FAHPBEP: A Fuzzy Analytic Hierarchy Process Framework in Text Classification

Razieh Asgarnezhad[1], Seyed Amirhassan Monadjemi[2*], Mohammadreza Soltanaghaei[3]
1- Department of Computer Engineering, Isfahan (Khorasgan) Branch, Islamic Azad University, Isfahan, Iran.
Email: r.asgarnezhad@khuisf.ac.ir
2- Faculty of Computer Engineering, University of Isfahan, Isfahan, Iran.
Email: monadjemi@eng.ui.ac.ir (Corresponding author)
3- Department of Computer Engineering, Isfahan (Khorasgan) Branch, Islamic Azad University, Isfahan, Iran.
Email: soltan@khuisf.ac.ir

Received: April 2020 Revised: July 2020 Accepted: August 2020

ABSTRACT:
With the availability of websites and the growth of comments, reviews of user-generated content are published on the Internet. Sentiment Classification is one of the most common problems in text mining, which applies to categorize reviews into positive and negative classes. Pre-processing has an important role when these textual contexts are employed by machine learning techniques. Without efficient pre-processing methods, unreliable results will be achieved. This research probes to investigate the performance of pre-processing for the Sentiment Classification problem on three popular datasets. We suggest a high-performance framework to enhance classification performance. First, features of user's opinions are extracted based on three methods: (1) Backward Feature Selection; (2) High Correlation Filter; and (3) Low Variance Filter. Second, the error rate of the primary classification for each method is calculated through the perceptron. Finally, the best method is selected through the fuzzy analytic hierarchy process. This framework is beneficial for companies to observe people's comments about their brands and for many other applications. The current authors have provided further evidence to confirm the superiority of the proposed framework. The obtained results indicate that on average this proposed framework outperformed its counterparts. This framework yields 90.63 precision, 90.89 accuracy, 91.27 recall, and 91.05% f-measure.

KEYWORDS: Data Mining, Sentiment Classification, Feature Selection, Fuzzy Analytic Hierarchy Process, Perceptron Neural Network.

1. INTRODUCTION

Millions of people in the world intensity employ websites like Twitter to publish reviews and tell others what are they thinking. It is widespread interest due to the explosion of data on the Internet. For example, huge comments are published every minute on websites like Twitter, and it has more than several billion views per day. Consequently, finding appropriate information from a huge content of reviews is considered as a challenging task. The problem becomes more sophisticated when these reviews are given in natural language. There is a big gap between reviews in natural language in one hand and computer language, which is described by zero and one on the other. Hence, it considers obtaining beneficial information from the reviews [1].

Sentiment Classification (SC) problem analyzes the unstructured data. It automatically obtains the opinion from websites and classifies the polarity of text in terms of positive (good) and negative classes. It causes a make-decision problem to be done automatically. Consequently, a suitable framework is very important to manage this problem and capture all helpful information.

SC is an absorbing field in text mining. The reviews from the unstructured data on the Internet extracted are classified as positive, negative, or neutral. Therefore, a fitting framework is very important to facilitate the capture of all useful information. In this context, the three levels of document, sentence, and feature are of concern. The classes of the first two are defined respectively. In the feature-level, the class of each feature of reviews are determined.

Most of the methods apply SC approaches, which in practice are classified into Machine Learning (ML), Lexicon-based, and Hybrid approaches. Regrettably, the lexicon-based approaches cannot determine the opinion words with domain and context specification orientations. The main advantage of these approaches ratio to its counterparts is to support finding domain and context-specific opinion words utilizing a domain

Doi: https://doi.org/10.29252/mjee.14.3.14
How to cite this paper:R. Asgarnezhad, S.A. Monadjemi, and M. Soltanaghaei, **“FAHPBEP: A Fuzzy Analytic Hierarchy Process Framework in Text Classification”**, *Majlesi Journal of Electrical Engineering*, Vol.14, No. 3, pp. 111-123, 2020.

corpus. Hybrid approaches combined the benefits of both approaches to progress the performance of SC.

These contributions of this paper are indicated as follows:

- Proposing a hybrid framework
- Appling three feature selection methods, Backward Feature Selection (BFS), High Correlation Filter (HCF), and Low Variance Filter (LVF) with the perceptron observer for improving classification efficiency
- Providing the Analytic Hierarchy Process (AHP) selector based on the evaluation of error and decision indices for choosing the best feature extraction method
- Reducing input data dimensions and appropriate cover multi-conditional decision challenge on three datasets

The remaining of this article are organized as follows. Sec. 2 and Sec. 3 provide background and related works, respectively. The proposed model is introduced in Sec. 4. Experimental results are exposed in Sec. 5. Eventually, the article ends with a conclusion in Sec. 6.

2. BACKGROUND

Here, a taxonomy of SC approaches is of concern (Fig. 1). The ML approaches applied supervised, unsupervised, and semi-supervised methods and employed linguistic features. The lexicon-based approaches are classified into corpus-based and dictionary-based approaches. The main advantage of them is to support determining domain and context-specific opinion words utilizing a domain corpus. In lexicon-based approaches, a document is divided by aggregating the sentiment orientation of all available words. A document with more positive words is labeled as positive; whereas, a document with more negative words is labeled as negative. Hybrid approaches combined the advantages of both approaches to improve the performance of SC.

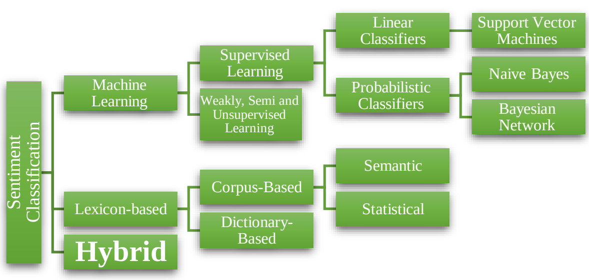

Fig. 1. Sentiment Classification approaches.

In 2015, we compared the validity of supervised and unsupervised approaches [2]. High accuracy from supervised ML algorithms and producing resistance for lexicon-based approaches were obtained. Here, explanations of some classifiers are of concern:

- Naïve Bayes (NB)

NB obtained reasonable accuracy. It is simple and is assumed as independent features. Also, it mainly is used when the size of the training set is not vast. Here, (1) is applied [3] to calculate the probability of event A in column A, provided that class C holds:

$$P(K=A\,|\,C)=\frac{1}{\sqrt{2\pi\sigma^2_{K=C}}}e^{-\frac{A-\mu_{K=C}}{2\sigma^2_{K=C}}} \tag{1}$$

Where, $\mu k=c$ is the column K mean, while the row belongs to the class C and $\sigma^2_{k=c}$ is the variance of the kth therein, and no input classification is required. An example is presented to explain the Bayes Continuous Decider, where, there exist four features with positive or negative classes.

