

Improved ICHI square feature selection method for Arabic classifiers

Hadeel N. Alshaer, Mohammed A. Otair, Laith Abualigah

Faculty of Computer Sciences and Informatics, Amman Arab University, Jordan

Article Info

Article history:

Received Jan 23, 2020

Revised Mar 3, 2020

Accepted Mar 20, 2020

Keywords:

Classification algorithms
Feature selection methods
Text classification

ABSTRACT

Feature selection problem is one of the main important problems in the text and data mining domain. This paper presents a comparative study of feature selection methods for Arabic text classification. Five of the feature selection methods were selected: ICHI square, CHI square, Information Gain, Mutual Information and Wrapper. It was tested with five classification algorithms: Bayes Net, Naive Bayes, Random Forest, Decision Tree and Artificial Neural Networks. In addition, Data Collection was used in Arabic consisting of 9055 documents, which were compared by four criteria: Precision, Recall, F-measure and Time to build model. The results showed that the improved ICHI feature selection got almost all the best results in comparison with other methods.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Laith Abualigah,
Faculty of Computer Sciences and Informatics,
Amman Arab University,
Jordan Street–Mubis, Amman-Jordan.
Email: lx.89@yahoo.com

1. INTRODUCTION

Research and development in the science of information retrieval and development in the technology used has increased in many applications in various data such as texts, images, sound ... etc. These data are textual in different languages, and here we speak the Arabic language as the Arabic language flourished in the field of information retrieval and specifically in the field of classification [1]. This boom has led researchers to research and develop properties that improve results and improve efficiency. Information Retrieval (IR) is a field of computer science of great importance in our time because of the increasing volume of information [2]. This information may need to be arranged and classified so that it can be easily retrieved. Text classification (TC) a process that has been emerged in an important manner in various fields, especially in areas on the Internet [3].

Text mining is a textual analysis of data in natural language text and seeks to extract useful information from textual data [4]. In addition, text mining helps organizations extract valuable ideas from document content. Text mining process used to increase the efficiency of the text retrieval process by discovering patterns in the text and the relationships between them to help the retrieval of texts correctly [5]. The most important applications on Text Mining are: IR, Information Extraction, Classification and Natural Language Processing (NLP). Meaning it extracts useful information from a large amount of data to replace data and information search problems [6].

There are many processes that can be searched from the data collection used through the pre-processing to the classification process as pre-processing contains several steps [7, 8], including: Tokenization, Normalization, Stopword removal and Stemming. There are several algorithms in these

processes. In addition to this there is a stage called feature selection step and there are several algorithms that do this process and, in this research, will talk about some algorithms, namely: CHI square, ICHI square, Information Gain, Mutual Information and Wrapper. Where they will be compared in their use with the classification process and using several algorithms classification: Bayes Net, Naive Bayes, Random Forest, Decision Tree and Artificial Neural Networks. The results will be compared based on several criteria: Precision, Recall, F-measure and Time to build model. These algorithms will be tested using the pre-processing step, Tokenization, Normalization, Stopword removal and Stemming with feature selection step and without using the pre-processing step. These processes are used in data mining applications and examples of important applications in data mining: Natural Language Processing (NLP), Classification and Information Extraction [9, 10].

Feature selection problem is one of the main important problems in the text and data mining domain. Feature selection is a common problem used in many data mining areas; it is used to eliminate the irrelevant feature for reducing the number of selected features and increasing the computational time. This paper presents a comparative study of feature selection methods for Arabic text classification. Five of the feature selection methods were selected: ICHI square, CHI square, Information Gain, Mutual Information and Wrapper. It was tested with five classification algorithms: Bayes Net, Naive Bayes, Random Forest, Decision Tree and Artificial Neural Networks. Besides, Data Collection was used in Arabic consisting of 9055 documents, which were compared by four criteria: Precision, Recall, F-measure and Time to build model. The results showed that the improved ICHI Square Feature Selection Method for Arabic Classifiers got better results in almost all test cases in this paper. Furthermore, the improved ICHI Square Feature Selection Method for Arabic Classifiers is compared to other well-known feature selection methods and classification algorithm and it shows its superior ability in solving the feature selection problems. The rest of this paper is organized as Section 2 presents the literature review for more related works, the text preprocessing, the feature selection methods, and the evaluation criteria. Section 3 presents experiments results and analysis. Finally, Section 4 presents conclusions and future research.

2. LITERATURE REVIEW

2.1. Data collection

Data collection has been compiled from several Arab newspapers and magazines. It consists of 9055 documents classified into 12 categories: Art & Culture, Economy, Education, Engineering, Law, Local, Medicine, Miscellaneous, Politics, Religion, Science and Technological, Table 1 shows the data collection in detail. Each document consists of 75-130 words or more [10].

Table 1. Arabic data collection

| No. | Class Name | No. of Document |
|-------|---------------|-----------------|
| 1 | Art & Culture | 700 |
| 2 | Economy | 1082 |
| 3 | Education | 550 |
| 4 | Engineering | 570 |
| 5 | Law | 800 |
| 6 | Local | 450 |
| 7 | Medicine | 600 |
| 8 | Miscellaneous | 460 |
| 9 | Politics | 1145 |
| 10 | Religion | 1078 |
| 11 | Science | 770 |
| 12 | Technological | 850 |
| Total | | 9055 |

2.2. Pre-processing

2.2.1. Overview

Pre-processing is a very important step in data mining operations as it supports the results and increases their efficiency so it is very important to go through this step before any process. In this research, we will use four processes of pre-processing to improve the classification process and these processes are: Tokenization, Normalization, Stopword removal and Stemming. Figure 1 shows the steps of this phase from data collection to classification.

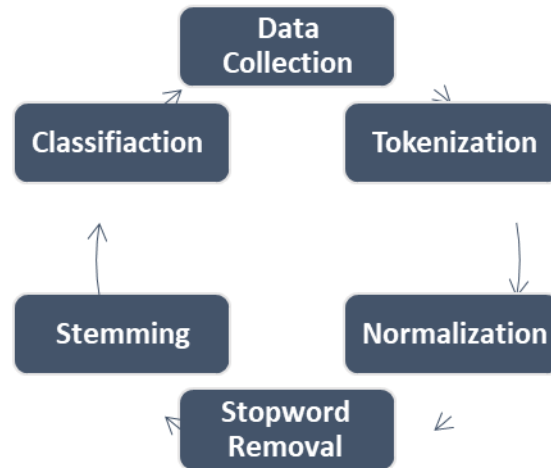


Figure 1. Pre-processing life cycle

2.2.2. Tokenization

At this phase, the sentences are cut into a series of words, depending on the white space between the words, to facilitate the compilation of words belonging to the same category. They are segmented into a series of keywords, phrases, etc., where some characters and symbols such as punctuation in the Tokenization process are ignored. This series becomes input into another process such as data analysis and data mining. Tokenization is also a language type and this process is very important in data mining tools [11].

