Arabic Text Classification using Feature-Reduction Techniques for Detecting Violence on Social Media

Hissah ALSaif[1], Taghreed Alotaibi[2]
College of Computer and Information Sciences, Shaqra University,Riyadh, Saudi Arabia

***Abstract*—With the current increase in the number of online users, there has been a concomitant increase in the amount of data shared online. Techniques for discovering knowledge from these data can provide us with valuable information when it comes to detecting different problems, including violence. Violence is one of the significant problems humanity has faced in recent years all over the world, and this is especially a problem in Arabic countries. To address this issue, this research focuses on detecting violence-related tweets to help in solving this problem. Text mining is an important technique that can be used to find and predict information from text. In this study, a text classification model is built for detecting violence in Arabic dialects on Twitter using different feature-reduction approaches. The experiment comprises bagging, K-nearest neighbors (KNN), and Bayesian boosting using different extraction features, namely, root-based stemming, light stemming, and n-grams. In addition, the study used the following feature-reduction techniques: support vector machine (SVM), Chi-squared (CHI), the Gini index, correlation, rules, information gain (IG), deviation, symmetrical uncertainty, and the IG ratio. The experiment showed that the bagging with tri-gram approach has the highest accuracy at 86.61%, and a combination of IG with SVM from reduction features registers an accuracy of 90.59%.**

Keywords—Violence; text mining; classification; feature-reduction techniques; Arabic; Twitter posts

I. Introduction

One of the most important elements in living a normal and stable life is living in peace. Individuals can reach this level of peace when they are removed from any manifestations of violence and persecution, and this will enhance their ability to be successful members of the community. In recent years, the rate of violence has been increasing all over the world, and this is an indicator of danger that communities should be aware of to take the necessary steps to avoid or reduce it. There are different of types of acts that can be categorized as violence, including physical, sexual, and psychological violence.

Due to the evolution of technologies and the increasing number of internet users around the world, especially when it comes to using social networks, social media are becoming a significant environment for studying phenomena related to violence; this is because social network users publish events rapidly, and some of them use social media sites to voice complaints and ask for help. From this perspective, this present research focus on studying violence in the Kingdom of Saudi Arabia (KSA) regarding the increased number of violence cases, where the Ministry of Labor and Social Development received more than 11,000 reports in a year [1]. This research is concentrate on Twitter to collect dataset due to the rising number of Saudi Twitter users, there were about 2.4 million active Twitter users in 2013, representing 41% of individuals online [2] [3]. To study this usage, four regions were choosing in the KSA with the highest populations to collect tweets from Twitter users.

This research is using text mining, which an important topic regarding the knowledge it uncovers. Text mining, defined as a technique of extracting valuable information from text, consists of different processes, specifically, information retrieval, categorization, extraction, textual analysis, and visualization [4][5]. After information extraction, the biggest challenge faced in text classification is the huge number of features that must consider. To solve this problem, feature-reducing techniques are used for selecting the appropriate subset. Feature reduction is a critical step that can positively or negatively affect the accuracy of the classifier according to the selected approaches [6][7]. This study investigated different feature-reduction methods and their effects on accuracy. Moreover, it used machine learning techniques, where machine learning algorithms can be categorized into two approaches—supervised and unsupervised. In the supervised approach, the data are labeled and the algorithm works to predict the output; in contrast, in the unsupervised approach, there is no labeling process for data, and the algorithm works to build structures to understand the data [8]. This study used the following supervised machine learning algorithms: bagging based on support vector machine (SVM), K-nearest neighbors (KNN), and Bayesian boosting using complement naïve Bayes (CNB) with different stemming and n-gram techniques, in addition to different feature-reduction techniques.

This research has the following objectives:

- To detect violence and find the most common cities with high rates of violence in Saudi Arabia;
- To compare the stemming and n-gram techniques to find the best one for the Arabic language;
- To evaluate some reduction features to find the best one for Arabic; and
- To evaluate some ensemble methods of classification algorithms.

This paper is organized as follows: Section 2 gives a general perspective on the Arabic language, while Section 3 presents an overview of violence. Section 4 briefly presents similar studies on text mining and feature-reduction using Arabic. Section 5 illustrates the machine learning algorithms used in this study, and Section 6 describes the methodology of the work in detail. Section 7 shows the performance measurement used. Section 8 illustrates the results of the research, and finally, Section 9 presents the conclusion of the study.

II. Arabic

Arabic is one of the most common languages currently in use, with more than 315 million people around the world who speak this language. Moreover, Arabic is the official language of 17 Arabic countries [9], and it is used in other Islamic countries as a second language. Arabic language can be classified into three types, namely, the original version of Arabic, called Classical Arabic (CA); Modern Standard Arabic (MSA); and dialectal Arabic (DA). CA is the basis for Arabic, and it is the version used in the Holy Qur'an; In contrast, MSA is currently used in the media, education, and formal speech. Finally, because Arabic is spoken in different countries, this has allowed for the emergence of different Arabic dialects, which represent spoken rather than written language [10]. The dialects can be classified depending on the countries in which they are spoken; for example, the Gulf dialect is used in Saudi Arabia, Kuwait, Qatar, Bahrain, and the United Arab Emirates. Moreover, while there are different Arabic dialects in different countries, there are also multiple dialects spoken in the same country; for instance, in Saudi Arabia, there are speakers of the Najd, Hijaz, and Qassim dialects, which are used in informal daily communication.

Arabic consists of 28 letters, which are أ ب ت ث ج ح خ د ذ ر ز س ش ص د ض ظ ع غ ف ق ك ل م ن و ي, and there are additional vowels letters ا و ي . In contrast to many other languages, Arabic is written from right to left; all the letters are lower case, and there is no capitalization. There are some challenges in working with Arabic. For example, a single letter in Arabic can have different styles depending on its position in the word. The letter (صـ) is written in one way if it comes at the beginning of a word, such as (صباح) “morning,” (ـصـ), another way if it comes in the middle of the word, such as (العصفور) “bird,” and yet another way if it comes at the end of the word (ص), such as (خاص) “special” [10][11].

In Arabic, there are some diacritics, which called Harakat that can be used with words. These can change the meaning of the word; for example, مُعَلم means “teacher” and مَعْلَم means “landmark” [10][11]. In addition, Arabic words can be conjugated differently for feminine or masculine meanings. For numbers of persons, Arabic uses the three following cases: singular for one person, dual for two people, and plural for more than two people; this is shown in Table I.

Arabic contains many words that have more than one meaning depending on the context of the sentence, such as the examples shown in Table II. In addition, there are different synonyms for the same term, for example, "زلة" "هفوة", ،"سقطة" "كبوة"، "عثرة"" all mean doing something wrong [11].

