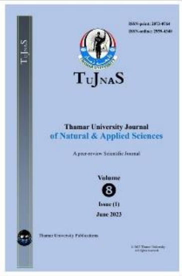
Thamar University Journal of Natural & Applied Sciences **8**(1) (2023) 26 – 38

TuJnaS
ISSN-Print: 2073-0764
ISSN-Online: 2959-4340

Thamar University Journal of Natural & Applied Sciences (TuJnaS)

Journal website:
www.tu.edu.ye/journals/index.php/TUJNAS/index



ORIGINAL ARTICLE

Hybrid Filter-Genetic Feature Selection Method For Arabic Sentiment Analysis

Muneer A.S. Hazaa[1]* and Saleh Ahmed Ali Hussein Salah[2]*

Affliations:
[1] Faculty of Computer And Information Systems ,Thamar University, Dhamar 87246, Yemen

[2] Yemen Academy For Graduate Studies Sana'a, Yemen

Corresponding Authors:
Muneer A.S. Hazaa, email: muneer_hazaa@yahoo.com and Saleh A. A. H. Salah, email: salehhazzeb@gmail.com

Received: Mar 12, 2023,

Accepted Date: May 16, 2023,

Online Date: Jun 12, 2023

Published: Jun 13, 2023

DOI:
https://doi.org/10.59167/tujnas.v8i1.1487

Abstract

The dramatic increase in user comments describing their feelings about products, services, and events brings sentiment analysis to the forefront as a way to monitor public opinion about products and events. Feature selection is an important subtask of sentiment analysis, which aims to improve the performance of learning algorithms and reduce the dimensionality of a problem. Feature selection is an important subtask of sentiment analysis, as it can improve the performance of learning algorithms while reducing the dimensionality of a problem. Moreover, the high-dimensional feature spaces caused by the morphological richness of Arabic motivate further research in this area. In this paper, a hybrid filter-based and genetic feature selection algorithm is proposed using four machine learning algorithms, namely decision tree, Naive-Bayes, K-NN and meta-ensemble methods. The performance of the proposed algorithm is compared with the performance of baseline models. A wide range of experiments are conducted on two standard Arabic datasets. The experimental results clearly show that the improved methods outperform the other baseline models for Arabic sentiment analysis. The results show that the improved models outperform traditional approaches in terms of classification accuracy, with a 5% increase in the macro average of F1.

Keywords

Machin Learning; Sentiment Analysis; Opinion Mining; Feature Selection; Arabic

1. Introduction

Given the large amount of online opinions and reviews provided by social media users about strategies, services, and products, understanding the extremely important information in social media content is of great value to many interested groups such as customers, business owners, and stakeholders. People use social media for a variety of purposes, including expressing their views on products and policies, leading various parties such as customers, businesses, and governments to begin analyzing those opinions. In fact, customer opinions and ratings play an important role in decision-making processes. In particular, decision makers take into account the experiences of their peers when making decisions, which consume valuable resources such as time and/or money. Identifying and analyzing customer opinions in an efficient and correct way to understand both current and potential customer needs has become a critical challenge for market-oriented product design.

Sentiment analysis is about identifying and analyzing explicit or implicit feelings and emotions expressed in posts on social media [1, 2]. Sentiment analysis is an important research area in the field of natural language processing (NLP). Sentiment analysis is a special type of text classification in which the general sentiment expressed in a review is classified into either positive or negative classes. As with any classification task, there are various methods and approaches for sentiment classification and opinion mining. Most of these methods and approaches can be divided into two main methods: supervised [3-7] and unsupervised learning approaches [1, 8, 9]. In supervised machine learning, sentiment corpora are used to train classifiers. Unsupervised approaches, also known as lexicon-based approaches, estimate the sentiment polarity of a text based on the subjective alignment of words or phrases [1, 8].

