weights.

Figure 1. Feature selection methods taxonomy



2. RELATED WORK

Current methods for predicting and optimizing sentiments typically include feature selection technique that helps to reduce the number of attributes that will be stored in the database to provide irrelevant or repetitive features to provide progressively useful and effective results. Researcher Zdzislaw Pawlak in 1989 represents the frame of rough set theory, which can develop concept of approximation with fragmented data. There are many examples of such concept in accessible data and relationships between each other, for example, indiscernibility, set of approximation, deduction, and dependency. (Moustakidis & Theocharis, 2010)(Papakostas et al., 2011)

As per the paper concern, the proposed algorithm is an ensemble based methodology. So the objective of ensemble based methodology is to exploit the computational effectiveness of the channel model and the best possible execution of the covering methodologies. In which, the proposed commitment system is the channel model, which depends on data hypothesis. Common data as one of channel strategies is utilized to quantify the nature of features by surveying the relationship and excess of highlight, which has a strong hypothetical establishment. Until now, the combination of shared data, “significance” and “repetition” is broadly utilized for the purpose of feature selection. For instance, Peng et al. utilized the shared data as a measurement to gauge the connection between the features and class of examination named minimum Redundancy-Maximum Relevancy (mRMR).(Peng et al., 2005) Notwithstanding, their algorithms that was proposed is just appropriate for cooperation between two dependent or independent variables. To distinguish more confused variable communications, a few arrangements have been proposed. For further instances, Bennasar et al. proposed two new features selection techniques dependent on data hypothesis: Joint Mutual Information Maximization (JMIM) and Normalized Joint Mutual Information Maximization (NJMIM).(Bennasar et al., 2015) These two techniques are intended to address the issue of picking excess and unimportant highlights in certain situation. In additional instances, Vinh et al. proposed

a fundamental methodology for inferring new higher-dimensional machine learning based feature determination approaches by loosening up the recognized suspicions, and efficiently examined the issues of utilizing high-request conditions for common data based features selection.(Vinh et al., 2016) Feature selection dependent on the important repetition compromise measures has become a mainstream technique in the field of information mining. Be that as it may, the current features selection algorithm dependent on shared data despite everything have a few impediments on normal practice of feature selection. This sort feature selection techniques have the issue of overestimation or underestimation of significance of feature. To beat these significance, Che et al. proposed a novel shared data include choice technique dependent on the standardization of the most extreme pertinence and least regular repetition (N-MRMCR-MI).(Che et al., 2017)

Hu and Yao proposed estimates of rough sets held in complete and fragmented data frames to serve as the basis for three-way choices of rough sets.(Q. Hu et al., 2010) To deal with an inappropriate data frameworks researcher Feizi et al. present an increasingly generalized approach that consider the incomplete information of system.(Feizi-Derakhshi & Ghaemi, 2014) The calculation of rules and determining features are two important uses of rough sets. Balan EV.et al. introduced rule induction for each section of the acceptance of model in detail.(Balan et al., 2015) Halim Z. et al. carried out the rule induction in the absence of key feature values in the data frame and presented the use of research on sentiment analysis as a method to dissolve the proximity of human tracking in escort promotions extracted from the open web.(Halim et al., 2017)

Traditional strategies have not attentive on evaluation of human tracking as a textual sign and have instead focused on other visual cues (for example proximity to tattoos in related pictures) or printed cues (explicit style of ad structure, keywords, etc.). They applied two probing models commonly referenced to assumptions: the Netflix and Stanford models and train binary and categorical (multiclass) sentiment analysis models that use the escort review data extracted from the open network. Demonstrations of individual models and exploratory research led them to develop two models of group perceptions that effectively differentiate humans after evaluating 53% of those compared to a lot of 38,563 ads given to the company as DARPA MEMEX project. Hu, Z. et al.(Z. Hu et al., 2015) examine the effects of the selection of features in the perception test of the Chinese online review. From the outset, grams of n-char-grams and n-POS-grams are chosen as potential features for sentiment analysis. At that point, the improved document frequency strategy is used to include the feature subset, and the Boolean weighting technique is obtained to determine the feature weight. Chi-square test is performed to assess the significance of test results. The results suggest that 4-POS-gram as a test of perception of Chinese online review achieves greater accuracy as features. In addition, improved document iteration achieves a significant improvement in analysis of sentiments of Chinese online reviews.

Until this point in time, there have been many applied and developed instances of FOA. For example, Haindl et al. proposed a discrete form or binary version of the FOA to understand and solve the discrete issues.(Haindl et al., 2006) Yang et al. utilized FOA to locate the most intelligent response for the multidimensional rucksack issue utilizing a low number of emphases and low computational exertion.(Yang et al., 2015) Also, FOA is a developmental calculation that has been applied to include choice. Ghaemi and Feizi-Derakhshi utilized FOA to adapt to discrete hunt space issues like element choice and accomplished acceptable outcomes.(Ghaemi & Feizi-Derakhshi, 2016) What's more, he consolidated FOA with an angle technique to improve the fuzzy c-means (FCM) algorithm.(Chaghari et al., 2018)

The assessment of sentiments is essentially an evaluation of opinion mining that is used to extract a consistent review of individuals on any subject, occasion, or product. A practical way to present the results of this task of sentiment analysis or classification is step wise mining. In general, this classification is binary, whether positive or negative, about the review point, opportunity, or individual products. In general, the examination of the sentiment analysis and opinion mining is negotiable; however, in 2014 Walaa Medhat characterized them to be slightly different. (Medhat et al., 2014) Analysis of sentiments is used when a conclusion must be communicated in a file

or material, although opinion mining is used to extract people's feelings for research. Sentiment analysis research examines material perception and then accepts its polarity of sentiments. Pak A. and Paroubek P. proposed a three-step methodology for examination of sentiments that includes (i) corpus classification, (ii) corpus analysis and (iii) classifier training.(Pak & Paroubek, 2010) For the most part, the focus is on the classification of information for consideration in light of the fact that there are still no reference datasets for the problem of inclined arrangements. The majority of the dataset relies on reviews taken from micro blogging sites such as IMDB, Twitter and Amazon. com. IMDB has a film reviews; Twitter has review customers on small-scale opportunities in event courses, while Amazon has a wide variety of products. The most important and earliest step in the spread of emotion is extraction or choice, where some features of the material are chosen to break the tilt of the chosen material or file.