TABLE I. Arabic Cases for Person Number

Verb	Singular	Dual	Plural
لعب (played)	لعب	لعبا	لعبوا

TABLE II. Examples of Different Word Meanings Depending on the Context of the Sentence

Meaning	Example	Word
Walking	" وَجَاءَتْ سَيَّاره"	سياره
Car	ركبت سيارة	

All Arabic words rely on a root, which can be defined as the main part for any words that cannot be removed without losing the meaning of the word. There are more than 11,000 roots in Arabic [12]. Roots can be connected to affixes; these can come at the beginning of the word (prefixes), end of the word (suffixes), or middle of the word (infixes). To illustrate this in details, see the example of the word (لعب) in Table III [13][10]. All these features of Arabic show the challenges researchers face when data mining information written in Arabic.

TABLE III. Affixes in Arabic Words

Suffix	Infix	Prefix	Word
-	-	م	ملعب (Stadium)
-	ا	أ	ألعاب (Toys)

- Violence

Violence is one of the biggest problems society faces today. The number of victims of violence is showing an increasing trend around the world. The World Health Organization showed that about 10 million persons lose their life as a result of violence in the age range of 15–44 years [14]. Violence can come in many forms: It can be defined as the premeditated use of force, either physical or emotional, toward the self, other individuals, or a group in such a way that causes harm, injury, or fatality or is likely to do so [14]. From this definition, it is clear that violence can be classified into different types as physical, sexual, or psychological violence [14]. Physical violence consists of any physical harm of people, such as causing injury, beating, or killing them. Sexual violence includes any forcible sexual acts against children, women, or men [14]. Finally, psychological violence includes any intentional action against a person to harm him or her psychologically by coercion or threats[15]. According to the Ministry of Justice, in the KSA, the number of domestic violence cases in 2014 were about 34.3% in Makkah and 23.3% in Riyadh, followed by 12.9% and 5.6% in the Eastern Region of KSA and Asir, respectively [16].

III. Related Work

Text classification is an excellent technique when it comes to discovering new knowledge from text. However, most related research has focused on text mining in English, and thus, there is still a need for more research connected to Arabic, especially DA. In addition, some research has tried to use extraction and selection features for improving the text classification performance. To our knowledge, the present study is first research on detecting violence in Saudi dialect, with the use to feature extraction and selection techniques.

In [17], the authors studied sexual violence on Twitter. The dataset consisted of 700,000 extracted tweets using the #MeToo hashtag in 2017. The study evaluated different algorithms, such as a Convolutional Neural Network (CNN) and Multilayer Perceptron (MLP), for classifying tweets based on the place and offender. The results showed that the CNN outperformed the other algorithms by 83%. In addition, Subramani et al. [18], extracted Facebook posts and comments as their experimental dataset. The researchers aimed to classify intent for domestic violence discourse on social media. The dataset contained 8,856 posts and 28,873 comments. Four models were applied, namely, SVM, Naïve Bayes (NB), Decision Tree (DT), and KNN with two reduction features: Linguistic Inquiry and Word Count (LIWC) and CHI. The results showed that the accuracy of NB was highest, at 82%, followed by KNN. In [19], the researchers detected religious hate speech on Twitter in Arabic. The data, which were collected using Twitter's Application Programming Interface (API), comprised 600 tweets as the training set and 6,000 tweets as the testing set. They used three different classification models, namely, a lexicon-based model, an N-gram-based model using logistic regression and a deep neural network using Gated Recurrent Units (GRU). For reduction features they used CHI, Pointwise Mutual Information (PMI) and Bi-Normal Separation (BNS). Tweets were classified using help from crowdsourcing workers. The results showed that the GRU-based system with a simple Recurrent Neural Network (RNN) provided the best results, with 79% accuracy.

In [20], a classifier was built to study six Arabic dialects, namely, Shami, Iraqi, Gulf, Sudanese, Moroccan, and Egyptian. The study used 2,000 tweets from Twitter as dataset, while the algorithms used were the NB, rule-based, and DT algorithms. Data collection was done using the Topsy website and Twitter API. As result, the accuracies for the classifiers were 71.18% for Ripper, 57.43% for DT, and 71.09% for NB, but the classifier had difficulty differentiating the Sudanese, Egyptian, and Gulf dialects.

In Raheel and Dichy's [21] research, they studied which types of feature extraction are best for Arabic languages. They chose Arabic articles covering seven categories as their dataset, which comprised 7,034 articles. They classified the dataset into five versions; in each version of the dataset they use one of the following types of feature extraction: 3-gram, 4-gram, root and lemma. In addition, they used 10-fold cross-validation, and for the algorithms, they used SVM and NB networks with two selection features—IG and CHI For measuring weight, they used term frequency–inverse document frequency (TF-IDF). In each test, they increased the number of features from 400 to 2,000. The results showed that SVM recorded higher accuracy values using the 3-gram approach, at 92.28% for IG and 92.41% for CHI.

In [22], the researchers implemented a method based on two stages for reducing the dimensionality of the features. First, they used IG for ranking the importance of features; then, they applied the Principal Component Analysis (PCA) and Genetic Algorithm techniques for reducing the feature number. The experiments were conducted on two well-known public datasets written in English with two models, namely the C4.5 DT and KNN classifier. The results revealed good improvements in classification accuracy. The experiments in study [23] were conducted on a public dataset consist of 937 reviews in English. The researchers evaluated four classification methods, including SVM, logistic regression, bagging, and Bayesian boosting with PCA as reduction techniques for decreasing the dimensionality of the dataset and improving the classification accuracy. The results revealed better performance in classification with PCA. In [24], the researchers performed text mining on an Arabic dataset from King Abdulaziz City for Science and Technology (KACST), covering five fields with about 2,243 texts from the Islamic topics of Feqh, Tafseer, Lughah, Aqeedah, and Hadeet. The researchers employed three schemas for representation—Boolean, LTC, and TF-IDF. The CHI and weirdness coefficients were applied as feature selection, with NB and KNN as classifiers. The study used a different number of features each time, starting from 10 and ending at 300. As a result, NB with CHI squared showed high accuracy, which was 89.25% with LTC. In contrast, the weirdness coefficient showed high accuracy with NB and the Boolean scheme, at 85.22%.

In [25], the researchers focused on comparing various types of NB approaches, such as multivariate guess Naïve Bayes (MGNB), Flexible Bayes (FB), and Multinomial Naïve Bayes (MNB). For feature extraction, different techniques were used, such as the light stemmer and term-based n-gram. For feature reduction, several approaches, such as CHI, Odds Ratio (OR), Mutual Information (MI), and the Galavotti, Sebatiani, Simi (GSS) coefficient, were employed. The results revealed that FB had the best performance among the classifiers.

In [26], the researchers built a large corpus for Arabic using different categories of topics from Saudi newspapers, forums, writers, the Saudi Press Agency, Topics in Islam, websites, and Arabic poems. The study experiment was performed using SVM, NB, KNN, MLP neural networks, C4.5, and C5.0. As a result, SVM registered a high accuracy. For feature selection, the researchers used CHI, GSS, the Darmstadt Indexing Approach (DIA), IG, MI, NG, Goh and Low (NGL) coefficient and Relevancy Score (RS), OR and none. The best result were showed by GSS, none, and RS. The researchers also studied the performance using the following weighting terms, TF-IDF, LTC, TFC, Boolean, relative frequency, and entropy. The LTC came first with the best result followed by Boolean and TFC.

