# Sentiment Analysis of Twitter Data for Saudi Universities

Meshrif Alruily and Osama R. Shahin

***Abstract*—With the tremendous increase in web technologies, many people are conveying and expressing their feelings through social media. These data collections, called big data, can be of high value to institutions, governments, and researchers if analyzed and interpreted as well. Most of the studies on sentiment analysis have been performed for the English language and few of these works focus on Arabic. Further, these studies rarely focus on colloquialisms, and no research was carried out on Twitter data related to Saudi universities. Therefore, this paper aims to develop a sentiment analysis system for analyzing Tweets generated by Saudi Twitter users about Saudi universities. There are two proposed different models used for Twitter-based sentiment analysis. The first model begins with Tweet collection followed by preprocessing. Latent Dirichlet Allocation, used here, eliminated Tweets that were not expressing sentiments; therefore, the document will be reduced, which affects system accuracy. Finally, Support-Vector Machine (SVM) chosen to perform Tweet classification. The second model is based on using different classifier models, such as Naïve Bayes (NB), Stochastic Gradient Descent (SGD), Sequential Minimal Optimization (SMO), K-nearest neighbors Classifier (KNN), Random Forest, Multilayer Perceptrons using Deeplearning, (MPD), and Sentiment Score Calculation (SSC).**

***Index Terms*—Sentiment analysis, latent dirichlet allocation, word embeddings support-vector machine.**

## I. Introduction

The importance of data in this age is increasing significantly. The size of data traded daily is doubling exponentially, and the size of data traded in 2020 is expected to reach 44ZB [1]. Individuals are the primary and essential source of these data. Social networks like Twitter and Facebook became well-linked ways in which of communication across communities. Internet Users have to express their view and feeling on a topic on such social media like Twitter, Facebook, and Blogs. Since Twitter was established in 2006, it has allowed users to express opinions and communicate feelings in the style of SMS. Twitter is one of the most important platforms, full of sentiments and emotion. It's a micro blogging web site with Tweets; all Tweets have "140 characters or fewer." The 140 million active users on Twitter could produce 1 billion Tweets every 72 hours [1]. This makes it a major example of "big data." On Twitter, nearly 6.5 million Arab users create more than 10.8 million Tweets each day. In addition, many studies for monitoring real-time events using social media data have been performed; for example, using Twitter streams or messages for traffic detection, tracking flu infections, and monitoring food-borne outbreaks [2]-[5]. The Arab world has been affected by these recent technological developments. For instance, the total number of Twitter users in Arab countries now stands at more than 11 million, with 27.4 million Tweets per day. Moreover, more than half of these Tweets are from users in Saudi Arabia and Egypt, accounting for 30% and 20% of all Tweets, respectively. The most active users are from Saudi Arabia, followed by Egypt, Algeria and the United Arab Emirates.

Big data is a term that describes the large size of data– whether structured (i.e., data within databases) or unstructured (i.e., data on the web) to obtain important information that can be used for decision making and analysis in case of sentiment analysis (SA) [6]. This field is used to derive meaning from the text. Analysis of Arabic is very complex compared to English due to the following difficulties:

1) Some words such as (غير، مش، لكن), can change the meaning of what comes after them.

2) The roots of Arabic words can contain multiple context-based forms, such as (يتشاجر, شجرات, أشجار)

3) Arabic has the same spelling, but in a different sense, it depends on its diacritics, such as (يَخْرُجْ) which means get out and (يُخرِجْ)which means extract.

4) Each country/part of the province has an Arabic dialect, which makes analyzing sentiments with different dialects complex.

### A. General Methods for Data Analysis

Data analyses are often achieved by passing four major steps. It begins with distinguishing keywords to evaluate individuals' sentiments. Then there's the assembly stage to gather targeted Tweets. These bulk feeds will be held in a data set. The subsequent step is to preprocess Tweets that may take away all irrelevant content and only retain the Arabic text. The word filtering step will remove all words that don't have an effect on the text's meaning. The next stage is to categorize the content into negative and positive. These stages are going to be represented in additional detail within the following sub-sections.

i) Collecting

The first step in analyzing sentiment is to gather Tweets by identifying a specific keyword to recover all the Tweets associated with word. There are different sources that can be used to gather the need Tweets [6].

ii) Preprocessing

Preprocess is a procedure that used to remove useless information, such as URLs, labels, images, and English characters in case of research dealing only with Arabic

Manuscript received September 7, 2019; revised December 11, 2019.

Meshrif Alruily is with the College of Computer and Information Sciences, Jouf University, Saudi Arabia (e-mail: mfalruily@ju.edu.sa).