- Maximum Entropy (ME)

Unlike NB, ME is assumed as dependent features [1]. This technique estimates $P(c\,|\,d)$ in (2):

$$P_{ME}(c\,|\,d)=\frac{1}{Z(d)}e^{\sum_i \lambda_{i,c}F_{i,c}(d,c)} \tag{2}$$

Where, Z(d) is a normalization function and $F_{i,c}$ is a function for feature F_i and class *c*, as in Eq. (3):

$$F_{i,c}(d,c')=\begin{cases}1 & n_i(d)>0 \;\; and \;\; c'=c\\ 0 & otherwise\end{cases} \tag{3}$$

- Neural Network (NN)

Here, description of the perceptron classifier is of concern: If *m* is the count of the chosen features and the dataset is named P, each user named P_i would have been assigned to the *m* features, and if any connection attribute *x* is considered, there are variables x_1 to x_m for each connection. These inputs are samples of the training network. It is the training method with a supervisor because the network is trained through samples with the correct output (Fig. 2).

Single-layer perceptron learning:

1-First, the values of the free parameters were considered randomly.

2- First iteration: P_1 is employed to the input of the network and obtained the output of y_1. The error of this step (e_1) is defined as a function of t_1 and y_1, where is named as a function of error. For example, $e_1=t_1-y_1$. If $e_1=0$, it means that the network has performed correctly at this stage. However, if $e_1\neq 0$, it means that the network is not acting properly and it needs to change the

free parameters of the network to reduce e_1. The above step is named the first iteration.
3-First period: Then the second repetition is performed until the r^{th} repetition like the first repetition, etc. The sum of these iterations (from 1 to R) is named a period. In each of the above iterations, it expects the free network parameters to be close to their ideal state. This means that by employing any of the p_k s to this network, $e_k = 0$.
4-Then the second period runs as in the previous period and continue to run the periods until all errors are zero in a complete period. Once reaching this stage, it will no longer change the free parameters.

In most cases, it is impossible and time-consuming to reach such a stage. Hence, a specific criterion for the achievement of the steps is considered, and the error that is less than a certain limit and acceptable ignore. If the error is less than the mentioned limit, it is named a converged network. Sometimes it prefers to define two conditions that, if each of these two conditions is satisfied, the process of repeating the periods will end. The conditions are: the convergence of the network and reaching the number of periods to a certain number [4].
5-Here, define the error function in the nth repetition as $e(n)=t(n)-y(n)$. The effect of the error on the nth repetition is considered to correct the free parameters in the $(n+1)^{th}$ repetition as follows, (4):

$$w_i(n+1)=w_i(n)+\Delta_i(n) \tag{4}$$

Where, $\Delta_i(n)=-\mu e(n)X_i(n)$, and η is a positive coefficient named the learning rate. Thus, if $e(n)=0$, there is no change in the free parameters of the next iteration, but if $e(n)\neq 0$, it tries to change the free parameters in such a way that the error decreases in the next iteration.

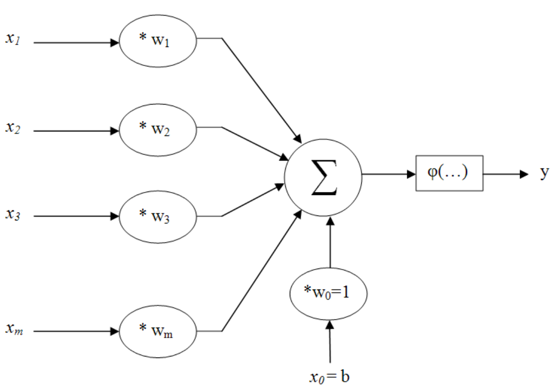

Fig. 2. Single-layer perceptron [4].

- Support Vector Machine (SVM)

In this structure, first the attribute table is converted into a set of data points $\{(x_1,c_1),(x_2,c_2),...,(x_n,c_n)\}$, and then, these are divided into two classes $c_i=\{-1,1\}$. Each x_i is a p-dimensional vector of real numbers, which are the same properties extracted from the previous step.

Linear classification methods aim to classify data though producing a hyperplane, which is a linear equation). The SVM classification determines the best hyperplane that classifies data from two classes with maximum margin. A picture of a data set belonging to two classes, which selects the best hyperplane for separating them is exposed in Fig. 3. In this form, the data is two-dimensional, that is, each data consists of only two variables [5].

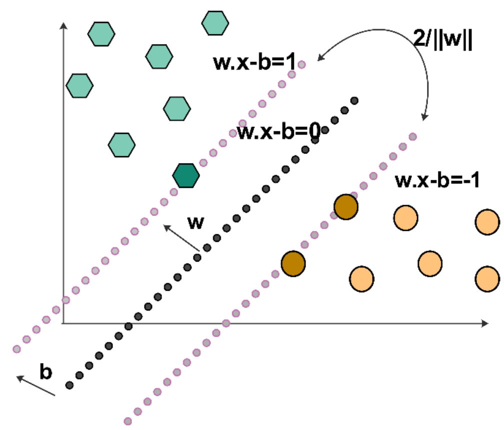

Fig. 3. Hyperplane with maximum separator boundary with separating boundaries for classification.

How the separator hyperplane is produced through the SVM?

This section explains in detail how to produce a separator hyperplane. An accurate picture of how the separator hyperplane is produced through the SVM is exposed in Fig. 4.

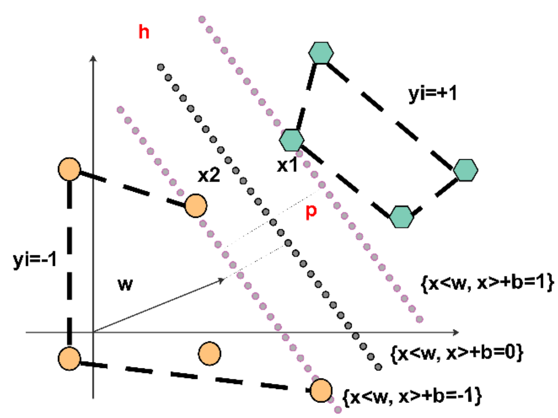

Fig. 4. How to build a separating hyperplane between two data classes in two-dimensional space.

First, consider a convex hull around the points of each class. In Fig. 4, the convex hull is drawn around the points related to class -1 and class +1. Line P is the line that shows the closest distance between two convex hulls. *h*, which is the separating hyperplane, is a line that splits P and is vertical to it. *b* is the width of the source for the hyperplane with the maximum separation limit. If *b* is ignored, the solutions are the only hyperplane that go beyond the source. The vertical distance of the hyperplane to the source is achieved through dividing the absolute value of the parameter *b* by the length *w*.

The basic idea is to choose the proper separator. It refers to the separator that is farthest from the neighboring points on both floors. This answer has the highest boundary with points on two different floors and can be bounded by two parallel hyperplanes that pass through at least one of the floor points. These vectors are named support vectors. The mathematical equations for these two parallel hyperplanes are of concern (5) and (6):

$$w.x - b = 1 \quad (5)$$

$$w.x - b = -1 \quad (6)$$

It is remarkable to remark that if the training data are linearly separable, the two boundary hyperplanes can be chosen in such a way that there is no data between them, and then, the distance between the two parallel hyperplanes can be maximized. Applying geometric theorems, the distance between the two hyperplanes is $|\frac{w}{2}|$, so $|w|$ has to be minimized. It is also necessary to prevent data points from being placed within the boundary, for which a mathematical constraint is added to the formal definition. For each *i*, it is ensured through employing the following constraints that no point is placed on the boundary. For data related to the first and second floors, (7) and (8) are of concern, respectively.

$$w.x_i - b \geq 1 \quad (7)$$

$$w.x_i - b \leq -1 \quad (8)$$

The following constraint can be shown as follows, (9):

$$c_i(w.x_i - b) \leq 1 \qquad 1 \leq i \leq n \quad (9)$$

3. LITERATURE REVIEW

Here, the summary of the major resources is described from 2002 to 2020:

The first article in [6] applied three ML methods consist of NB, SVM, and ME on the Movie Dataset. Next in [7], an opinion mining method based on semantic orientation is proposed with the possibility of being presented in 5 states and then graded.