For short text documents, Faqeeh et al. [27] implemented a classifier that studied 1,000 Facebook comments in Arabic and English. They chose and filtered comments related to weather and food manually, then applied four algorithms—SVM, NB, KNN, and DT. As a result, the accuracy for the Arabic dataset was higher than that for English, and the SVM classifier showed a better result than the other classifiers did. Moreover, in [28], the authors performed text mining over an Arabic dataset. The dataset contained 1,000 documents from five fields (health, politics, economics, technology, and sport). They used 5-fold cross-validation because of the limitation of their devices. For stemming, they applied a light stemmer and root-based, and for algorithms, they chose NB, SVM, KNN, DT, and decision table. They established three versions of the dataset—one with no stemming, one with a root-based stemmer, and one with a light stemmer separately. As a result, SVM with 10 light stemmers recorded the best accuracy, at 98.20%.

Another study categorized documents according to their contents into four classes, namely, sport, politics, economics, and arts [29]. The dataset consisted of more than 3,000 documents collected from different websites. The classification was evaluated by the KNN algorithm with four types of similarity measures—cosine, Jaccard, dice, and I_{new}—in terms of execution time and accuracy. The results indicated the superiority of I_{new} over the other similarity measures.

A further study considered the influence of Singular Value Decomposition (SVD) on a large dataset containing 4,000 documents in Arabic [30]. Experiments were applied using seven known algorithms, specifically, NB, KNN, Neural Networks, Cosine Similarity, Random Forest, SVM, and DT. The results showed an improvement in the classification process, where the accuracy of the classification increased from 67.25% to 82.50%.Abuhaiba et al. [31] performed a study comparing the performances between a single classification algorithm and a combination of different algorithms for categorizing Arabic text. The study used text documents from three news networks with different categories. In addition, they built four models for improving the performance, which were bagging, stacking, AdaBoost, and fixed combining rules. For stemming, they tested each model using light and root-based stemming, and for the weighting term, they used TF-IDF. For the last model, the researchers used the average, product, majority, minimum, and maximum rules. As a result, classifier with the majority voting rule registered higher accuracy, at 94.5%; this was superior to that of each single classifier, which were the radial basis function network (RBFN), Nearest-Neighbor-Like, NB, SVM, C4.5, KNN, and Decision Stump. In contrast, the stack model with five classifiers had even higher accuracy, at 99.5%, compared with each single classifier, which were NB, learning vector quantization (LVQ), C4.5, KNN, and decision stump. Then, the AdaBoost model with 10 iterations outperformed the C4.5 classifier, recording 99.5% accuracy. Finally, the bagging model with 10 iterations had 99.4% accuracy, which was higher than that of DT. Tables IV and V summarize the previews mentioned related works.

TABLE IV. COMPARISION OF SIMILAR STUDIES FEATURES EXTRACTION METHODS

Study	Dataset Source	Dataset Language	Features Extraction Methods
Raheel and Dichy [21]	Articles	Modern Arabic	N-gram, Root-based stemming.
Harun Uguz [22]	Articles	English	-
Al-Thubaity et al.[24]	Articles	Modern Arabic	-
Kadhim and Omar [25]	Articles	Modern Arabic	Light- stemming, N-gram
Khorsheed and Al-Thubaity. [26]	Newspapers	Modern Arabic	-
Vinodhini and Chndrasekarn[23]	Reviews	English	-
Faqeeh et al.[27]	Facebook comments	Arabic, English	Light stemming, Root-based stemming.
Hmeidi et al. [28]	Articles	Modern Arabic	Light -stemming, Root-based stemming
Alhutaish and Omar [29]	Articles	Modern Arabic	Light stemming, N-gram
Al-Anzi , Dia AbuZeina [30]	Articles	Modern Arabic	-
Abuhaiba and Dawoud [31]	Articles	Modern Arabic	Light stemming, Root-based stemming
Al-Walaie and Khan. [20]	Twitter post	Dialect Arabic	-
Albadi et al. [19]	Twitter post	Modern Arabic	N-gram.
Subramani et al. [18].	Facebook comments	English	-
Khatua et. Al. [17]	Twitter post	English	-

TABLE V. COMPARISION OF SIMILAR STUDIES FEATURES REDUCTION METHODS AND CLASSIFICATION ALGORITHMS

Study	Features Reduction Methods	Classification Algorithms
Raheel and Dichy [21]	IG, CHI	SVM, Naïve Bayesian Networks
Harun Uguz [22]	PCA, IG, Genetic algorithm	C4.5 DT, KNN
Al-Thubaity et al.[24]	CHI, Weirdness coefficient (W)	NB., KNN.
Kadhim and Omar[25]	CHI, OR, MI, GSS	MGNB, FB, MNB.
Khorsheed and Al-Thubaity. [26]	CHI, GSS, DIA, IG, MI, NGL, RS, OddsR,	SVM, NB, KNN, MLP, C4.5 DT, C5.0 DT.
Vinodhini and Chandrasekaran[23]	PCA	SVM, Logistic Regression, Bagging, Bayesian boosting
Faqeeh et al.[27]	-	SVM, NB, KNN ,DT
Hmeidi et al. [28]	-	NB. SVM, KNN, J84 DT, Decision Table
Alhutaish and Omar [29]	-	KNN
Al-Anzi , Dia AbuZeina [30]	SVD	NB, KNN, NN, Cosine similarity, Random Forest, SVM, DT
Abuhaiba and Dawoud [31]	-	NB, SVM, C4.5, KNN, Decision Stump, LVQ, RBFN.
Al-Walaie and Khan. [20]	-	NB, Ripper, DT
Albadi et al. [19]	CHI, PMI, BNS	Logistic regression, SVM.
Subramani et al. [18].	CHI, LIWC dimensions	NB, DT, KNN, SVM
Khatua et. Al. [17]	-	CNN ,MLP

IV. Machine Learning

Machine learning algorithms can be classified into the two following types: eager learners and lazy learners. Eager learners construct models using training datasets before they obtain new objects and then use the model to classify the new objects. In this type of learning, a single global hypothesis is constructed that encompasses the entire dataset. In contrast, lazy learners store data for later use, when there is a new object; they are also called instance-based learners. They compare each new instance of data with instances in the training sets that stored memory; this allows them to measure the similarity [32][33]. In this study, two ensemble methods are used to build models, including bagging and Bayesian boosting, as well as KNN, which is one of the most widely used classifiers. In this section, these algorithms are explained briefly.

A. K-Nearest Neighbors (KNN)

KNN is a simple algorithm that can be used to construct many local hypotheses for each new object in a dataset using the lazy learning approach. This approach assigns objects based on the majority of voting of its neighbors. KNN is slow in classifying data, requires a large amount of memory space, and is computationally expensive because a hypothesis must be constructed for each new object. There are many measurements used to determine the distance between objects in space. One popular measure in information retrieval is cosine similarity. The basic idea of cosine similarity is measuring the angle between two objects represented as vectors in the space. The following equation represent cosine similarity [34][35]:

$$cosine(x, y) = \frac{x.y}{||x||||y||} \tag{1}$$

B. Bagging

Bagging is a well-known ensemble method that creates multiple independent hypotheses, then calculates aggregated results that lead to better performance. Bagging is based on the bootstrapping method, which selects a sample set randomly with replacement. This allows the creation of a more accurate and robust model and avoidance of overfitting [34].

