An Effective Method of Feature Selection in Persian Text for Improving the Accuracy of Detecting Request in Persian Messages on Telegram

Zahra Khalifeh Zadeh
Faculty of Computer Engineering, Yazd University, Iran
Zahra.kh2005@gmail.com
Mohammad Ali Zare Chahooki*
Faculty of Computer Engineering, Yazd University, Iran
chahooki@yazd.ac.ir

Received: 01/Aug/2020 Revised: 02/Nov/2020 Accepted: 02/Jan/2021

Abstract

In recent years, data received from social media has increased exponentially. They have become valuable sources of information for many analysts and businesses to expand their business. Automatic document classification is an essential step in extracting knowledge from these sources of information. In automatic text classification, words are assessed as a set of features. Selecting useful features from each text reduces the size of the feature vector and improves classification performance. Many algorithms have been applied for the automatic classification of text. Although all the methods proposed for other languages are applicable and comparable, studies on classification and feature selection in the Persian text have not been sufficiently carried out. The present research is conducted in Persian, and the introduction of a Persian dataset is a part of its innovation. In the present article, an innovative approach is presented to improve the performance of Persian text classification. The authors extracted 85,000 Persian messages from the Idekav-system, which is a Telegram search engine. The new idea presented in this paper to process and classify this textual data is on the basis of the feature vector expansion by adding some selective features using the most extensively used feature selection methods based on Local and Global filters. The new feature vector is then filtered by applying the secondary feature selection. The secondary feature selection phase selects more appropriate features among those added from the first step to enhance the effect of applying wrapper methods on classification performance. In the third step, the combined filter-based methods and the combination of the results of different learning algorithms have been used to achieve higher accuracy. At the end of the three selection stages, a method was proposed that increased accuracy up to 0.945 and reduced training time and calculations in the Persian dataset.

Keywords: Feature Selection*;* Text Mining*;* Classification Accuracy*;* Machine Learning*;* Ensemble Classifier*.*

1- Introduction

Nowadays, the rapid progress and easy access to Internet technologies, multimedia, and social networks have drastically changed and affected human life. In addition to facilitating individual communication, social networks also serve as channels of communication between companies and customers [1], [2]. Social networks have a considerable impact on the potential value of businesses [3]. They are widespread and highly regarded among users. Thus, virtual societies have become valuable sources of political, social, and commercial information. Social networks are utilized in many businesses to provide services and interact effectively with customers. Therefore, the knowledge extracted from social networks such as Facebook, Twitter, Telegram, and other social networks is valuable for marketing and data mining companies [4]-[6].

Telegram is a messaging service with many users from different countries [7]. The number of monthly active users of Telegram in October 2019 is 300 million worldwide [8]. Moreover, 60% of Iranians use Telegram [9], and it has become a popular and extensively used social network in various fields such as the development of certain Internet businesses and contains valuable information. Telegram data possesses hidden knowledge, the extraction of which is extremely useful. The request-type messages that are exchanged among Telegram users are among these data with hidden knowledge. In Telegram, a message can be sent containing a request for help to buy a house or a product, etc. If this request is identified and sent to the owners of related jobs, it will promote business development.

In this research, the authors are dealing with Telegram text data, and it is necessary to process and classify the text to

* Corresponding Author

identify the hidden requests in these text documents. The task of classifying text is to classify a document into a predefined category [2]. In the present study, each Telegram message is considered as a document. Each category or class is a Boolean value that indicates whether the message is a request or not. A major problem with text classification is the increase in the size of the data being processed or the feature space's high dimensions [10], [11]. One solution is to reduce features using feature selection methods [10]-[15].

Feature selection involves selecting a set of relevant and informative features to build a predictive model with maximum efficiency [2]. Future selection (FS) has played an important role in machine learning and data science [16]-[19]. Although there are many and comprehensive methods for FS, it is an open and NP-Complete problem due to its complexity. There is no definitive and single solution. Nowadays, there are many articles on new FS methods, each with its advantages and disadvantages. Some of them combined three main categories of FS methods, filter, wrapper, and embedded [10], [12], [16], [20], to increase performance and accuracy. Some FS methods are designed to be applied in various fields, and some are designed for a specific issue [18].

However, the feature selection for the Persian text has not been sufficiently investigated. In addition to traditional methods, addressing this type of data requires more advanced techniques. Because this data has certain words and features, previous methods did not consider which. Due to this dataset's nature, a feature selection method is required to select the appropriate features for the Persian dataset of Telegram. Due to the use of learning algorithms, the wrapper and embedded selection methods show better performance than filters. However, filters are faster because they are independent of learning algorithms [12], [18], [20], [21]. In this study, the most extensively used methods were applied based on local and global filters, and wrapper methods described in Section 2 were used to take advantage of filter speed and wrapper accuracy at the same time. Local and global filtering methods are applied as pre-processing in the wrapper method and reduce the feature. The combination of these methods with the proposed approach in the present study led to high accuracy. No such combination has been developed in Telegram's Persian data so far. When the number of main features is very significant, filter and wrapper methods are a powerful combination for selecting the optimal subsets of features. The combination of the two methods can overcome the disadvantages caused by each. Combining these methods reduces the calculation time and calculates the relationships between the features [22].

The authors have proposed a combination method for selecting the most relevant features and optimizing the classifier parameters to achieve higher classification accuracy in the Persian text on Telegram with high dimensions. Although the accuracy obtained by applying the most used employed filter and wrapper methods was acceptable, these proposed combination approaches were used with ensemble methods to increase the accuracy, and another method was suggested to provide higher accuracy. Ensemble data mining methods are frequently used to improve classification performance and are also known as classifier combination. This method is not suitable for high-dimensional datasets [23]. In the present investigation, an ensemble method was proposed for classifying high-dimensional data. In the proposed method, the authors use the combined output of the most broadly used methods based on local and global filters as input and pre-processing features and reduce the main features' space. Each generated feature subset is then trained by a learning algorithm, and the results of each classifier are combined with a majority vote.

The performance of our proposed approaches has been evaluated with a Persian dataset. This dataset has been extracted from the Idekav-system 1of Yazd University, which is a Telegram search engine. Millions of messages are monitored daily on Idekav-system. Many of these messages exchanged among Telegram users are request-type messages. Request-type messages create many opportunities for monetization and are attractive to many businesses. The authors identify these requests and send them to business owners who can respond to them. Responding to these requests solves many users' problems and helps them quickly access their requests that lead to business development, and marketing is done by saving time and money. Therefore, request identification is an important issue that should be addressed further concerning the expansion of social networks.

In this regard, the authors suggest a combined machine learning method for the feature selection process in Persian texts of Telegram messengers using three feature selection techniques. As previously mentioned, FS has not been sufficiently investigated for the Persian text and no such combinations have been made for the Persian data. Furthermore, the data in this study is different from other messengers, both in terms of language and gender. Therefore, the second innovation of this research is identifying the requests of the Persian messages of Telegram. The authors performed many experiments to prove the method's validity with a different number of samples and features and analyzed the results. Therefore, the main part of this article is summarized as follows:

- In feature selection, a combined approach was offered based on local and global filters, which is useful for evaluating selected features and improving the efficiency of the training and testing phases.
- Empowering the selection of features for request identification in Persian messages on Telegram by

1 Idekav.com

introducing an approach of selecting the combined feature of filter and wrapper. This method uses the advantage of filters' speed and wrapper accuracy to increase accuracy.

- The authors offered a new method that combines the benefits of selecting features and ensemble classifiers to improve performance and accuracy of the classification in Persian text.
- In order to increase the performance of the classification in the Persian text dataset, an ensemble approach was introduced by combining the results of multiple classifications (SVM, NB, DT, MLP) using a majority voting based on average probabilities.
- This proposal has been compared with 85,000 samples using the methods available in a test platform consisting of a Persian dataset. Experimental results show that the proposed solution maintains Micro−F1, Macro−F1, and RMSE criteria at acceptable levels.

2- A Review of Combined Research in the Field of Feature Selection

Some recent studies have shown that combining feature selection methods can improve classification performance. In these combined methods, it was concluded that one method's performance might be inadequate as an individual, but its combination with other methods provides high efficiency. In general, feature selection methods are divided into three main categories: wrapper, embedded, and filter [10], [12], [16], [20], and their combination can be used to increase performance.