As part of sentiment analysis, feature selection is an important task that can significantly improve the performance of sentiment analysis [3, 5, 10, 11]. Feature selection is an important step to shorten the processing time and improve the analysis accuracy by reducing the feature sizes. In general, feature selection methods can be divided into two types: statistical methods and meta-heuristic methods. Statistical methods usually consist of two steps, feature evaluation and selection of the best subset. Meta-heuristics based feature selection methods are divided into four steps: Generation of initial subsets, evaluation of subsets, generation of next iteration, stopping criteria, and verification of results. Statistical methods are time-saving and provide faster results, while meta-heuristic feature selection is power-efficient as it can significantly improve accuracy. To combine the strengths of these approaches, a filter-based guided meta-heuristic feature-based method is introduced in this paper to improve the time and performance efficiency of the meta-heuristic method. The key ideas are based on reducing the search space for meta-heuristic feature selection. In the new paper, we propose two layers of feature selection methods that combine a genetic algorithm with traditional feature selection methods. In addition, this paper evaluates four machine learning algorithms - decision tree, Naive-Bayes, K-NN, and meta-ensemble - with the proposed feature selection method.

The remainder of this paper is organized as follows. Section 2 presents related work, while Section 3 describes the methodology as well as the classification modules and feature selection methods. In Section 4, we present the experimental setup. Section 5 discusses the experimental results. Finally, we conclude our work and discuss future research directions in section 6.

2. Related Work

This section provides a review of techniques proposed for analysis sentiment and Arabic subjectivity including feature selection and extraction and sentiment analysis. Abbasi, *et al.* [12] proposed a system for sentiment analysis task in a multi-language web forum at the document level. The system depends on an Entropy-Weighted Genetic Algorithm (EWGA) to choose the best features, and the Support Vectore Machine (SVM) with the linear kernel for the sentiment classification. Their method tries to find an over- lap between language-independent features, including syntactic and stylistic features and This research applied developed the Entropy Weighted Genetic Algorithm (EWGA) for efficient feature selection in order to improve accuracy and identify key features for each sentiment class. This research study applied developed the Entropy Weighted Genetic Algorithm (EWGA) for efficient feature selection in order to improve accuracy and identify key features for each sentiment class.

Another group of researchers Rushdi-Saleh, *et al.*, [13] focused on investigating two machine learning (ML) classifiers, Naive Bayes and Support Vector Machine(SVM), with two different weighting schemes (term frequency and term frequency-inverse document frequency) and three n-gram models. The effect of using the stem of the Arabic work has also been studied with different n-gramme models in Arabic and English corpora. The Arabic corpora consist of 500 reviews (250 positive and 250 negative). The results show that SVM with the tri-gramme model and without stemming achieved an F-value of 90% in the Arabic corpus and 86.9% in the English corpus.

Al-Moslmi *et al*., [5] and Omar *et al*., [3, 4] present detailed study concerning the effect of the feature selection technique on Arabic and Malay sentiment analysis. They present an empirical comparison of seven feature selection methods (Information Gain, Principal Components Analysis, ReliefF, Gini Index, Uncertainty, Chi-squared, and Support Vector

Machines), and three classifiers (SVM, Naïve Bayes, and K-nearest neighbor).

In [14] present two different hierarchical classifiers and compared their accuracy with the flat classifiers using three different classifiers namely Support Vector Machines (SVM), Naïve Bayes, (NB), and K-nearest neighbor (KNN) . The results showed that, in general, hierarchical classifiers gave significant improvements over flat classifiers.

3. Methodology

In this work, we create a unified framework that includes all the tasks required for sentiment analysis. This modular method allows us to study different approaches to Arabic sentiment analysis, focusing on feature selection. The proposed framework consists of different phases ranging from preprocessing, data representation, feature selection, and sentiment analysis, as shown in Figure 1. The main goal of this work is to develop two layers of feature selection methods that combine a genetic algorithm with traditional feature selection methods. In addition, these models use an ensemble machine learning method where a meta-classifier is used to combine the results of the basic machine learning method.