In this study, SVM is used as a meta-algorithm in bagging because it is an effective method in the classification task. This is due to the complexity of hypotheses, depending on the size of the margin, unlike other methods, in which the difficulty is associated with the number of features in the document [36].

C. Bayesian Boosting

Bayesian boosting is an iterative method in which each model determines which dataset will be used in the next iteration for model building. It uses bootstrapping, but unlike in bagging, the sample selection is based on weight. This allows building a more accurate model and reducing noise [34].

In this study, CNB is used as a meta-algorithm in boosting. CNB is a fast and simple algorithm that is widely used in classification problems. It is considered a highly accurate algorithm because it avoids the overfitting problem found in the original version of NB [34][33].

V. Study Methodology

To implement models for this study, the following steps were followed.

A. Data Collection

As mentioned before, Twitter is used to be the dataset source for this study because it is one of the most widely used applications in the KSA. Tweets from 2017–2018 from four regions in the KSA with higher populations were collected. The regions were Riyadh, Makkah, the Eastern Region, and Asir. Each region comprises different cities. The dataset consists of 6,500 violence tweets and 6,000 tweets from various fields written in colloquial Arabic in the Saudi Dialect. To collect the data, the Twitter API is used to obtain tweets from the four identified regions, and the data is extracted to an Excel file for processing. As a result, the tweets are labeled according to three classifications, namely, physical violence, psychological violence, and no violence. The final dataset consisted of 119,296 features with 32,143 unique features.

B. Data Preprocessing

To make the dataset more consistent before it was used on the models, the following important preprocessing steps are performed:

- In Arabic, some characters in a word can be written in different ways, such as اكل، أكل, which means “ate.” Because this research is concentrating on a Saudi colloquial dialect, there are some users writing tweets with different spelling problems. To overcome this problem, some letters are normalized { أ، إ، آ } to { ا }, {ى} to { ي}, {ؤ} to {و}, and {ة} to {ه};
- Removing English characters, punctuation, and symbols like @, #, and _, which are found in Twitter hashtags, and replaced them with spaces to avoid confusion between letters;
- Removing some links (URLs), diacritics, numbers, stop words, and any word with two letters or less because such words do not have meaning; and
- This paper used TF-IDF to produce a composite weight for every term.

C. Feature Extraction

To simplify the processing step for machine learning models, this study needs to apply feature extraction. It used two leading approaches, namely, light stemming and root-based stemming. Light stemming works by removing the prefixes and suffixes of words, while root-based stemming returns the word to its root [31][37]. In addition, the research applied bag of words (BOW), bi-gram, tri-gram, and a combination of the latter two. In bow, a list of words is built with their occurrences [32], while bi-gram and tri-gram focus on splitting words regarding the value of n into two and three words [21].

D. Feature Reduction

On completion of the extraction process, the number of features extracted was huge. A large number and variety of features leads to a low level of performance in the classification due to the presence of many words that have no value. To overcome this issue, various feature-reduction techniques are applied for selecting the most appropriate and relevant subset based on specific measurements. The main goal of feature reduction is improving the overall performance of classification, in addition to improving the speed of the processing and memory usage [6][7]. In this section, some feature-reduction techniques are described briefly.

1) Rule-based feature reduction: In rule-based feature reduction, the weight of features is calculated regarding the label feature. It builds a single rule for each feature, then calculates the errors. After that, it considers the feature with the highest weight as the pertinent feature [38].

2) CHI: CHI weight is a popular statistical method that uses CHI for calculating the weight of features regarding the attribute's class. Features with higher weight are considered more relevant. If CHI has a large value, then the feature is important for the category. This approach is used for measuring the difference between the expected and observed numbers of times the event occurs. The following formula is used for calculating the CHI statistic [39][40].

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

where A is the number of occurrences for t with c, N is the number of documents, D is the number that c and t neither occurred, B is the number of occurrences for t only without c, and C is the number of occurrences of c only without t.

3) Symmetrical uncertainty: To calculate the feature weight, the symmetrical uncertainty is calculated regarding class and label features. This calculates between the feature and target class, and it is used to calculate the efficiency of features. The following formula can be used to calculate symmetric uncertainty [41][42]:

$$SU(X,Y) = 2\left(\frac{IG\,(X|Y)}{H(X)+H(Y)}\right) \tag{3}$$

It is clear that symmetrical uncertainty is based on the IG of features. H(X) and H(Y) are the entropy values of features X and Y, respectively. This approach normalizes features with a large number of different values to the range [0,1], which means that when SU = 0, the features are not related to each other, but when SU = 1, knowledge of one feature allows predicting the value of the other.

4) Deviation: Using deviation-reduction methods, features are weighted based on normalized standard deviation. Features with a higher weight are considered relevant. The standard deviation illustrates how much variance from the mean is present. A low value of standard variance indicates that a feature is extremely close to the mean, whereas a higher value represents a large contrast. The formula for standard deviation is the variance square root [43].

5) Correlation: Correlation is another statistical operation used for understanding the relationship between features. For example, if we have A and B features and want to know how close they are to each other, this can be determined by calculating correlation coefficient. This takes a value in the range of –1 to 1. If the value is positive, this means that when A has large value, it is related to a large value of B. In contrast, a negative value of correlation shows an inverse relationship. The correlation coefficient can be calculated using the following formula [44]:

$$r_{A,B} = \frac{\sum_{i=1}^{n}(a_i\, b_i) - n\,\bar{A}\ \bar{B}}{n\,\sigma_A \sigma_B} \tag{4}$$

where the number of tuples is n, while a_i and b_i represent the values of A and B in each tuple. In addition, σ_A, σ_B and $\bar{A}\,, \bar{B}$ are the standard deviation and mean values for A and B, respectively.

6) Gini index: The Gini Index is another statistical method that calculates the feature weight regarding the class label using the impurity of features [44].

7) Information Gain (IG): The IG is a popular measure that was introduced by J.R. Quinlan for determining the extent to which a specific feature gives informative value about a class. It ranks scores for each feature, then selects the highest score, while lower scores are removed. ID3 uses IG as a split measure for selecting the attribute with the highest information gain value [33][34].

8) Information gain ratio: The simple form (IG) can work for most cases, but it works by choosing features with large values as root nodes. The IG by ratio is a new version of IG that is modified to reduce this bias toward choosing features with large values by normalizing IG. It observes the number of branches that will occur and takes this into account before making the split [45].

9) SVM: SVM is used as feature selection for selecting the most relevant features based on setting the hyperplane coefficients, which are calculated by an SVM as the weights of features [33].