Combined Filter and Wrapper-based Feature Selection

In some studies, some techniques have been proposed that combine a filter and a wrapper method. Feature selection and model learning are made simultaneously by embedded methods. Wrapper methods use a learning algorithm to evaluate the subset of features, which increases performance [12], [18], [20], [21]. However, these algorithms require a great deal of time to be fully processed, and their main problem is to create an additional calculation cost [21]. For this reason, they are not directly preferred for text classification [21]. Some studies, e.g., Wah et al. [24], have compared filter and wrapper feature selection methods to maximize the accuracy of the classifier; and in some other studies, the FS methods, which are a combination of filter-based local methods and wrapper-based methods, have been investigated by Uysal [25]. In one category, wrappers can be applied in two areas: forward and backward methods. Xie et al. [26] presented a combined FS method that utilizes the benefits of filter and wrapper methods to select the optimal feature subset from the set of main features. They combined improved F-score, a filter evaluation criterion, with a wrapper evaluation system named Sequential Forward Search (SFS) to find a subset of the optimal feature in the FS process. The results revealed that the features decreased, and the classification accuracy increased. In this study, the authors applied SFS and a combination of filtering methods as pre-processing was used to increase training speed.

Combined Filter-based Feature Selection

Filter methods select features based on a pre-processing step and independent of the learning algorithm [18], [20]; and for this reason, they are straightforward and fast in terms of computation [12], [21]; hence they work well for high-dimensional data; although wrapper methods are highly time-consuming for high-dimensional data and provide acceptable accuracy in practice [21]. Furthermore, despite filtering methods, wrapper and embedded methods require frequent classifier interaction, which increases execution time [12]; therefore, filtering methods are more efficient.

Filtering methods have two categories, local and global [12], [21]. In some studies, global methods have been named as corpus-based, and local methods have been called class-based [21]. In the study of Ogura et al. [27], filter-based feature selection methods are divided into two categories of one-sided and two-sided based on their characteristics. Popular feature selection methods [12], [14] include: document frequency [21], information gain [21], [27], [28], Gini index [27], and distinguishing feature selector [12]. Odds ratio [12], [28], [29] and Correlation coefficient [12], [27], [29] are commonly used local selection methods. In the present study, a comprehensive study was performed on the most widely used filter-based FS methods, and then a brief description of the mathematical contexts of these methods was presented in Section 3.

Filter-based methods have been applied in many studies. Sometimes these methods are used individually and sometimes in combination with non-filter methods. In [12], [14], [25], filter-based methods have been combined with wrapper methods; they have also been employed with Principal Component Analysis (PCA) and genetic algorithms. Uysal and Gunal [30] suggested a filter-based probabilistic feature selection method called Distinguishing Feature Selector for text classification. BİRİCİK et al. [15] showed that chi-squared feature selection methods and correlation coefficient produce a subset of better features.

Recently, Uysal [12] combined the power of a filter-based global feature selection method and a one-sided local selection method called IGFSS1 to improve FS by applying filtering methods. The results indicated that this combined method's performance was better than the individual performance of the methods. This proposed method was not

1. Improved Global Feature Selection Scheme (IGFSS)

suitable for unbalanced data with a large number of classes. In order to solve this problem, Agnihotri et al. [14] proposed the VGFSS1 method, which is a combination of a global and odds ratio method. The idea is to build a set of final features that show each class based on the distribution of terms in the classes. By comparing and experimentally evaluating both local and global methods, Melo et al. [31] showed that local feature selection performed better than global [31]; additionally, in some investigations, local methods produced better results for lower feature value and global methods for higher feature value [21]; therefore in this study, a combination of global and local filter-based methods are also used to get better results; in other words, the global and local scores are employed directly in the feature ranking [14]. The references [25], [32], [33] are among the studies that have independently used filtering methods in combination.

Combined Feature Selection and Ensemble Classifier

The ensemble method is a machine learning technique that combines the results of several basic classifiers and increases accuracy [34]. For example, Bolon-Canedo et al. [35] provided a combination of classifiers and filters. The results revealed that the proposed method performed better in most cases and reduced the number of features by more than 80%. In the present study, using ensemble methods and combined filtering methods instead of individual filtering methods, an approach was proposed that provided the most accuracy for the dataset applied in this study. In other studies, filter compounds (combined filter methods) were not utilized as input and pre-processing of learning algorithms.

Ensemble methods are popular in machine learning research and pattern recognition. The purpose of ensemble methods is to combine the decisions of a set of weak learning algorithms or base learners to increase the accuracy and strength of the developed classified model. The generalizability of ensemble methods is better than that of single base learners. Ensemble methods can be divided into two categories: dependent and independent [36]. Voting is the simplest and most extensively used form of combining basic learning algorithms. There are several methods to combine the output of basic classification algorithms. These combined methods include majority voting, weight majority voting, combination Law of Naïve Bayes, behavioral knowledge space method, and probabilistic approximation [36]. In the present study, the authors used majority voting in the third proposed method by applying the combined method of local and global filters.

In this article, the combined or two-step methods were used to achieve reduced dimensions. Also, the most widely used methods of local and global filters were combined; the combined methods reduced the feature and increased accuracy. The combination of combined filter and wrapper methods was applied, and better accuracy was obtained. Moreover, in order to increase accuracy, the first proposed combination methods were applied using ensemble methods, which significantly increased the accuracy. The following is a description of these proposed methods.

1. Variable Global Feature Selection Scheme (VGFSS)

3- Proposed Method

Filter-based feature selection methods provide us important features by scoring each feature in a dataset. These methods are independent of classification algorithms. Filter-based methods inherently use statistical tests on a dataset, and the ranking of features is the main criterion in selecting the features. The authors determine a threshold experimentally. All features are scored and removed if they are less than the threshold value. Due to the simplicity of these methods, they can be extensively used for practical applications involving large amounts of data [12], [21]. Combining the output of these filter-based methods can increase accuracy. This combined method is referred to as the first proposed method.

The wrapper selection methods have less simplicity and speed compared to filters. However, these methods have higher accuracy than filters due to the application of learning algorithms. These methods can be combined to take advantage of both speed and accuracy. In the combined method, the features obtained using filter-based methods as the pre-processing step for wrapper methods can be applied. Wrapper methods are not suitable for large amounts of data due to their low speed. However, the use of filters as pre-processing can be appropriate. This combined method is called the second proposed method.

In combined methods, different classifications or learning algorithms are used to evaluate the accuracy. In the present study, ensemble methods are applied as the third proposed method, which is a combination of these algorithms. However, in order to increase the accuracy, the output of the methods in the first proposed method is used as input for this method. SVM, NB, MLP, and DT are of the algorithms used in this method, which are broadly used in text classification studies and feature selection.

In this section, some of the most extensively applied methods for selecting features are described based on local and global filters. The authors use the outputs related to these filtering methods for FS. In some sections, the output of these methods is combined, which include IG, GI, DF, CC, OR, DFS. By combining these methods, appropriate results are obtained, which include an increase in accuracy. Some of these output features are common features with high scores and, therefore, can be considered as selected features.