Authors in [8] classified tweets and messages distribution into positive, negative, and neutral classes. The extracted features included n-gram, lexicon, Part of Speech, and Micro-Blogging. Authors in [9] suggested a remarkable approach for many applications in Sentiment Analysis (SA). Their results indicated that the approach can improve the accuracy of classification. The highest accuracies belong to the Stanford dataset and Twitter-Sanders-Apple (TSA). In 2015, researchers [10] investigated the behavior of NB and SVM classifiers using a different pre-processing scheme. It reveals that the best accuracy is achieved by bigrams.

In 2013, ensemble classifier system was proposed by researchers in [11] on Twitter datasets. Ensemble and boosting were used as the base classifiers. Their results revealed that multiple classifier systems improve the performance of individual classifiers. In 2018, authors [12] focused on Text analysis and applied the Bag of Word to select the best feature by deleting insignificant words. Next, weighing was applied based on their frequency; thus, the words with more weight are selected as the proper features. Following this, the sentences are categorized into positive, negative, or neutral classes. Although this proposed method selects the effective features that may contribute to reducing the data volume and search space, it is possible that not all more frequent words are effective and essential, likewise, the less frequent but important words in the text are not selected as the effective feature through this method.

In 2018, a fuzzy data-based method was proposed in [13] that requires the comparison of the words of each phrase with other default words of the database, according to which the words are weighted. This method is more flexible in determining the subsequent polarity implicated in the words and sentences because it is applied in fuzzy domains. Instead of categorizing the sentences into positive or negative classes, the authors could be classified into more extended domains.

In 2018, the present researchers in [14] proposed a model named SFT for Twitter SC in 2018. The goal of our model was to investigate the role of weighting feature techniques in SC using supervised methods on the Twitter data set. The applied classifier in the current article is based on the SFT model in our previous article. In 2019, a twitter SA is run on the data concerning 2016 and 2017 in [15], where the Arabic language was subject to study with classification on a two-point and a five-point ordinal scale. The same procedure in the SA method was adopted for English in 2016 and on Arabic in 2017. Their results showed that NNs are efficient methods to be applied in analyzing and classifying sentiments. The SVM and Feature selection methods when combined, provide an appropriate method for sentiment and opinions analysis. Hence, the NN methods are applied in the current study.

The major resources employed in the current paper are listed in Table 1.

Table 1. The comparison of the available findings for Sentiment Classification.

Ref.	Applied dataset	Method	Findings
[8]	Twitter	Lexicon	A=75
[12]	Movie Review	Bag-of-n-grams, bag-of-audio-words, bag-of-visual-words	A=68.6
[13]	A semantic network of	OntoSenticNet	-
[15]	Twitter	SVM	A=70
[9]	Sanders Stanford Obama-McCain Debate Health Care Reform	NN, SVM, random forest, linear regression	A=79.1 P=77.4 R=80.8
[11]	Twitter	Ensemble, boosting	A=86.1
[16]	A collection of reviews about hotel	SVM, Fuzzy Domain Ontology	A=82.7 P=74.1 R=60.5
[17]	TripAdvisor, Facebook, and Twitter	Fuzzy Ontology, Bi-directional Long Short-Term Memory	A=84 P=88 R=86
[18]	Twitter	NB, logistic regression	A=84.9
Our	**Movie review and Twitter**	**Hybrid**	**A=90.8 P=90.6 R=91.2 F=91.05**

Note: A=Accuracy, P=Precision, R=Recall, and F=F-measure

4. PROPOSED FRAMEWORK

The structure of the suggested framework is shown in Fig. 5. Our framework has two feature selection and classification units. In the feature selection unit, the operation is performed by three feature selection methods, HCF, BFS, and LVF through the perceptron supervisor. Then, the best feature selection method is chosen based on the different error index through the Fuzzy AHP. Selected indices are provided to the classification unit and this unit is applied to the SVM to classify the user's opinions.

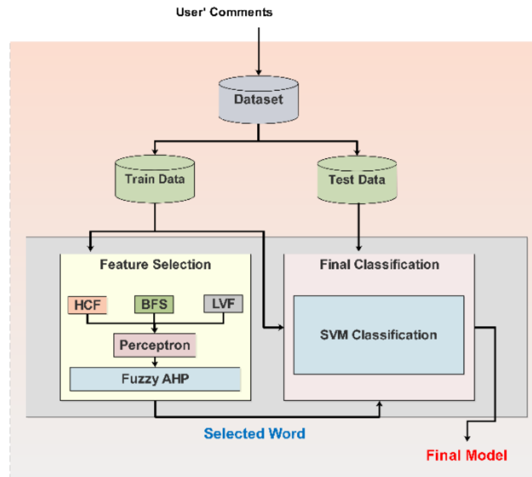

Fig. 5. The structure of the FAHPBEP framework.

4.1. Data Pre-processing

The sentences that convey opinions are decomposed into words for early editions.

A text is included in sentences, some of which contain opinions stated in words. To decompose the text, the opposite process is run.

Stop words like a, about, all, am, did, has, have, etc. are commonly employed in English, and lack contribution in recognizing relevant words.

To identify and record word frequency, they must be stemmed, which provides the word conversion into their simplest alternative. The more accurate this process, the more precise the similar words recognition.

For the goal, the three Term Frequency (TF), Inverse Document Frequency (IDF), and Inverse Class Frequency (ICF) criteria are performed and defined in [19], [20], and [21].

TF: A term's frequent appearance in all the comments (i.e. the total TF (I, d) count is the frequency of term I in Document d) that is defined through TF.

IDF: The word count regularly and ordinarily employed in a text is defined through IDF, calculated through (10).

$$IDF_i = \log \frac{N}{DF_i} \quad (10)$$

Where, N is the total comment count and DF_i is the comment including the word *i* count.

ICF: The count of regularly and ordinarily employed words in a class is defined through ICT, calculated through (11):

$$ICF_i = \log \frac{N}{CF_i} \quad (11)$$

Where, C is the class count, and CF_i is the class count including the word *i* count.

4.2. The FAHPBEP Framework

- Important features selection

One of the most important and fundamental stages that leads to increasing the speed and accuracy of the final model is the extraction of important features. In this stage, first, the data is extracted and then, feature selection is performed through three methods, BFS, HCF, and LVF. Then, the error rate of the initial classification for each method is obtained through perceptron and the best method is selected using fuzzy AHP decision-making.

- Perceptron classifier

As mentioned in Sec. 2, the perceptron classifier was applied in the present study.