Osama R. Shahin is with College of Science & Arts, Jouf University, Saudi Arabia, he was with Physic and Mathematic Dept., Faculty of Engineering, Helwan University, Egypt (e-mail: osama.r.shaheen@gmail.com).

doi: 10.18178/ijmlc.2020.10.1.892

sentiments [7], [8]. Different steps in preprocessing can be summarized as follows:

iii) Removing Non-Arabic Tweets

Removing all Tweets which are Non-Arabic in nature. Twitter is allowed in more than 60 languages. However, this work is currently focused on Arabic Tweets only.

iv) Removing URLs

URLs do not have much information about feelings and thus are removed.

v) Removing Numbers and Punctuation

Numbers and punctuation are useless when measuring feeling. Thus, these are removed to optimize the Tweet content.

vi) Removing repeated characters

Removing repeated characters from words simplifies the process of identifying and recognizing the word. For example, "عبيبير" will be transformed into "عبير ";here, the character "ي" was repeated three times.

vii) Stop-Word Removal

Stop words occur in each training set of positive and negative sentiments, thus adding more ambiguity in the classification model. They do not carry any information about emotions. Create a list of stop words like, " ",هي" ",هم" على" ",في" ",عند" ," etc., and ignore them while recording the sentiment.

viii) Character Normalization

Normalization is the technique of substituting similar characters that can be used mutually through one of them. Also remove letters that do not carry any information. This process can be done through the following:

a. Eliminate short vowels such as (ً,,ِ).

b. Eliminate any special character and from the word such as (= + $ ~).

c. Eliminate " أ"",ا ," and "آ" with stick alf " ؤ" ",ا " with "و," and " ئ" and "ىء " with"ي" regardless position of the hamza ""ء

in the word.

d. Replace the final character "ى" with "ي"

e. Replace the final character "ة" with "ه"

ix) Filtering

The filtering phase removes all words that have no effect on the meaning of the statement, such as spelling, recurring messages or letters, and stop words.

### B. Classifications

This represents the final stage where all positive and negative Tweets will be categorized by a predefined classifier. This stage has been necessary for classifying Tweet sentiments. Tweet sentiments can usually be classified into two categories, negative and positive. They are generally classified with supervised methods (such as Support-Vector Machine (SVM), Linear Regression) and unsupervised methods (such as K-means, Clustering). After training the model, it will be able to classify the polarity of the new data set into positive and negative. The supervised method gives more accurate results since it is trained on a huge database.

## II. Related Work

Social media have become significantly important platforms for natural language processing research. Sentiment analysis on Twitter data has become an active research field in many languages as well as for real-time event extraction. To the authors' knowledge, no research has been carried out on Twitter data related to Saudi universities. In the following section, the most important research on Arabic Sentiment Analysis (ASA) will be presented. In general, most of the research described below has dealt with ASA from two main aspects: first, by using machine learning techniques (supervised techniques), and second, by using unsupervised techniques. Some of them used semi-supervised methods, and the rest used hybrid methods.

A machine learning approach, using SVM, was introduced in [7], [8]. This work depended on subjectivity sentiment classification at the sentence level. The dataset collected in [8] reached 11,918, and it was collected from different social media sources. M. Elarnaoty *et al.* [9] dealt the sentiment analysis in Arabic news articles by three approaches (supervised, semi-supervised, and a combination of the first two approaches). The proven improvement of the algorithm performance, due to the preprocessing step that used at the beginning of the algorithm, such as removing stop words, stemmer, and n-grams combination, would be established in [10]. In this work, a sentiment classification was made up using SVM, NB and KNN. They used two corpora: one of them is OCA mentioned in [11], and the other is developed by the authors. The last corpus was captured from reviews of the two domains dataset (sports and movies). The authors in [12] dealt with 3,090 Arabic sentiments collected from different social networks. This work was performed using ACLWSDS, which was developed in [13]. Returning to Abdul-Majeed [14] this work presented a large scale, multi-genre sentiment lexicon, formed from of 224,564 records covering multiple Arabic dialects. Nabil *et al.* [15] collect a dataset of about 10,006 Arabic Tweets. This approach treated the problem by creating four sentiment classes. Like the work suggested in this paper, [15] used many machine learning algorithms, such as (SVM, KNN, BNB, MBN, and stochastic gradient descent (SGD)). Abu Hammad and El-Halees [16] collecting 2,848 Arabic opinions from different websites. They proposed a supervised algorithm for detecting Arabic opinions by using SVM, NB, and KNN. In [17], authors proposed a hybrid approach for sentiment classification combined from Sentiment Orientation with a machine learning classifier. Alhazmi and Salim [18] proposed a supervised approach to extract personal opinions from Arabic Tweets. In this work, a sentiment classification process was performed by using SVM, NB, and KNN classifiers.