3. FEATURE SELECTION METHODS

The feature extraction approaches are used to extract valuable features from high-dimensional datasets. To predict or optimize sentiments extracting a hidden character window from available word or sentence required. Table 1, presents about the degree of sentiments with its characteristics.

Table 1. The degree of sentiments and its characteristics

S. No.	Dataset level	Delimiter	Depth of Granularity	Multiplicity of sentiments	Interpretation of Sentiments
1.	Text Records	'\n' newline character	Overall opinion at upper level	Single opinion of multiple entities	Overall sentiment of on document
2.	Sentence or Paragraph	"" Strings	Factual polarity of individual sentences	Multiple opinions of multiple entities	Subjectivity classification
3.	Entity or aspect level	Space character or named entities	At finest level words are the target entities	Single opinion of single entity	Two-tuple as <Sentiment, target>
4.	Character level	Special symbols and space characters are omitted	Micro level of character embedding	Multiple opinions about single word entity	Morphological extraction of words

Finally, the algorithm of classifier tries to assign selected features to the target class. Feature selection is also a procedure of exclusion unwanted features and data from dataset to improve the order of execution. Figure 1 presents about the scientific classification of several strategies for the feature selection based on Artificial Neural Networks (ANN).(Feizi-Derakhshi & Ghaemi, 2014) Three ways of feature selection approaches includes; the filter based methodology was probably the earliest method used to determine major feature subsets selection. The filter approach uses information properties and views instead of learning algorithm (for example, separation estimation, iteration calculation, etc.). For the most part, it validates two techniques for ranking the factors by assessing their importance and selecting a subset of features.(Blitzer et al., 2007) The latter methodology is the wrapper approach that uses learning algorithm while searching for a suitable subset of features. The selected classifier is a piece of feature selection technique; the wrapping method is the best choice of features when considering the accuracy of classification as part of its evaluation capability to discover the fitness of the model. Wrapping methods, when contrasted with filter techniques, are

extra time consuming for datasets with a large amount of features. In the embedded approach, the algorithm is used to detect the best features and based on the algorithm the features are chosen and then a classifier is used to assess its performance. Contrast with wrapping technique, filter technique is faster and more direct. Table 2 presents these methods with their respective techniques and frequency. However, in some past cases it has higher performance that grabs the attention of researchers for analysis through features selection.

Table 2. Feature Selection methods with their respective techniques and frequency

S. No.	Method Type	Feature Selection Technique	Frequency
1.	Filter	Correlation	6
		Correlation-based feature selection	4
		Relief	3
		Gain information	1
2.	Wrapper	Forward Selection	18
		Stepwise Regression	6
		Backward Selection/elimination	5
		Genetic algorithm	5
		Hill Climbing	4
		Gray relational analysis	4
		Exhaustive search	3
		Random search	2
		Best first Search	2
		Fuzzy logic	1
3.	Embedded	Mutual information	1
		Feature selection using clustering	1

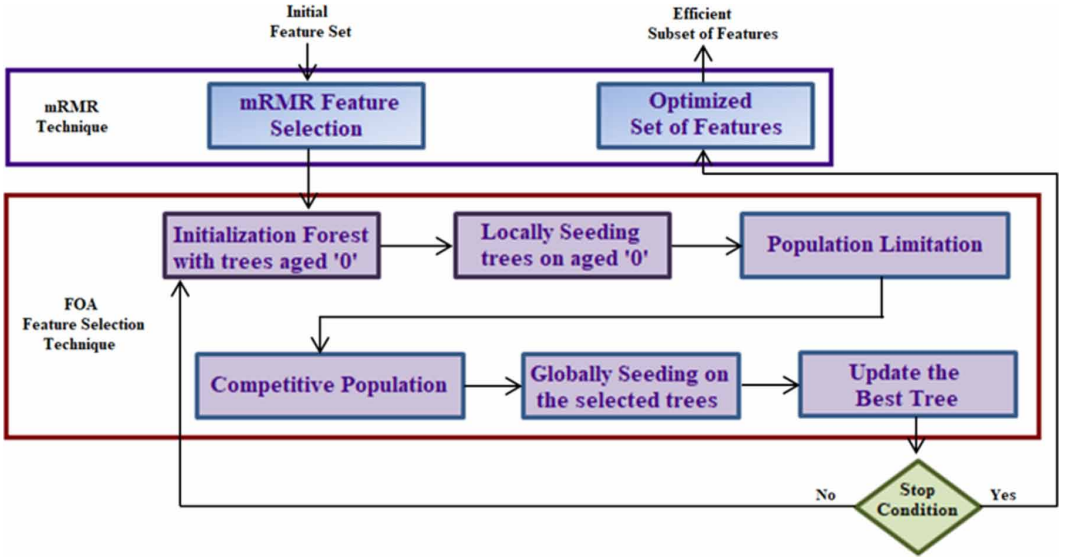
4. PROPOSED ALGORITHMS FOR SELECTION OF OPTIONAL FEATURES

The FOA algorithm is used to select the subsets of features and the data pre-preparation is skilled using the strategy of minimum redundancy and maximum relevance (mRMR). It is used to eliminate unwanted and non-essential features from the dataset before algorithm is used to feature selection. The proposed model hybrid mRMR and FOA to provide the best results. Figure 2 presents about the proposed working process of feature selection.