2.2.3. Normalization

This phase converts the set of words and sentences into more complex and more precise sentences, so that the process of processing is easy and data handling easier. Normalization improves text matching and improves retrieval by taking synonyms for meaningful words to produce the best results and higher efficiency. For example, in this process the un-dotted Arabic letter (ع) replace to the dotted letter (ع) [12].

2.2.4. Stopword removal

This phase removes the stopword from the document where the stopword has no meaning like pronouns, vowels and the like. That is, at this point stopword removal, words and symbols that do not affect the meaning are removed and thus improve the results of the retrieval and classification processes. Examples are icons (@, #, !, \$, %, &, *, ^, +, -, ...). The preposition (... عن في، على، في)، and the adverbs (... تحت، فوق، [13].

2.2.5. Stemming

The stemming process is important to find the root of the word, to reduce the size of the document by removing the Prefix and Suffixes in addition to removing the increase characters such as (أ، و، ع). The process of stemming is by pre-processing the information retrieval and also facilitates searches in search engines thus increasing the results [14].

2.3. Feature selection methods

2.3.1. Overview

The feature selection process is important to reduce the amount of unwanted data and to maintain the most important data. The features are defined according to predefined criteria and are a pre-processing stage and are important in improving the classification process. Figure 2 shows the stages in the classification process using feature selection methods.

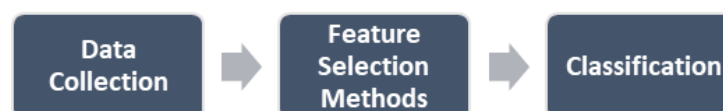


Figure 1. Feature selection life cycle

2.3.2. CHI square

Is a statistical method by selecting random data and depends on two independent variables or variables from the data sample and is considered as a feature selection method. CHI square is an important pre-processing in the classification process and CHI square is used in the classification system [15].

2.3.3. Improved CHI square (ICHI)

ICHI is a development of the CHI square method and is done by improving the basic properties where ICHI square is an extension of the original CHI square method. ICHI square has been used with Chinese language. The results have proved effective in using ICHI square with Chinese language. [16] In addition, ICHI square was used in Arabic using several algorithms: Bayes Net, Naïve Bayes, Naïve Bayes Multinomial, Random Forest, Decision Tree and Artificial Neural Networks. The results proved effective in using Arabic [17].

2.3.4. Information gain (IG)

Information Gain is based on the measurement of the number of bits gained by the information obtained by predicting the category by the presence of the word or not in the document [18].

2.3.5. Mutual information (MI)

Mutual information is based on the fact that information exchanged is a measure of dependence between variables, since the information exchanged is a measure of information that is (possibly multi-dimensional). Where the amount of information obtained is determined by one or more random variables [19].

2.3.6. Wrapper

Wrapper Algorithm is based on the use of learning algorithms where it selects relevant features based on the performance of learning algorithms. In addition to that the process of training the model with different features, or selecting the features that lead to the best results out of the model and be an example of a complex model of several features related to each other [20].

2.4. Arabic text classifiers

2.4.1. Overview

In order to carry out the classification process, we need algorithms that perform the classification process through the classification process through the training process based on a data set categorized into the testing process of a new text and distinguish it into which category belongs. As the classification process is supervised learning and is set to input. Figure 3 shows Text Classification system.

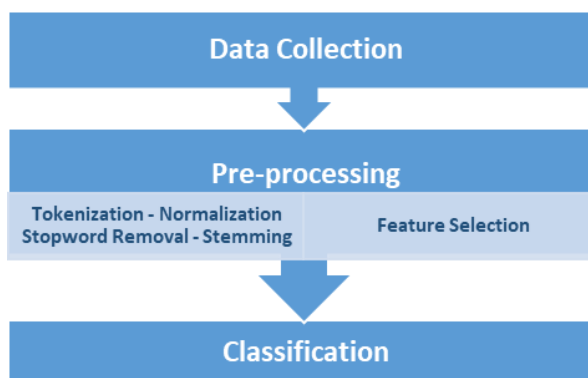


Figure 2. Classification system

2.4.2. Bayes net (BN)

Bayes Net classifier is a model that reflects situations that are part of the Bayes Net classifier and how they relate to each other. This model may be related to any entity in this world and can be represented by BN classifier and all entities occur frequently when another entity exists. This model is very important

because it helps us to predict the results of the entity in this world and it is easy to represent the entity through this model [21].

2.4.3. Naïve bayes (NB)

Naive Bayes classifier is based on Bayes Theorem, where NB is easy to build and very useful with large data. Despite the ease of building the NB model it outperforms many of the workbooks in the classification process [11]. In the NB analysis process, the final classification is produced by combining all sources of information. The model is created based on training data to classify the document into any category [21].

2.4.4. Random forest (RF)

Random Forest Classifier is a set of classification algorithms that are widely used in many applications, especially with the large data set due to its characteristic properties. The Random Forest algorithm is used in many applications, including: Network intrusion detection, Email spam detection, gene classification, Credit card fraud detection, and Text classification [22].

2.4.5. Decision tree (DT)

Decision Tree is an algorithm that compiles and is also used in the field of extracting data in various fields on the Internet. Also used in data mining algorithms, in addition, the Decision Tree algorithm is used in Arabic Text Classification [23]. The Decision Tree with the CHI square as a feature selection method was used with data collection in Arabic language, and Decision Tree was used with the ICHI square as a feature selection method. Results showed that the ICHI square surpassed the regular CHI square [24].

2.4.6. Artificial neural networks (ANNs)

Artificial Neural Networks is an algorithm that is used for classification. In addition, its branch of artificial intelligence, where different sets of functions have been studied from training and learning to the testing process. [25] ANNs has been successfully applied to problems in: pattern classification, function approximation, optimization, and pattern matching [26].