Naaima Boudad *et al.* [19], compared three different classification approaches, such as unsupervised, supervised, and hybrid. Samar Al-Saqqa *et al.* [20] proposed an approach to classifying sentiment polarity based on voting algorithm. In this work, four classifiers have been proposed, i.e., KNN, SVM, Decision Trees, and NB. The dataset used in this algorithm was composed of three different datasets with a total of 18,948 Arabic reviews and Tweets.

Ref. [21] proposed a literature of ASA based on the corpus approach. This work has been carried out on the works of classical Arabic dialects and modern Arabic in

which sentiment are analyzed. Another survey that presented a complete outline of ASA was explained in [22]. This study attempts to show all the gaps and obstacles in ASA, which can be avoided later in recent studies.

## III. Proposed Model

There are two proposed different models used for Twitter-based sentiment analysis. The first model began with collecting Tweets, followed by performing the preprocessing step. The corresponding sentiment for the trained Tweets will be determined. The transliteration process will facilitate dealing with Arabic Tweets. Latent Dirichlet Allocation used here eliminated Tweets that did not discuss any feeling, reducing document space and affecting system accuracy. Word embedding, or word encoding, is a mathematical method that converts words into a mathematical object representation, called a vector, where each dimension's value corresponds to a specific feature. Finally, a SVM chosen to classify Tweets. A schematic diagram for the proposed model is shown in Fig. 1.

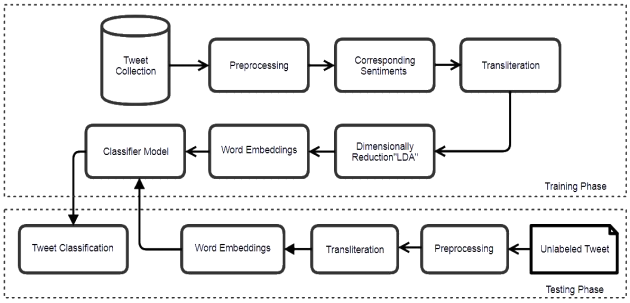


Fig. 1. Phases of the first proposed system.

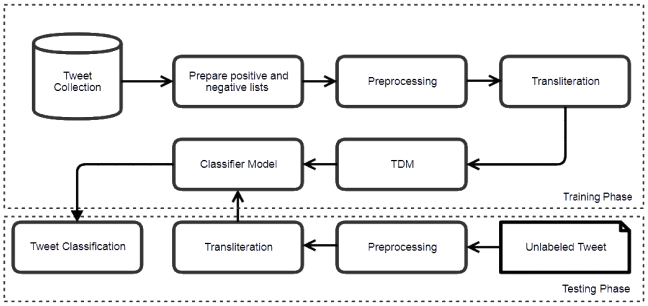


Fig. 2. Phases of the second proposed system.

The second model is based on using different classifier models. As mentioned before, this proposed model started with Tweet collection followed by the preprocessing step. After that, two word lists will be determined, one of them concerning by all available positive sentiment and other belongs to negative sentiments. The Tweet collection dataset and the two previously mentioned lists will be transliterated. Term-document matrix will summarize all collected Tweets followed by a positive and negative word count for each comment. Finally, different classifier models, such as Naïve Bayes, Stochastic Gradient Descent (SGD), Sequential Minimal Optimization (SMO), K-nearest neighbors Classifier (KNN), Random Forest, Multilayer Perceptrons using Deep learning, (MPD), and Sentiment Score Calculation (SSC) will be used to classify the sentiments into positive and negative. A schematic diagram for this proposed model depicted in Fig. 2.

First, the following sub-sections explained the common phases in the two proposed methods; after that, each system will be explained separately.

### A. Data Collection

The corpora are collected from Twitter for 22 Saudi universities. These corpora are classified into two sets; the first is the corpus for training, and the second is the corpus used for testing. Initially, all sentiments from each university were collected separately into 22 Excel files. Then, this data arranged as separate text files with UTF-8 encoding through designing a small micro in Excel. From the 600K collected comments, and after removing all topics which were irrelevant to analyzing societal feelings toward universities by using dimensional reduction, more than 60K comments were classified as positive or negative. For the Word2Vec corpus, we collected a big text corpus of 12.5 megabytes to build the required Continuous Bag of Words model. For the proposed model, the collected dataset was classified into a training set and test set, where 15% (882 negatives, 882 positives) for test and 85% (5,000 positives, 5,000 negative) for training. The comment statistics are summarized in Table I.

TABLE I: Dataset Statistics

| Dataset | Number of Tweets | |
|---|---|---|
| | Positive | Negative |
| Total comments in the training set | 5000 | 5000 |
| Total comments in the testing set | 882 | 882 |
| Total words | 60323 | 72643 |
| Avg. words in each comment | 10.25 | 12.35 |

### B. Preprocessing

The preprocessing step removes Non-Arabic Tweets, URL, numbers, repeated characters, and stop words. The stemming process will be performed in this phase.