Information Gain (IG): IG is one of the FS methods used in text classification, which utilizes a global filter-based approach [37]. IG is a method for evaluating entropy-based features [38] and is widely used in statistics and machine learning [21]. The higher the entropy is, the more information about the feature is obtained [37].

$$IG(t) = -\sum_{i=1}^{M} P(C_i)\log P(C_i) + P(t)\sum_{i=1}^{M} P(C_i|t)\log P(C_i|t) + P(\bar{t})\sum_{i=1}^{M} P(C_i|\bar{t})\log P(C_i|\bar{t}) \quad (1)$$

Gini Index (GI): GHI is a method to select global features for text classification. It was first used in DT algorithms, and then an improved form of this algorithm was proposed for FS in the text. It is a supervised method with a simpler calculation compared to IG [12], [39], [40].

$$GI(t) = \sum_{i=1}^{M} P(t|C_i)^2 P(C_i|t)^2 \quad (2)$$

Distinguishing Feature Selector (DFS): DFS is one of the most recent and appropriate FS methods for text classification, which has been proposed by Uysal and Gunal [41].

$$DFS(t) = \sum_{i=1}^{M} \frac{P(C_i|t)}{P(\bar{t}|C_i)+P(t|\bar{C}_i)+1} \quad (3)$$

Document Frequency (DF): This method scores the features according to the number of views in the document [40]. DF defines the document label based on the highest frequency term, and it is the simplest global feature selection [40], [42].

$$DF(a_j) = N.p(a_j) \quad (4)$$

Correlation Coefficient (CC): This method is a type of chi-square and can be seen as a one-side chi-square. This FS method selects terms with the highest value of cc as a feature. CC is an FS method based on the local filter [15], [29].

$$cc(t, c_i) = \frac{\sqrt{N}[P(t,c_i)P(\bar{t},\bar{c_i}) - P(t,\bar{c_i})P(\bar{t},c_i)]}{\sqrt{P(t)P(\bar{t})P(c_i)P(\bar{c_i})}} \approx \frac{\sqrt{N}(AD-CB)}{\sqrt{(A+C)\times(B+D)\times(A+B)\times(C+D)}} \quad (5)$$

Odds Ratio (OR): The OR criterion is a filter-based local method, which obtains the membership of a special class with nominator and obtains the non-membership with a denominator. Membership and non-membership scores are normalized by dividing them over each other; in order to obtain the highest score from the formula, the nominator and denominator values must be maximized and minimized, respectively [12], [39], [40].

$$OR(t|C_i) = log\frac{P(t|C_i)[1-P(t|\bar{C_i})]}{[1-P(t|C_i)]P(t|\bar{C_i})} \quad (6)$$

3-1- Process of Implementing the Proposed Method

In this section, the details, parameters, and steps of the proposed method are described in a step-by-step manner. In Fig. 1, the general steps of the proposed method are presented.

The proposed method in the present article uses the output of the methods described earlier. The authors use the output of these methods in three ways, the result of which is to provide three proposed methods. In all three methods, individual filter methods are applied for the initial FS process. Each of the above methods selects different features. These features are considered to be the most important features according to their computational formula. The output of each proposed method is used by applying learning algorithms and evaluation criteria to determine the final features. A threshold is determined experimentally to determine the number of final features and is specified at each step. After determining the output features of each proposed method, if it meets the threshold criterion, it is determined as the output feature and sent for the purpose of classifying the text. The rest of the features are removed from the features matrix. The following figures demonstrate the three proposed methods. The following algorithm is used in these proposed methods for FS.

Data production and labeling: The authors used real-time textual data extracted from the Idekav-system for test and evaluation. These data are Persian Telegram messages and a label is considered for each message. Then the first 80% and 20% of the dataset are considered as training and test, respectively.

Pre-processing: The output of the previous step is a set of text documents that need to be pre-processed. Pre-processing converts textual contents into numbers and includes the tokenization, stop words, stemming, and weighting phases. The text in this study is a number of messages and each message is a sentence. In pre-processing, each sentence is broken down into a number of words and each word is a feature. The number of obtained features is significant. In order to remove insignificant and redundant features, pre-processing steps, such as deleting stop words, must be applied. In this step, a bag of word (BOW) was created and a feature vector was formed for each sentence. If there is a feature in a sentence, the corresponding entry in the feature vector gets a value equal to one. In the absence of that feature, the corresponding entry is equal to zero. The final feature vector of a matrix consists of zeros and ones. After the pre-processing steps, the total number of features extracted from the original text is equal to 6754. The authors also perform the steps of constructing a feature vector or matrix for feature selection methods.

In this research, each of the filter-based feature selection methods is applied to select the optimal set from 6754 features extracted from the original text. The selected feature sets of each local and global method are created separately; hence, the results in this section are shown separately for each method per feature matrix. The feature vector is used for learning methods as input, and each feature indicates the presence or absence of a word. The feature vector is considered as a matrix. First, the term class matrixes are extracted from the main dataset. The term class is a matrix, the columns and rows of which represent terms and classes, respectively. Each cell of this matrix contains the number of documents that contain a term such as t in a class such as c. The calculation of this matrix is necessary for all other feature selection methods.

After selecting the optimal feature subset as shown in Fig. 1, from the original matrix, the columns should be selected for which the corresponding feature is selected and a new matrix should be created. This new matrix will actually be the input to the combined feature selection methods.
The following are the algorithms used for these proposed combined FS methods.

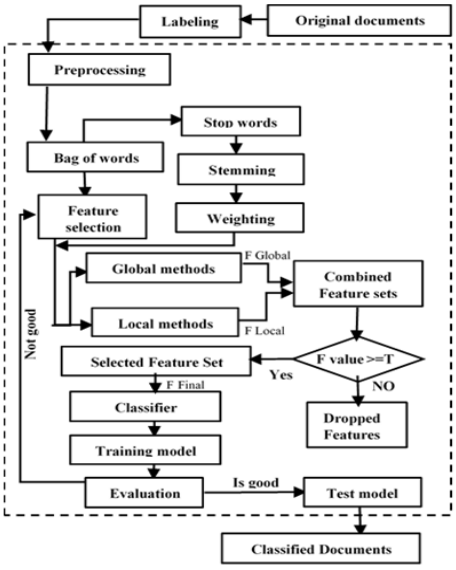

Fig. 1. General diagram of the research method
(In this diagram, a combination of local and global filtering methods is used as input to learning methods)

Algorithm 1. Scoring features with filter methods
Step 1- Set of main features: The set of features obtained from the preprocessing stage is shown with F = {F_1, F_2, …, F_N} and defined as a set of all the main features that in this set, N = 6754. The set C_i also denotes the positive and negative classes, which are equal to C_i = {C_1, C_2}.

Table 1. Explain the parameters used to select the feature vector in the proposed method

Parameter	Description	Collection
F	Set of main features without feature reduction	**F ={F_1, F_2, … , F_N }**
F Local	Selected features by local feature selection method	**F Local={F_1, F_2, … , F_L }, l < n**
F Global	Selected features by global feature selection method	**F Global={F_1, F_2, … , F_G }, g < n**
F Final	Final selected features by combining local and global features	**F Final={F_1, F_2, … , F_M }, m < l +g**
C_i	Selected feature class	**C_i = { C_1,C_2}**
T	Threshold value for selected features	**F value < T→ Dropped Features F value > T→ Feature Selection**
F value	The value of the selected feature	

Step 2- Feature selection: The number of features in set F are considerably large, that makes the implementation of learning methods time consuming. For this reason, the dimensions of the features must be reduced through some ways. The filter-based algorithms are examples of the most extensively used methods for reducing dimensions. For these feature vectors, both local and global feature selection methods are performed. In the feature selection steps, a series of parameters are applied, which are explained in Table 1. Also, in Fig. 2, the step of selecting a combined feature (combined FS) is shown in more detail. The steps of the feature selection method used in this article are as follows:
Step 3- Feature selection with local methods: In this step, using the local feature selection methods OR and CC, the features of each C_i class in set F are given a score. These methods are described in Section 3. These scores indicate the difference between terms or features in a dataset. In the next step, all the terms are arranged in descending order according to the score they gained in the feature selection stage. Then, *L* features with the highest score in the feature set are selected as the final features. The value of *L* is a definite number that is usually obtained experimentally. F Local ⊆ F is selected as a set of locally selected features. F Local = {F_1, F_2, …, F_L} contains *L* number of features.
Step 4- Feature selection with global methods: In this step, using the global feature selection methods GI, IG, DF and DFS, the features of each C_i class in the set F are given a score. These methods are described in Section 3. Then, as in the previous step, the features are arranged in descending order of scores. F Global ⊆ F is defined as a set of selected global features. F Global = {F_1, F_2, … , F_G} contains a number of *G* features.
Algorithm 2. Combining local and global output features
Step 1- Combining feature sets: From steps 3 and 4 in Algorithm 1, two feature sets were obtained, each of which has a specified value. In this step, the combination of the set of output features arranged from each of the local methods with the set of output features arranged from each of the global methods are performed. From these two sets, the features with higher scores are selected.
Step 2- If the F value feature is greater than the threshold T, that feature is selected, otherwise it is removed from the feature set. The value of T is determined experimentally.
Algorithm 3. Selecting the final set of the first proposed method
Step 1 - Selected Features: From Algorithm 2, a set of final features F Final = {F_1, F_2, …, F_M} is obtained, that F Final ⊆ (F Global ∪ F Local). The F Final feature set possesses a set value in which the best features with higher values are selected.
The combined algorithm in this research is shown in Fig. 2. The details of selecting the feature of the previous steps are shown in this figure with different colors. Local methods are shown in yellow and the output is the F Local features. Global methods are illustrated in blue and the output is the F Global features. After combining these two feature sets, the F Final set is obtained in green color, which selects the features with the F value > T feature. The value of T in this figure represents a threshold for the selected features that have been determined experimentally. This step is shown in more detail in Fig. 1.