- AHP decider

In the fuzzy AHP process, a problem with several indices is divided into a hierarchy of levels. In this process, the decision indices are Normal Mean Square Error (NMSE), Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE), Mean Absolute Square Error (MASE), and Root Mean Square Error (RMSE). The structure of the fuzzy hierarchy model for choosing the best feature selection method is shown in Fig. 6, where it is observed that a high level is the principal goal of the decision-making process. The second level is assigned as the fundamental and main indices, where it may be divided into minor and secondary indices in the next level. The last level is asserted decision alternatives. The goal of this stage is choosing the best feature selection method, where it is placed in the first level. To reach this goal, five important error indices are applied, where these are positioned in the second level of the model. Also, three feature selection methods are applied as the decision alternatives in the third level.

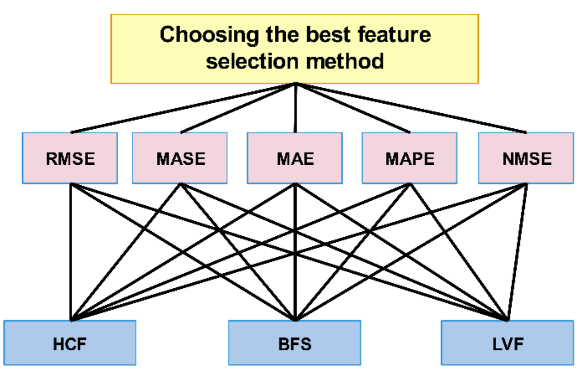

Fig. 6. The structure of the fuzzy hierarchy model for choosing the best feature selection method.

Paired comparisons: Here, several comparisons among different decision alternatives and based on each index and its importance are enacted. After the design of the problem hierarchy, the sets of the matrix are produced through the importance of the indices. To reach the aim, comparisons between the decision elements, that are paired comparisons are enacted. Integer scales as the importance of preferential between the two decision elements are devoted to comparisons. This stage id is generated through comparisons among alternatives and the i^{th} indices concerning alternatives and the j^{th} indices. The valuation of the indices is exposed in Table 2.

The second level of the hierarchy is the main criteria. Each of the main criteria is determined based on the priority goal. The criteria should be compared based on the goal pairwise. According to the valuation of the indices, their structure of priority concerning each other is specified. The table structure of the indices evaluation is as the following:

The table structure is square, where there are indices in each row and column. An intersection of each row and column of this table, a fuzzy value like x would be in term of Table 2, where the x^{-1} value is placed in the intersection of each row and column of this index. The diameter of this table that specifies the priority of each element concerning itself has a value of 1. The priority of the criteria concerning each other and based on target is specified in Table 3.

Table 2. The preferential valuation of the indices concerning each other.

Membership function	Domain	Triangular fuzzy scale	The comparison status of *i* concerning *j*	Fuzzy integer
1	1	(1,1,1)	Exactly identical	1
$\frac{3-x}{3-1}$	$1 \le x \le 3$	(1,1,3)	Equal importance	$\tilde{1}$
$\frac{x-1}{3-1}$	$1 \le x \le 3$	(1,3,5)	Relative more importance	$\tilde{3}$
$\frac{5-x}{5-3}$	$3 \le x \le 5$			
$\frac{x-3}{5-3}$	$3 \le x \le 5$	(3,5,7)	More importance	$\tilde{5}$
$\frac{7-x}{7-5}$	$5 \le x \le 7$			
$\frac{x-5}{7-5}$	$5 \le x \le 7$	(5,7,9)	Very more importance	$\tilde{7}$
$\frac{9-x}{9-7}$	$7 \le x \le 9$			
$\frac{x-7}{9-7}$	$7 \le x \le 9$	(7,9,9)	Absolute importance	$\tilde{7}$

Table 3. The preferential valuation of the indices.

NMSE	MAPE	MAE	MASE	RMSE	
$\tilde{2}$	$\tilde{3}$	$\tilde{5}$	$\tilde{3}$	1	**RMSE**
$\tilde{2}^{-1}$	$\tilde{1}$	$\tilde{2}$	1	$\tilde{3}^{-1}$	**MASE**
$\tilde{3}^{-1}$	$\tilde{2}^{-1}$	1	$\tilde{2}^{-1}$	$\tilde{5}^{-1}$	**MAE**
$\tilde{2}^{-1}$	1	$\tilde{2}$	$\tilde{1}$	$\tilde{3}^{-1}$	**MAPE**
1	$\tilde{2}$	$\tilde{3}$	$\tilde{2}$	$\tilde{3}^{-1}$	**NMSE**

The paired scale matrix of the criteria concerning the 1^{st}, 2^{nd}, 3^{rd}, 4^{th}, and 5^{th} indices would be of concern in Table 4 to Table 8 in the same manner, respectively.

Table 4. The paired scale of the criteria concerning the RMSE index.

BFS	LVF	HCF	RMSE
$\tilde{7}$	$\tilde{2}$	1	**HCF**
$\tilde{5}$	1	$\tilde{2}^{-1}$	**LVF**
1	$\tilde{5}^{-1}$	$\tilde{7}^{-1}$	**BFS**

Table 5. The paired scale of the criteria concerning the MASE index.

BFS	LVF	HCF	MASE
$\tilde{1}$	$\tilde{2}^{-1}$	1	**HCF**
$\tilde{1}$	1	$\tilde{2}$	**LVF**
1	$\tilde{1}$	$\tilde{1}$	**BFS**

Table 6. The paired scale of the criteria concerning the MAE index.

BFS	LVF	HCF	MAE
$\tilde{2}^{-1}$	$\tilde{3}^{-1}$	1	**HCF**
$\tilde{2}$	1	$\tilde{3}$	**LVF**
1	$\tilde{2}^{-1}$	$\tilde{2}$	**BFS**

Table 7. The paired scale of the criteria concerning the MAPE index.

BFS	LVF	HCF	MAPE
$\tilde{2}^{-1}$	$\tilde{2}^{-1}$	1	**HCF**
$\tilde{1}$	1	$\tilde{2}$	**LVF**
1	$\tilde{1}$	$\tilde{2}$	**BFS**

Table 8. The paired scale of the criteria concerning the NMSE index.

BFS	LVF	HCF	NMSE
$\tilde{3}$	$\tilde{2}$	1	**HCF**
$\tilde{2}$	1	$\tilde{2}^{-1}$	**LVF**
1	$\tilde{2}^{-1}$	$\tilde{3}^{-1}$	**BFS**

According to the obtained fuzzy structure, these would be converted to Table 9 to Table 14.

Table 9. The fuzzy valuation of the criteria.

NMSE	MAPE	MAE	MASE	RMSE	
(1,2,4)	(1,3,5)	(3,5,7)	(1,3,5)	(1,1,1)	**RMSE**
$(1,\frac{1}{2},\frac{1}{4})$	(1,1,3)	(1,2,4)	(1,1,1)	$(1,\frac{1}{3},\frac{1}{5})$	**MASE**
$(1,\frac{1}{3},\frac{1}{5})$	$(1,\frac{1}{2},\frac{1}{4})$	(1,1,1)	$(1,\frac{1}{2},\frac{1}{4})$	$(\frac{1}{3},\frac{1}{5},\frac{1}{7})$	**MAE**
$(1,\frac{1}{2},\frac{1}{4})$	(1,1,1)	(1,2,4)	(1,1,3)	$(1,\frac{1}{3},\frac{1}{5})$	**MAPE**
(1,1,1)	(1,2,4)	(1,3,5)	(1,2,4)	$(1,\frac{1}{3},\frac{1}{5})$	**NMSE**

Table 10. The fuzzy paired scale matrix of the criteria concerning the RMSE index.