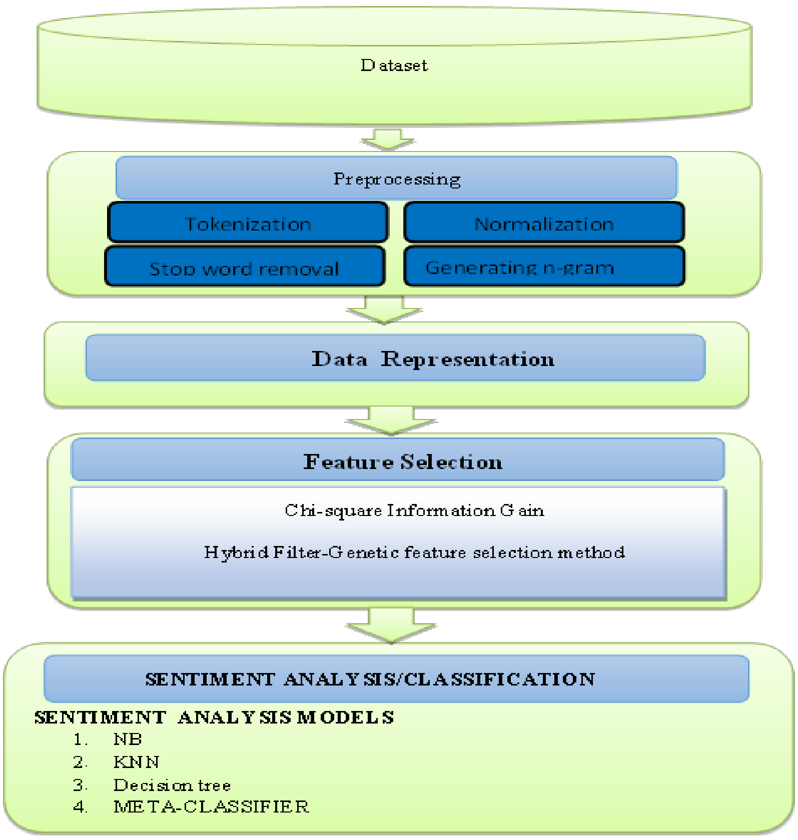


Figure 1. The framework of the implementation of the Arabic sentiment analysis/classification models.

A. *Pre-Processing*

Reviews usually include noisy data, recurring characters, spelling errors, and Html codes. Therefore, it is crucial to pre-process the data before analysis them using machine learning approaches. In the pre-processing phase, several NLP techniques were applied. The pre-processing here consists of four steps, namely: 1) tokenization; 2) normalization; and 3) stop word removal 4) Generating N-Gram Elements. In normalization, the raw data collected from social media is not 100% pure, so it contains a lot of noise. In this step, first, any non-Arabic words or characters that may belong to HTML, links, or the programming language code have been removed from the text. In tokenization, each review is spilt into sentences and words. It converted to a bag of words representation. Tokenization depends on the use of punctuation marks and white spaces in delimiting word boundaries (or main tokens). All texts including reviews contain many stop words such as pronouns, prepositions, and conjunctions. These words are removed as they do not provide worthless information about class or cluster of any text. in generating n-gram elements, each review is represented as a bag of both bigrams and unigrams which are useful to train a sentiment classifier (Bollegala et al. 2013). Based on that, a review was modeled as a bag of words to obtain unigrams and bigrams from each sentence.

B. *Data Reprsentation Phase*

Text representation focuses on how to convert pre-processed text to build the feature vectors and how to assign weights to the elements of the vector. In this paper, dataset is represented as a matrix (table) in which the columns represent the n-grams unigram and bigram in the reviews and the values represent their frequencies in a review. Each row is used to represent a review. Therefore, each review $\mathrm{R_i}$ is represented as$\mathrm{R_i} = (\mathrm{a_{i1}} \mathrm{a_{i2}}, \dots\dots, \mathrm{a_{im}})$, where $\mathrm{a_{ij}}$ is the frequency of n-gram $\mathrm{g_j}$in the review $\mathrm{R_i}$. This value can be calculated in various ways.