2.5. Evaluation measure

2.5.1. Overview

After the classification process shows us many results and until the results are read and compared, there must be criteria for this comparison. These results are compared to four criteria: Precision, Recall, F-measure and Time to build model.

2.5.2. Precision

Precision is a positive predictive value which is a part of the cases retrieved relevant from all cases retrieved from the results of the process [26, 27, 28]. Where precision is represented by the following equation:

$$Precision = \frac{|{\text{Relevant Document}} \cap {\text{Retrieved Document}}|}{|{\text{Retrieved Document}}|} \quad (1)$$

2.5.3. Recall

The recall is described as the sensitivity measure, which is the part of the relevant cases of the total number of recovered cases.

$$Recall = \frac{|{\text{Relevant Document}} \cap {\text{Retrieved Document}}|}{|{\text{Relevant Document}}|} \quad (2)$$

2.5.4. F-measure

The F-measure is a measure of the accuracy of the classifier being tested, which is a precision and recall ratio.

$$F - measure = 2 * \frac{(Precision * Recall)}{(Precision + Recall)} \quad (3)$$

2.5.5. Time

It is the time it takes to build the data model, which is also the time used in the process of analyzing the data and calculating the processes performed by the model.

3. EXPERIMENTAL RESULTS AND ANALYSIS

3.1. Results without natural language processing

3.1.1. Bayes net (BN)

Table 2 shows bayes net classifier results without pre-processing using several feature selection algorithms. The results showed that in the case of Precision was the highest result of classifier when used with ICHI square as a feature selection method. Moreover, that in the case of Recall was the highest results of the classifier when used with ICHI square as a feature selection method, but in the case of F-measure was the highest results of the classifier when used with ICHI square as a feature selection method and in the case of Time was the less time to build model of the classifier when used with Information Gain as a feature selection method. Figure 4 shows diagram Bayes Net results without pre-processing.

Table 2. Bayes net results without pre-processing

| Algorithm | Precision | Recall | F-measure | Time |
|------------|-----------|--------|-----------|-------|
| ICHI | 0.976 | 0.940 | 0.957 | 25.6 |
| CHI | 0.858 | 0.910 | 0.883 | 18.20 |
| IG | 0.851 | 0.899 | 0.874 | 13.3 |
| MI | 0.819 | 0.873 | 0.845 | 22.5 |
| Wrapper | 0.832 | 0.888 | 0.859 | 20.7 |
| Average | 0.8672 | 0.902 | 0.883 | 20.06 |
| Without FS | 0.805 | 0.871 | 0.836 | 16.7 |

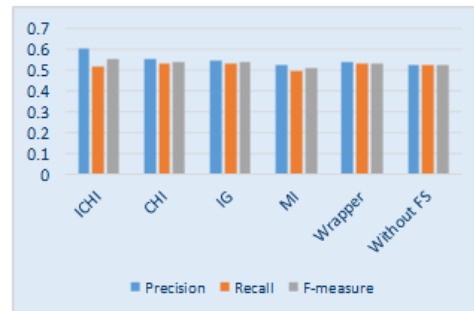


Figure 4. Bayes net results without pre-processing

3.1.2. Naïve bayes (NB)

Table 3 shows naïve bayes classifier results without pre-processing using several feature selection algorithms. The results showed that in the case of Precision was the highest result of classifier when used with ICHI square as a feature selection method. Moreover, that in the case of Recall was the highest results of the classifier when used with ICHI square as a feature selection method, but in the case of F-measure was the highest results of the classifier when used with ICHI square as a feature selection method and in the case of Time was the less time to build model of the classifier when used with Without FS as a feature election method. Figure 5 shows diagram Naïve Bayes results without pre-processing.

Table 3. Naive bayes results without pre-processing

| Algorithm | Precision | Recall | F-measure | Time |
|------------|-----------|--------|-----------|-------|
| ICHI | 0.976 | 0.954 | 0.964 | 18.20 |
| CHI | 0.926 | 0.950 | 0.937 | 4.11 |
| IG | 0.922 | 0.952 | 0.938 | 9.5 |
| MI | 0.913 | 0.945 | 0.928 | 8.4 |
| Wrapper | 0.919 | 0.950 | 0.934 | 9.2 |
| Average | 0.931 | 0.950 | 0.940 | 9.88 |
| Without FS | 0.913 | 0.951 | 0.931 | 3.16 |

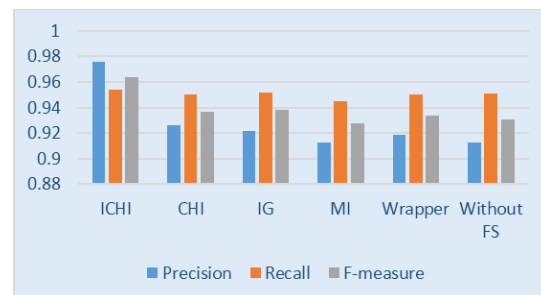


Figure 5. Naive bayes results without pre-processing

3.1.3. Random forest (RF)

Table 4 shows random forest classifier results without pre-processing using several feature selection algorithms. The results showed that in the case of Precision was the highest result of classifier when used with ICHI square as a feature selection method. Moreover, that in the case of Recall was the highest results of the classifier when used with ICHI square as a feature selection method, but in the case of F-measure was the highest results of the classifier when used with ICHI square as a feature selection method and in the case of Time was the less time to build model of the classifier when used with Wrapper as a feature selection method. Figure 6 shows diagram Random Forest results without pre-processing.

