

Arabic Sentiment Analysis of Users' Opinions of Governmental Mobile Applications

Mohammed Hadwan^{1,2,3,*}, Mohammed A. Al-Hagery⁴, Mohammed Al-Sarem⁵ and Faisal Saeed^{5,6}

¹Department of Information Technology, College of Computer, Qassim University, Buraydah, 51452, Saudi Arabia

²Department of Computer Science, College of Applied Sciences, Taiz University, Taiz, Yemen

³Intelligent Analytic Group, College of Computer, Qassim University, Buraydah, 51452, Saudi Arabia

⁴Department of Computer Science, College of Computer, Qassim University, Buraydah, 51452, Saudi Arabia

⁵College of Computer Science and Engineering, Taibah University, Medina, 42353, Saudi Arabia

⁶Institute for Artificial Intelligence and Big Data, Universiti Malaysia Kelantan, Pengkalan Chepa, Kota Bharu, 16100, Kelantan, Malaysia

*Corresponding Author: Mohammed Hadwan. Email: m.hadwan@qu.edu.sa

Received: 14 January 2022; Accepted: 08 March 2022

Abstract: Different types of pandemics that have appeared from time to time have changed many aspects of daily life. Some governments encourage their citizens to use certain applications to help control the spread of disease and to deliver other services during lockdown. The Saudi government has launched several mobile apps to control the pandemic and have made these apps available through Google Play and the app store. A huge number of reviews are written daily by users to express their opinions, which include significant information to improve these applications. The manual processing and extracting of information from users' reviews is an extremely difficult and time-consuming task. Therefore, the use of intelligent methods is necessary to analyse users' reviews and extract issues that can help in improving these apps. This research aims to support the efforts made by the Saudi government for its citizens and residents by analysing the opinions of people in Saudi Arabia that can be found as reviews on Google Play and the app store using sentiment analysis and machine learning methods. To the best of our knowledge, this is the first study to explore users' opinions about governmental apps in Saudi Arabia. The findings of this analysis will help government officers make the right decisions to improve the quality of the provided services and help application developers improve these applications by fixing potential issues that cannot be identified during application testing phases. A new dataset used for this research includes 8000 user reviews gathered from social media, Google Play and the app store. Different methods are applied to the dataset, and the results show that the k nearest neighbourhood (KNN) method generates the highest accuracy compared to other implemented methods.

Keywords: Arabic sentiment analysis; software quality; user satisfaction; improving online governmental services; machine learning; intelligent systems; mobile app



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1 Introduction

The Arabic language is a morphologically complex language used by more than 422 million people, both native and non-native speakers. Currently, the huge volumes of Arabic reviews that reflect users' opinions about mobile applications (apps) are increasing dramatically. This increase makes it difficult to analyse and extract important information manually, especially for popular apps that have a large number of reviews [1]. To track users' opinions and behaviour towards certain apps, sentiment analysis and machine learning play important roles. Sentiment analysis is a branch of natural language processing that analyses users' attitudes about a specific topic, service, or product to abstract important information expressed in text. According to [2], "Sentiment analysis is a fundamental natural language processing task to automatically analyze raw text and infer from it semantic meaning that focuses on the author's attitude towards the written text". The development of machine learning techniques helps in addressing the issue of opinions, but mostly for reviews written in the English language [3]. However, the Arabic language represents an important area for researchers to investigate the ways that machine learning techniques and sentiment analysis can help to automatically obtain accurate information. Users depend heavily on reviews and ratings before downloading apps, especially when there are other similar options, which makes the analysis of users' opinions vital to application owners and developers [4].

Important information about bugs, improvements, and users' feelings and experiences are available in user reviews [5]. App owners, developers, companies, and governments use the information available in user reviews to better understand users' needs [5]. Manual analysis of user reviews seems to be an unachievable task due to the large number of available reviews; therefore, automated extraction is a good option. The nature of review text, different uses of words, the use of slang and idioms, the different structures of languages, fake reviews, and opinion spam are among the challenges for automated analysis [5].