C. *Feature Selection Phase*

The size of features generated after the preprocessing phase in a text mining task is relatively huge (i.e., the curse of dimensionality). The process of selecting discriminating independent features is critical to any text mining that hopes to be effective in classification. Most of these features are not informative or cannot provide high discrimination. In general, several thousands of features are obtained, only an extremely small percentage of these features convey valuable information towards sentiment analysis objective. Subsequently, we typically need an algorithm that compresses the obtained feature vector and reduces its dimension. This work proposes two layers feature selection that uses filter-based methods (information gain and chi-squared) and genetic algorithm as feature reduction techniques.

Layer 1: Scoring Features Using Filtering Based Methods:

The first layer uses filter-based feature selection methods to score individual features according to their discriminative ability, i.e., their capability of separating the classes. To this end, A feature with a high-ranking value indicates higher discrimination of this feature compared to other categories and means that the feature contains information potentially useful for classification. Based on the ranking value they obtain; features are then returned in an ordered list where they appear in descending order of relevance. A subset of top n ranked features is selected. This work uses the following most widely used feature selection which proves to be efficient for sentiment analysis:

Information Gain.

The information gain (IG) method is used in the ranking and selecting the most relevant features. Information gain (IG) measures the relevance of a feature to a class. Information gain (IG) is a very popular method in feature selection (FS). It is used as a measure for feature goodness in the area of machine learning (ML). IG calculates the information amount that is present in or absent from a feature. The value obtained in the calculation of IG for each attribute is useful in determining the correct classification of any class.

$$\mathrm{IG}(t) = -\sum_{i=1}^{|c|} p(c_i)\log p(c_i) + p(t)\sum_{i=1}^{|c|} p(c_i|t)\log p(c_i|t) + p(\bar{t})\sum_{i=1}^{|c|} p(c_i|\bar{t})\log p(c_i|\bar{t}) \tag{1}$$

Where $p(c_i)$denotes the probability that class $c_i$occurs; $p(t)$ denotes the probability that term t occurs, and $p(\bar{t})$ denotes the probability that word $\bar{t}$ does not occur.

Chi-Squared Statistic (χ^2): χ^2 is one of the most widely used filter-based feature selection algorithms. The $\chi 2$ value for each feature t in class c is calculated by the following equation:

$$\chi^2(c,t) = \frac{N \times (\text{AD-BC})}{(A+C)(B+C)(A+B)(C+D)} \tag{2}$$

$$\chi_{\max}{}^2(t) = \max_i (\chi^2(t, c_i)) \tag{3}$$

Where N is the total count of training reviews while A is the number of Arabic reviews have class c and contains features t, and B is the number of Arabic reviews that do not belong to class c but contains feature t. Meanwhile, C is the number of Arabic reviews that do not belong to class c and do not contain feature t. D is the number of Arabic reviews that do not belong to class c and do not contain term t.

Layer 2: Selecting Final Feature Subset

A GA is a heuristic search algorithm that mimics the natural evolution process of man. Given the top n selected features from the filter-based method, the proposed hybrid meta-heuristic of GAs is applied on these selected features to determine the minimal subset of features, for which the different classes are best distinguished during sentiment analysis. GA has five important steps which include Generation of Initial Population, fitness evaluation, selection mechanisms, genetic operators and criteria to stop the GA. The following describes the main steps of the GA algorithm and how it combined with the Filter based method.

a). **Population initialization (selects n chromosomes randomly)**: The initial population here is an n chromosome, each Chromosome has Length m which consists of a vector of chromosomes (genotype), which is a Boolean vector indicating if a feature should be included or not. The Population Size is the number of chromosomes (individuals) in the population, while Chromosome Length (Genome Length) is the number of bits (genes) in each chromosome. The GA is iteratively performed over several generations and reproduction is performed to obtain individuals that are a best fit for a certain environment.