### C. Transliteration

TABLE II: Transliteration Process

| | arabic_text | sents | arabic_text_t |
|---|---|---|---|
| 1 | مش مظبوط | Negative | mW mZbwT |
| 2 | كسول | Negative | kswl |
| 3 | مهمل | Negative | mhml |
| 4 | جاب العيد | Negative | jab al3yd |
| 5 | خايس | Negative | Kays |
| 6 | خربان | Negative | Krban |
| 7 | تمام | Positive | tmam |
| 8 | ممتاز | Positive | mmtaz |
| 9 | اخرطي | Negative | aKrTy |
| 10 | عدل كلامه | Positive | 3dl klamh |
| 11 | غير جيد | Negative | Gyr jyd |
| 12 | ليس صحيح | Negative | lys S7y7 |
| 13 | مش عاوزينه | Negative | mW 3awzynh |
| 14 | ممتاز | Positive | mmtaz |
| 15 | ماينفعش | Negative | maynf3W |
| 16 | ياريت | Positive | yaryt |
| 17 | ضربة معلم | Positive | Drb0 m3lm |
| 18 | كلام فاضى | Negative | klam faDA |
| 19 | طب مش يحل الاول | Negative | Tb mW y7l alawl |
| 20 | مش عاوزينه | Negative | mW 3awzynh |

Transliteration is a process for converting Arabic text into

equivalent transliterated Arabic to simplify the classification process. Some of the transliteration results are shown in Table II.

TABLE III: CORRESPONDING SENTIMENTS

| | arabic_text | sents | arabic_text_transliteration |
|---|---|---|---|
| 1 | احتجاج طلاب خريج كل | Negative | a7tjaj Tlab Kryj kl |
| 2 | انت يامهد مفسد كاذب محارب سعود كار وطن | Negative | ant yamhd mfsd kaib m7arb s3wd kar wTn |
| 3 | اكتملت فرح طلع عند اختبار بماد عْثيث احد اه جامع طيب لعين ...ما | Negative | ma aktmlt fr7 Tl3 3nd aktbar bmad GUyU a7d ah jam3 Ty... |
| 4 | لل حرام انا سجلت جامع جد عش مواصل بعيد مرا عن حين نقبلت ه... | Negative | ll 7ram ana sjlt jam3 jd 3W mwa5l b3yd mra 3n 7yn nQblt... |
| 5 | جامع طيب تظلم اهل جامع خيب لل حرام بن بدرسو سن اخرت يقعد... | Negative | jam3 Tyb tZlm ahl jam3 Kyb ll 7ram bn ydrsw sn aKrt yQ3d... |
| 6 | ان تم ايقاف قبول مثل كثير مجال | Negative | an tm ayQaf Qbwl mUl kUyr mjal |
| 7 | قم قهر ان مال دخل | Negative | Qm Qhr an mal dKl |
| 8 | جامع طيب تظلم اهل لل بيحاسب مساول تسبب برفع نسب بيحاسب... | Negative | jam3 Tyb tZlm ahl ll by7asb msawl tsbb brf3 nsb by7asb ... |
| 9 | مو غرررررريب جامع طيب مت هالظلم | Negative | mw Grrrrryb jam3 Tyb mt halZlm |
| 10 | لل عون ظلم عين عينك | Negative | ll 3wn Zlm 3yn 3ynk |
| 11 | جامع طيب فرع علا تظلم ابنا | Negative | jam3 Tyb fr3 3la tZlm abnaa |
| 12 | لل عون ظلم عين عينك | Negative | ll 3wn Zlm 3yn 3ynk |
| 13 | قم قهر ان مال دخل | Negative | Qm Qhr an mal dKl |
| 14 | انت يامهد مفسد كاذب محارب سعود كار وطن | Negative | ant yamhd mfsd kaib m7arb s3wd kar wTn |
| 15 | جامع باج زم تتير عجب | Positive | jam3 ba7 zm tUyr 3jb |
| 16 | اذا انت اهل جد بيتك جد خلك جامع جد افضل بكثير لل يحفظك | Positive | aia ant ahl jd bytk jd Klk jam3 jd afDl bkUyr ll y7fZk |
| 17 | استودعت لل حد سجلت جامع باج شا لل بارب يوفقهم جميعا بارب.. | Positive | astwd3t ll 7d sjlt jam3 ba7 Waa ll yarb ywfQhm jmy3a yar... |
| 18 | بارك لل فيك ولك | Positive | bark ll fyk wlk |
| 19 | لل حمد تم تخصص جامع باج كل هندس قسم هندس معمار اسال ل.. | Positive | ll 7md tm tKSS jam3 ba7 kl hnds Qsm hnds m3mar asal ll ... |
| 20 | شكرا جامع ام قر شكرا مدرب راائع | Positive | Wkra jam3 am Qr Wkra mdrb raa3 |

### 1) First proposed system

After explaining the common phases in the two proposed systems, we now explain the first proposed system in detail. The sentiment value is added manually beside the comments in the Tweets collection. The result of this process is shown in Table III.