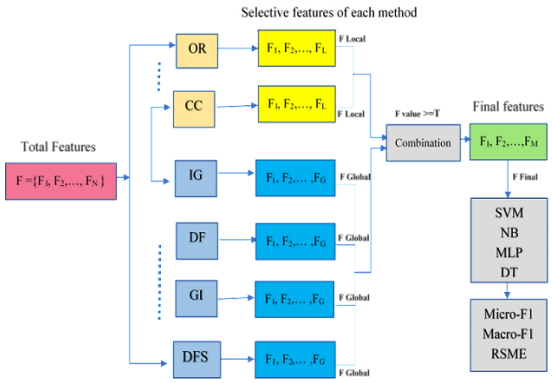

Fig. 2. Details of the steps to combine the feature set obtained from the feature selection methods

Step 2- Evaluation: Finally, the classification algorithms take the F Final set. After passing the training model, if the evaluation is acceptable, the documents will be classified.

The performance of classifications can be measured using learning methods. As shown in Fig. 2, four learning methods SVM, NB, MLP and DT were used in the present investigation. Micro-F1, Macro-F1 and RSME evaluation criteria were also applied. According to Fig. 1, if the evaluation results were not good enough, the feature selection step can be retrieved and the selection criteria be changed to reach the new set and the desired result.

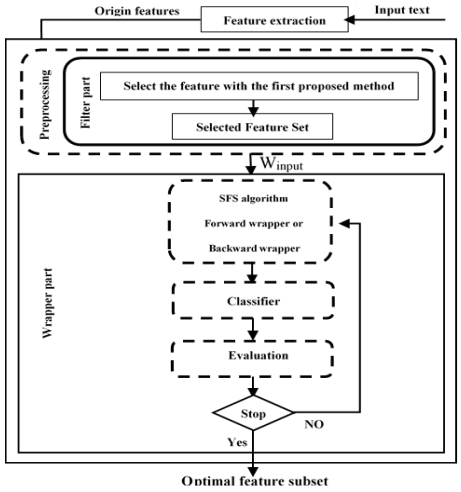

Fig. 3. General diagram of the second proposed method
(In this method, the output of a combination of local and global methods is used as input for learning-based methods)

Therefore, a new set of features have been obtained that includes the features with highest scores from the combination of local and global methods. Now, this new set can be tested using common learning methods in text classifications and feature selection and then the results can be compared. The experimental details of these steps are given in Section 4. This set of features (the feature set obtained in the first proposed method) is also used for the second proposed method (Fig. 3) and the third proposed method (Fig. 4).

Algorithm 4. Selecting the final set of the second proposed method

Step 1: The W_{input} dataset is equal to the output features of algorithm 3 or the F Final set of the step 4 of this algorithm.

Step 2: With T-threshold, determine the number of input features.

Step 3: Send the specified features as the input to the wrapper method.

Step 4: If the number of selected features is less than half of the T-threshold, continue to select the feature with the wrapper method.

Step 5: Repeat step 1 for all the features of the combined set, which are obtained from local and global methods.

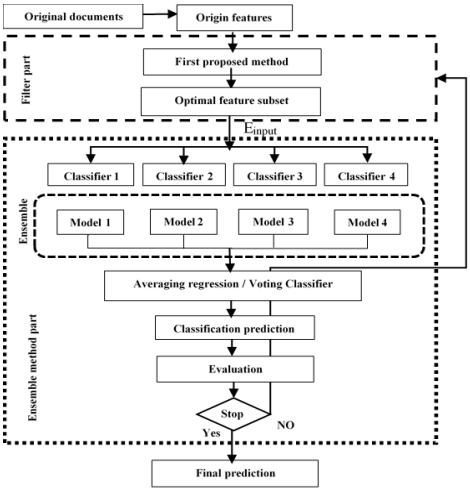

Fig. 4. General diagram of the third proposed method
(In this method, the output of a combination of local and global methods is used as input for ensemble learning methods)

Algorithm 5. Selecting the final set of the third proposed method

Step 1: The E_{input} dataset is equal to the output features of algorithm 3 or the F Final set of the step 4 of this algorithm.

Step 2: With T-threshold, determine the number of input features.

Step 3: Send the specified features to the classifiers as input.

Step 4: Train each classifier for each model.

Step 5: Get the output of step 4 with majority voting or average regression.

In this article, these algorithms were used to provide a model for classification and selecting features in the Persian text, using local and global filter feature selection methods, wrapper, and combining different classifiers. As shown in the figures, the pre-processing steps were performed before any action. Instead of using individual filter methods, a combination of local and global methods was applied as pre-processing for wrappers and combining of the classifiers. By following the suggested methods,

high accuracy was achieved. In the next section, the dataset used and the evaluation results will be presented.

4- Evaluation Results

In this section, the proposed FS models will be practically evaluated using two sets of real data (the Persian dataset of the Idekav-system and the famous Reuters dataset). The use of two datasets attempts to evaluate the performance of extensively used machine learning algorithms in FS fields. In the following, the dataset and the various widely used criteria will be presented, which are used in the text classification and FS studies.

4-1- Data and Evaluation Criteria

Persian Dataset of Telegram: In order to detect requests, it is required to check user messages. Therefore, the relevant documents with the highest and lowest scales are extracted; however, since numerical rankings cannot be applied equally to all phrases and sentences that are a part of the review, filter methods are used based on the characteristics of the studied language. The authors carry out the pre-processing steps, such as deleting the stop words of the Persian language, stemming, etc., on the phrases, and then use the matrix of the obtained features to calculate the score of that phrase. The Idekav dataset is applied.

The extracted dataset from the Idekav-system, which is a Telegram search engine, was applied to validate the proposed algorithm. This system includes many messages from the Telegram social network in Persian, which are regularly updated. The Telegram messenger is quite popular and beneficial among its users. These text messages have several different topics that can be used in different areas, such as data mining, opinion mining, and request identification. In the present investigation, the data used was extracted and processed by ourselves. The process of collecting and preparing this data was performed by seven senior and doctoral students of Yazd University. Training the work steps began with an explanatory session on labeling methods and rules. Each person received a username and password. People entered the Idekav-system and used interrogative keywords such as "how," "who," "I'm a buyer," and "I need" in the search field to find users' questions. They labeled sentences and messages according to the defined rules. For labeling, at first, it should be determined whether the message is a request type message or not. The authors label a message that contains a request with a positive label, and if there is no request, with it is labeled with a negative label. These explanations relate to the discussion of request identification or question identification. After the labeling process, the obtained file included 85748 records, each of which expressed a text message. The specifications of each message were shown in 14 columns. The columns indicate the characteristics of the text of the message, the message length, the positive and negative or neutral labels by the first and second person, the group ID, the group name, the number of group members, the user ID that sent the message, the message type and the sending time of the message, respectively. The dataset collection process started on February 7, 2017, and ended on March 29, 2017. Each message is labeled by two people to ensure that the labeling is correct. Labeling was performed in several different time stages. Before the last step, the statistics are reported as follows:

U1: 448, U2: 439, U3: 15185, U4: 14289, U5: 20462, U6: 9942, U7: 14569.

The obtained final statistic, which was recorded in an excel file, includes 85748 records. In the present investigation, because the data is inherently random, 80% of the primary data was used as a training data set, and the remaining 20% was applied as a test data set. Cross-validation was used for training and testing purposes.