b). **Fitness computation**: Fitness is the key component of the genetic algorithm and is used to calculate how well a chromosome is suited to the environment and only chromosomes with high fitness will survive over time. First, the algorithm evaluates each chromosome Cu. The fitness of each chromosome in this work is evaluated using MI-based fitness function using Equation:

$$\mathrm{MI}\ (\mathrm{f}, \mathrm{c}) = \log \frac{\mathrm{p}(\mathrm{f} \wedge \mathrm{c}\,)}{\mathrm{p}(\mathrm{f}) \times \mathrm{p}\ (\mathrm{c})} \tag{4}$$

c). **Selection operation:** The algorithm then sorts the n chromosomes in the descending order of their fitness values. The algorithm then chooses the n/2 number of best chromosomes from the initial population of n chromosomes using the fitness function.

c) **Crossover operation**: The algorithm sorts the |n/2| chromosomes in descending order according to their fitness values. All chromosomes participate in the crossover operation pair by pair since for a crossover operation we need a pair of chromosomes. Then the algorithm applies the twin removal operation on the new population made of the offspring chromosomes.

d) **Mutation operation:** The basic idea of the mutation operation is to randomly change some of the chromosomes to explore different solutions. While adding random changes to the chromosomes, the algorithm uses a probabilistic approach where a chromosome with a low fitness has a high probability of getting a random change, and vice versa.

e) Stopping Criteria: After a certain number of iteration where steps a) to d) are repeated, the best chromosome in the last iteration is selected. The best chromosome consists of several selected features

D. Classification Phase

In this section, we have briefly described the classification approaches used in this work. These approaches have been used to classify the comments or reviews written in Arabic as positive and negative classes. The following subsection briefly describes the classification approaches used in this paper.

1) Decision Tree

A Decision tree is a supervised hierarchical machine learning model for inducing a decision tree from training data. The decision tree a predictive model which is a mapping from features of an item to classification about its target value. In the decision tree structures, leaves represent classifications, non-leaf nodes are features, and branches represent conjunctions of features that lead to the classifications The classes are represented by the leafs. In this work, the decision tree classifier J48 is used. J48 is a decision tree classifier in which an attribute is selected based on information gain from the training data to build each node of the tree. The selected attributes effectively split a set of training data into subsets enriched in one class or the other. It is mostly used because of its simplicity in explanation and interpretation. However, to find the best ordering features, all available features must be ranked. Therefore, entropy-based measure such as information gain based on the input training set S and a single feature F, with following equation:

$$\mathrm{InformationGain}(S, F) = \mathrm{Entropy}(S) - \mathrm{Average\ entropy}\ (S, F). \qquad (5)$$

The average entropy is defined by the following formula,

$$\mathrm{Average\ entropy}\ (S, F) = \sum_i \frac{|s_i|}{s}\ \mathrm{Entropy}(S_i) \qquad (6)$$

2) Naive Bayes (NB) Classifier

The Naive Bayes (NB) algorithm is a widely used algorithm for sentiment analysis. Given a feature vector table, the algorithm computes the posterior probability that the document belongs to different classes and assigns it to the class with the highest posterior probability. The major advantage of NB sentinement analysis algorithms is that they are easy to implement, often they have superior performance.

3) *K-Nearest Neighbour(K- NN)*

The K-nearest neighbor (KNN) is a typical example-based classifier. Given a test document d, the system finds the K-nearest neighbors among training documents. The similarity score of each nearest neighbor's document to the test document is used as the weight of the classes in the neighbor's document. The weighted sum in KNN categorization can be written as in Equation 3.10:

$$\text{score}(d, t_i) = \sum_{d_j=KNN(d)} sim(d, d_j)\, \delta(d_j, c_i) \qquad (7)$$

Where KNN *(d)* indicates the set of K- nearest neighbours of document *d.* If d_jBelongs toc_i, $\delta(d_j, c_i)$ equals 1, or other-wise 0. For test document$d$, it should belong to the class that has the highest resulting weighted sum.