Phase 1: Dataset

The dataset is derived from Amazon that has been developed by John Blitzer and Mark Dredges to analysis the sentiments of customers used for classification purpose.(Gill et al., 2017) This dataset contains a variety of products and customers reviews, including electronic devices, kitchen appliances, books, DVDs and so on. All reviews are in a raw structure with HTML tags of review content, review identification number, date, and title, rating, product, and the location of customer. A product review with a star rating has five stars. The star rating review is turned into a positive review, when they have multiple stars or negative, they have fewer than three stars, and the rest of the audit has been terminated due to their ambiguous polarity. This research has used electronic products and books review datasets for the classification purpose. To investigate these customer reviews, some of the per-processing steps have been taken are as follows:

Figure 2. Proposed model for feature selection



- Review identification number, review text content and label were separated from HTML tags. The content of the review is the essential material for the examination.
- The entire contents were changed and accent marks were emptied to bring the lower case.
- Stop words, Short length words (i.e. less than 3 letters) and alphanumeric characters were removed.
- Steaming was implemented using the Porter steamer strategy.[1]

The input data to the algorithm must be in the form of feature vector type. After pre-processing, we get the word of words (BOW) that helps to forms feature vector. The pre-processing step has been performed on online MATLAB platform.(Halim et al., 2017)

Phase 2: Selection of features subset based on mRMR and Forest Optimization Algorithms (mRMR + FOA)

The BOW or feature vector, which was created in phase 1, has an exceptionally large size, even though it still have lots of unnecessary features. These insignificant features can affect the accuracy of the system and the larger size of the feature vector can extends the computational time. To maintain a strategic distance from this, we must include techniques for feature subsets selection in this feature vector to achieve an increasingly useful and effective set of features.

4.1 Selection of mRMR feature subset

mRMR is used to reduce the size of the feature vector that are capable to produces increasingly accurate results.(Ghaemi & Feizi-Derakhshi, 2014) Results and examinations have indicated that the MIQ (Mutual Information Quotient) rule is suitable for use with discrete information outside of the two mRMR criteria. Set of features obtained through the use of MIQ, and then sent to the FOA to obtain the best subset of features. This subset of features will give high calcification performance.

mRMR depend on common data that measure the data shared between two features. The data shared between two variables A and B is represented as the joint probability distribution of two variables A and B.

The shared data of a feature and class is known as the importance of the feature with the class. It refers to the shared data of a component with various features as iteration of the feature. mRMR for any feature 'k' is expressed as follows:

$$mRMR_k = \frac{Significance_k}{Iteration\, y_k}$$

This ratio of mRMR underlines that low iteration and high significance indicate that the component is not fundamentally related to different features and is highly dependent to the class.

4.2 Forest Optimization Algorithm (FOA)

The forest optimization algorithm (FOA) is a transformational algorithm inspired by some trees that survive from other trees based on their superior survival status or high fitness values. The purpose of the FOA is to deal with search space problems. FOA is used for various tasks but in this paper it is used for the feature selection process. The FOA has three primary stages, to be specific, (i) seeding trees locally, (ii) trees population limitation and (iii) seeding trees in an globally way.

The FOA for the most part begins with the population initialization, which accumulate forests that depend on algorithm. Each tree is characterized by the arrangement of each feature. A tree, regardless of the estimate of variable, is a section that presents the age of the tree concerned. Initially the age of a tree is set to '0'. After the trees have been initialized, the local seeding operator will produce or seed new trees from trees aged 0 and add new trees to the forest. At that point, all trees, with the exception of new ones, grow old and grow in age by '1'. Next, trees in the forest have control over the number of inhabitants and some trees will be excluded from the forest and will shape the population requesting for a global seeding phase. In the phase of global seeding, a level of competitive population is chosen to move away into the forest. The global seeding phase adds some potential new responses to the forest to eliminate ideal local states. Currently, forest trees are deployed for their fitness estimation and the tree with the highest fitness estimate is selected as the best tree and its age is set to 0 to maintain a strategic distance from maturity and shortly excluding the best tree from the forest tree (due to the local seeding phase extends the lifespan of all trees, including the best tree). These stages will continue until the final paradigm is completed.

```
Algorithm: Forest Optimization Algorithm (FOA)
Step 1:
Population initialization that creates forest in the algorithm
Step 2:
Forest initialization with tree aged '0'
Local trees seeding on aged '0'
Step 3:
Pre-defined population limitation,
if parameter "Age" >"Lifetime" process terminated
Step 4:
Global seeding for selected trees from candidate population
Step 5:
After best tree selection, fitness value is used for selecting the
best tree.
Step 6:
Finalizing process if the termination criteria achieved
```

FOA has been used to cover feature subset selection. This is encouraged by the process of seeding trees in a forest.(Huang et al., 2019) There are trees in the forests that have been survive for decades and some are due to a finite period. The survivability of a tree depends on the conditions of the area in

which they have been seeded. The FOA simulates the characteristic process of seed dispersal, which is essentially of two types, the local and the global dispersion processes of the seeds. In local seed dispersal, the seeds simply fall under trees and begin to develop, while in the spread of global seeds the seeds are carried to places that are away through animals, winds, or streams of water. Local seed spreading is known as the "local seeding" and the global seed spreading as "global seeding" in FOA.

The underlying population of trees is a matrix framework where each column presents of a possible solution and is known as a "tree". A single tree is a $1\times N_{var}$ dimensional. Monitoring the age of each tree makes it a $1\times\left(N_{var}+1\right)$ dimensional vector. This $\left(N+1\right)$ dimensional vector is our feature vector.

The fundamental progress phases involved with the forest optimization algorithm is examined below.

4.2.1 Initialize trees

Every tree from the forest represents the variable value. Regardless of the variable, each tree has a section identified with the 'Age' of that tree. In the beginning, the 'age' of each tree is set to '0' for each newly created tree, resulting in local seeding or global seeding. After each local seeding phase, the age of the trees increases, with the exception of new trees produced in the local seeding phase, by '1'. This growth is later used as a control tool in the amount of trees in the forest in the age of trees. Equation_1 shows a tree for the emission location of the N_{var} -dimensions, where the estimation variable and the "age" section indicate the age of the tree concerned.