BFS	LVF	HCF	RMSE
(5,7,9)	(1,2,4)	(1,1,1)	**HCF**
(3,5,7)	(1,1,1)	$(1,\frac{1}{2},\frac{1}{4})$	**LVF**
(1,1,1)	$(\frac{1}{3},\frac{1}{5},\frac{1}{7})$	$(\frac{1}{5},\frac{1}{7},\frac{1}{9})$	**BFS**

Table 11. The fuzzy paired scale matrix of the criteria concerning the MASE index.

BFS	LVF	HCF	MASE
(1,1,3)	$(1,\frac{1}{2},\frac{1}{4})$	(1,1,1)	**HCF**
(1,1,3)	(1,1,1)	(1,2,4)	**LVF**
(1,1,1)	(1,1,3)	(1,1,3)	**BFS**

Table 12. The fuzzy paired scale matrix of the criteria concerning the MAE index.

BFS	LVF	HCF	MAE
$(1,\frac{1}{2},\frac{1}{4})$	$(1,\frac{1}{3},\frac{1}{5})$	(1,1,1)	**HCF**
(1,2,4)	(1,1,1)	(1,3,5)	**LVF**
(1,1,1)	$(1,\frac{1}{2},\frac{1}{4})$	(1,2,4)	**BFS**

Table 13. The fuzzy paired scale matrix of the criteria concerning the MAPE index.

BFS	LVF	HCF	MAPE
$(1,\frac{1}{2},\frac{1}{4})$	$(1,\frac{1}{2},\frac{1}{4})$	(1,1,1)	**HCF**
(1,1,3)	(1,1,1)	(1,2,4)	**LVF**
(1,1,1)	(1,1,3)	(1,2,4)	**BFS**

Table 14. The fuzzy paired scale matrix of the criteria concerning the NMSE index.

BFS	LVF	HCF	NMSE
(1,3,5)	(1,2,4)	(1,1,1)	**HCF**
(1,2,4)	(1,1,1)	$(1,\frac{1}{2},\frac{1}{4})$	**LVF**
(1,1,1)	$(1,\frac{1}{2},\frac{1}{4})$	$(1,\frac{1}{3},\frac{1}{5})$	**BFS**

Determining the weight of the indices concerning each other: The weight vectors are induced through calculations. The formation of the fuzzy valuation matrix is as the following:

At first, some of M and M^{-1} values are calculated and the valuation table of the indices is tabulated in Table 15.

Table 15. The valuation table of the indices.

$\sum M$	**NMSE**	**MAPE**	**MAE**	**MASE**	**RMSE**	
(7,19,22)	(1,2,4)	(1,3,5)	(3,5,7)	(1,3,5)	(1,1,1)	**RMSE**
(5,4.8,8.7)	$(1,\frac{1}{2},\frac{1}{4})$	(1,1,3)	(1,2,4)	(1,1,1)	$(1,\frac{1}{3},\frac{1}{5})$	**MASE**
(4.3,2.5,1.8	$(1,\frac{1}{3},\frac{1}{5})$	$(1,\frac{1}{2},\frac{1}{4})$	(1,1,1)	$(1,\frac{1}{2},\frac{1}{4})$	$(\frac{1}{3},\frac{1}{5},\frac{1}{7})$	**MAE**
(5,4.8,8.5)	$(1,\frac{1}{2},\frac{1}{4})$	(1,1,1)	(1,2,4)	(1,1,3)	$(1,\frac{1}{3},\frac{1}{5})$	**MAPE**
(5,8.3,14.2)	(1,1,1)	(1,2,4)	(1,3,5)	(1,2,4)	$(1,\frac{1}{3},\frac{1}{5})$	**NMSE**

Here, the sum of M values is of concern, (12):

$$\sum_{i=1}^{n}\sum_{j=1}^{m} M_{i,j} = (26.3, 39.4, 54.2) \tag{12}$$

The inverse of the sum of M values would be equal the (13):

$$\left[\sum_{i=1}^{n}\sum_{j=1}^{m} M_{i,j}\right]^{-1} = (0.018, 0.025, 0.038) \tag{13}$$

To calculating the S set, the following procedures should be of concern:

$$S_1 = (7,19,22)\times(0.018,0.025,0.038) = (0.126,0.475,0.836)$$
$$S_2 = (5,4.8,8.7)\times(0.018,0.025,0.038) = (0.09,0.12,0.330)$$
$$S_3 = (4.3,2.5,1.8)\times(0.018,0.025,0.038) = (0.077,0.062,0.065)$$
$$S_4 = (5,4.8,8.5)\times(0.018,0.025,0.038) = (0.09,0.12,0.323)$$
$$S_5 = (5,8.3,14.2)\times(0.018,0.025,0.038) = (0.09,0.20,0.539)$$

Following the previous stage, the greatest degree of the set S concerning each other is calculated as the following:

$$V(M_2 \geq M_1) = \begin{cases} 1 & m_2 \geq m_1 \\ 0 & l_1 \geq u_2 \\ \dfrac{l_1 - u_2}{(m_2 - u_2) - (m_1 - l_1)} & otherwise \end{cases}$$

$$V(S_1 \geq S_2) = 1,\ V(S_1 \geq S_3) = 1,\ V(S_1 \geq S_4) = 1,\ V(S_1 \geq S_5) = 1$$
$$V(S_2 \geq S_1) = \frac{l_1 - u_2}{(m_2 - u_2) - (m_1 - l_1)} = 0.36$$
$$V(S_2 \geq S_3) = 1,\ V(S_2 \geq S_4) = 1,\ V(S_2 \geq S_5) = 0.75$$
$$V(S_3 \geq S_1) = 0,\ V(S_3 \geq S_2) = 0.85,\ V(S_3 \geq S_4) = 0.91,\ V(S_3 \geq S_5) = 0.26$$
$$V(S_4 \geq S_1) = 0.36,\ V(S_4 \geq S_2) = 1,\ V(S_4 \geq S_3) = 1,\ V(S_4 \geq S_5) = 0.14$$
$$V(S_5 \geq S_1) = 0.59,\ V(S_5 \geq S_2) = 1,\ V(S_5 \geq S_3) = 1,\ V(S_5 \geq S_4) = 1$$

And last, the weight of each criterion concerning each other is calculated. The value of the calculated weight and the normal weight of each index is specified in Table 16.

Table ۱6. The weight of the indices concerning each other.

Normalized weight	**Non-normalized weight**	**Criteria**
0.74	1	**RMSE**
0.18	0.36	**MASE**
0.12	0.26	**MAE**
0.06	0.14	**MAPE**
0.34	0.59	**NMSE**

For all elements of the paired scale matrix, the weight of the criteria concerning the indices is calculated and the weight matrix is produced for each element in the likewise manner. The weight tables of the indices are exposed in Table 17 to Table 21.

Table 17. The weight of the indices concerning the RMSE index.

Normalized weight	**Non-normalized weight**	**RMSE**
0.53	1	**HCF**
0.47	0.87	**LVF**
0	0	**BFS**

Table 18. The weight of the indices concerning the MASE index.