Evaluation Criteria: Selecting the right criterion for evaluation is considerably important. Common evaluation criteria applied in text classification for evaluating the performance of learning algorithms are divided into two categories: internal and external. Internal criteria include similarity measurements. Accuracy [21], precision, recall, and F-measure [12], [14] have been identified as external criteria. The authors used Micro−F1, Macro−F1, and RMSE to evaluate the performance of the proposed methods. These criteria are regularly applied to measure the performance of classification methods. Depending on a classification model and a test dataset, the performance of the model in the test dataset can be measured based on these criteria.

Micro-averaged calculations give each document equal weight. However, macro-averaged provides equal weight to each category [43]. In the Micro−F1 calculation equation, p is the value of precision and r is the value of recall for all classification decisions in the whole dataset [12], [14], [25], [30], [39].

$$Micro - F1 = \frac{2 \times p \times r}{p + r} \tag{7}$$

The Macro−F1 calculation equation is the average calculation of each specific class. Where, p is the precision value and r is the recall value of class k [12],[14],[25],[30],[39]. The F1 score is the harmonic mean of precision and recall. Balancing precision and recall performance in optimizing the classifier is performed by its assistance [44].

$$Macro - F1 = \frac{\sum_{k=1}^{C} F_k}{C} . \quad F_k = \frac{2 \times p_k \times r_k}{p_k + r_k} \tag{8}$$

The RMSE (Root Mean Square Error) is the standard deviation of the remainder in the data. RMSE demonstrates the proximity of the predicted values to the

actual values; therefore, a lower value of RMSE indicates that the model performance is appropriate [45], [46].

$$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (9)$$

All experiments were performed with the Python language using the machine configuration as follows: OS: 64-bit Windows 10, CPU Speed: 2.60 GHz, Processor: Intel Core i7-3720QM, RAM: 24GB .The authors also used the Scikit-Learn library for machine learning to train the text classification model.

4-2- First Proposed Method (Combined Local and Global Filters)

The performance of the first proposed method is presented in the following figures. SVM was the first type of machine learning algorithm used. Fig. 5 indicates the Micro−F1 criteria for individual filter methods using this algorithm. Among individual methods, CC had a higher accuracy value.

Fig.6 presents the Micro−F1 criterion for combined filter methods using this algorithm. The highest accuracy obtained for this algorithm was equal to 0.844, which was related to the CC&DF combined method. The accuracy of this algorithm without selecting the feature was equal to 0.696. The accuracy values obtained with these algorithms in the first proposed method had a higher percentage of increase compared to the accuracy values without FS. According to the obtained results, SVM and NB learning methods performed better than other methods in the Persian dataset.

In combined methods, an optimal feature subset was obtained, which included 300 features. The results of this algorithm have been compared with the average results of other machine learning algorithms in Fig. 7.

In the first proposed method, the SVM classifier showed a more considerate increase in accuracy; hence, different kernels were obtained from this Kernel-based learner. The kernel types are linear, polynomial, sigmoid, and Radial Base Functions (RBF). Different kernels select different features, which also changes the amount of accuracy. The different kernels' results are presented in Fig. 8. In the Linear SVM, the CC&DF combination method had an accuracy of 0.846 for 300 features, which was higher than other kernels.

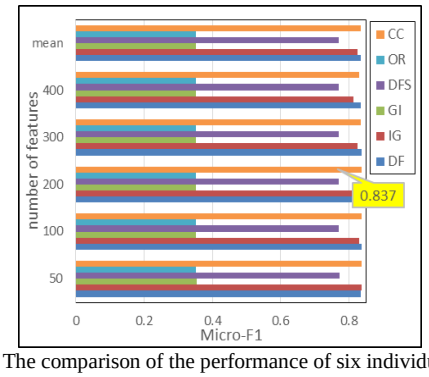

Fig. 5. The comparison of the performance of six individual filter methods for the different number of features and average of features using SVM classifier and Micro-F1 criterion.

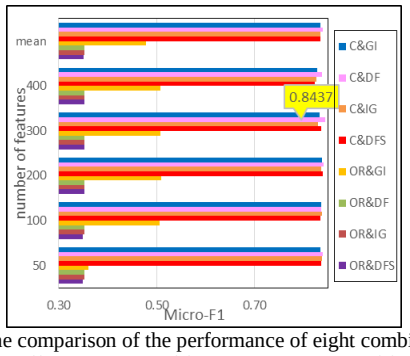

Fig. 6. The comparison of the performance of eight combined filter methods for different number of features and average of features using SVM classifier and Micro-F1 Criterion

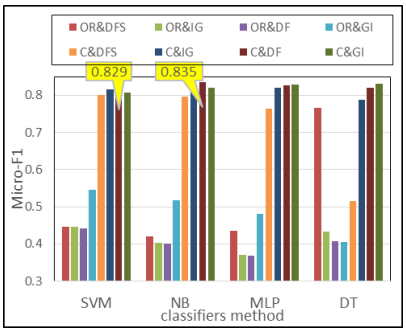

Fig. 7. The comparison of the average of SVM, AND, MLP, DT classifiers, and Micro-F1 criterion for eight combined methods of the filter.

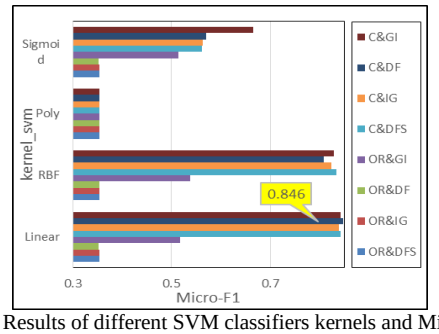

Fig. 8. Results of different SVM classifiers kernels and Micro-F1 criterion for eight combined filter methods

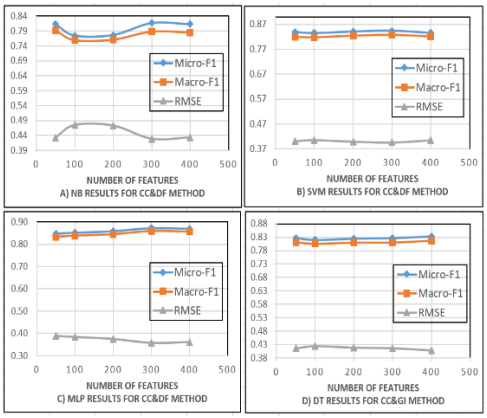

Fig. 9. Results of Micro−F1, Macro−F1, and RMSE criteria in the SVM, NB, MLP, and DT classifiers for the combined filter methods with the best results

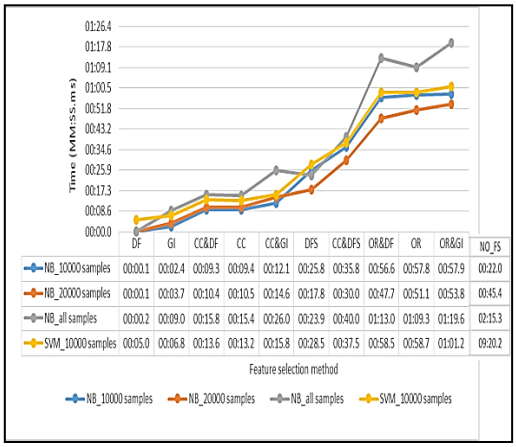

Fig. 10. Comparison of results of the times of the proposed methods when changing the processed dataset with NB and SVM classification algorithms with and without feature selection

Fig. 9 shows the values measured for different evaluation criteria. These diagrams are related to the combined methods with the highest accuracy in each learning algorithm. In NB and MLP, as in SVM, the CC+DF combination method presents a better result. In the following diagrams, the different criteria of this combined method are indicated. In the proposed method, a series of new features are obtained for each FS method, some of which may have selected common features. In Fig. 9, it is shown that the combination of CC and DF had a higher degree of accuracy.

The main focus of the present research was on increasing the accuracy, and it is shown in the results diagram that the reduction of feature using the proposed method led to an increase in the accuracy. However, some experiments have been performed to compare the timing of classification algorithms with and without feature selection. In some methods such as NB and SVM, the claim of time reduction as a result of feature reduction (its importance in the prediction stage by all means) was proved by performing the experiments.