A tree can also be considered in the equation as having a length of $1\times\left(N_{var}+1\right)$ according to equation 1, where N_{var} is the parameter that presents of the dimensions of the problems. At the beginning, all parameters are available in the vector. Parameters presences are mentioned by 1 and absence by 0 as per equation_1.

$$Tree=\begin{bmatrix} Age & v_1 & v_2 & v_3 & \cdots & v_{N_{var}} \\ 0 & 1 & 1 & 1 & \cdots & 1 \end{bmatrix}$$

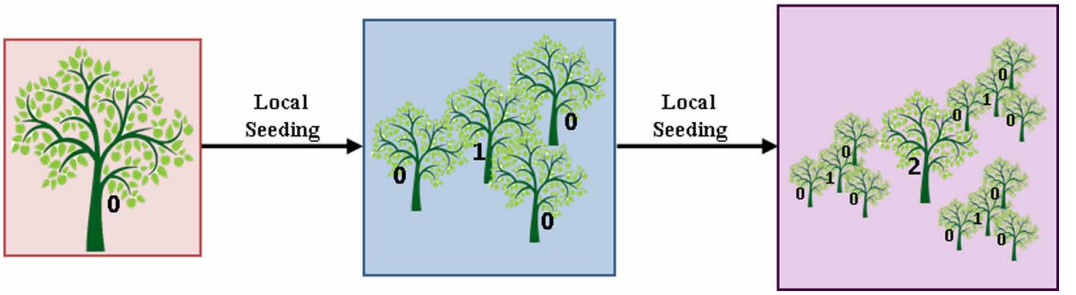
The most extreme permitted age of a tree is a predefined parameter and is called the "lifetime" parameter. The "lifetime" parameter must be resolved at the beginning of the algorithm. When the "age" of a tree is raised to the "lifetime" parameter, the tree is removed from the forest and added to the competitive population. In the event that we choose a big number for this parameter, each step on algorithm only expands the age of the trees and the forest will be filled with older trees, which do not participate in the local seeding stage. Furthermore, if we choose a small value for this parameter, the trees will aged very soon and are excluded at the beginning of the challenge. Therefore, this parameter should give a decent probability of local search.

4.2.2 Local seeding of the trees

In the environment, while the tree seeding procedure begins, some seeds simply come near the trees and after a while they turn into young trees. Currently, the challenge between trees begins and those trees become the winners of this challenge to bear better growth conditions, such as adequate sunlight and better area. The local seeding of the trees tries to simulate this method of nature. This operator is built in trees with age '0' and includes some neighbors of each tree in the forest. In Figure 3 represents two steps forwards of this operator in one tree. The numbers composed within the trees in Figure 3 represent the "age" estimate of the related trees; which looks like a zero for a newly created tree. Local planting is performed on trees with an age of 0, believed to extend from 1 in addition to new trees produced at this time.

Expanding of the age of trees as a means of controlling the number of trees in the forest and the effects of this time: a tree is promising, the algorithm procedure restores the age of that tree to '0' and, as a result, it conceives will include neighbors of large trees in the forests through the local seeding phase. In addition, non-promising trees get aged with promotion of each iteration and, in the die after some iteration.

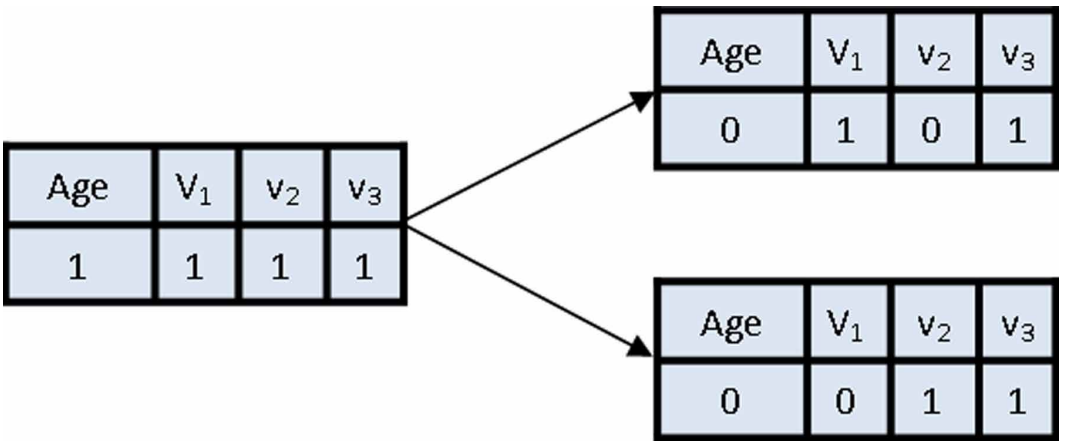
Figure 3. Example of local seeding of tree for first 2 iterations



The seeds that fall on the ground near the trees and then become trees as neighbors are considered a parameter of this algorithm and are called "local planting changes" abbreviate "LSC". The estimate of this parameter is 3 as Figure 4. As a result of this, performing local seeding operator in a tree with age 0 will distribute 3 new trees. This parameter problem must be solved by the dimension component. The local seeding operator adds several trees to the forest, so there should be a restriction on the number of trees. This control ends in the next step of the algorithm.

This process is done several times for each tree in the local seeding phase. Local seeding is applied to each tree within the forests. Within adjacent recurrence cycles, access to properly maintained trees may decrease as the age of all trees increases in the hope of new ones.

Figure 4. Tree parameters before and after local seeding



4.2.3 Population Limiting

The number of trees in the forests should be limited to avoid endless development of forests. There are 2 parameters that limit the population of trees in forest: the parameters "Lifetime" and "area boundary". From the beginning, trees whose age exceeds parameter "lifetime" are removed from the forest and will shape the competitive population. The second restriction is the "area boundary", in which, after the placement of the trees is indicated by an estimate of their fitness, if the number of trees exceeds the area limit, additional trees are expelled. From the forest and added to the competitive population. Forest limitation is another parameter and is called "area limitation".

4.2.4 Global Seeding Of The Trees

There are many types of trees in the forest and many animals and birds feed on the seeds and products of these trees. Therefore, the seeds of trees are appropriated throughout the forest and later, the important place of trees becomes more extensive. Similarly, other common processes such as ingestion and progress of water help spread seeds throughout the forest and ensure dominance of different types of trees in different districts. The global seeding phase attempts to simulate the appropriation of tree seeds in the forest.