Table 4. Random forest results without pre-processing

| Algorithm | Precision | Recall | F-measure | Time |
|------------|-----------|--------|-----------|-------|
| ICHI | 0.947 | 0.942 | 0.944 | 94.3 |
| CHI | 0.938 | 0.805 | 0.866 | 66.5 |
| IG | 0.931 | 0.899 | 0.914 | 69.1 |
| MI | 0.917 | 0.849 | 0.881 | 70.3 |
| Wrapper | 0.929 | 0.851 | 0.888 | 66.4 |
| Average | 0.932 | 0.869 | 0.899 | 73.3 |
| Without FS | 0.915 | 0.844 | 0.878 | 71.48 |

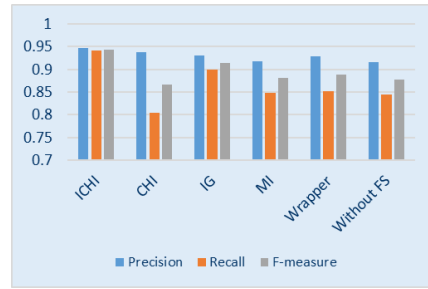


Figure 6. Random forest results without pre-processing

3.1.4. Decision tree (DT)

Table 5 shows decision tree classifier results without pre-processing using several feature selection algorithms. The results showed that in the case of Precision was the highest result of classifier when used with ICHI square as a feature selection method. Moreover, that in the case of Recall was the highest results of the classifier when used with CHI square as a feature selection method, but in the case of F-measure was the highest results of the classifier when used with ICHI square as a feature selection method and in the case of Time was the less time to build model of the classifier when used with Wrapper as a feature selection method. Figure 7 shows diagram Decision Tree results without pre-processing.

Table 5. Decision tree results without pre-processing

| Algorithm | Precision | Recall | F-measure | Time |
|------------|-----------|--------|-----------|------|
| ICHI | 0.604 | 0.518 | 0.557 | 4.40 |
| CHI | 0.552 | 0.537 | 0.544 | 2.51 |
| IG | 0.547 | 0.533 | 0.539 | 3.7 |
| MI | 0.529 | 0.500 | 0.514 | 8.9 |
| Wrapper | 0.539 | 0.530 | 0.534 | 6.3 |
| Average | 0.554 | 0.523 | 0.538 | 5.16 |
| Without FS | 0.528 | 0.523 | 0.525 | 1.01 |



Figure 7. Decision tree results without pre-processing

3.1.5. Artificial neural networks (ANNs)

Table 6 shows artificial neural networks classifier results without pre-processing using several feature selection algorithms. The results showed that in the case of Precision was the highest result of classifier when used with ICHI square as a feature selection method. Moreover, that in the case of Recall was the highest results of the classifier when used with CHI square as a feature selection method, but in the case of F-measure was the highest results of the classifier when used with CHI square as a feature selection method and in the case of Time was the less time to build model of the classifier when used with Without FS as a feature selection method. Figure 8 shows diagram ANNs results without pre-processing.

Table 6. ANNs results without pre-processing

| Algorithm | Precision | Recall | F-measure | Time |
|------------|-----------|--------|-----------|-------|
| ICHI | 0.713 | 0.492 | 0.582 | 44.17 |
| CHI | 0.699 | 0.542 | 0.611 | 43.7 |
| IG | 0.635 | 0.523 | 0.573 | 50.1 |
| MI | 0.590 | 0.460 | 0.516 | 59.9 |
| Wrapper | 0.601 | 0.499 | 0.545 | 49.1 |
| Average | 0.647 | 0.503 | 0.565 | 49.4 |
| Without FS | 0.586 | 0.452 | 0.510 | 36.34 |

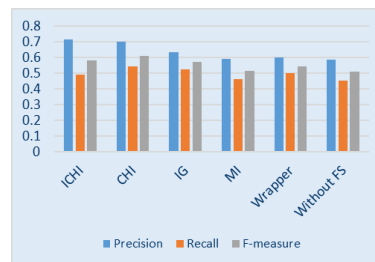


Figure 8. ANNs results without pre-processing

3.2. Results with natural language processing

3.2.1. Bayes net (BN)

Table 7 shows bayes net classifier results with pre-processing using several feature selection algorithms. The results showed that in the case of Precision was the highest result of classifier when used with ICHI square as a feature selection method. Moreover, that in the case of Recall was the highest results of the classifier when used with ICHI square as a feature selection method, but in the case of F-measure was the highest results of the classifier when used with ICHI square as a feature selection method and in the case of Time was the less time to build model of the classifier when used with Wrapper as a feature selection method. Figure 9 shows diagram Bayes Net results with pre-processing.

Table 7. Bayes net results with pre-processing

| Algorithm | Precision | Recall | F-measure | Time |
|------------|-----------|--------|-----------|------|
| ICHI | 0.883 | 0.922 | 0.902 | 20.3 |
| CHI | 0.851 | 0.888 | 0.869 | 22.5 |
| IG | 0.844 | 0.891 | 0.866 | 27.1 |
| MI | 0.790 | 0.801 | 0.795 | 21.1 |
| Wrapper | 0.839 | 0.870 | 0.854 | 17.2 |
| Average | 0.841 | 0.874 | 0.857 | 21.6 |
| Without FS | 0.828 | 0.893 | 0.859 | 16.7 |



Figure 9. Bayes net results with pre-processing

3.2.2. Naïve bayes (NB)

Table 8 shows naïve bayes classifier results with pre-processing using several feature selection algorithms. The results showed that in the case of Precision was the highest result of classifier when used with CHI square as a feature selection method. Moreover, that in the case of Recall was the highest results of the classifier when used with ICHI square as a feature selection method, but in the case of F-measure was the highest results of the classifier when used with ICHI square as a feature selection method and in the case of Time was the less time to build model of the classifier when used with Without FS as a feature selection method. Figure 10 shows diagram Naïve Bayes results with pre-processing.

Table 8. Naive bayes results with pre-processing

| Algorithm | Precision | Recall | F-measure | Time |
|------------|-----------|--------|-----------|------|
| CHI | 0.944 | 0.966 | 0.954 | 5.5 |
| ICHI | 0.953 | 0.956 | 0.954 | 9.11 |
| IG | 0.929 | 0.950 | 0.939 | 11.2 |
| MI | 0.914 | 0.945 | 0.929 | 13.7 |
| Wrapper | 0.920 | 0.949 | 0.934 | 17.1 |
| Average | 0.932 | 0.953 | 0.942 | 11.3 |
| Without FS | 0.905 | 0.944 | 0.924 | 3.11 |



Figure 10. Naive bayes results with pre-processing

3.2.3. Random forest (RF)

Table 9 shows random forest classifier results with pre-processing using several feature selection algorithms. The results showed that in the case of Precision was the highest result of classifier when used with ICHI square as a feature selection method. Moreover, that in the case of Recall was the highest results of the classifier when used with ICHI square as a feature selection method, but in the case of F-measure was the highest results of the classifier when used with ICHI square as a feature selection method and in the case of Time was the less time to build model of the classifier when used with Without FS as a feature selection method. Figure 11 shows diagram Random Forest results with pre-processing.