The results of comparing the time of classification algorithms with and without feature selection is indicated in Fig. 10 for more accurate methods. In this figure, it is shown that using the proposed feature selection methods has reduced the training time. In particular, in SVM, it was shown that in most cases, the use of feature selection methods increased accuracy and reduced the training time. In some cases, an increase in time has been experienced. However, due to the percentage of increase in accuracy and the percentage of reduction of feature (as seen in Table 2), the increase in time is insignificant and negligible. In Fig. 10, SVM results with 10,000 samples of datasets and NB results with different numbers of samples (10,000 samples, 20,000 samples, and all samples) of datasets are shown. As demonstrated in the figure, as the amount of processed data increases, the time has increased accordingly in most methods. A column called NO_FS has been added to the diagram to indicate time for algorithms without feature reduction. As can be seen, there was a significant reduction in time in SVM by applying feature selection methods.

4-3- Comparison with Previous Studies

Because the database has been created by the authors, it is not possible to compare this database with the methods of other articles .In this section, the authors used another dataset, Reuters-21578, which has been applied in many studies related to text classification and FS, to evaluate and compare with previous works. This dataset has also been used in a study conducted by Uysal [12].

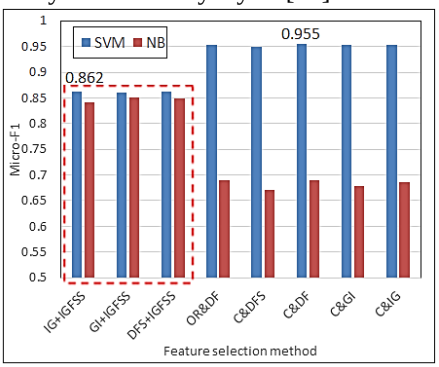

Fig. 11. Comparison of the proposed filter-based FS method of this article with the FS methods of previous articles with non-Persian data Reuters-21578 for SVM and NB classifiers and Micro-F1 criteria

The proposed methods in the present article were implemented on this dataset, and the results are shown Fig. 11; similar to Uysal's study, Micro-F1 and Macro-F1 criteria were used in the present research for evaluation. For individual methods using SVM, the highest accuracy is related to CC, which indicates the correlation coefficient and has a value of 0.954. Among the combined methods using SVM, the CC&DF method shows a better result compared to the other methods and has a value of 0.955. In Uysal's study, the highest result with SVM was equal to 0.862, which is lower than the highest accuracy (0.955) used in the first proposed method of this study. In Fig. 11, a comparison has been made between the combined

method, the first proposed method, and the proposed method of Usyal's study .In this diagram, it can be seen that in both studies, the use of SVM led to better results than NB. In Fig. 11, the first three pairs of bar graphs are related to the results of Usyal's method, and the next five bar graphs are related to the proposed method of the present research.

4-4- Second Proposed Method (Combined Filter and Wrapper Methods)

In this method, before assigning the main features to the wrapper methods, they must pass through the filter of the first proposed method, and then these more optimal features to be assigned to the wrapper methods as input features. The highest result is related to the combination of SFS and filter methods that include CC. Moreover, its combination with CC, which is a local method, shows a better result among individual methods. Among the combined methods, the combination of the SFS and FCC+DFS method provides a better result.

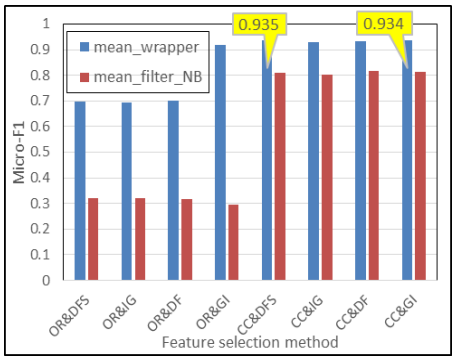

Fig. 12. Comparison of the average of combining the wrapper method and combined filter methods in 10,000 samples using NB classifier and Micro-F1 criteria

In the first proposed combined method, FCC+DFS gives an excellent result of 0.848. The combination of this method and the wrapper method has an accuracy equal to 0.937, which indicates the efficiency of the second proposed method and a higher degree of accuracy for the FS process. These results are obtained for 10,000 samples with 50 features. In Fig. 12, the comparison of the average number of different features in the combined wrapper and filter methods is presented.

4-5- Third Proposed Method (Use of Ensemble Learning Methods)

In this method, the output of the first proposed method was applied as input. In the third method, a combination of classifiers or learning algorithms was employed. In Fig. 13, the results of combining the output of the first proposed method and ensemble methods are shown. The results revealed that the combined filtering methods produced a higher result compared to the individual methods of filtering in this proposed method. Among the first proposed methods, the $E_{CC\&DF}$ combination method has a better result than the other combinations. The output of F_{CC+DF} is the input of ensemble methods, and the combined method of $E_{CC\&DF}$ was obtained. In this proposed method, other classifiers, such as random forest or KNN, can be used, and the results can be compared with the current results. The authors applied the same classifiers in the first proposed method.

In Fig. 14 the comparison of this proposed method is presented using different features and an average of features. On average, the combined method of $E_{CC\&IG}$, which is a combination of ensemble methods and CC&IG, provides better results than other methods. Moreover, the use of combined methods with CC as an input in ensemble methods leads to better results.

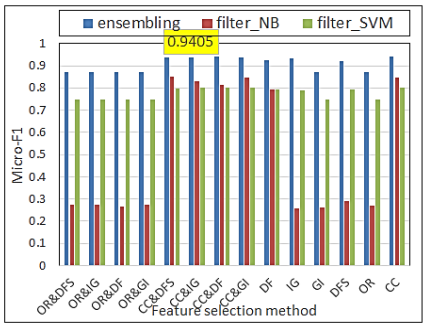

Fig. 13. Comparison of the combination of ensemble methods (combining the results of SVM, NB, MLP, and DT classifiers) and individual and combined filter methods in 50 features and 10,000 samples in SVM and NB classifiers with Micro-F1 criteria

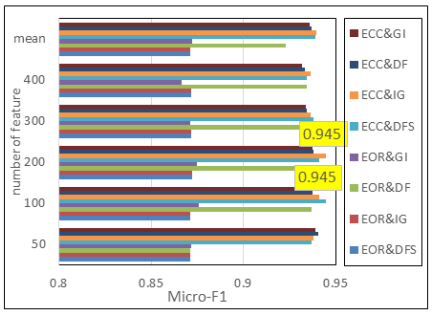

Fig. 14. Comparison of the average of combining the ensemble methods (combining the results of SVM, NB, MLP, and DT classifiers) and eight combined filter methods in 10,000 samples with Micro-F1 criteria

4-6- Comparison of the Three Proposed Methods

In the present study, three proposed methods were presented, which were described in the previous sections. In this section, all three proposed methods were compared. The results of performance evaluation and comparison are presented in Fig. 15. In this diagram, three methods are displayed all-in-one view, and it is observed that on average, the third proposed method, which is a combination of filter and ensemble

methods, outperforms the first and second proposed methods. Then the second proposed method, which is a combination of the wrapper and filter methods (the first proposed method), shows better results compared to the first proposed method. Moreover, combining the individual CC method with all the proposed methods possesses a higher value. Therefore, the combinations of this method, which has a high value, frequently show a favorable result. Now, by combining each of these combined methods with wrapper and ensemble methods, these combinations produced better results. The high degree of accuracy in CC-containing methods is because the correlation coefficient exactly selects the words that indicate membership in a classification. Therefore, the use of CC in FS leads to a significant improvement in classification performance. There is a relationship between the size of feature set, performance, and finding the size of the optimal feature set where there is the performance peak. Moreover, with the increase in the number of samples, this increase in performance has risen. The accuracy has been increased with an increase in the number of features until reaching the optimal subset. However, after the optimal set, the performance has decreased with an increasing number of features. Regularly, combining several FS methods may be better than a single method if each FS method shows unique scoring behavior and relatively high performance.