4) *Stacking Classifier Combination : Meta-Classification*

The stacking classifier combination is an ensemble machine learning technique to combine multiple classification models via a meta-classifier. The base classification models are trained based on the complete or percentage of the training set; then, the meta-classifier is trained based on the outputs of the base classification models in the ensemble. The output for all base classifiers is considered as a new feature for meta-learning. The classification model used for this purpose is Naïve Bayes.The stacking combination involves two phases. The first phase involves the construction of a set of base-level classifiers (individual classifiers). The second phase involves the combination of the output of the base-level classifiers into a meta-level classifier. When a meta-classifier is used to combine the classifiers, the outputs of all the labels of the classes of the participating classifiers will be used as features for meta-learning.

hybrid approach

A hybrid approach that was crafted to improve the performance measures of sentiment analysis for the Arabic Language. This study focused on tweets sentiment classification for Egyptian dialect. Arabic is one of the widely used languages on the web [12]. Many researchers have worked on Arabic language sentiment analysis on different data sets with different tools and algorithms [13].

Following steps were carried out by the researcher for the implementation of the hybrid technique:

- Step 1: The features to be used by the machine learning approach are identified and separated.

- Step 2: The annotated corpus to be used for training and validation of the best classifier at different corpus sizes is built by the system.

- Step 3: Sentiment lexicon of different sizes is built using the annotated corpus

- Step 4: Theses different approaches are combined and tested for better and optimized results

- Step 5: Straight forward and simple method is crafted to detect negations in the hybrid approach

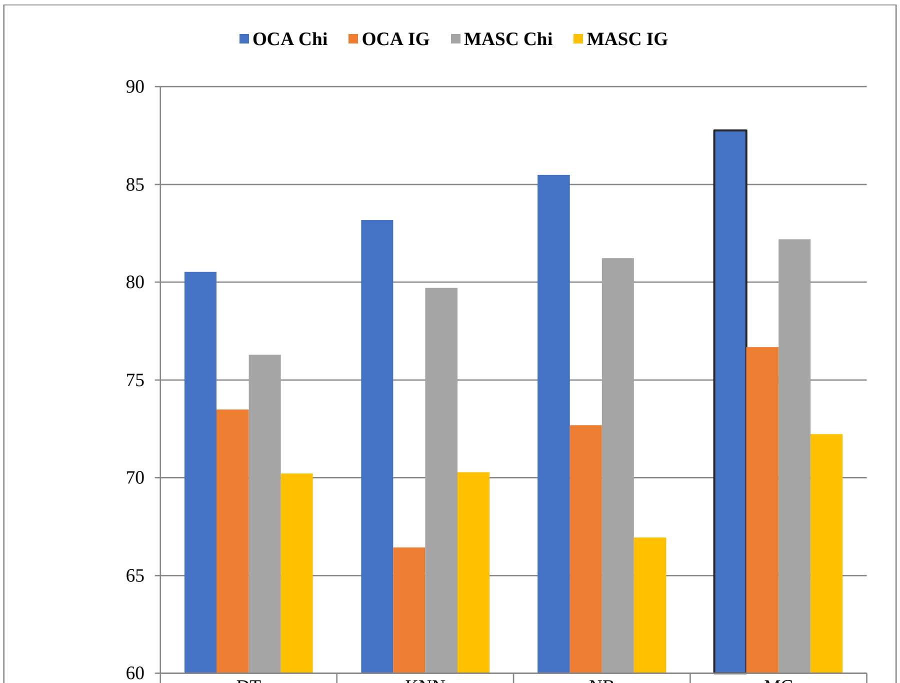
The results obtained by this study using hybrid approach showed better performance than other sentence-level classification systems.