Normalized weight	**Non-normalized weight**	**MASE**
0.17	0.27	**HCF**
0.50	0.78	**LVF**
0.33	0.52	**BFS**

Table 19. The weight of the indices concerning the MAE index.

Normalized weight	**Non-normalized weight**	**MAE**
0.12	0.15	**HCF**
0.51	0.67	**LVF**
0.37	0.48	**BFS**

Table 20. The weight of the indices concerning the MAPE index.

Normalized weight	**Non-normalized weight**	**MAPE**
0.23	0.36	**HCF**
0.45	0.71	**LVF**
0.32	0.49	**BFS**

Table 21. The weight of the indices concerning the NMSE index.

Normalized weight	Non-normalized weight	NMSE
0.50	0.76	**HCF**
0.41	0.63	**LVF**
0.08	0.12	**BFS**

Following these calculations, with regarding the indices and their weights, the score of each criterion can be calculated. The values of 5 indices are of concern, Table 22.

Table 22. The values of the indices.

NMSE	MAPE	MAE	MASE	RMSE	Value
0.50	0.23	0.12	0.17	0.53	**HCF**
0.41	0.45	0.51	0.50	0.47	**LVF**
0.08	0.32	0.37	0.33	0	**BFS**

As observed in Table 22, the importance coefficients and the score of each index are calculated in Table 23.

Table 23. The valuation table of the indices.

Score	NMSE	MAPE	MAE	MASE	RMSE	
	0.34	**0.06**	**0.12**	**0.18**	**0.74**	**Weight**
0.61	0.17	0.01	0.01	0.03	0.39	**HCF**
0.67	0.14	0.03	0.06	0.09	0.35	**LVF**
0.51	0.03	0.38	0.04	0.06	0	**BFS**

Thus, BFS is selected as a better predictor because it has of lower error score in this example.

- Classification through SVM

As mentioned in Sec. 2, the SVM classifier was applied for the final classification in the present study.

4.3. Dataset Description

The following three datasets are employed in evaluating the proposed framework performance:

1-The Cornell movie review [22], including the Polarity Movie Dataset (PMD) of 1000 positive and negative reviews.

2-This dataset is a subset of Sanders Analytics (TSA2) and consists of 479 tweets.

3-This dataset is also a subset of Sanders Analytics (TSA3) and contains 988 tweets [23].

5. RESULTS AND DISCUSSION

MATLAB S/W is applied to assess the efficiency of the suggested framework. The K-fold technique with 10 classes is employed to improve the accuracy of evaluation [24]. Windows 10 is the test environment.

5.1. The Compared Methods

The compared methods are summarized as follows:

HybBSRL: In the hybrid classification method, several classifiers such as NN, SVM, random forest, and linear regression were employed. Feature selection was performed and the error rate of each classifier was calculated and the best output was selected as a final model [9].

BECM: In the ensemble classification methods, a system based on classifiers was proposed. The applied classifiers used boosting. After the feature selection, the error rate of the classifiers was calculated and the final model was extracted [11].

5.2. Evaluation Indices

Correctness and Error are the indices here [24] [25]. The accuracy, precision, recall, f-measure, Mean of Square Error (MSE), and Mean of Prediction Error Deviation error indices are introduced, (14) to (19):

$$Accuracy = \frac{T_p + T_n}{T_p + T_n + F_p + F_n} \tag{14}$$

$$Precision = \frac{T_p}{T_p + F_p} \tag{15}$$

$$Recall = \frac{T_p}{T_p + F_n} \tag{16}$$

$$F-measure = \frac{2.\,precision.recall}{precision + recall} \tag{17}$$

$$MSE = \frac{\sum_{i=1}^{n}(y_i - \overline{y_i})^2}{n} \tag{18}$$

$$SSE = \sum_{i=1}^{n}(y_i - \overline{y_i})^2 \tag{19}$$

5.3. Evaluation of Our Framework

The achieved results are revealed as follows:

- The First dataset

Its precision value is corresponded with HybBSRL and BECM in 10-fold cross-validation exposed (Fig. 7), where its higher precision is evident. Also, the comparison of its accuracy in 10-fold cross-validation is exposed (Fig. 8), where its higher accuracy is evident. The best accuracy rate obtained using HybBSRL is 89.37% and using BECM is 87.86%; whereas, our framework obtained an accuracy of 91.95%. It appears that accuracy is the most outstanding index in each method. According to the precision and accuracy, the recall of FAHPBEP with HybBSRL and BECM methods in 10-fold cross-validation the have same holds (Fig. 9), and it is evident that this framework provides higher and more precise feature extraction. The overall f-measure rate obtained using HybBSRL is 89% and using BECM is 86.88%; whereas, our framework obtained f-measure rate of 91.05% (Table 24). The details of this comparison are to error indices of FAHPBEP framework, HybBSRL, BECM methods in 10-fold cross-validation tabulated in Table 25.

- The second dataset

Its precision value in 10-fold cross-validation is shown in Fig. 10, where it is revealed that this framework is a more precise method than the other two. The comparison of its accuracy in 10-fold cross-validation is shown (Fig. 11) and its higher accuracy is evident. We achieved an accuracy rate of 91.93%; however, the best accuracy rate obtained using HybBSRL and BECM is 90.63% and 88.19%, respectively. According to the precision and accuracy, the recall of FAHPBEP with HybBSRL and BECM methods in 10-fold cross-validation have the same holds (Fig. 12), and it is evident that this framework provides higher and more precise feature extraction. The overall f-measure rate obtained using HybBSRL is 88.73% and using BECM is 86.89%; whereas, our framework obtained f-measure rate of 90.88% (Table 26). The details of this comparison are to error indices of FAHPBEP framework, HybBSRL, BECM methods in 10-fold cross-validation tabulated in Table 27.

- The third dataset

Its precision value compared with HybBSRL and BECM in 10-fold cross-validation is shown in Fig. 13, where it revealed that this framework is a more precise method than the other two. The comparison of its accuracy in 10-fold cross-validation is shown in Fig. 14, where its higher accuracy is evident. We achieved an accuracy rate of 90.75%, while the best accuracy rate obtained using HybBSRL and BECM is 88.92% and 87.48%, respectively. According to the precision and accuracy, recall methods in 10-fold cross-validation have the same holds (Fig. 15) and, it is evident that this framework yields higher and more precise feature extraction. By analyzing the mentioned three methods, the outperformance of this framework is evident. The overall f-measure rate obtained of HybBSRL is 87.91% and using BECM is 85.87%; whereas, our framework obtained f-measure rate of 90.01% (Table 28). The details of this comparison as to error indices of FAHPBEP framework, HybBSRL, BECM methods in 10-fold cross-validation are tabulated in Table 29.

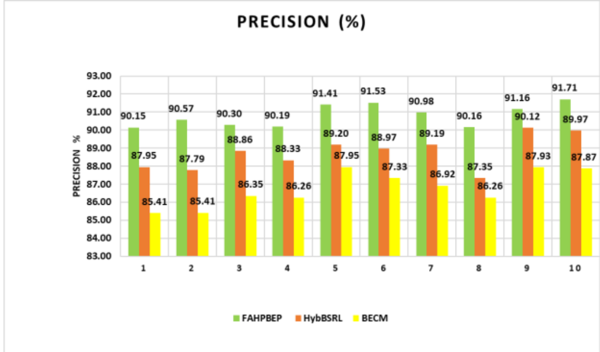

Fig. 7. The precision value of the FAHPBEP compared to the two methods on the PMD dataset.