The global seeding operator is performed at a predefined level of the requested population of the previous phase. This is another parameter of rate algorithm that must be characterized first and is called "transfer rate". The global seeding operator is working as: from the beginning, the trees in the competitive population are selected. At that point, a part of each tree's variable is chosen arbitrarily. This time, the estimate of each variable chosen is interacting with another incentive produced at random in the range of the respective variables. In this sense, the entire search space is considered and not a restricted area. Therefore, with age 0 a tree is added to the forest. This operator performs global search in the search space. The quantity of factors whose properties will be changed is another parameter of algorithm and is called "Global Seeding Changes" abbreviate "GSC".

4.2.5 Fitness Function

The accuracy of the classification form the classifier was selected according to the fitness function. To validate the process of feature selection, classification accuracy is an operative degree and is defined as;

$$CA = \frac{the\,number\,of\,correct\,classification}{the\,number\,of\,all\,samples\,of\,the\,dataset} \times 100\%$$

4.2.6 Updating The Best So Far Tree

At this phase of FOA algorithm, according to the estimated fitness value arranging the trees, the tree with the highest fitness estimate has been selected as the best tree. Then, the age of the best tree will be set to 0 to maintain a strategic distance from the maturity of the best tree as a result of the local seeding phase. At this time, it will be possible for the best tree to be locally organized by the local seeding operator.

4.2.6 Stop Condition

Alike other evolutionary algorithms, three break conditions considered as follows:

- Predefined iteration number
- Monitoring of any variation in the fitness value of the best tree for some iterations
- Achieve predefined level of accuracy

Algorithm. Proposed Forest Optimization Algorithm (LSC, GSC, lifetime, area_limit, transfer_rate)

Input variables		:	LSC, GSC, lifetime, area_limit, transfer_rate
Output		:	Optimized includes subsets with high fitness values
Step 1:	Set the forest at random with age 0 or 1		
Step 2:	Each tree has a vector X in (D + 1)		
Step 3:	The Age = 0 of all trees in the forest		
	while	Condition of detention is not met at that time, seeding for the tree is done with age = 0.	
	for	i = 0;	/* local seeding changes*/
	do		
		Variables are arbitrarily selected from the chosen tree	
		Upgraded the attribute Age from 0 to 1	
	end *for*		
	Age of trees increased by 1. The population constraint was distributed. Perform global seeding changes. Some measure of competitive population is selected		
	for /*each selected tree*/		
	Global seeding changes value of selected tree is randomly chosen. Change the age value from 0 to 1 or vice versa.		
	end *for*		
	Step 3.1	Best tree is updated	
		Trees are ordered by the fitness value. Best trees parameters are set as 0.	
	end *while*		
Step 4:	Return best tree, including the chosen features		

4.3 Ensemble-Based Classification Technique

Three fundamental classifiers are used to ensemble model the sentiment classification. K-Nearest Neighbor (k-NN), Naïve Bayes (NB) and Support Vector Machine (SVM). All these supervised machine learning classification algorithms are used in a similar topology of multi-classification frameworks (ensemble). In our proposed approach, similar classifiers would improve classification accuracy for the task of sentiment analysis.

5. RESULTS AND ANALYSIS

In the proposed approach, classification performance is improved by using classifier algorithm. The dataset used appears in Table 3. All the used datasets used in this research paper are taken from the UC Irvine Machine Learning Repository (UCI) repository. At the first step, FOA is applied to 10 classification datasets and obtained results are compared with the mostly used classifier such as NB, SVM, and KNN.

5.1 Classification accuracy without FOA

The classification accuracy obtained from the Naïve Bayes, Support vector machine and k-nearest neighbors in the dataset mentioned above in Table 4. (Here k-NN is used with an estimate of k = 1)

Table 3. Reference classification datasets from UCI repository

S. No.	Dataset Name	No. of Instances	Area	No. of Attributes	No. of Classes	Type of attributes
1.	Breast Cancer Wisconsin (Original)	699	Life	10	2 for benign, 4 for malignant	Integer
2.	Hepatitis	155	Life	19	2	Categorical, Integer, Real
3.	Lung Cancer	32	Life	56	4	Integer
4.	Iris	150	Life	4	3	Real
5.	Statlog (German Credit Data)	1000	Financial	20	2	Categorical, Integer
6.	Chess (King-Rook vs. King-Pawn)	3196	Game	36	2	Categorical
7.	Drug consumption	1885	Social	32	7	Real
8.	Bank Marketing	45211	Business	17	2	Real
9.	Alcohol QCM Sensor	125	Computer	8	5	Real
10.	Teaching Assistant Evaluation	151	Others	5	3	Categorical, Integer

Table 4. Classification accuracy of NB, SVM and k-NN (k = 1) on UCI datasets

Dataset Name	No. of Instances	No. of Attributes	NB(%)	SVM(%)	k-NN (k=1) (%)
Breast Cancer Wisconsin (Original)	699	10	86.56	79.25	82.34
Hepatitis	155	19	56.67	53.40	52.63
Lung Cancer	32	56	51.34	39.85	37.78
Iris	150	4	96.25	82.34	95.45
Statlog (German Credit Data)	1000	20	58.89	54.23	61.70
Chess (King-Rook vs. King-Pawn)	3196	36	47.51	51.88	39.35
Drug consumption	1885	32	37.45	38.76	47.53
Bank Marketing	45211	17	64.21	68.90	72.23
Alcohol QCM Sensor	125	8	86.75	92.71	90.67
Teaching Assistant Evaluation	151	5	90.40	89.20	87.56

5.2 Classification Accuracy with FOA

The classification accuracy obtained from the FOA is shown in Table 5. For this simple FOA used to select the features. The FOA provides promising results in these reference datasets. The amount of features reduced by FOA is better when characterization accuracy is improved compared to single features. Similarly Table 6 is used to compare the classification accuracy with and without FOA on the selected datasets from the UCI repository listed in Table 3.