4. Experimental Setting

Several experiments will be carried out to assess the proposed Arabic sentiment analysis models. First, many of experiments were carried out to measure the performance of the traditional machine learning models namely decision tree (DT), K-nearest neighbor classifier (KNN), Naive Bayes (NB) and meta-classification ensemble machine learning method with traditional feature selection methods namely information gain and chi-squared for Arabic opinion and sentiment analysis. Next, several experiments were performed to evaluate the proposed filter-based Directed Genetic optimized feature selection method with all classification models. All experiments are carried out using two datasets (1) opinion corpus for Arabic (OCA) [13]. The corpus comprises 500 reviews retrieved from a web page and blogs. From the 500 reviews in the OCA, 250 are listed as positive opinion reviews while 250 are negative opinion reviews. (2) Multi-domain Arabic Sentiment Corpus (MASC) [1]. The total number of reviews in the corpus is 8,860 reviews. The total positive reviews amount to 5408; while the total negative documents amount to 3453. The standard classification measurement precision, recall, and F-measure are used to assess the proposed model. Finally, all the experiments are performed on both corpora which are divided into 90% for training the proposed model, while 10% used for testing.

5. Results and Discussion

First, several experiments are conducted to evaluates the four baseline Arabic sentiment analysis models (decision tree (DT), K-nearest neighbor classifier (KNN), Naive Bayes (NB) and meta-classification ensemble method) along with Information gain and Chi-square. Figure 2, the performance (F-measure) the best results obtained by baseline Arabic sentiment analysis models with Information Gain (IG) and Chi-square (CHI) feature selection method on OCA and MASC. It can be observed that the enhanced meta-classification ensemble model (e-MT) outperforms other baseline classifiers with two feature selection methods. Furthermore, the meta-classification ensemble model achieved the best performance results with all datasets. The meta-classifier combines the strength of its individuals (base-classifiers). It expected when several individual classifiers agree on classifying most of the cases and only disagree with small cases (when one of them becomes wrong), then combining these classifiers always achieves higher results. Besides, combining the decision of several single classifiers which achieve a high result, better than individual classifier (base-classifier).



	DT	KNN	NB	MC
OCA Chi	80.52	83.18	85.49	87.76
OCA IG	73.48	66.43	72.68	76.68
MASC Chi	76.29	79.7	81.23	82.19
MASC IG	70.21	70.28	66.94	72.22

Figure 2. Performance (F-measure) the best results obtained by baseline Arabic sentiment analysis models with IG and CHI on OCA and MASC.

Second, several experiments are conducted to study the effect of the proposed two-layer filter-based Genetic (TF_GenFS) feature selection method on the four Arabic sentiment analysis models. Table 1 shows the results of four enhanced classification models on both opinion corpuses for Arabic (OCA) and Multi-domain Arabic Sentiment Corpus (MASC). The results obtained show that the results of all classification models are significantly improved with the proposed two-layer filter-based Genetic (TF_GenFS) feature selection method. It can also be observed that the enhanced meta-classification ensemble model (MT) with the proposed TF_GenFS feature selection outperforms other classification models. These findings reveal that the classifier combination method with the proposed TF_GenFS feature selection method is the most suitable technique for Arabic sentiment analysis.

Table 1: Performance (F-measure) of enhanced Arabic sentiment analysis models on OCA and MASC.

	OCA	**MASC**
DT+TF-GenFS	84.96	80.26
KNN+TF-GenFS	89.15	83.61
NB+TF-GenFS	90.43	85.49
MC+TF-GenFS	93.74	87.01

6. Conclusions

This paper empirically evaluates four machine learning methods namely decision tree (DT), K-nearest neighbor classifier (KNN), Naive Bayes (NB) and meta-classification ensemble method with information gain and chi-squared traditional feature selection methods for Arabic opinion and sentiment analysis task. In addition, this paper introduces enhanced Arabic opinion and sentiment analysis models based on a two-layer filter based Genetic feature selection method. This paper demonstrates that using the two-layer filter based Genetic feature selection method improve the performance of all four machine learning methods for Arabic sentiment classification. Experimental results demonstrate these findings reveal that the classifier combination method with the proposed feature selection method is the most suitable technique for Arabic sentiment analysis.