Table 9. Random forest results with pre-processing

| Algorithm | Precision | Recall | F-measure | Time |
|------------|-----------|--------|-----------|------|
| ICHI | 0.958 | 0.868 | 0.911 | 70.1 |
| CHI | 0.917 | 0.796 | 0.852 | 91.2 |
| IG | 0.910 | 0.789 | 0.845 | 91.3 |
| MI | 0.851 | 0.771 | 0.809 | 84.1 |
| Wrapper | 0.891 | 0.781 | 0.832 | 71.9 |
| Average | 0.905 | 0.801 | 0.849 | 81.7 |
| Without FS | 0.849 | 0.765 | 0.804 | 61.3 |

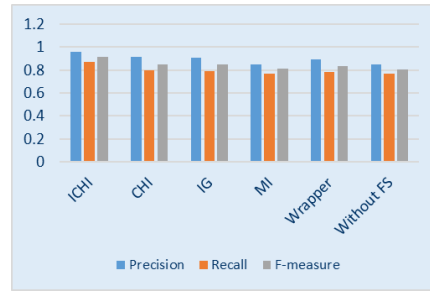


Figure 11. Random forest results with pre-processing

3.2.4. Decision tree (DT)

Table 10 shows decision tree classifier results with pre-processing using several feature selection algorithms. The results showed that in the case of Precision was the highest result of classifier when used with ICHI square as a feature selection method. Moreover, that in the case of Recall was the highest results of the classifier when used with ICHI square as a feature selection method, but in the case of F-measure was the highest results of the classifier when used with ICHI square as a feature selection method and in the case of Time was the less time to build model of the classifier when used with Without FS as a feature selection method. Figure 12 shows diagram Decision Tree results with pre-processing.

Table 10. Decision tree results with pre-processing

| Algorithm | Precision | Recall | F-measure | Time |
|------------|-----------|--------|-----------|------|
| ICHI | 0.577 | 0.539 | 0.557 | 3.1 |
| CHI | 0.523 | 0.502 | 0.512 | 3.34 |
| IG | 0.521 | 0.503 | 0.512 | 4.1 |
| MI | 0.515 | 0.500 | 0.507 | 5.4 |
| Wrapper | 0.519 | 0.502 | 0.510 | 8.2 |
| Average | 0.531 | 0.509 | 0.519 | 4.82 |
| Without FS | 0.516 | 0.503 | 0.509 | 1.05 |

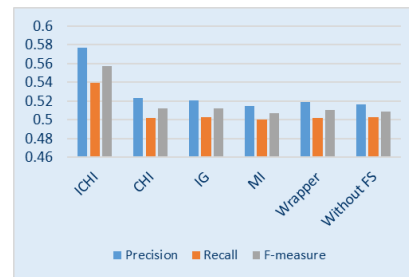


Figure 12. Decision tree results with pre-processing

3.2.5. Artificial neural networks (ANNs)

Table 11 shows artificial neural networks classifier results with pre-processing using several feature selection algorithms. The results showed that in the case of Precision was the highest result of classifier when used with ICHI square as a feature selection method. Moreover, that in the case of Recall was the highest results of the classifier when used with ICHI square as a feature selection method, but in the case of F-measure was the highest results of the classifier when used with ICHI square as a feature selection method and in the case of Time was the less time to build model of the classifier when used with Mutual Information as a feature selection method. Figure 13 shows diagram ANNs results with pre-processing.

Table 11. ANNs results with pre-processing

| Algorithm | Precision | Recall | F-measure | Time |
|------------|-----------|--------|-----------|------|
| ICHI | 0.719 | 0.560 | 0.629 | 51.2 |
| CHI | 0.672 | 0.467 | 0.551 | 39.1 |
| IG | 0.666 | 0.559 | 0.607 | 29.1 |
| MI | 0.599 | 0.501 | 0.545 | 21.4 |
| Wrapper | 0.651 | 0.541 | 0.591 | 39.2 |
| Average | 0.661 | 0.525 | 0.585 | 35.9 |
| Without FS | 0.580 | 0.506 | 0.540 | 39.3 |

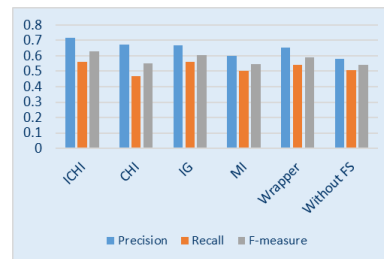


Figure13. ANNs results with pre-processing

3.3. Discussion the results without pre-processing

3.3.1. Results based on avg. precision

Table 12 and Figure 14 shows results based on avg. precision without pre-processing. We found that when we compare the state of Precision and also without pre-processing, the best results are as follows:

- Bayes Net Classifier when used with ICHI square as a feature selection method.
- Naïve Bayes Classifier when used with ICHI square as a feature selection method.
- Random Forest Classifier when used with ICHI square as a feature selection method.
- Decision Tree Classifier when used with ICHI square as a feature selection method.
- Artificial Neural Networks Classifier when used with ICHI square as a feature selection method.

Table 12. Results based on avg. precision without pre-processing

| Algorithm | ICHI | CHI | IG | MI | Wrapper | Without FS |
|-----------|-------|-------|-------|-------|---------|------------|
| BN | 0.976 | 0.858 | 0.851 | 0.819 | 0.832 | 0.805 |
| NB | 0.976 | 0.926 | 0.922 | 0.913 | 0.919 | 0.913 |
| RF | 0.947 | 0.938 | 0.931 | 0.917 | 0.929 | 0.915 |
| DT | 0.604 | 0.552 | 0.547 | 0.529 | 0.539 | 0.528 |
| ANNs | 0.713 | 0.699 | 0.635 | 0.590 | 0.601 | 0.586 |

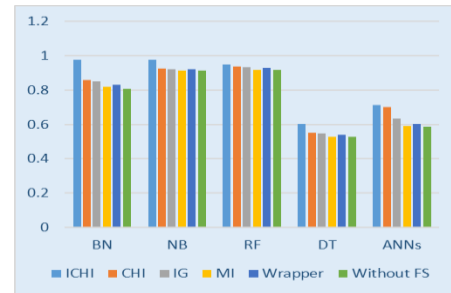


Figure 14. Results based on avg. precision without pre-processing

3.3.2. Results based on avg. recall

Table 13 and Figure 15 shows results based on avg. recall without pre-processing. We found that when we compare the state of Recall and also without pre-processing, the best results are as follows:

- Bayes Net Classifier when used with ICHI square as a feature selection method.
- Naïve Bayes Classifier when used with ICHI square as a feature selection method.
- Random Forest Classifier when used with ICHI square as a feature selection method.
- Decision Tree Classifier when used with CHI square as a feature selection method.
- Artificial Neural Networks Classifier when used with CHI square as a feature selection method.