VI. Performance Measurement

When using different classifiers, there should be an assessment of how the classifier performs when it is predicting class labels. To accomplish this, there are different measurements for evaluation, which are known as recall, precision, and accuracy, also called the recognition rate [44]. In this study, three of these measurements—accuracy, recall, and precision are choosing to be evaluated. Accuracy can be defined as the ratio of the number of tuples a classifier predicted correctly divided by the total number of tuples. Recall is the number of true-positive tuples divided by the sum of true-positive and false-negative tuples. Finally, the precision is the total of true-positive tuples divided by the sum of true-positive and false-negative tuples. Their equations are as follows [27], [44]:

$$Accuracy\ (a) = \frac{TP+TN}{TP+FP+TN+FN}, \tag{5}$$

$$Precision\ (p) = \frac{TP}{TP+FP}, \tag{6}$$

$$Recall\ (r) = \frac{TP}{TP+FN} \tag{7}$$

where TP is the number of positive tuples predicted correctly by the classifier, TN is the number of negative tuples predicted correctly by the classifier, FN is the number of positive tuples the classifier predicted incorrectly, and FP is the number of negative tuples the classifier predicted incorrectly. In addition, the study used 10-fold cross-validation to test the model. The data were divided randomly into 10 folds, and in each iteration, one of these folds was used as the test set and the others as the training partition [44].

VII. Experimental Results

The experiments were implemented using the RapidMiner tool. RapidMiner is an open-source tool written in Java that provides a wide range of text-mining techniques and machine learning algorithms. For the experiments, a 64-bit Windows 10 laptop devices are used with 16 GB of memory and Intel core i7.

A. Dataset Analysis

After collecting the datasets, they are analyzed depending on the types of violence and the user's location. As a result, some tweets did not have a clear location, so they are excluded. The results in Fig. 1 show that the most tweets were collected from Riyadh, at 54.57%, followed by 34.59% from Makkah and 14.67% and 5.17% for the eastern region and Asir, respectively. The high number of tweets from Riyadh can be explained that this city has one of the highest number of Twitter users in the world [46]. In contrast, the psychological violence was found in 65.95% of the tweets, which was higher than the rates of physical and sexual violence, at 30.96% and 3.09% respectively (see Fig. 2). Psychological violence was higher because people find it easier to act on others via coercion or threats. In contrast, people in the Saudi community find it difficult to disclose any information related to sexual violence they faced to other people.

B. Classifier Accuracy

In this section, bagging SVM, KNN, and Bayesian boosting are applied for constructing models with different feature-extraction techniques that was described in section VII to evaluate their performance. Tables VI, VII and VIII represent the results for accuracy, recall and precision that were explained in the previous section.

In KNN, Choosing the appropriate value for K in datasets has a significant influence on the accuracy of classifications. In the experiments, the value of K was 5 because the initial experiments indicated that it would provide better results. In bagging SVM and Bayesian boosting, the number of iterations was set to 10. The results showed a greater superiority of bagging SVM over the other algorithms because it was used for SVM as a meta-algorithm that has been proved to be effective in the classification of texts.

The experiments clearly showed that the tri-gram approach had the best performance, at 86.61%, 82.80%, and 77.09% for bagging SVM, Bayesian boosting, and KNN, respectively. This was because most words in Arabic have a triangular root, which the tri-gram approach contributes to extracting. In contrast, BOW has shown the worst results because it contains a large number of features. Furthermore, light stemming had better accuracy than root-based stemming because it maintained words' meanings. Bi-gram obtained good results in general, and it was better than tri-gram with KNN because KNN is influenced by the number of features, which was only 837 for bi-gram, while tri-gram consisted of 8,960 features. Combining bi-gram and tri-gram gave poor results, contrary to expectations.

C. Baseline and Ensemble Methods Comparison

In this section, baseline methods of SVM, CNB and ensemble methods with bagging SVM and Bayesian boosting are compared to evaluate the effective of ensemble methods. Fig. 3 represents a comparison between the baseline methods of SVM and CNB and ensemble methods with bagging SVM and Bayesian boosting in term of accuracy with Tri-gram technique.

The results showed that the bagging SVM improved the performance of baseline SVM well, while the performance of Bayesian boosting decreased significantly, contrary to expectations.

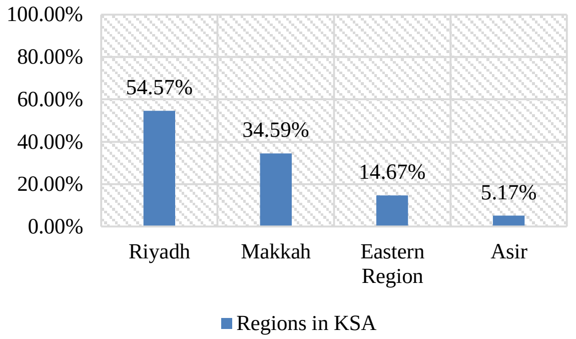

Fig. 1. Percentages of Violence Tweets in Four Regions in the KSA.

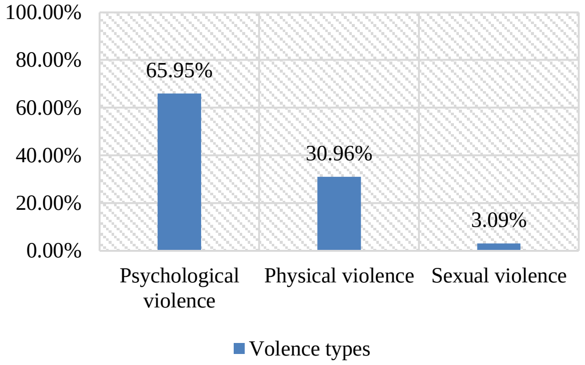

Fig. 2. Violance Tweets Types in Percentages in the KSA.

TABLE VI. Effects of Features Extraction Techniques on Bayesian Boosting

Bayesian Boosting	**Features**	**Accuracy**	**Recall**	**Precision**
BOW	32143	68.18	74.37	71.94
Light stemming	16633	69.65	69.67	71.11
Root-based stemming	6677	69.66	69.73	76.28
Bi-gram	837	79.23	79.15	79.13
Tri-gram	8960	82.80	82.76	82.77
Tri-gram + Bi-gram	9506	74.27	74.21	74.37

TABLE VII. EFFECTS OF FEATURES-EXTRACTION TECHNIQUES ON KNN

KNN	Features	Accuracy	Recall	Precision
BOW	32143	68.89	68.82	70.62
Light stemming	16633	69.59	69.59	70.02
Root-based stemming	6677	64.13	64.15	64.43
Bi-gram	837	79.56	79.49	79.53
Tri-gram	8960	77.09	75.09	75.09
Tri-gram + Bi-gram	9506	68.38	68.36	68.44

TABLE VIII. EFFECTS OF FEATURE-EXTRACTION TECHNIQUES ON BAGGING SVM

Bagging	Features	Accuracy	Recall	Precision
BOW	32143	74.89	74.85	75.58
Light stemming	16633	75.3	75.21	75.61
Root-based stemming	6677	72.58	72.52	73.04
Bi-gram	837	85.81	85.76	85.76
Tri-gram	8960	86.61	86.61	86.62
Tri-gram + Bi-gram	9506	76.57	76.60	76.89