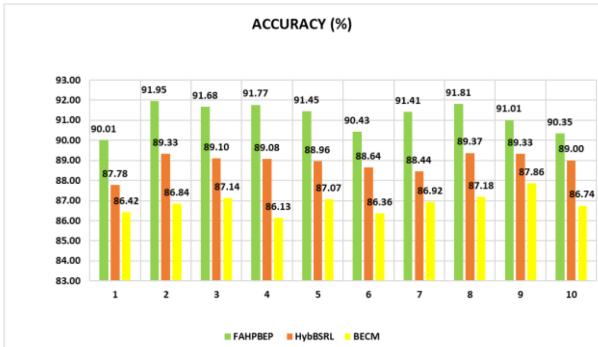

Fig. 8. The accuracy value of the FAHPBEP compared to the two methods on the PMD dataset.

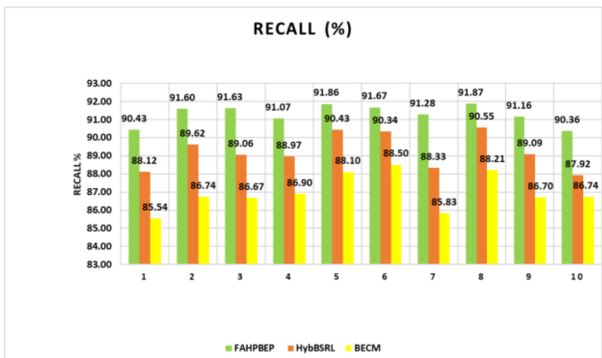

Fig. 9. The recall value of the FAHPBEP compared to the two methods on the PMD dataset.

Table 24. The overall f-measure of our framework compared to the two methods.

Overall f-measure	Methods
91.05	FAHPBEP
89.00	HybBSRL
86.88	BECM

Table 25. Error Indices comparison.

SSE	MSE	Forecasting methods
391	0.0987	FAHPBEP
432	0.1058	HybBSRL
475	0.1143	BECM

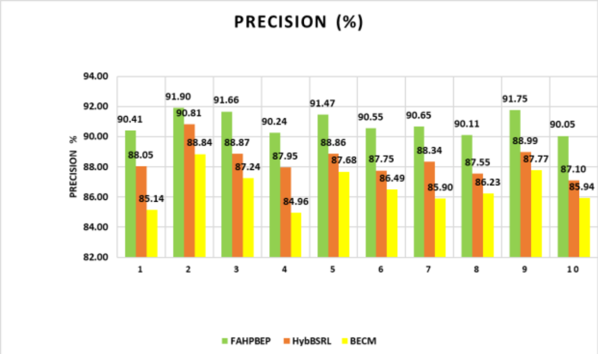

Fig. 10. The precision value of the FAHPBEP compared to the two methods on the TSA2 dataset.

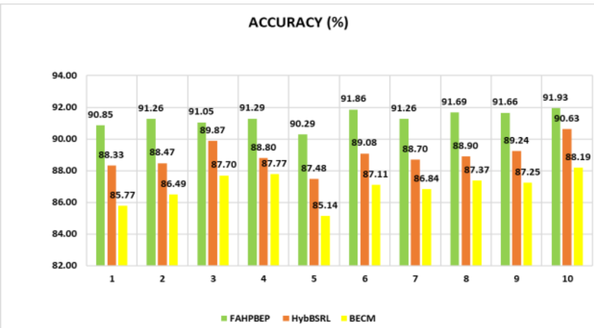

Fig. 11. The accuracy value of the FAHPBEP compared to the two methods on the TSA2 dataset.

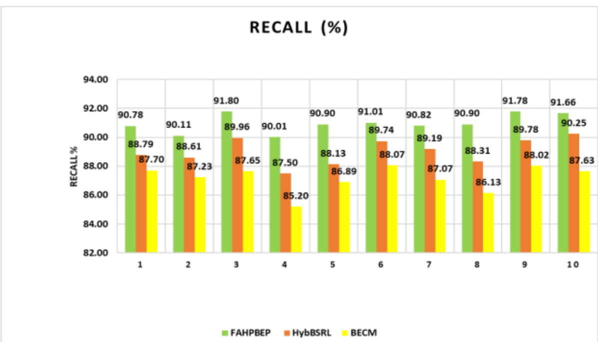

Fig. 12. The recall value of the FAHPBEP compared to the two methods on the TSA2 dataset.

Table 26. The overall f-measure of our framework compared to the two methods.

Overall f-measure	Methods
90.88	FAHPBEP
88.73	HybBSRL
86.89	BECM

Table 27. Error Indices comparison.

SSE	MSE	Forecasting methods
382	0.0991	FAHPBEP
419	0.1003	HybBSRL
462	0.1115	BECM

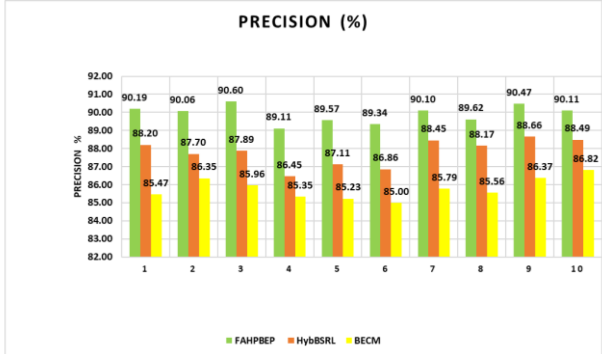

Fig. 13. The precision value of the FAHPBEP compared to the two methods on the TSA3 dataset.

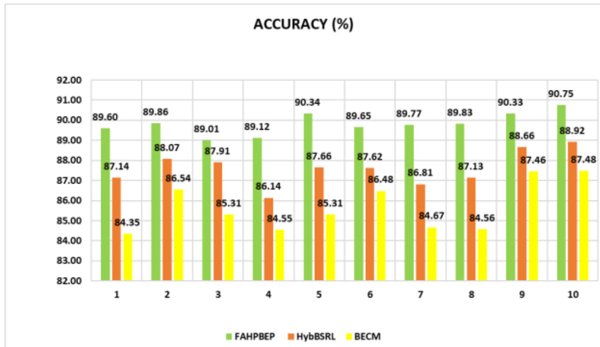

Fig. 14. The accuracy value of the FAHPBEP compared to the two methods on the TSA3 dataset.

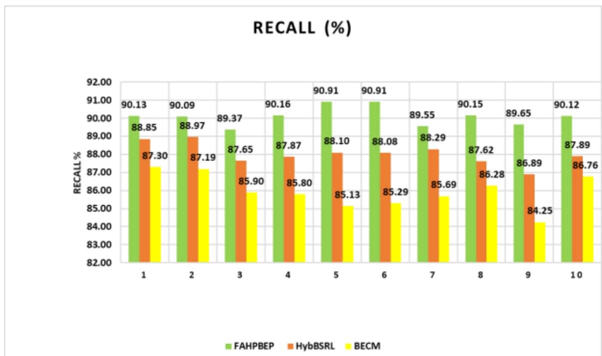

Fig. 15. The recall value of the FAHPBEP compared to the two methods on the TSA3 dataset.