#### a) Dimensional reduction

Dimensional reduction methods organize the Tweets into a number of a certain topics. The different topics of Tweets may include the Tweet sentiments toward 22 Saudi universities, which are of interest in this research. The other Tweets are related to other topics such as job opportunities, student activities held at the university, etc. We will reduce the data in subsequent phases. In this work, the Latent Dirichlet Allocation (LDA) technique will be used in dimensional reduction for the feature vectors in Arabic Tweet. The basic idea of the LDA is that the documents could be represented as random distribution over other topics, where each topic has a specific value derived from a specific probability distribution. It is considered a probabilistic model, specifying a given corpus [23].

#### b) Word embedding

After reducing the sentiment topics, word embedding will take place. Word embedding, or word encoding, is a mathematical method for understanding the text rather than just counting keywords. It is used to convert words into a mathematical object representation, called a vector, where each dimension's value corresponds to a specific feature [24]. Word2Vec and the TF-IDF (Term Frequency-Inverse Document Frequency) represent the most important methods that are used in modern search engines. Some examples of Word2Vec corpus depicted in Fig. 3.

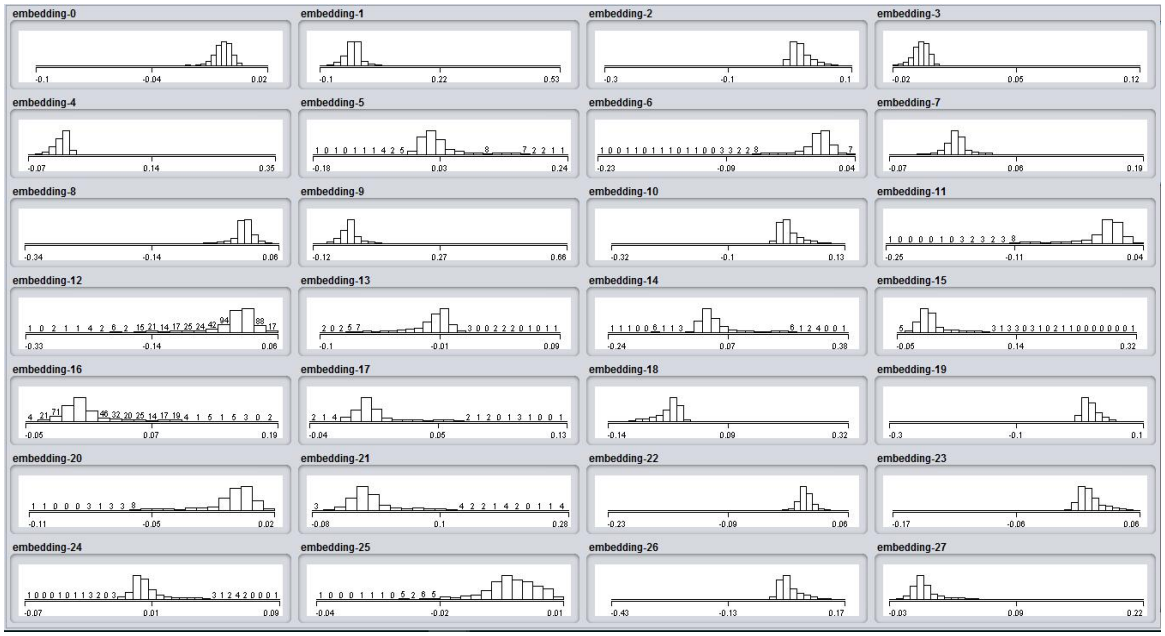


Fig. 3. Results of Word2Vec process.

Term-Frequency (TF) measures the number of times that a word occurs in a document, while Inverse Document Frequency (IDF) determines the significance of this word [25]. *TF* and *IDF* can be calculated using the following equations:

$$tf(t,d) = \sum_{t' \in d} f_{t',d} \quad (1)$$

$$idf(t,d) = log\frac{N}{n_t} \quad (2)$$

where $t'$ represents the appearance of the topic $t$ in a document $d$, and $N$ represent the total number of documents in the corpus. Finally, $n_t$represent the number of documents.