Table 13. Results based on avg. recall without pre-processing

| Algorithm | ICHI | CHI | IG | MI | Wrapper | Without FS |
|-----------|-------|-------|-------|-------|---------|------------|
| BN | 0.940 | 0.910 | 0.899 | 0.873 | 0.888 | 0.871 |
| NB | 0.954 | 0.950 | 0.952 | 0.945 | 0.950 | 0.951 |
| RF | 0.942 | 0.805 | 0.899 | 0.849 | 0.851 | 0.841 |
| DT | 0.518 | 0.537 | 0.533 | 0.500 | 0.530 | 0.523 |
| ANNs | 0.492 | 0.542 | 0.523 | 0.460 | 0.499 | 0.523 |

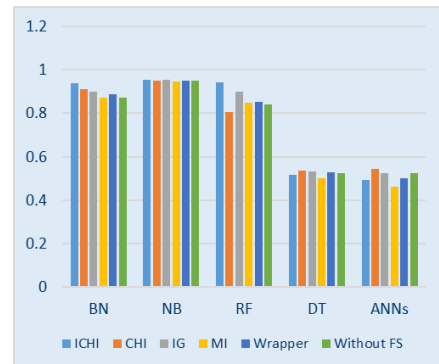


Figure 15. Results based on avg. recall without pre-processing

3.3.3. Results based on avg. f-measure

Table 14 and Figure 16 shows results based on avg. f-measure without pre-processing. We found that when we compare the state of F-measure and also without pre-processing, the best results are as follows:

- Bayes Net Classifier when used with ICHI square as a feature selection method.
- Naïve Bayes Classifier when used with ICHI square as a feature selection method.
- Random Forest Classifier when used with ICHI square as a feature selection method.
- Decision Tree Classifier when used with ICHI square as a feature selection method.

- Artificial Neural Networks Classifier when used with ICHI square as a feature selection method.

Table14. Results based on avg. F-measure without pre-processing

| Algorithm | ICHI | CHI | IG | MI | Wrapper | Without FS |
|-----------|-------|-------|-------|-------|---------|------------|
| BN | 0.957 | 0.883 | 0.874 | 0.845 | 0.859 | 0.836 |
| NB | 0.964 | 0.937 | 0.938 | 0.928 | 0.934 | 0.931 |
| RF | 0.944 | 0.866 | 0.914 | 0.881 | 0.888 | 0.878 |
| DT | 0.557 | 0.544 | 0.539 | 0.514 | 0.534 | 0.525 |
| ANNs | 0.582 | 0.611 | 0.573 | 0.516 | 0.545 | 0.510 |

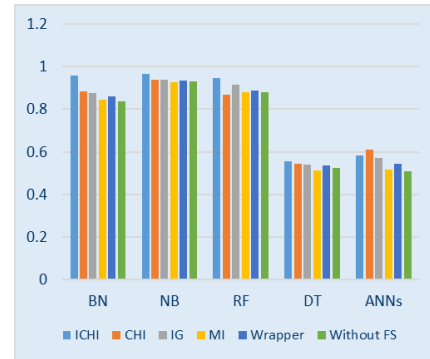


Figure 16. Results based on avg. f-measure without pre-processing

3.3.4. Results based on avg. time

Table 15 and Figure 17 shows results based on avg. time without pre-processing. We found that when we compare the state of Time and also without pre-processing, the best results with less time are as follows:

- Bayes Net Classifier when used with Information Gain as a feature selection method.
- Naïve Bayes Classifier when used without Feature Selection.
- Random Forest Classifier when used with Wrapper as a feature selection method.
- Decision Tree Classifier when used without Feature Selection.
- Artificial Neural Networks Classifier when used without Feature Selection.

Table15. Results based on avg. time without pre-processing

| Algorithm | ICHI | CHI | IG | MI | Wrapper | Without FS |
|-----------|-------|-------|------|------|---------|------------|
| BN | 25.6 | 18.20 | 13.3 | 22.5 | 20.7 | 16.7 |
| NB | 18.20 | 4.11 | 9.5 | 8.4 | 9.2 | 3.16 |
| RF | 94.3 | 66.5 | 69.1 | 70.3 | 66.2 | 71.48 |
| DT | 4.40 | 2.51 | 3.7 | 8.9 | 6.3 | 1.01 |
| ANNs | 44.17 | 43.7 | 50.1 | 59.9 | 49.1 | 36.34 |



Figure17. Results based on avg. time without pre-processing

3.4. Discussion the results with pre-processing

3.4.1. Results based on avg. precision

Table 16 and Figure 17 shows results based on avg. precision with pre-processing. We found that when we compare the state of Precision and also with pre-processing, the best results are as follows:

- Bayes Net Classifier when used with ICHI square as a feature selection method.
- Naïve Bayes Classifier when used with CHI square as a feature selection method.
- Random Forest Classifier when used with ICHI square as a feature selection method.
- Decision Tree Classifier when used with ICHI square as a feature selection method.
- Artificial Neural Networks Classifier when used with ICHI square as a feature selection method.

Table16. Results based on avg. precision with pre-processing

| Algorithm | ICHI | CHI | IG | MI | Wrapper | Without FS |
|-----------|-------|-------|-------|-------|---------|------------|
| BN | 0.883 | 0.851 | 0.844 | 0.790 | 0.839 | 0.828 |
| NB | 0.944 | 0.953 | 0.929 | 0.914 | 0.920 | 0.905 |
| RF | 0.958 | 0.917 | 0.910 | 0.851 | 0.891 | 0.905 |
| DT | 0.577 | 0.523 | 0.521 | 0.515 | 0.519 | 0.516 |
| ANNs | 0.719 | 0.672 | 0.666 | 0.599 | 0.651 | 0.580 |



Figure18. Results based on avg. precision with pre-processing

3.4.2. Results based on avg. recall

Table 17 and Figure 19 shows results based on avg. recall with pre-processing. We found that when we compare the state of Recall and also with pre-processing, the best results are as follows:

- Bayes Net Classifier when used with ICHI square as a feature selection method.
- Naïve Bayes Classifier when used with ICHI square as a feature selection method.
- Random Forest Classifier when used with ICHI square as a feature selection method.
- Decision Tree Classifier when used with ICHI square as a feature selection method.
- Artificial Neural Networks Classifier when used with ICHI square as a feature selection method.