Table 28. The overall f-measure of our framework compared to the two methods.

Overall f-measure	Methods
90.01	FAHPBEP
87.91	HybBSRL
85.78	BECM

Table 29. Error Indices comparison.

SSE	MSE	Forecasting methods
402	0.0998	FAHPBEP
468	0.1103	HybBSRL
493	0.1181	BECM

6. CONCLUSION

With the rapid growth of web sites and the Internet, people can express and share their reviews and comments simply. It prompted a large number of comments about products, services, etc. It affected the brand of companies significantly. Negative and positive reviews are published fast by utilizing social platforms like Twitter. Companies need to investigate their big data and lead the strategies based on the revealed findings. SC is increasingly becoming a vital factor in this field that companies can be focused on. Hence, important information can be extracted from these reviews. The current authors believe that this paper is appropriate to other social media analyses. This study suggests a hybrid framework, FAHPBEP. Here, the features of the user's opinions are extracted based on three methods: BFS, HCF, and LVF. The error rate of the primary classification for each method is calculated through the NN. Finally, the best method is selected through fuzzy AHP. For evaluation, the three PMD, TSA2, and TSA3 datasets are applied. This FAHPBEP framework as to its accuracy, precision, recall, and error indices is compared with HybBSRL and BECM. The obtained results indicate that on average this proposed framework outperformed its counterparts.

In this context, considering ensemble methods can be suggested as future work.

7. ACKNOWLEDGMENT

We thank the editor and all anonymous reviewers for their constructive comments.

REFERENCES

[1] Alsaeedi and M. Z. Khan, **"A Study on Sentiment Analysis Techniques of Twitter Data,"** *International Journal of Advanced Computer Science and Applications*, Vol. 10, No. 2, pp. 361-374, 2019.

[2] R. Asgarnezhad and K. Mohebbi, **"A comparative classification of approaches and applications in opinion mining,"** *International Academic Journal of Science and Engineering*, Vol. 2, No. 1, pp. 68.80, 2015.

[3] S. R. Ahmad, M. Z. M. Rodzi, N. S. S. Nurhafizeh, M. M. Yusop, and S. Ismail, **"A Review of Feature Selection and Sentiment Analysis Technique in Issues of Propaganda,"** *International Journal of Advanced Computer Science and Applications*, Vol. 10, No. 11, pp. 240-245, 2019.

[4] J. J. Shynk, **"Performance surfaces of a single-layer perceptron,"** *IEEE Transactions on Neural Networks*, Vol. 1, No. 3, pp. 268-274, 1990.

[5] D. Michie, D. J. Spiegelhalter, and C. Taylor, **"Machine learning,"** *Neural and Statistical Classification*, Vol. 13, pp.1-298, 1994.

[6] Pang, L. Lee, and S. Vaithyanathan, **"Thumbs up? sentiment classification using machine learning techniques,"** in *Proc. of the ACL-02 conference on Empirical methods in natural language processing*, Vol. 10, 2002, pp. 79-86.

[7] P. Chaovalit and L. Zhou. **"Movie review mining: A comparison between supervised and unsupervised classification approaches,"** *in Proc. of the 38th annual Hawaii international conf. on system sciences*, IEEE, 2005, pp.1-9.

[8] E. Kouloumpis, T. Wilson, and J. Moore, **"Twitter sentiment analysis: The good the bad and the omg!,"** *in Fifth International AAAI conf. on weblogs and social media*, 2011, pp.538-341.

[9] N. F. Da Silva, E. R. Hruschka, and E. R. Hruschka Jr., **"Tweet sentiment analysis with classifier ensembles,"** *Decision Support Systems*, Vol. 66, pp. 170-179, 2014.

[10] G. Tripathi and S. Naganna, **"Feature selection and classification approach for sentiment analysis,"** *Machine Learning and Applications: An International Journal*, Vol. 2, pp. 1-16, 2015.

[11] A. Hassan, A. Abbasi, and D. Zeng, **"Twitter sentiment analysis: A bootstrap ensemble framework,"** *in International Conf. on Social Computing*, 2013, pp. 357-364.

[12] N. Cummins, S. Amiriparian, S. Ottl, M. Gerczuk, M. Schmitt, and B. Schuller, **"Multimodal Bag-of-Words for cross domains sentiment analysis,"** *in IEEE International Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2018, pp. 4954-4958.

[13] M. Dragoni, S. Poria, and E. Cambria, **"OntoSenticNet: A commonsense ontology for sentiment analysis,"** *IEEE Intelligent Systems*, Vol. 33, No. 3, pp. 77-85, 2018.

[14] R. Asgarnezhad, S. A. Monadjemi, M. Soltanaghaei, and A. Bagheri, **"SFT: A model for sentiment classification using supervised methods in Twitter,"** *Journal of Theoretical & Applied Information Technology*, Vol. 96, No. 8, pp. 2242-2251, 2018.

[15] S. Rosenthal, N. Farra, P. Nakov, **"SemEval-2017 task 4: Sentiment analysis in Twitter,"** *in 11th international workshop on semantic evaluation (SemEval-2017)*, 2019, pp. 502–518.

[16] F. Ali, K. S. Kwak, and Y. G. Kim, "Opinion mining based on fuzzy domain ontology and Support Vector Machine: A proposal to automate online review classification," *Applied Soft Computing*, Vol. 47, pp. 235-250, 2016.

[17] F. Ali, S. El-Sappagh, and D. Kwak, "Fuzzy ontology and LSTM-based text mining: a transportation network monitoring system for assisting travel," *Sensors*, Vol. 19, No. 2, pp. 1-23, 2019.

[18] M. B. Alvi, N. M. Mahoto, M. A. Unar, and M. A. Shaikh, **"An Effective Framework for Tweet Level Sentiment Classification using Recursive Text Pre-Processing Approach,"** *International Journal of Advanced Computer Science and Applications*, Vol. 10, No. 6, pp. 572-581, 2019.

[19] D. Manning, P. Raghavan, and H. Schütze, **"Introduction to information retrieval,"** Cambridge university press, 2008.

[20] Q. Nguyen, D. Q. Nguyen, T. Vu, and S. B. Pham, **"Sentiment classification on polarity reviews: an empirical study using rating-based features,"** *in 5th*

Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, 2014, pp. 128–135.

[21] A. Nawaz, S. Asghar, and S. H. A. Naqvi, **“A segregational approach for determining aspect sentiments in social media analysis,”** *The Journal of Supercomputing*, Vol. 75, No. 5, pp. 2584-2602, 2019.

[22] Movie review data set, Available: http://www.cs.cornell.edu/people/pabo/movie-review-data/.

[23] Dataset available at http://www.sananalytics.com

[24] J. Han and M. Kamber, **“Data mining: concepts and techniques,”** Morgan Kaufmann Publishers–An Imprint of Elsevier, Vol. 500, pp. 105-150, 2006.

[25] B. McDonald, D. Waugaman, and C. Kettleborough, **“A statistical analysis of a packed tower dehumidifier,”** *Drying Technology*, Vol. 10, No. 1, pp. 223-237, 1992.