#### c) Classification model

The classification model is considered the main step in the proposed system. Such a system must distinguish sentiments into two categories, positive and negative, effectively. There are many automated learning algorithms to build such model. Here SVM was chosen to perform this task. SVM is a supervised learning approach in the machine learning field, where SVM is used in classification tasks. The SVM predicts that the negative class is present when

$w^T x + b$ is negative. Likewise, it predicts that the positive class is present when $w^T x + b$ is positive [26], where the linear function of the SVM can be written as

$$w^T x + b = b + \sum_{i=1}^{m} \alpha_i x^T x^{(i)} \quad (3)$$

where $x^{(i)}$ and $\alpha$ are a training example and a vector of coefficients. This adaptation allows us to replace $\boldsymbol{x}$ by the output of a given feature function $\phi(x)$ and the dot product with a function $k(x,x^{(i)}) = \phi(x) \cdot \phi(x^{(i)})$.

The most important step for performing the proposed model is to build a classification system that can effectively distinguish opinions into two categories, positive and negative, which are described in the training phase. Many automated learning algorithms can be used to build this model [34]. As mentioned before, features extracted from a Tweet group are divided into training and testing sets. The training set will be passed through the SVM to record its results. Then, the classifier model will be used over the set of testing data to verify the accuracy and performance of the proposed system.

*2) Second proposed system*

After collecting all the Tweets, and completing all required preprocessing steps, two word lists that indicate positive and negative sentiments are grouped into two lists, the Negative Word List (NWL), and the Positive Word List. These words come from two sources, the first of which are important words within the various topics that result from the application of LDA. The second source is a manual process in which positive and negative expressions are selected and chosen. An example of these sentiments is given in Table IV.

TABLE IV: SAMPLE OF POSITIVE AND NWLS

| | NWL | PWL |
|---|---|---|
| 1 | احتجاج | بطل |
| 2 | ما في فايدة | مشكورين |
| 3 | إحباط | شكرا جزيلا |
| 4 | فساد | الحمد لله |
| 5 | حسبي الله | يسر |
| 6 | ظلم | شكرا جدا |
| 7 | فجور | حبيب و غالي |
| 8 | مشكلة | بالتوفيق |
| 9 | فساد واضح | مفيد |
| 10 | جامعة خطية | ممتاز |
| 11 | كلب | رائع |
| 12 | كاذب ومحارب | مبارك |
| 13 | ايقاف | تبارك |
| 14 | القهر و الظلم | تمام |
| 15 | شايف نفسه | عذب الكلام |
| 16 | تظلم | جيد |
| 17 | كاره للوطن | سعادة |
| 18 | الله يحرقهم | رحيم |
| 19 | حرام عليكم | بخير |
| 20 | الظلم ظلمات | سعيد |

*a) Term-Document Matrix (TDM)*

Each comment in the Tweet Dataset will be split into a Bag of Words. Then the number of positive and negative words in each document are counted. Each comment with a correspondence number of positive and negative Tweets will be arranged in Term-Document Matrix (TDM). TDM describe the frequency of predefined positive and negative sentiment in each comment.

*b) Classifier models*

The second model, based on using different classifier models such as Naïve Bayes (NB), SGD, Sequential Minimal Optimization (SMO), K-nearest neighbors Classifier (KNN), Random Forest, MPD, and SSC, will be used to classify the sentiments as positive or negative. In the following section, the results of each classifier will be explained.

## IV. EXPERIMENTAL RESULTS

To train the SVM model, we used the RTextTools library to train and test our model with a linear kernel; refer to (3). The results obtained can be summarized by SVM for a sample testing set, shown in Table V.

TABLE V: SAMPLE OF SVM RESULTS

| | SVM_LABEL | SVM_PROB | data | myterms |
|---|---|---|---|---|
| 1 | Negative | 0.9901023 | lys jyd | ليس جيد |
| 2 | Negative | 0.9547313 | mW makbol | مش مقبول |
| 3 | Positive | 0.9812272 | mmtaz | ممتاز |

Here, the confusion matrix was used to prove the efficiency of the proposed topic classification systems. The confusion matrix is a relation between the various terminologies used to determine the system performance such as *TP, TN, FP,* and *FN*. After the essential terminologies used in the confusion matrix were defined, the performance metrics used to evaluate text classification can be specified using four metrics: precision, recall, accuracy, and F1 score. Precision can be computed as follows:

$$\text{Precision} = \frac{TP}{TP+FP} \quad (4)$$

Recall metric can be defined as the number of TP divided by the sum of the real positive sentiments, and it is can be given by:

$$\text{Recall} = \frac{TP}{TP+FN} \quad (5)$$

The accuracy evaluates the efficiency of the classification process at all, and it is computed as

$$\text{Accuracy} = \frac{TN+TP}{TN+TP+FP+FN} \quad (6)$$

Finally, the F1 score is considered weighted by recall and precision, and it is given by:

$$F1 - \text{Score} = \frac{2 \times \text{Preision} \times \text{Recall}}{\text{Preision} + \text{Recall}} \quad (7)$$

All the obtained results from this work can be summarized in Fig. 4.