Table 17. Results based on avg. recall with pre-processing

| Algorithm | ICHI | CHI | IG | MI | Wrapper | Without FS |
|-----------|-------|-------|-------|-------|---------|------------|
| BN | 0.922 | 0.888 | 0.891 | 0.801 | 0.870 | 0.893 |
| NB | 0.966 | 0.956 | 0.950 | 0.945 | 0.949 | 0.944 |
| RF | 0.868 | 0.796 | 0.789 | 0.771 | 0.781 | 0.765 |
| DT | 0.539 | 0.502 | 0.503 | 0.500 | 0.502 | 0.503 |
| ANNs | 0.560 | 0.467 | 0.559 | 0.501 | 0.541 | 0.506 |

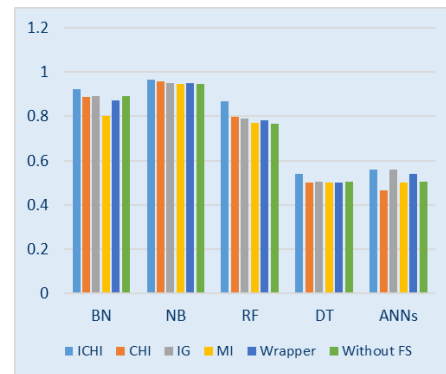


Figure19. Results based on avg. recall with pre-processing

3.4.3. Results based on avg. f-measure

Table 18 and Figure 20 shows results based on avg. f-measure with pre-processing. We found that when we compare the state of F-measure and also with pre-processing, the best results are as follows:

- Bayes Net Classifier when used with ICHI square as a feature selection method.
- Naïve Bayes Classifier when used with ICHI square and CHI square as a feature selection method.
- Random Forest Classifier when used with ICHI square as a feature selection method.
- Decision Tree Classifier when used with ICHI square as a feature selection method.
- Artificial Neural Networks Classifier when used with ICHI square as a feature selection method.

Table 18. Results based on avg. f-measure with pre-processing

| Algorithm | ICHI | CHI | IG | MI | Wrapper | Without FS |
|-----------|-------|-------|-------|-------|---------|------------|
| BN | 0.902 | 0.869 | 0.866 | 0.795 | 0.854 | 0.859 |
| NB | 0.954 | 0.954 | 0.939 | 0.929 | 0.934 | 0.924 |
| RF | 0.911 | 0.852 | 0.845 | 0.809 | 0.832 | 0.804 |
| DT | 0.557 | 0.512 | 0.520 | 0.507 | 0.510 | 0.509 |
| ANNs | 0.629 | 0.551 | 0.607 | 0.545 | 0.591 | 0.540 |

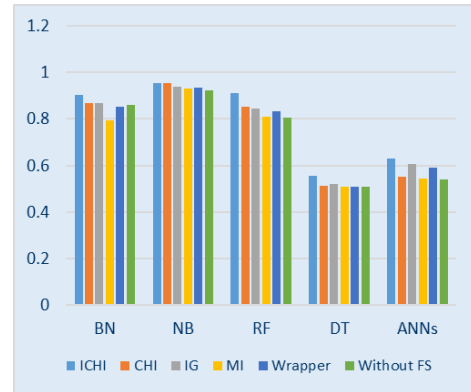


Figure 20. Results based on avg. f-measure with pre-processing

3.4.4. Results based on avg. time

Table 19 and Figure 21 shows results based on avg. time with pre-processing. We found that when we compare the state of Time and also with pre-processing, the best results with less time are as follows:

- Bayes Net Classifier when used with Wrapper as a feature selection method.
- Naïve Bayes Classifier when used without Feature Selection.
- Random Forest Classifier when used without Feature Selection.
- Decision Tree Classifier when used without Feature Selection.
- Artificial Neural Networks Classifier when used with Mutual Information as a feature selection method.

Table 19. Results based on avg. time with pre-processing

| Algorithm | CHI | ICHI | IG | MI | Wrapper | Without FS |
|-----------|------|------|------|------|---------|------------|
| BN | 20.3 | 22.5 | 27.1 | 21.1 | 17.2 | 21.6 |
| NB | 5.5 | 9.11 | 11.2 | 13.7 | 17.1 | 3.11 |
| RF | 70.1 | 91.2 | 91.3 | 84.1 | 71.9 | 61.3 |
| DT | 3.1 | 3.34 | 4.1 | 5.4 | 8.2 | 1.05 |
| ANNs | 51.2 | 39.1 | 29.1 | 21.4 | 39.2 | 39.3 |

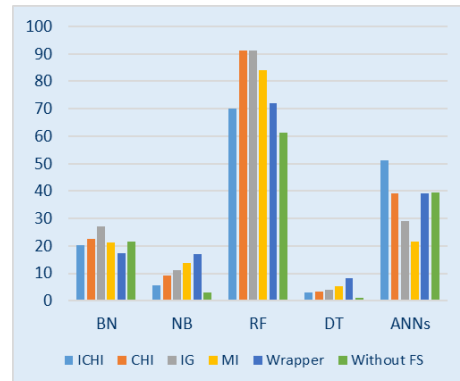


Figure 21. Results based on avg. time with pre-processing

4. CONCLUSION

Several feature methods were tested: ICHI square, CHI square, Information Gain, Mutual Information and Wrapper. By testing them with five of the classification algorithms: Bayes Net, Naïve Bayes, Random Forest, Decision Tree and Artificial Neural Networks. The testing process was done without pre-processing and with used pre-processing. The results were compared with four performance measure: Precision, Recall, F-measure and Time to build model. We found that the best feature selection method when testing without using pre-processing, in the case of Precision was the best result for ICHI square, and so is the case for Recall and F-measure. While the time needed to build model was the best result without used feature selection. In Addition, we found that the best feature selection method when testing with using pre-processing, in the case of Precision was the best result for ICHI square, and so is the case for Recall and F-measure. While the time needed to build model was the best result without used feature selection. We conclude from the results that the ICHI square as a feature selection method for use with the Arabic language is superior to the feature selection methods, which were tested together in the same environment.