References

[1] Al-Moslmi, T., Albared, M., Al-Shabi, A., Omar, N., Abdullah, S. (2018) Arabic senti-lexicon: Constructing publicly available language resources for Arabic sentiment analysis, *Journal of information science* **44**: 345-362.

[2] Yue, L., Chen, W., Li, X., Zuo, W., Yin, M. (2019) A survey of sentiment analysis in social media, *Knowledge and Information Systems* 1-47.

[3] Omar, N., Albared, M., Al-Moslmi, T., Al-Shabi, A. (2014) A comparative study of feature selection and machine learning algorithms for Arabic sentiment classification, Asia information retrieval symposium, Springer, pp. 429-443.

[4] Omar, N., Albared, M., Al-Shabi, A.Q., Al-Moslmi, T. (2013) Ensemble of classification algorithms for subjectivity and sentiment analysis of Arabic customers' reviews, *International Journal of Advancements in Computing Technology* **5**: 77.

[5] Al-Moslmi, T., Gaber, S., Al-Shabi, A., Albared, M., Omar, N. (2015) Feature selection methods effects on machine learning approaches in malay sentiment analysis, Proc. 1st ICRIL-Int. Conf. Inno. Sci. Technol.(IICIST), pp. 1-2.

[6] Tuhin, R.A., Paul, B.K., Nawrine, F., Akter, M., Das, A.K. (2019) An Automated System of Sentiment Analysis from Bangla Text using Supervised Learning Techniques, 2019 IEEE 4th International Conference on Computer and Communication Systems (ICCCS), IEEE, pp. 360-364.

[7] Silva, N.F.F.D., Coletta, L.F., Hruschka, E.R. (2016) A survey and comparative study of tweet sentiment analysis via semi-supervised learning, *ACM Computing Surveys (CSUR)* **49**: 1-26.

[8] Khoo, C.S., Johnkhan, S.B. (2018) Lexicon-based sentiment analysis: Comparative evaluation of six sentiment lexicons, *Journal of Information Science* **44**: 491-511.

[9] Al-Saffar, A., Awang, S., Tao, H., Omar, N., Al-Saiagh, W., Al-bared, M. (2018) Malay sentiment analysis based on combined classification approaches and Senti-lexicon algorithm, *PloS one* **13**: e0194852.

[10] Madasu, A., Elango, S. (2020) Efficient feature selection techniques for sentiment analysis, *Multimedia Tools and Applications* **79**: 6313-6335.

[11] Gokalp, O., Tasci, E., Ugur, A. (2020) A novel wrapper feature selection algorithm based on iterated greedy metaheuristic for sentiment classification, *Expert Systems with Applications* **146**: 113176.

[12] Abbasi, A., Chen, H., Salem, A. (2008) Sentiment analysis in multiple languages: Feature selection for opinion classification in web forums, *ACM Transactions on Information Systems (TOIS)* **26**: 1-34.

[13] Rushdi-Saleh, M., Martín-Valdivia, M.T., Ureña-López, L.A., Perea-Ortega, J.M. (2011) OCA: Opinion corpus for Arabic, *Journal of the American Society for Information Science and Technology* **62**: 2045-2054.

[14] Al-Ayyoub, M., Nuseir, A., Kanaan, G., Al-Shalabi, R. (2016) Hierar-chical classifiers for multi-way sentiment analysis of arabic reviews, *International Journal of Advanced Computer Science and Applications (IJACSA)* **7**: 531-539.