Table 5. Classification accuracy using FOA on UCI datasets

Dataset Name	No. of Instances	No. of Attributes	NB(%)	SVM(%)	k-NN (k=1) (%)
Breast Cancer Wisconsin (Original)	699	10	88.10	81.33	84.45
Hepatitis	155	19	57.90	52.45	62.61
Lung Cancer	32	56	56.34	42.55	40.77
Iris	150	4	96.15	86.92	95.20
Statlog (German Credit Data)	1000	20	61.00	55.43	68.87
Chess (King-Rook vs. King-Pawn)	3196	36	48.34	54.87	41.10
Drug consumption	1885	32	41.85	46.91	52.76
Bank Marketing	45211	17	65.15	72.84	75.25
Alcohol QCM Sensor	125	8	87.03	92.17	91.60
Teaching Assistant Evaluation	151	5	91.73	90.12	87.43

5.3 mRMR Based Feature Selection Using FOA

For feature subset selection mRMR is applied to all these datasets to improve the classification results with lower numbers features. Dataset with their specific number of features and chosen features that use the mRMR method are shown in Table 7.

The amounts of features mentioned above are chosen based on their superior order of classification accuracy. The accuracy of the classification is obtained using two distinct arbitrary initial populations of 25 and 50 and which appear in Tables 8 and Table 9.

A comparative examination has been performed in the reference dataset as compare to the subsequent effect of mRMR + FOA and the algorithm of claimants such as PSO and GA. Table 10 shows the resulting effect of the Naïve Bayes classifier in the three evolutionary algorithm used with the mRMR method.(Blitzer et al., 2007) The results obtained for Breast Cancer, Hepatitis and Chess (King-Rook vs. King-Pawn) datasets improve by 4%, 7% and 9% respectively.

5.4 mRMR Based Sentiment Classification Using FOA

To classifying the sentiments, in this research a product review dataset taken from Amazon and created by John Blitzer and Mark Dredze has been used.(Blitzer et al., 2007) It includes the review for four different types of items that considers books, DVDs, electronic appliances, and kitchen products. This research covers the sentiments classification results based on the Blitzer electronic products review

Table 6. Comparison of classification accuracy with and without FOA on UCI datasets

Dataset Name	NB w/o FOA(%)	NB FOA (%)	SVM w/o FOA(%)	SVM FOA(%)	k-NN (k=1) w/o FOA(%)	k-NN (k=1) FOA (%)
Breast Cancer Wisconsin (Original)	86.56	88.10	79.25	81.33	82.34	84.45
Hepatitis	56.67	57.90	53.40	52.45	52.63	62.61
Lung Cancer	51.34	56.34	39.85	42.55	37.78	40.77
Iris	96.25	96.15	82.34	86.92	95.45	95.20
Statlog (German Credit Data)	58.89	61.00	54.23	55.43	61.70	68.87
Chess (King-Rook vs. King-Pawn)	47.51	48.34	51.88	54.87	39.35	41.10
Drug consumption	37.45	41.85	38.76	46.91	47.53	52.76
Bank Marketing	64.21	65.15	68.90	72.84	72.23	75.25
Alcohol QCM Sensor	86.75	87.03	92.71	92.17	90.67	91.60
Teaching Assistant Evaluation	90.40	91.73	89.20	90.12	87.56	87.43

Table 7. Selected features using mRMR technique

Dataset Name	No. of Instances	No. of Attributes	mRMR attributes	NB(%)	SVM(%)	k-NN (k=1) (%)
Breast Cancer Wisconsin (Original)	699	10	6	93.44	95.85	90.11
Hepatitis	155	19	11	81.14	86.05	78.30
Lung Cancer	32	56	15	78.20	82.55	70.72
Iris	150	4	3	97.62	89.72	96.29
Statlog (German Credit Data)	1000	20	10	81.50	72.43	78.81
Chess (King-Rook vs. King-Pawn)	3196	36	14	78.38	74.82	71.19
Drug consumption	1885	32	14	61.81	76.21	62.86
Bank Marketing	45211	17	9	78.25	82.81	85.65
Alcohol QCM Sensor	125	8	4	95.63	96.27	94.91
Teaching Assistant Evaluation	151	5	3	96.70	96.62	92.29

data set. A set of total 7580 components of the electronic products review dataset extracted after data pre-processing phase. These 7580 attributes are reduced to 1050 after the use of the mRMR strategy. The classification accuracy obtained was less than 80% without applying mRMR. Improved results can be seen in Tables 10 and Table 11.

Table 8. Classification accuracy of mRMR-FOA with a population size of 25 on UCI datasets

Dataset `Name	No. of Instances	No. of Attributes	mRMR attributes	mRMR-FOA attributes	No. of Runs	NB (%)	SVM(%)	k-NN (k=1) (%)
Breast Cancer Wisconsin (Original)	699	10	6	5	4080	94.04	96.15	91.23
Hepatitis	155	19	11	9	4080	84.45	88.25	81.80
Lung Cancer	32	56	15	13	4080	80.00	83.20	72.12
Iris	150	4	3	3	4080	97.87	90.65	97.23
Statlog (German Credit Data)	1000	20	10	7	4080	82.25	73.21	79.67
Chess (King-Rook vs. King-Pawn)	3196	36	14	11	4080	79.13	75.87	72.24
Drug consumption	1885	32	14	12	4080	62.70	78.45	66.20
Bank Marketing	45211	17	9	7	40 80	79.95	83.27	85.97
Alcohol QCM Sensor	125	8	4	4	4080	96.12	97.98	96.27
Teaching Assistant Evaluation	151	5	3	3	4080	97.67	97.07	93.65

Table 9. Classification accuracy of mRMR-FOA with a population size of 50 on UCI datasets

Dataset Name	No. of Instances	No. of Attributes	mRMR attributes	mRMR-FOA attributes	No. of Runs	NB (%)	SVM(%)	k-NN (k=1) (%)
Breast Cancer Wisconsin (Original)	699	10	6	5	4080	94.34	96.45	91.23
Hepatitis	155	19	11	9	4080	84.45	88.25	81.80
Lung Cancer	32	56	15	13	4080	80.20	83.45	72.12
Iris	150	4	3	3	4080	97.87	90.65	97.23
Statlog (German Credit Data)	1000	20	10	7	4080	82.25	73.21	79.67
Chess (King-Rook vs. King-Pawn)	3196	36	14	11	4080	79.89	76.34	72.24
Drug consumption	1885	32	14	12	4080	62.70	78.45	66.20
Bank Marketing	45211	17	9	7	4080	79.95	83.27	86.09
Alcohol QCM Sensor	125	8	4	4	4080	96.76	97.98	96.19
Teaching Assistant Evaluation	151	5	3	3	4080	97.67	97.07	93.83