REFERENCES

- [1] MZH, J., Hossen, A., Hossen, J., Raja, J. E., Thangavel, B., & Sayeed, S., "AUTO-CDD: automatic cleaning dirty data using machine learning techniques," *Telkommnika*, vol. 17, no. 4, pp. 2076-2086, 2019.
- [2] Al Sbou, A. M., "A survey of arabic text classification models," *International Journal of Electrical & Computer Engineering*, vol. 8, no. 6, pp. 4352-4355, 2018.
- [3] Abualigah, L. M. Q., *Feature selection and enhanced krill herd algorithm for text document clustering*. Berlin: Springer, pp. 1-165, 2019.
- [4] A. Souiri. A, Al Achhab. M, Elmohajir. B. E, Zbakh. A., "Neural network dealing with Arabic language," *International Journal of Informatics and Communication Technology*, vol. 9, no. 2, pp. 73-82, 2020.
- [5] Thilahar, R., & Sivaramakrishnan, R., "Fuzzy neuro-genetic approach for feature selection and image classification in augmented reality systems," *International Journal of Robotics and Automation (IJRA)*, vol. 8, no. 3, pp. 194-204, 2019.
- [6] Abualigah, L. M., & Khader, A. T., "Unsupervised text feature selection technique based on hybrid particle swarm optimization algorithm with genetic operators for the text clustering," *The Journal of Supercomputing*, vol. 73, no. 11, pp. 4773-4795, 2017.
- [7] Sarkar, A., "Multilayer neural network synchronized secured session key based encryption in wireless communication," *International Journal of Artificial Intelligence*, vol. 8, no. 1, pp. 44-53, 2019.
- [8] Venkatesh, R. K., "Legal Documents Clustering and Summarization using Hierarchical Latent Dirichlet Allocation," *International Journal of Artificial Intelligence (IJ-AI)*, vol. 2, no. 1, 2019.
- [9] Abualigah, L. M., Khader, A. T., & Hanandeh, E. S., "Hybrid clustering analysis using improved krill herd algorithm," *Applied Intelligence*, vol. 48, no. 11, pp. 4047-4071, 2018.
- [10] Abualigah, L. M., Khader, A. T., & Hanandeh, E. S., "A combination of objective functions and hybrid Krill herd algorithm for text document clustering analysis," *Engineering Applications of Artificial Intelligence*, vol. 73, pp. 111-125, 2018.
- [11] Aliwy, A., "Tokenization as Preprocessing for Arabic Tagging System," *International Journal of Information and Education Technology*, vol. 2, no. 4, pp. 348, 2012
- [12] Alshaer, H., Alzawahrah, B. & Otair, M., "Arabic text classification using bayes classifiers," *International Journal of Information Systems and Computer Sciences*, 2017.
- [13] Ayedh, A., *et al*, "The effect of preprocessing on Arabic document categorization," *Algorithms*, vol. 9, no. 2, pp. 27, 2016.
- [14] Otair, M., "Comparative Analysis of Arabic Stemming Algorithms," *International Journal of Managing Information Technology (IJMIT)*, vol. 5, no. 2, pp. 1-13, 2013.
- [15] Mesleh, A., "Chi Square Feature Extraction Based Svms Arabic Language Text Categorization System," *Journal of Computer Science*, vol. 3, no. 6, pp. 430-435, 2007.
- [16] Zheng, W., Wang, H., & Wu, D., "An improved CHI feature selection method for Chinese text classification," *Rev. T c. Ing. Univ. Zulia*, vol. 39, no 5, 117-122, 2016.
- [17] Alshaer, H., "The Effect of Improved CHI Square (As a Feature Selection Method) on Arabic Text Classifiers," *Thesis of Master's Degree in Computer Science*, 2018.
- [18] Mesleh, A., "Feature sub-set selection metrics for Arabic text classification," *A Journal of Pattern Recognition Letters*, vol. 32, no 14, pp. 1922-1929, 2011.
- [19] Sebastiani, F., "Machine learning in automated text categorization," *ACM Comput. Surv.*, vol. 34, no. 1, pp. 1-47, 2002.
- [20] Aboud, N., & Benhilima, L., "Review on wrapper feature selection approaches," *International Conference on Engineering & MIS (ICEMIS)*, pp. 1-5, 2016
- [21] Alshaer, H., Alzawahrah, B. & Otair, M., "Arabic Text Classification Using Bayes Classifiers," *International Journal of Information Systems and Computer Sciences*, 2017.
- [22] Zakariah, M., "Classification of large datasets using Random Forest Algorithm in various applications: Survey," *International Journal of Engineering and Innovative Technology (IJEIT)*, vol. 4, no. 3, 2014.
- [23] Saad, M., & Ashour, W., "Arabic text classification using decision trees," *12th international workshop on computer science and information technologies*, 2010.
- [24] Bahassine, S., Madani, A., & Kissi, M., "An improved Chi-sqaure feature selection for Arabic text classification using decision tree," *11th International Conference on Intelligent Systems*, pp. 1-5, 2016.
- [25] Saravanan, K., & Sasithra, S., "Review on Classification based on Artificial Neural Networks," *International Journal of Ambient Systems and Applications (IJASA)*, vol 2, pp. 11-18, 2014.
- [26] Parekh, R., Yang, J., & Honavar, V., "Constructive Neural-Network Learning Algorithms for Pattern Classification," *IEEE Transactions on Neural Networks*, vol. 11, no. 2, pp. 436-451, 2000.
- [27] Abualigah, L. M., Khader, A. T., Hanandeh, E. S., & Gandomi, A. H., "A novel hybridization strategy for krill herd algorithm applied to clustering techniques," *Applied Soft Computing*, vol. 60, pp. 423-435, 2017.
- [28] Abualigah, L. M., Khader, A. T., & Hanandeh, E. S., "A novel weighting scheme applied to improve the text document clustering techniques," In *Innovative Computing, Optimization and Its Applications* Springer, Cham, pp. 305-320, 2018.