As a consequence of COVID-19, the government of the Kingdom of Saudi Arabia (KSA) has developed many mobile applications (apps) and made them compulsory (Absher, Sehah, Tabaod, Tataman, Tawakalna, Es'efni). These apps are available in Google Play and the app store and on the official governmental website for users to download and review. Due to the huge number of users, thousands of positive, negative, and neutral reviews are reported regularly based on users' experiences. These reviews are important for improving the apps, but it is time-consuming to examine each reviews independently.

This research aims to help and support the great efforts made by the Saudi government for its citizens and residents by analysing the opinions of people in Saudi Arabia on social media and other platforms (Google Play and the app store) regarding these efforts using intelligent methods such as sentiment analysis and machine learning methods. This analysis will support government officers in decision-making to improve the quality of the provided services and will help application developers improve the developed applications and fix any potential issues.

2 Literature Review

Sentiment analysis in general and Arabic sentiment analysis (ASA) specifically are among the hottest research topics for the scientific community. Many general systematic survey papers have been published recently that surveyed ASA in general, such as [6–9], which covered different aspects related to ASA. Researchers in [10,11] discussed the challenges and issues facing ASA. In [12–15], the authors reviewed the tools and approaches of ASA available in the literature. Twitter has received attention from researchers, as in [16,17]. Some of the research on ASA is dedicated to analysing opinions in

social media (Facebook, Twitter, YouTube), as in [17–21]. Many available datasets for testing proposed methods of ASA for standard Arabic can be found in [19,20,22]. Like other languages, the Arabic language has many local dialects that people use instead of standard Arabic to express their ideas and opinions. Therefore, many researchers have studied ASA for their local dialects, such as the Moroccan [23], Algerian [4], Saudi [24], Jordanian [25], Tunisian [26], Egyptian [27], Iraqi [28], and Yemeni [29] dialects. Opinions and reviews in Google Play and the app store have motivated many researchers to analyse users' opinions of the different mobile applications available for download on these platforms [3,5,30,31]. Some researchers study sentiment analysis at the word level [17,32], while others consider the sentence level, as in [33,34].

For ASA, different popular ML classifiers have been used, such as naïve Bayes [25,28,35,36], K-means, [37,38], K-nearest neighbour (KNN) [18,21,35,39], support vector machine (SVM) [21], [4,25,40], and decision tree (DTree) [28,39,41].

In [4], the researchers analysed Algerian reviews in application stores using ASA. Two approaches were utilized for the analysis: (i) the automatic approach based on machine learning and (ii) the lexicon-based approach. For evaluation, 50000 reviews were collected from popular Algerian applications in the Google Play store. The obtained results were promising as they achieved an accuracy of 80% using the lexicon-based approach and 72% for SVM on dialect reviews.

In [42], the authors discussed sentiment analysis applications that used efficient classification techniques in different domains, such as marketing, health care, and education. With regard to sentiment analysis applications, the issue has been noted that these applications can be removed using mobile phones.

A deep learning model was implemented in [30] on a refined dataset to study the various intricate details of the underlying data. A few machine learning techniques have been implemented, such as naïve Bayes and XGBoost, in addition to deep learning classifiers and multilayer perceptron classifiers (MLPs). Furthermore, functional layers of Keras have been implemented to combine all the features of mobile apps, such as text reviews, numerical features, the total number of reviews, and categorical columns. The performance was evaluated based on the metrics of accuracy and area under the receiver operator characteristic (ROC) curve.

In [43], using textual reviews on Google Play, the authors proposed a system to determine the polarity of sentiments that can be performed on mobile devices. A client server-based system architecture was used where the server performed training and classification tasks, while the clients were mobile devices that performed sentiment analysis tasks that could be run on small-resource mobile devices. Naïve Bayes was used for the developed application, and a linear support vector machine was used for comparison. The accuracy of the naïve Bayes classifier was 83.87%, while the accuracy of SVM was 89.49%. It was reported that the use of semantic handling contributed to reducing the accuracy of the classifiers.