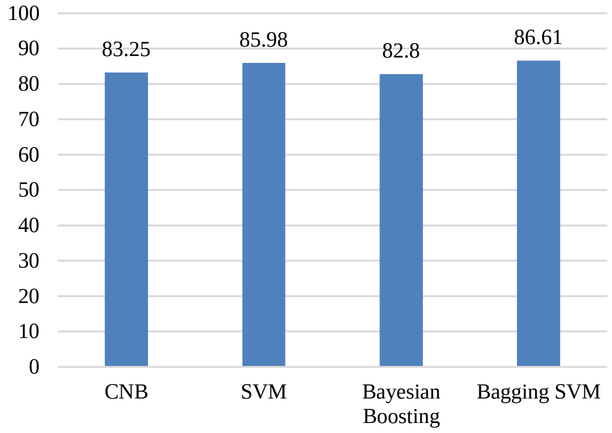

Fig. 3. Results of Ensemble and Baseline Methods.

D. Feature-Reduction Performance

In this section, nine feature-reduction techniques are evaluated to show how they affected bagging SVM. Different number of features are used for each technique at 1,000, 1,500, 2,000, 2,500, 3,000, 3,500 and 4,000 features to see how the number of features affected the performance of each type of feature-reduction techniques. Table IX illustrates the effectiveness of the feature-reduction techniques in term of accuracy, recall, and precision with bagging SVM.

SVM weighting provided the best accuracy, recall, and precision among the investigated methods, confirming SVM's high capabilities in weighting. The highest score was recorded at 90.41%, when the number of features was 3,000. The rule technique was the worst among them, giving between 60.24% and 78.31%; moreover, the result indicated that the accuracy was positively correlated with the increase in the number of features that means the weakness of its reduction capabilities. It is clear from the results that CHI, deviation, and correlation had a negative effect on the performance, significantly reducing the accuracy of bagging SVM. The results revealed that the performances of the Gini Index and IG were somewhat similar, reaching the highest level at 3,500 features, with 87.68% and 87.92% for the Gini Index and IG, respectively. The IG ratio resulted in a dramatic declined in accuracy to between 83.30% and 86.34%. The last type of reduction method, symmetrical uncertainty, provided reasonable performance with few features at 1,500 and 2,000. Fig. 4 shows how the number of features affects the accuracy of each type.

Based on the findings, the best results were obtained from the SVM, IG, and Gini Index approaches. These approaches are combined in sequence for evaluating the effects of each combination on the accuracy, recall, and precision of bagging SVM. The value of the number of features on the first side of the sequence has been set to a value that obtained higher accuracy in previous experiments. The results of the first side of the feature-reduction technique will be used on second side. When SVM was the first side, the number of features were 3,000, while for IG and the Gini Index, the values were from 2,500 to 2,000, 1,500, and 1,000. In contrast, when the IG and Gini Index were on the first side, the number of features were 3,500, while in SVM, the value was from 3,000 to 2,500, 2,000, 1,500, and 1,000. Table X illustrates the effectiveness of combinations the feature-reduction techniques in term of accuracy, recall, and precision with bagging SVM.

The overall results of integrating SVM with IG or the Gini Index showed a good influence of SVM on accuracy, recall, and precision. However, the performance of the SVM alone was still better than when integrated with other approaches, except when integrated with IG in the sequence of IG followed by SVM at 90.59% with 1,500 features. In contrast, the performance of merging IG and the Gini Index gave results close to the performance of each one alone, at accuracies of between 86% and 87%. Fig. 5 represents the accuracy of bagging SVM when combinations of feature-reduction techniques were applied.

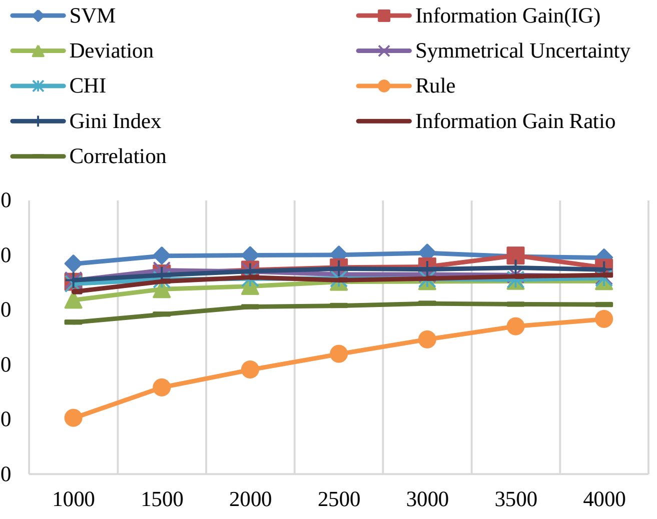

Fig. 4. Effect of K-Features on Bagging SVM Accuracy.

TABLE IX. Effects of Feature-Reduction Techniques on Bagging SVM Accuracy

Features Reduction	Features	Accuracy	Recall	Precision
Information Gain (IG)	1000	85.09	85.12	**85.49**
	1500	86.63	86.65	**86.81**
	2000	87.30	87.31	**87.72**
	2500	87.73	87.73	**87.78**
	3000	87.37	78.36	**87.38**
	3500	87.92	87.91	**87.93**
	4000	87.69	87.70	**87.72**
Information Gain Raito	1000	83.30	83.37	**84.29**
	1500	85.21	85.24	**85.60**
	2000	85.94	85.98	**86.27**
	2500	85.43	85.44	**85.57**
	3000	85.74	85.75	**85.85**
	3500	86.12	86.12	**86.17**
	4000	86.34	86.34	**86.39**
SVM	1000	88.42	88.44	**88.60**
	1500	89.87	89.89	**89.94**
	2000	89.96	89.96	**89.99**
	2500	90.05	90.05	**90.09**
	3000	90.41	90.41	**90.43**
	3500	89.72	89.72	**89.74**
	4000	89.53	89.53	**89.55**
Rule-based	1000	60.24	60.36	**61.26**
	1500	65.81	65.89	**66.45**
	2000	69.07	69.12	**70.25**
	2500	71.92	71.94	**72.08**
	3000	74.60	74.62	**74.82**
	3500	77	77.02	**77.31**
	4000	78.31	78.32	**78.44**
Symmetrical Uncertainty	1000	85.38	85.41	**85.73**
	1500	87.21	87.24	**87.47**
	2000	87.01	87.01	**87.11**
	2500	86.48	86.49	**86.60**
	3000	86.48	86.48	**86.59**
	3500	86.37	86.38	**86.44**
	4000	86.10	86.10	**86.17**
Deviation	1000	81.81	81.83	**82.17**
	1500	83.77	83.77	**83.86**
	2000	84.31	84.31	**84.37**
	2500	85.12	85.11	**85.16**
	3000	85.21	84.86	**85.22**
	3500	85.26	85.25	**85.26**
	4000	85.22	85.22	**85.23**
CHI	1000	84.76	84.73	**85.13**
	1500	85.53	85.55	**85.74**
	2000	85.72	85.73	**85.81**
	2500	85.70	85.70	**85.80**
	3000	85.65	85.65	**85.69**
	3500	85.53	85.53	**85.55**
	4000	85.81	85.50	**85.83**
Gini Index	1000	85.41	85.44	**85.78**
	1500	86.36	86.37	**86.54**
	2000	87.04	87.05	**87.14**
	2500	87.55	87.55	**87.61**
	3000	87.42	87.42	**87.46**
	3500	87.68	87.68	**87.71**
	4000	87.35	87.35	**87.37**
Correlation	1000	77.71	77.58	**80.50**
	1500	79.17	79.06	**81.62**
	2000	80.54	80.43	**82.75**
	2500	80.76	80.66	**82.88**
	3000	81.16	81.07	**83.02**
	3500	81.02	80.93	**82.02**
	4000	80.95	80.87	**82.59**