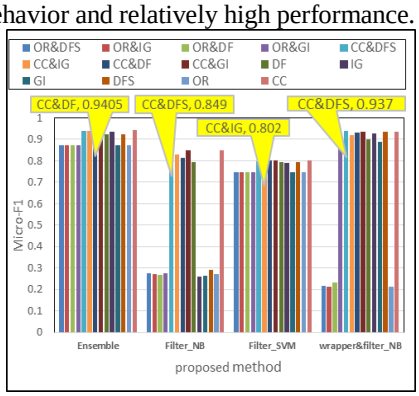

Fig. 15. Comparison of the combined ensemble methods (combining the results of SVM, NB, MLP, and DT classifiers), individual and combined filter methods (with SVM and NB classifiers), wrapper combinations in 10,000 samples, and 50 features with Micro-F1 criteria

In this figure, the results for 10,000 samples are presented. With this number of samples, the subset of the optimal feature possesses 50 features. With more samples, it is also true that the third proposed method shows a higher result than the second and first proposed methods but leads to an increase in the number of features in the optimal feature subset.

The results of the present research revealed that the combination of methods based on local and global filters showed a better classification performance than individual methods. These combinations reduced the dimensions of the feature space by producing the optimal subset of all the important and efficient features, leading to an increase in accuracy. On the other hand, the algorithms that select the best subset of the features are extremely optimal in terms of time and reduced the computational time. In other words, the learning model is obtained with higher generalizability, which shows the importance of FS in the classification. With these combined methods, the most optimal subset of the feature can be obtained. At each step, the FS operation is performed, and its related learning model is trained. The authors will continue these steps until reaching the best feature reduction rate and classification accuracy.

Table 2. The percentage of increase in accuracy and percentage of feature reduction of the proposed methods with Micro-F1 criterion and number of different training samples

Accuracy Without Feature Reduction	Accuracy with Feature Reduction	Percentage of Increase in Accuracy	Classifier	Better Method	Optimal Features	Percentage of Feature Reduction	Number of Samples
First Proposed Method							
0.69	0.84	21.74%	SVM	CC	200	-96.67%	all samples
0.69	0.843	**22.17%**	SVM	CC&DF	**300**	-95.00%	all samples
0.74	0.83	12.16%	NB	IG	50	-99.17%	all samples
0.74	0.82	10.81%	NB	CC&DF	300	-95.00%	all samples
0.856	0.864	1%	MLP	DF	300	-95.00%	all samples
0.856	0.872	1.87%	MLP	CC&DF	300	-95.00%	all samples
0.838	0.83	-1%	DT	IG	400	-93.33%	all samples
0.838	0.832	-0.72%	DT	CC&GI	400	-93.33%	all samples
0.602	0.848	40.86%	NB	CC	50	-99.17%	10000
0.602	0.849	**41.03%**	NB	CC&DFS	**50**	-99.17%	10000
0.345	0.777	125.22%	SVM	DFS	100	-98.33%	20000
0.345	0.776	124.93%	SVM	CC&DF	100	-98.33%	20000
Second Proposed Method							
0.602	0.937	55.65%	NB	CC&DFS	50	-99.17%	10000
Third Proposed Method							
0.746	0.94	26.01%	SVM	CC&DF	50	-99.17%	10000
0.602	0.94	**56.15%**	NB	CC&DF	**50**	-99.17%	10000
0.746	0.871	16.76%	SVM	CC&DF	200	-96.67%	all samples
0.602	0.871	**44.68%**	NB	CC&DF	**200**	-96.67%	all samples

The highest increase in accuracy was related to 50 features in 10,000 samples. In 20,000 samples, the maximum increase in accuracy was related to 100 features. In all samples, the maximum increase in accuracy was in 300 features, and this set was considered to be the optimal subset. Therefore, it can be concluded that as the volume of processed data increases, the optimal feature subset also rises. The percentage of increase in some cases is considerably small and also in some cases, is significant, and there is no incremental percentage in other cases; however, it is noteworthy that the reduction of training time has been significant due to the feature reduction.

Table 2 also indicates the percentage of increase in accuracy and the percentage of feature reduction of the proposed methods with the Micro-F1 criterion and the number of different training samples (increase in the amount of data processed). In this table, the results of the proposed algorithms are shown when changing the database.

5- Conclusions and Future Work

In the present article, three proposed methods were suggested to increase the accuracy of request identification in Persian messages on Telegram. In the first method, which was a combination of local and global filter-based methods, the CC&DF combination method increased the accuracy up to 0.844. This value

is related to the SVM classifier, which showed a better result than other classifiers. It is the reason that the authors have calculated the different kernels, among which the linear kernel showed a better result. The optimal feature subset in this method included 300 features. Based on the results obtained, the proposed combined methods considerably increased the accuracy, and the computation time was reduced. Accuracy and calculation time are effective criteria in machine learning methods. Wrapper algorithms have more accuracy than filter methods; however, their implementation of high-dimensional data takes much time to calculate. Therefore, the first proposed method of this research was applied as pre-processing for these methods, and the data dimensions were significantly reduced. Furthermore, this combined method was better than the first proposed method by providing an accuracy of 0.937. In the third proposed method, the output of the first proposed method was used as the input of ensemble methods. Then, the classifiers used in the first method were combined, and the result was better compared to the first and second proposed methods. The accuracy of the third proposed method was equal to 0.945. The authors applied Micro−F1, Macro−F1, and RMSE criteria to evaluate the performance of the proposed methods.

In the future, other ensemble classifiers, such as Random Forest, AdaBoost classifier, etc., will be evaluated. A combination of other filter and wrapper-based, as well as embedded methods will be used and the results will be compared with the results of the present study. Data on other social media can also be applied. It is also possible to use the proposed methods to select important features of bourse signals and improve business development by increasing the prediction accuracy.

References

[1] W. Y. Wang, D. J. Pauleen, and T. Zhang. "How social media applications affect B2B communication and improve business performance in SMEs". Industrial Marketing Management, vol. 54, pp. 4–14, 2016.

[2] E. Omer, "Using machine learning to identify jihadist messages on Twitter". M.S Theses, Dept. Information Technology, Uppsala Univ., Sweden, 2015.

[3] J. Surma and A. Furmanek. "Improving marketing response by data mining in social network ", in 2010 International Conference on Advances in Social Networks Analysis and Mining, 2010, pp. 446–451.

[4] W. He, S. Zha, and L. Li. "Social media competitive analysis and text mining: A case study in the pizza industry". International Journal of Information Management, vol. 33, no. 3, pp. 464–472, Jun. 2013.

[5] H. A. Vamerzani and M. Khademi. "Exploring the Uses and Challenges of Big Data in Opinion Analysis," in Proceedings of the 7th Iranian Conference on Electrical and Electronics Engineering, Gonabad, Islamic Azad University of Gonabad, 2016.

[6] M. Kiani nejad, T. hashemi, and M. rashidi. " Text mining social networks for consumer brand feelings and desires," in Proceedings of the 6th International Conference on Economics, Management and Engineering Sciences, Belgium, International Center for Academic Communication, 2016.

[7] Iran Analytical News Agency, "In which countries do telegram messengers favor?", khabaronline.ir, July. 2, 2019. [Online]. Available: khabaronline.ir/news/1275665. [Accessed:4 Jan 2020].

[8] Wikipedia contributors, "Telegram (software)," Wikipedia, The Free Encyclopedia, 27 Dec 2019, 15:24 UTC. [Online].Available: https://b2n.ir/907494.[Accessed:4 Jan 2020].

[9] Economics News, "Latest statistics from the mostpopular social networks in Iran", eghtesadnews.com, April. 9, 2019. [Online]. Available: https://b2n.ir/661242. [Accessed:4 Jan 2020].

[10] M. Nekkaa and D. Boughaci. "Hybrid harmony search combined with stochastic local search for feature selection". Neural Processing Letters, vol. 44, no. 1, pp. 199–220, 2016.

[11] X. Deng, Y. Li, J. Weng, and J. Zhang. "Feature selection for text classification: A review". Multimedia Tools and Applications, vol. 78, no. 3, pp. 3797–3816, 2019.

[12] A. K. Uysal. "An improved global feature selection scheme for text classification". Expert systems with Applications, vol. 43, pp. 82–92, 2016.

[13] L. M. Abualigah, A. T. Khader, M. A. Al-Betar, and O. A. Alomari. "Text feature selection with a robust weight scheme and dynamic dimension reduction to text document clustering". Expert Systems with Applications, vol. 84, pp. 24–36, 2017.