With the aim of identifying the most relevant topics in a document, researchers in [44] used a sentiment analysis approach that included a lexicon-based model for specifying the set of emotions and a statistical methodology that was the target of the sentiments. In addition, a heuristic learning method was used to improve the initial knowledge considering users' feedback. The proposed sentiment analysis approach was integrated into an Android-based mobile app. It automatically assigned sentiments to pictures, taking into account the descriptions provided by the users.

In [45], a system was proposed to model mobile users' feedback behaviours to analyse sentiment trends. The dataset was collected from a popular Chinese mobile application called Toutiao. Few

analysis methods have been proposed for the sentiment of comments, and modelling algorithms have been proposed for feedback behaviours. A system called MoSa was built to identify several implicit behaviour models and hidden sentiment trends. This system and modelling methods provided empirical results to guide interaction design for the mobile internet, social networks, and blockchain-based crowdsourcing.

Naïve Bayes and support-vector machine supervised machine learning algorithms were utilized in [46] to predict sentiments and highly recommended brands based on Twitter tweets. Based on the experiment, the support-vector machine produced more accurate results than naïve Bayes.

In [47], the researchers adopted the keyword co-occurrence measure (SKCM) for Arabic enhanced sentiment analysis. They started with special pre-processing steps followed by SKCM to extract sentiment-based feature selection using an SVM classifier. The results were very promising for enhancing the accuracy of sentiment analysis. In [48], four machine learning techniques were implemented for three Arabic language corpora to increase the accuracy of opinions using rule-based feature selection approaches. The results showed that the proposed approaches yielded better results as different domains were used for the experiments to show the impact of the proposed model on several ML techniques.

In [49,50], the authors introduced a novel feature selection method with voting classifier algorithm to classify the CT images to determine whether COVID-19 is positive or negative. Their proposed voting classifier called PSO-Guided-WOA achieved the best results among other compared methods.

3 Research Methodology

The methodology followed in this research started with the data collection step, in which reviews were gathered from Google Play and the app store for selected governmental applications. Then, extensive pre-processing was implemented on the collected reviews. Subsequently, features were extracted and a corpus was built, followed by the exploration of several machine learning algorithms that were applied to obtain results and calculate accuracy. Fig. 1 shows the methodology followed in this research.

3.1 Dataset Collection

The dataset used in this research was collected from users' reviews of some Saudi governmental applications available on Google Play and the app store. These applications were Tawakkalna, Tetaman, Tabaud, Sehhaty, Mawid, and Sehhah. Only Arabic reviews were gathered. The total number of reviews was **8000** extracted from Google Play and the app store for the mentioned applications. After pre-processing the dataset, **7759** reviews were considered for analysis. Tab. 1 shows the statistics of the reviews based on their applications.

3.2 Dataset Pre-processing

The Arabic language involves many issues, such as morphological complexities and dialectal varieties. Thus, it requires progressive pre-processing and lexicon-building steps. The pre-processing steps are as follows.