The results obtained from the proposed topic classification algorithm are summarized in Table VI.

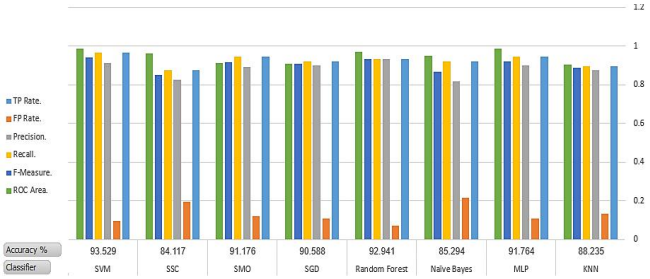


Fig. 4. Extraction evaluation metrics for the proposed approaches.

TABLE VI: EXPERIMENTAL RESULTS

| Classifier | TP Rate | FP Rate | Precision | Recall | F-Measure | ROC Area | Accuracy % |
|---|---|---|---|---|---|---|---|
| SVM | 0.966 | 0.096 | 0.913 | 0.966 | 0.939 | 0.985 | 93.529 |
| Naïve Bayes | 0.920 | 0.217 | 0.816 | 0.920 | 0.865 | 0.949 | 85.294 |
| SGD | 0.920 | 0.108 | 0.899 | 0.920 | 0.909 | 0.906 | 90.588 |
| SMO | 0.943 | 0.120 | 0.891 | 0.943 | 0.916 | 0.911 | 91.176 |
| KNN | 0.897 | 0.133 | 0.876 | 0.897 | 0.886 | 0.902 | 88.235 |
| Random Forest | 0.931 | 0.072 | 0.931 | 0.931 | 0.931 | 0.970 | 92.941 |
| MLP | 0.943 | 0.108 | 0.901 | 0.943 | 0.921 | 0.987 | 91.764 |
| SSC | 0.874 | 0.193 | 0.826 | 0.874 | 0.849 | 0.960 | 84.117 |

## V. CONCLUSION

In this paper, a system for analyzing Tweets generated by Saudi Twitter users about Saudi universities has been presented. There are two proposed different models for Twitter-based sentiment analysis. One of them depended on the use of SVM and the other method depended on using multiple classifier approaches. The obtained results show that the SVM outperformed all other classification models used in this work on the same dataset. In future work, adding a neutral sentiment class will be planned with an increased dataset size.

## CONFLICT OF INTEREST

No conflict of interest.

## AUTHOR CONTRIBUTIONS

Meshrif Alruily supervised the research and did the necessary work to achieve the research goals; he collected the required tweets from 22 Saudi universities. Osama Shahin analyzed and clean all tweets; also, he converted the tweets into separate text files. Both authors discussed the previous researches to find the appropriate way to build a classification model. Osama Shahin built the proposed classification models and he tested the model over new tweets to measure system performance and accuracy using Confusion Matrix. Both authors wrote the paper and they had approved the final version.

## ACKNOWLEDGMENT

We would like to thank Deanship of Scientific Research at Jouf University for financial support of this research project under grant No 39/186.

## REFERENCES

[1] J. Gantz and D. Reinsel, Extracting value from chaos," IDC's Digital Universe Study, vol. 1142, pp. 1–12, 2011.

[2] R. Kulkarni, S. Dhanawade, S. Raut, and D. S. Lavhakare, "Twitter stream analysis for traffic detection in real time," *International Journal of Advance Research, Ideas and Innovations in Technology*, vol. 2, 2016.

[3] E. D'Andrea, P. Ducange, B. Lazzerini, and F. Marcelloni, "Real-time detection of traffic from twitter stream analysis," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 4, pp 2269–2283, Aug. 2015.

[4] N. Elhadad, L. Gravano, D. Hsu *et al.* (2014). Information extraction from social media for public health. [Online]. Available: http://people.dbmi.columbia.edu/noemie/papers/14kdd_bloomberg.pdf

[5] A. Lamb, M. J. Paul, and M. Dredze, "Separating fact from fear: Tracking flu infections on twitter," Presented at NAACL, 2013.

[6] B. Liu, "Sentiment analysis and opinion mining," *Synthesis Lectures on Human Language Technologies*, vol. 5, no. 1, pp. 1–167, 2012.

[7] M. Abdul-Mageed, M. T. Diab, and M. Korayem, "Subjectivity and sentiment analysis of modern standard Arabic," in *Proc. the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, 2011, pp. 587–591.

[8] M. Abdul-Mageed, M. Diab, and K. S. Samar, "Subjectivity and sentiment analysis for Arabic social media," *Comput. Speech Lang.*, vol. 28, no. 1, pp. 20–37, 2014.