Table 10. Comparison of FOA with GA and PSO for NB on UCI datasets

Dataset Name	No. of Attributes	Accuracy of GA + mRMR	Accuracy of PSO + mRMR	Accuracy of FOA + mRMR
Breast Cancer Wisconsin (Original)	5	92.15	94.07	95.12
Hepatitis	9	84.17	89.65	92.20
Lung Cancer	13	80.20	83.45	72.12
Iris	3	97.87	90.65	97.23
Statlog (German Credit Data)	7	82.25	73.21	79.67
Chess (King-Rook vs. King-Pawn)	11	80.10	82.83	85.10
Drug consumption	12	62.70	78.45	66.20
Bank Marketing	7	75.91	84.12	87.30
Alcohol QCM Sensor	4	96.49	96.61	96.92
Teaching Assistant Evaluation	3	95.19	96.24	92.36

Table 11. Sentiment classification using mRMR-FOA for the products' reviews dataset (population size 40)

Number of instances	Number of Attributes	mRMR-FOA selected Attributes	No. of Runs	NB	SVM	k-NN (k=1)
100	7580	1050	1020	76.72%81.00%	82.17%87.82%	79.14%83.33%
150	7580	1050	1020	80.19%82.45%	89.15%91.06%	78.00%81.20%

For sentiment classification this research has used three classifications techniques k-NN, NB and SVM with FOA. Research has used datasets with three different numbers of instances to look at the collaboration of FOA with increase in number of instances.

5.5 Results from Ensemble of Classifiers

The set of three classifiers, which are NB, SVM and k-NN further enhances performance by selecting the result with the most notable classification accuracy. In all assessments of classification of the order of sentiments, SVM has played the best of the three considered classifiers. The results can be found in table 13.

6. DISCUSSIONS AND CONCLUSION

Hybridization of FOA with classifier techniques NB, SVM and k-NN have outperformed individual classifiers NB, SVM and k-NN, when applied to different reference classification datasets downloaded from the UCI repository. Following these improved results, the proposed strategy that ensemble mRMR with FOA was applied to the electronic products review dataset taken from Amazon. For the sentiment of classification, research has executed three classifiers, including SVM, to observe the accuracy of the classification. The experiment was completed in two different quantities with two initial population sizes. The expansion in the size of instances gradually increases the FOA calculation time. With each of the two, SVM takes extra time with FOA, although the computation time of K-NN and NB increases with the number of consecutive instances. When the mRMR (features

Table 12. Sentiment classification using mRMR-FOA for the products' reviews dataset (population size 20)

Number of instances	Number of Attributes	mRMR-FOA selected Attributes	No. of Runs	NB	SVM	k-NN (k=1)
100	7580	1050	1020	81.61%85.45%	83.81%88.47%	80.17%84.55%
150	7580	1050	1020	79.14%82.90%	76.19%83.35%	76.90%78.15%

Table 13. Outcomes from ensemble classifiers for products' reviews dataset

Population	Number of Attributes	mRMR-FOA selected Attributes	NB	SVM	*k*-NN (k=1)
40	7028	1050	81.00%	87.82%	83.33%
20			85.45%	88.47%	84.55%

subset selection) strategy is applied with FOA, the accuracy of the order of sentiment classification of the similar dataset is improved by 12-18%. The ensemble classifier NB, SVM and k-NN runs to optimize the final performance. The final conclusive results for the two unique populations 40 and 20 are 87% and 88% separately.

REFERENCES

Balan, E. V., Priyan, M. K., Gokulnath, C., & Devi, G. U. (2015). Fuzzy based intrusion detection systems in MANET. *Procedia Computer Science*, *50*, 109–114. doi:10.1016/j.procs.2015.04.071

Bennasar, M., Hicks, Y., & Setchi, R. (2015). Feature selection using joint mutual information maximisation. *Expert Systems with Applications*, *42*(22), 8520–8532. doi:10.1016/j.eswa.2015.07.007

Blitzer, J., Dredze, M., & Pereira, F. (2007). Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification. *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, 440–447.

Chaghari, A., Feizi-Derakhshi, M.-R., & Balafar, M.-A. (2018). Fuzzy clustering based on Forest optimization algorithm. *Journal of King Saud University-Computer and Information Sciences*, *30*(1), 25–32. doi:10.1016/j.jksuci.2016.09.005

Chang, C.-C. (2001). *A library for support vector machines.* Http://Www. Csie. Ntu. Edu. Tw/~ Cjlin/Libsvm

Che, J., Yang, Y., Li, L., Bai, X., Zhang, S., & Deng, C. (2017). Maximum relevance minimum common redundancy feature selection for nonlinear data. *Information Sciences*, *409*, 68–86. doi:10.1016/j.ins.2017.05.013

Feizi-Derakhshi, M.-R., & Ghaemi, M. (2014). Classifying different feature selection algorithms based on the search strategies. *International Conference on Machine Learning, Electrical and Mechanical Engineering*, 17–21.

Ghaemi, M., & Feizi-Derakhshi, M.-R. (2014). Forest optimization algorithm. *Expert Systems with Applications*, *41*(15), 6676–6687. doi:10.1016/j.eswa.2014.05.009

Ghaemi, M., & Feizi-Derakhshi, M.-R. (2016). Feature selection using forest optimization algorithm. *Pattern Recognition*, *60*, 121–129. doi:10.1016/j.patcog.2016.05.012

Gill, A. J., Hinrichs-Krapels, S., Blanke, T., Grant, J., Hedges, M., & Tanner, S. (2017). Insight workflow: Systematically combining human and computational methods to explore textual data. *Journal of the Association for Information Science and Technology*, *68*(7), 1671–1686. doi:10.1002/asi.23767

Haindl, M., Somol, P., Ververidis, D., & Kotropoulos, C. (2006). Feature selection based on mutual correlation. *Iberoamerican Congress on Pattern Recognition*, 569–577.