[14] D. Agnihotri, K. Verma, and P. Tripathi. "Variable global feature selection scheme for automatic classification of text documents". Expert Systems with Applications, vol. 81, pp. 268–281, 2017.

[15] G. BİRİCİK, B. Diri, and A. C. SÖNMEZ. "Abstract feature extraction for text classification". Turkish Journal of Electrical Engineering & Computer Sciences, vol. 20, no. Sup. 1, pp. 1137–1159, 2012.

[16] P. Lachheta and S. Bawa ."Combining synthetic minority oversampling technique and subset feature selection technique for class imbalance problem", in Proceedings of the International Conference on Advances in Information Communication Technology & Computing, 2016, p. 25.

[17] A. F. Sheta and A. Alamleh. "A professional comparison of c4. 5, mlp, svm for network intrusion detection based feature analysis", in The International Congress for global Science and Technology, 2015, vol. 47, p. 15.

[18] F. Aragón-Royón, A. Jiménez-Vílchez, A. Arauzo-Azofra, and J. M. Benítez. "FSinR: an exhaustive package for feature selection". arXiv preprint arXiv:2002.10330, 2020.

[19] A.-Z. Ala'M, A. A. Heidari, M. Habib, H. Faris, I. Aljarah, and M. A. Hassonah. "Salp Chain-Based Optimization of Support Vector Machines and Feature Weighting for Medical Diagnostic Information Systems", in Evolutionary Machine Learning Techniques, Springer, 2020, pp. 11–34.

[20] O. Stromann, A. Nascetti, O. Yousif, and Y. Ban. "Dimensionality Reduction and Feature Selection for Object-Based Land Cover Classification based on Sentinel-1 and

Sentinel-2 Time Series Using Google Earth Engine". Remote Sensing, vol. 12, no. 1, p. 76, 2020.
[21] D. Ö. Şahin and E. Kılıç. "Two new feature selection metrics for text classification". Automatika, vol. 60, no. 2, pp. 162–171, 2019.
[22] M. A. Hassonah, R. Al-Sayyed, A. Rodan, A.-Z. Ala'M, I. Aljarah, and H. Faris, "An efficient hybrid filter and evolutionary wrapper approach for sentiment analysis of various topics on Twitter". Knowledge-Based Systems, vol. 192, p. 105353, 2020.
[23] Y. Piao et al., "A new ensemble method with feature space partitioning for high-dimensional data classification". Mathematical Problems in Engineering, vol. 2015, 2015.
[24] Y. B. Wah, N. Ibrahim, H. A. Hamid, S. Abdul-Rahman, and S. Fong. "Feature Selection Methods: Case of Filter and Wrapper Approaches for Maximising Classification Accuracy. ". Pertanika Journal of Science & Technology, vol. 26, no. 1, 2018.
[25] A. K. Uysal. "On two-stage feature selection methods for text classification". IEEE Access, vol. 6, pp. 43233–43251, 2018.
[26] J. Xie and C. Wang. "Using support vector machines with a novel hybrid feature selection method for diagnosis of erythemato-squamous diseases". Expert Systems with Applications, vol. 38, no. 5, pp. 5809–5815, 2011.
[27] H. Ogura, H. Amano, and M. Kondo. "Distinctive characteristics of a metric using deviations from Poisson for feature selection". Expert Systems with Applications, vol. 37, no. 3, pp. 2273–2281, 2010.
[28] C. Huang, J. Zhu, Y. Liang, M. Yang, G. P. C. Fung, and J. Luo. "An efficient automatic multiple objectives optimization feature selection strategy for internet text classification". International Journal of Machine Learning and Cybernetics, vol. 10, no. 5, pp. 1151–1163, 2019.
[29] Z. Zheng and R. Srihari. "Optimally combining positive and negative features for text categorization", in ICML 2003 Workshop, 2003.
[30] A. K. Uysal and S. Gunal. "A novel probabilistic feature selection method for text classification". Knowledge-Based Systems, vol. 36, pp. 226–235, 2012.
[31] A. Melo and H. Paulheim. "Local and global feature selection for multilabel classification with binary relevance". Artificial intelligence review, vol. 51, no. 1, pp. 33–60, 2019.
[32] M. mojaveriyan, H. Ebrahimpour-Komleh, and S. jalaleddin Mousavirad. "Text Feature Selection using Document Frequency and Colonial Competitive Algorithm", in 8th National Conference on Data Mining, At Amirkabir University of Technology, Tehran, Iran, 2014.
[33] Ö. Uncu and I. B. Türkşen. "A novel feature selection approach: combining feature wrappers and filters". Information Sciences, vol. 177, no. 2, pp. 449–466, 2007.
[34] Y. Zhou, G. Cheng, S. Jiang, and M. Dai, "Building an Efficient Intrusion Detection System Based on Feature Selection and Ensemble Classifier". Computer Networks, p. 107247, 2020.
[35] V. Bolon-Canedo, N. Sanchez-Marono, and A. Alonso-Betanzos. "Feature selection and classification in multiple class datasets: An application to KDD Cup 99 dataset". Expert Systems with Applications, vol. 38, no. 5, pp. 5947–5957, 2011.
[36] A. Onan, S. Korukoğlu, and H. Bulut. "Ensemble of keyword extraction methods and classifiers in text classification". Expert Systems with Applications, vol. 57, pp. 232–247, 2016.
[37] K. Kurniabudi, A. Harris, and A. Rahim. "Seleksi Fitur Dengan Information Gain Untuk Meningkatkan Deteksi Serangan DDoS menggunakan Random Forest". Techno. Com, vol. 19, no. 1, pp. 56–66, 2020.
[38] T. Z. Win and N. S. M. Kham. "Information Gain Measured Feature Selection to Reduce High Dimensional Data", in Seventeenth International Conference on Computer Applications (ICCA 2019), 2019.
[39] B. Z. Abbasi, S. Hussain, S. Bibi, and M. A. Shah. "Impact of Membership and Non-membership Features on Classification Decision: An Empirical Study for Appraisal of Feature Selection Methods", in 2018 24th International Conference on Automation and Computing (ICAC), 2018, pp. 1–6.
[40] G. Kou, P. Yang, Y. Peng, F. Xiao, Y. Chen, and F. E. Alsaadi. "Evaluation of feature selection methods for text classification with small datasets using multiple criteria decision-making methods". Applied Soft Computing, vol. 86, p. 105836, 2020.
[41] A. K. Uysal and S. Gunal. "A novel probabilistic feature selection method for text classification". Knowledge-Based Systems, vol. 36, pp. 226–235, 2012.
[42] B. Tang, S. Kay, and H. He. "Toward optimal feature selection in naive Bayes for text categorization". IEEE transactions on knowledge and data engineering, vol. 28, no. 9, pp. 2508–2521, 2016.
[43] K. D. Rosa and J. Ellen. "Text classification methodologies applied to micro-text in military chat", in 2009 International Conference on Machine Learning and Applications, 2009, pp. 710–714.
[44] D. Sarkar. "Text Classification", in Text Analytics with Python, Springer, 2019, pp. 275–342.
[45] S. A. Verma, G. T. Thampi, and M. Rao. "Efficacy of a Classical and a Few Modified Machine Learning Algorithms in Forecasting Financial Time Series", in Internet of Things, Smart Computing and Technology: A Roadmap Ahead, Springer, 2020, pp. 3–30.
[46] M. Swamynathan. Mastering machine learning with python in six steps: A practical implementation guide to predictive data analytics using python. Apress, 2019.

Zahra Khalifeh Zadeh received the B.S degree in Computer Engineering from Shiraz University, Shiraz, Iran. She received her MSc degree in Computer Engineering (software) from Yazd University, Iran in 2020. Her research interests include Machine learning and Data mining.

Mohammad Ali Zare Chahooki received his BS in computer engineering from Shahid Beheshti University, Tehran, Iran in 2000 and his MS and PhD in software engineering from Tarbiat Modares University, Tehran, Iran in 2004 and 2013, respectively. Currently, he is an assistant professor at the department of Computer Engineering in Yazd University, Yazd, Iran. His research interests include Machine learning, Computer vision, and Software engineering.