[9] M. Elarnaoty, S. A. Rahman, and A. Fahmy, "A machine learning approach for opinion holder extraction in Arabic language," ArXivPrepr. ArXiv:12061011, 2012.

[10] A. Mountassir, H. Benbrahim, and I. Berrada, "An empirical study to address the problem of unbalanced data sets in sentiment classification," in *Proc. the IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2012, pp. 3298–3303.

[11] M. Rushdi-Saleh, M. T. Martín-Valdivia, L. A. Ureña-López, and J. M. Perea-Ortega, "OCA: Opinion corpus for Arabic," *Journal of the American Society for Information Science and Technology*, vol. 62, no. 10, pp. 2045–2054, 2011.

[12] H. A. Wahsheh, M. N. Al-Kabi, and I. M. Alsmadi, "SPAR: A system to detect spam in Arabic opinions," in *Proc. the IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT)*, 2013, pp. 1–6.

[13] H. A. Wahsheh, M. N. Al-Kabi, and I. M. Alsmadi, "A link and content hybrid approach for Arabic web spam detection," *Int. J. Intell. Syst. Appl.*, vol. 5, no. 1, p. 30, 2012.

[14] M. Abdul-Mageed, M. T. Diab, "SANA: A large scale multi-genre, multi-dialect lexicon for Arabic subjectivity and sentiment analysis," in *Proc. the LREC*, 2014, pp. 1162–1169.

[15] M. Nabil, M. Aly, and A. F. Atiya, "ASTD: Arabic sentiment tweets dataset," in *Proc.* the *2015 Conference on Empirical Methods in Natural Language Processing*, pp. 2515–2519.

[16] A. A. Hammad and A. El-Halees, "An approach for detecting spam in Arabic opinion reviews," *Int. Arab J. Inf. Technol.*, vol. 12, no. 1, 2015.

[17] N. El-Makky *et al.*, *Sentiment Analysis of Colloquial Arabic Tweets*, 2015.

[18] M. Alhazmi and N. Salim, "Arabic opinion target extraction from tweets," *ARPN J. Eng. Appl. Sci.*, vol. 10, no. 3, 2015.

[19] B. Naaima, R. Faizi, R. Oulad, H. Thami, and R. Chiheb, "Sentiment analysis in Arabic: A review of the literature," *Ain Shams Engineering Journal*, 2017.

[20] A.-S. Samar, N. Obeid, and A. Awajan. "Sentiment analysis for Arabic text using ensemble learning." in *Proc. 2018 IEEE/ACS 15th International Conference on Computer Systems and Applications (AICCSA)*, 2018, pp. 1–7.

[21] A. Anwar and N. Arıcı. "The corpus based approach to sentiment analysis in modern standard arabic and arabic dialects: A literature review," *Politeknik Dergisi*, 2018.

[22] A.-A. Mahmoud, A. A. Khamaiseh, Y. Jararweh, and M. N. Al-Kabi, "A comprehensive survey of arabic sentiment analysis," *Information Processing & Management*, vol. 56, no. 2, pp. 320–342, 2019.

[23] L. Na, L. Ying *et al.*, "Mixture of topic model for multi-document summarization," in *Proc. The 26th Chinese Control and Decision Conference*, pp. 5168–5172, 2014.

[24] M. Naili, A. H. Chaibi, and H. H. B. Ghezala, "Comparative study of word embedding methods in topic segmentation," *Procedia Computer Science*, vol. 112, pp. 340–349, 2017.

[25] S. Khaled, A. E. Hassanien, and F. Tolba, *Intelligent Natural Language Processing: Trends and Applications*, vol. 740, Springer, 2017.

[26] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep Learning*, Cambridge: MIT Press, vol. 1, 2016.

**Meshrif F. Alruily** is an associate professor at the Department of Computer Science in Jouf University, Saudi Arabia. He received his PhD in computer science from the University of De Montfort, UK, in 2012. He has published many conference papers and journal articles, such as in the European Conference on Artificial Intelligence (ECAI) and the Information Processing & Management Journal. His research interests are related to Arabic text mining, such as information extraction, summarization, text classification and clustering, and data analysis.

**Osama R. Shahin** obtained his BSc degree in communications & computer engineering from Helwan, Egypt, in 2001, MSc and Ph.D degrees in computer science and engineering from Manoufia University, in 2008 and 2015, respectively. He is an assistant professor in the Mathematics Dept. at the Faculty of Engineering in Mattaria, Helwan University, Egypt. His research is in fields of image processing, data mining, and information security. He has published many international papers, mostly in the areas of image processing and information security. Currently, he is an assistant professor at Jouf University, Suadi Arabia.