TABLE X. Effects of Feature-Reduction Combinations on Bagging SVM

Features Reduction	Features	Accuracy	Recall	Precision
SVM ⇒ IG	1000	86.05	86.08	**86.39**
	1500	88.33	88.34	**88.43**
	2000	89.31	89.31	**89.42**
	2500	89.84	89.84	**89.54**
IG ⇒ SVM	1000	89.19	89.22	**89.42**
	1500	90.59	90.60	**90.70**
	2000	90.30	90.30	**90.36**
	2500	89.75	89.76	**89.79**
	3000	88.97	88.97	**88.98**
SVM⇒ Gini Index	1000	86.15	86.18	**86.42**
	1500	88.30	88.30	**88.40**
	2000	89.29	89.30	**89.34**
	2500	89.75	89.76	**89.79**
Gini Index ⇒ SVM	1000	89.74	89.77	**89.99**
	1500	90.42	90.44	**90.54**
	2000	90.44	90.39	**90.50**
	2500	89.86	89.86	**89.88**
	3000	89.05	89.05	**89.08**
IG⇒ Gini Index	1000	86.15	86.18	**86.42**
	1500	88.30	88.30	**88.40**
	2000	87.04	87.05	**87.14**
	2500	87.54	87.55	**87.61**
	3000	87.54	87.54	**87.58**
Gini Index ⇒ IG	1000	86.09	86.12	**86.49**
	1500	86.63	86.65	**86.81**
	2000	87.30	87.31	**87.39**
	2500	87.68	87.69	**87.75**
	3000	87.44	87.44	**87.48**

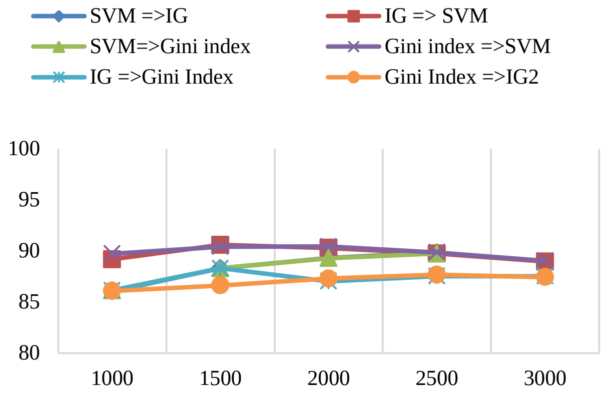

Fig. 5. Effect of K-Feature Reduction Combinations on Bagging SVM.

VIII. Conclusion

Due to the significant increase in violence in Arab societies, there is a need to discover and study them to find appropriate solutions. This research dealt with expressions of violence on social media in Saudi society.

Dataset collected from Twitter between 2017 and 2018 in colloquial Arabic from the four most densely populated administrative regions in Saudi Arabia, then the results classified. The findings indicated that the most violence is evident in Riyadh. In addition, number of classification algorithms of baseline and ensemble methods was compared in term of accuracy, recall and precision. In addition, multiple ways of extracting features, as well as ways of reducing the number of features are compared. The results showed the superiority of bagging SVM with tri-gram over other approaches at 86.61%; moreover, the combination of SVM weighting and SVM at classification contributed to higher performance at 90.59%.

In the future, the research will extend to cover other administrative regions in Saudi Arabia. In addition, the capabilities of the devices will expand to assess other types of reductions that this research could not implement because of the limitation in memory capabilities, such as PCA and SVD feature-reduction, in addition to extending experiments to other types of ensemble methods, such as stacking.

References

[1] M. Alshehri, "11 thousand cases of violence!," Saudi newspaper Okaz.

[2] "Twitter in the Arab Region." [Online]. Available: http://www.arabsocialmediareport.com/twitter/linechart.aspx.[Accessed: 30-Mar-2019].

[3] "These Are The Most Twitter-Crazy Countries In The World, Starting With Saudi Arabia (!?) | Business Insider." [Online]. Available: https://www.businessinsider.com.au/the-top-twitter-markets-in-the-world-2013-11. [Accessed: 16-Mar-2019].

[4] A.-H. Tan, "Text Mining: The state of the art and the challenges.," in Workshop on Knowledge Discovery from Advanced Databases (KDAD'99), 1999, p. 71–76.

[5] A. Hotho, A. Nürnberger, and G. Paaß, "A Brief Survey of Text Mining," Ldv Forum, vol. 20, no. 1, pp. 19–62, 2005.

[6] I. Guyon and A. Elisseeff, "An Introduction to Variable and Feature Selection," J. ofMachine Learn. Res., vol. 3, pp. 1157–1182, 2003.

[7] H. Liu and H. Motoda, "Feature Selection for Knowledge Discovery and Data Mining." , Kluwer Acadmic,2011.

[8] J. Maglogiannis, I., Karpouzis, K., Wallace, B.A., Soldatos, Emerging Artificial Intelligence Applications in Computer Engineering. 2007.

[9] "Summary by language size | Ethnologue." [Online]. Available: https://www.ethnologue.com/statistics/size. [Accessed: 12-Mar-2019].

[10] M. K. Saad, "Arabic Morphological Tools forText Mining," Int. Conf. Electr. Comput. Syst., pp. 1–6, 2010.

[11] R. Duwairi, "Arabic Text Categorization," Int. Arab J. Inf. Technol., vol. 4, 2007.

[12] M. Al-Zabidi, "The crown bride of the jewels dictionary." 1965.

[13] A. Farghaly and K. Shaalan, "Arabic Natural Language Processing," ACM Trans. Asian Lang. Inf. Process., vol. 8, no. 4, pp. 1–22, 2010.

[14] E. G. Krug, L. L. Dahlberg, J. A. Mercy, A. B. Zwi, and R. Lozano, "World Report on Violence and Health Geneva: World Health Organization, 2002. ISBN 92-4-154561-5.," Inj. Prev., 2002.

[15] "Psychological violence | European Institute for Gender Equality." [Online]. Available: https://eige.europa.eu/thesaurus/terms/1334. [Accessed: 15-Mar-2019].