Halim, Z., Atif, M., Rashid, A., & Edwin, C. A. (2017). Profiling players using real-world datasets: Clustering the data and correlating the results with the big-five personality traits. *IEEE Transactions on Affective Computing*.

Hu, Q., Che, X., Zhang, L., & Yu, D. (2010). Feature evaluation and selection based on neighborhood soft margin. *Neurocomputing*, *73*(10–12), 2114–2124. doi:10.1016/j.neucom.2010.02.007

Hu, Z., Hu, J., Ding, W., & Zheng, X. (2015). Review sentiment analysis based on deep learning. *2015 IEEE 12th International Conference on E-Business Engineering*, 87–94.

Huang, C., Zhu, J., Liang, Y., Yang, M., Fung, G. P. C., & Luo, J. (2019). An efficient automatic multiple objectives optimization feature selection strategy for internet text classification. *International Journal of Machine Learning and Cybernetics*, *10*(5), 1151–1163. doi:10.1007/s13042-018-0793-x

Liu, H., & Motoda, H. (1998). *Feature extraction, construction and selection: A data mining perspective* (Vol. 453). Springer Science & Business Media. doi:10.1007/978-1-4615-5725-8

Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, *5*(4), 1093–1113. doi:10.1016/j.asej.2014.04.011

Moustakidis, S. P., & Theocharis, J. B. (2010). SVM-FuzCoC: A novel SVM-based feature selection method using a fuzzy complementary criterion. *Pattern Recognition*, *43*(11), 3712–3729. doi:10.1016/j.patcog.2010.05.007

Pak, A., & Paroubek, P. (2010). Twitter as a corpus for sentiment analysis and opinion mining. *LREc*, *10*, 1320–1326.

Papakostas, G. A., Polydoros, A. S., Koulouriotis, D. E., & Tourassis, V. D. (2011). *Evolutionary feature subset selection for pattern recognition applications*. INTECH Open Access Publisher UK. doi:10.5772/15655

Parvathy, G., & Bindhu, J. S. (2016). A probabilistic generative model for mining cybercriminal network from online social media: A review. *International Journal of Computers and Applications*, *134*(14).

Peng, H., Long, F., & Ding, C. (2005). Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *27*(8), 1226–1238. doi:10.1109/TPAMI.2005.159 PMID:16119262

Stylios, G., Katsis, C. D., & Christodoulakis, D. (2014). Using Bio-inspired intelligence for Web opinion Mining. *International Journal of Computers and Applications*, *87*(5).

Vinh, N. X., Zhou, S., Chan, J., & Bailey, J. (2016). Can high-order dependencies improve mutual information based feature selection? *Pattern Recognition*, *53*, 46–58. doi:10.1016/j.patcog.2015.11.007

Yang, Q., Wang, M., Xiao, H., Yang, L., Zhu, B., Zhang, T., & Zeng, X. (2015). Feature selection using a combination of genetic algorithm and selection frequency curve analysis. *Chemometrics and Intelligent Laboratory Systems*, *148*, 106–114. doi:10.1016/j.chemolab.2015.09.007

Zhai, Y., Ong, Y.-S., & Tsang, I. W. (2014). The emerging" big dimensionality. *IEEE Computational Intelligence Magazine*, *9*(3), 14–26. doi:10.1109/MCI.2014.2326099

ENDNOTE

[1] MATLAB online platform: https://in.mathworks.com/products/matlab-online.html (accessed on 08 January 2020)

Parag Verma is working as Assistant Professor in Uttaranchal University, Dehradun. He is having a long experience of 8+ years in the field of industry and academics. He is having 15+ international papers in reputed conference and journals. He has contributed 2 books with publishers like Nerosa and Alpha, and currently working on 2 more accepted books. He had also contributed 4 chapters for reputed publishers. He is also editorial board members of many reputed conference and journals including Scopus. He is also guest editor of IJNCDS (Scopus Indexed Inderscience Journal) and many more. He is associated with many societies and organization for welfare of educationalist societies.

Ankur Dumka is working as Associate Professor in the Women Institute of Technology (Govt.), Dehradun. He is having a long experience of 10+ years in the field of industry and academics. He is associated with smart city Dehradun as academic expert committee member and co-ordinator in terms of projects related to IT. He is having 40 + international papers in reputed conference and journals. He has contributed 3 books with publishers like Taylor and Francis, IGI global etc. and currently working on 3 more accepted books. He had also contributed 14 chapters for reputed publishers. He is also editorial board members of many reputed conference and journals including Scopus and IEEE. He is also guest editor of IJNCDS (Scopus Indexed Inderscience Journal), guest editor for IJNCR (ACM digital library -IGI global journal) and many more. He is associated with many societies and organization for welfare of educationalist societies.

Anuj Bhardwaj, M. Tech. & Ph. D degree in computer science. He is currently a Professor of Computer Science & Engineering with Chandigarh University, Mohali, Punjab. He is having a long experience of 15+ years in the field of industry and academics. He is having 30+ international papers in reputed conference and journals. He has contributed 2 books with publishers like Nerosa and Alpha, and currently working on 2 more accepted books. He had also contributed 4 chapters for reputed publishers. His current research interests include pattern & character recognition, graphics & vision, artificial intelligence & neural networks and machine learning.

Alaknanda Ashok is working as a dean and professor, college of technology, G.B.Pant University of Agriculture and Technology, Pantnagar. She is having a large administrative experience as a director, women Institute of technology, and controller of examiner of Uttarakhand technical university. She is having more than 50 research papers in reputed journals and conferences of repute. She is having 3 patents published. She is having many research projects working under her supervision as a project in charge. She is also having many book chapters and currently working on proposed books with Taylor and Francis. She is also associated with many journals and conferences in the capacity of the editor, editorial board member, and convener.