[16] "Makkah region recorded the highest precentage of domestic violence with 34% - Saudi News -Okaz Newspaper." [Online]. Available: https://www.okaz.com.sa/article/927096. [Accessed: 15-Mar-2019].

[17] A. Khatua, E. Cambria, and A. Khatua, "Sounds of silence breakers: Exploring sexual violence on Twitter," in Proceedings of the 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2018, 2018, no. August, pp. 397–400.

[18] S. Subramani, H. Q. Vu, and H. Wang, "Intent Classification Using Feature Sets for Domestic Violence Discourse on Social Media," in Proceedings - 2017 4th Asia-Pacific World Congress on Computer Science and Engineering, APWC on CSE 2017, 2018, pp. 129–136.

[19] N. Albadi, M. Kurdi, and S. Mishra, "Are they Our Brothers? Analysis and Detection of Religious Hate Speech in the Arabic Twittersphere," in 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), 2018, pp. 69–76.

[20] M. A. Al-Walaie and M. B. Khan, "Arabic dialects classification using text mining techniques," in International Conference on Computer and Applications, (ICCA), 2017, pp. 325–329.

[21] S. Raheel and J. Dichy, "An empirical study on the feature's type effect on the automatic classification of Arabic documents," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 6008 LNCS, pp. 673–686, 2010.

[22] H. Uğuz, "A two-stage feature selection method for text categorization by using information gain, principal component analysis and genetic algorithm," Knowledge-Based Syst., vol. 24, no. 7, pp. 1024–1032, 2011.

[23] G. Vinodhini and R. M. Chandrasekaran, "Opinion mining using principal component analysis based ensemble model for e-commerce application," CSI Trans. ICT, vol. 2, no. 3, pp. 169–179, 2014.

[24] A. Al-Thubaity, A. Alanazi, I. Hazzaa, and H. Al-Tuwaijri, "Weirdness coefficient as a feature selection method for Arabic special domain text classification," in Proceedings - 2012 International Conference on Asian Language Processing, IALP 2012, 2012, pp. 69–72.

[25] M. H.Kadhim and N. Omar, "Bayesian Learning for Automatic Arabic Text Categorization," J. Next Gener. Inf. Technol., vol. 4, no. 3, pp. 1–8, 2013.

[26] M. S. Khorsheed and A. O. Al-Thubaity, "Comparative evaluation of text classification techniques using a large diverse Arabic dataset," Lang. Resour. Eval., vol. 47, no. 2, pp. 513–538, 2013.

[27] M. Faqeeh, N. Abdulla, M. Al-Ayyoub, Y. Jararweh, and M. Quwaider, "Cross-lingual short-text document classification for facebook comments," in Proceedings - 2014 International Conference on Future Internet of Things and Cloud, FiCloud 2014, 2014, pp. 573–578.

[28] I. Hmeidi, M. Al-Ayyoub, N. A. Abdulla, A. A. Almodawar, R. Abooraig, and N. A. Mahyoub, "Automatic Arabic text categorization: A comprehensive comparative study," J. Inf. Sci., vol. 41, no. 1, pp. 114–124, 2015.

[29] R. Alhutaish and N. Omar, "Arabic text classification using K-nearest neighbour algorithm," Int. Arab J. Inf. Technol., vol. 12, no. 2, pp. 190–195, 2015.

[30] F. S. Al-Anzi and D. AbuZeina, “Toward an enhanced Arabic text classification using cosine similarity and Latent Semantic Indexing,” J. King Saud Univ. - Comput. Inf. Sci., vol. 29, no. 2, pp. 189–195, 2017.

[31] I. S. I. Abuhaiba and H. M. Dawoud, “Combining Different Approaches to Improve Arabic Text Documents Classification,” Int. J. Intell. Syst. Appl., vol. 9, no. 4, pp. 39–52, 2017.

[32] I. H. Witten, Text mining, Practical handbook of Internet computing (CRC Press, Boca Raton). 2005.

[33] I. H. Witten, E. Frank, and M. a Hall, Data Mining:Practical Machine Learning Tools and Techniques second edition. 2011.

[34] J. Melton, S. Buxton, H. Samet, T. J. Teorey, S. S. Lightstone, T. P. Nadeau, J. Celko, G. Ralf, M. Schneider, J. Celko, E. Cox, T. Halpin, K. Evans, P. Hallock, B. Maclean, J. Melton, J. Melton, A. R. Simon, and M. Chisholm, Data Mining : Concepts and Techniques. 1999.

[35] B. Baharudin, L. H. Lee, and K. Khan, “A Review of Machine Learning Algorithms for Text-Documents Classification,” J. Adv. Inf. Technol., vol. 1, no. 1, 2010.

[36] T. Joachims, "Text Categorization with Support Vector Machines: Learning with Many Relevant Features Thorsten", Proceeding of the Tenth European Conference on Machine Learning, pp. 1–7, 1999.

[37] N. Alami, M. Meknassi, S. A. Ouatik, and N. Ennahnahi, “Impact of stemming on Arabic text summarization,” Colloq. Inf. Sci. Technol. Cist, pp. 338–343, 2017.

[38] “Weight by Rule - RapidMiner Documentation.” [Online]. Available: https://docs.rapidminer.com/latest/studio/operators/modeling/feature_weights/weight_by_rule.html. [Accessed: 26-Mar-2019].

[39] Y. Yang and J.P. Pedersen., “Feature selection in statistical learning of text categorization.,” Proc. Fourteenth Int. Conf. Mach. Learn. (ICML’97), 1997.

[40] H. Liu and R. Setiono, “Chi2: feature selection and discretization of numeric attributes,” in Proceedings of the IEEE 7th International Conference on Tools with Artificial Intelligence Chi2:, 2002, pp. 388–391.

[41] M. A. Hall and Li. A.Smith, “Practical Feature Subset Selection for Machine Learning ,” in proceedings of the 21st Australasian Computer Science Conference ACSC’98, 1998, vol. Volume 20, p. 586.

[42] S. I. Ali, “A Feature Subset Selection Method based on Conditional Mutual Information and Ant Colony Optimization,” Int. J. Comput. Appl., vol. 60, no. 11, pp. 5–10, 2012.

[43] “Weight by Deviation - RapidMiner Documentation.” [Online]. Available: https://docs.rapidminer.com/latest/studio/operators/modeling/feature_weights/weight_by_deviation.html. [Accessed: 26-Mar-2019].

[44] J. Han, M. Kamber, and J. Pei, Data Transformation by Normalization. 2011.

[45] “Weight by Information Gain Ratio - RapidMiner Documentation.” [Online]. Available: https://docs.rapidminer.com/latest/studio/operators/modeling/feature_weights/weight_by_information_gain_ratio.html. [Accessed: 26-Mar-2019].

[46] “The World’s Most Active Twitter City? You Won’t Guess It.” [Online]. Available: https://www.forbes.com/sites/victorlipman/2012/12/30/the-worlds-most-active-twitter-city-you-wont-guess-it/#4367128655c6. [Accessed: 16-Mar-2019].