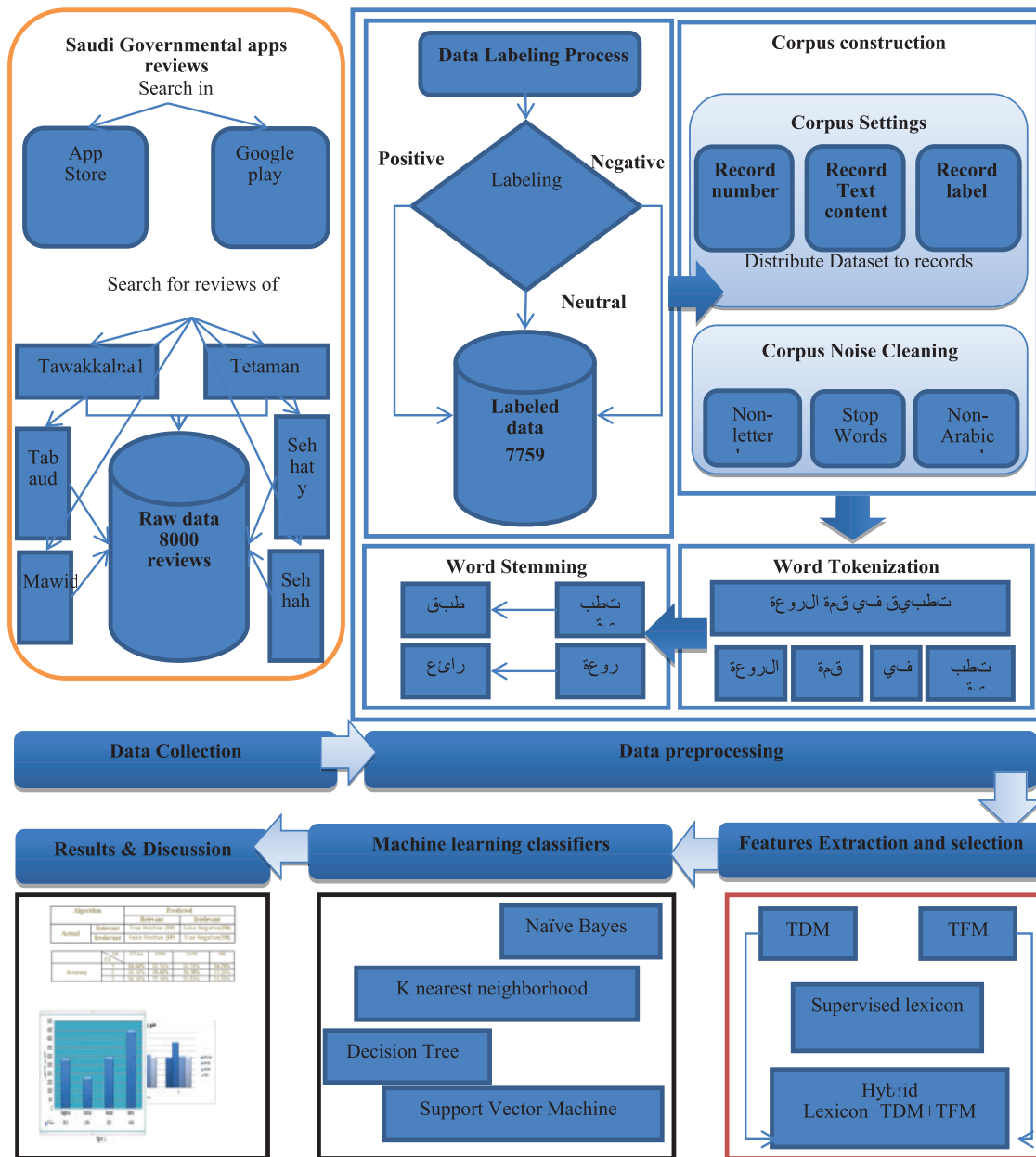


Figure 1: Research methodology

Table 1: Applications (Apps) review statistics

| Applications Name | Reviews source | No of reviews |
|---------------------|---------------------------------|---------------|
| Tawakkalna (توكلنا) | Google Play Store and App Store | 2500 |
| Tetaman (تظمن) | Google Play Store and App Store | 1500 |
| Tabaud (تباعده) | Google Play Store and App Store | 1000 |
| Sehhaty (صحتي) | Google Play Store and App Store | 1000 |
| Mawid (مواعد) | Google Play Store and App Store | 1000 |
| Sehhah (صحة) | Google Play Store and App Store | 1000 |

3.2.1 Dataset Labelling

The process of identifying raw data is called data labelling, where informative labels are added to the dataset to help machine learning algorithms in the learning process. Accurate data labelling is very important for machine learning algorithms to obtain highly accurate results. Polarity classification [51] is used in which labels are classified into three polarities: 1 for positive records, 0 for neutral records, and -1 for negative records. [Tab. 2](#) shows the results of the labelling stage. [Tab. 3](#) shows samples of tree portions from each corpus. [Fig. 2](#) shows dataset statistics.

Table 2: Dataset labelling

| | |
|------------------|-------|
| Total of records | 7759 |
| Negative records | 2913 |
| Positive records | 1834 |
| Neutral records | 3012 |
| Unique words | 16838 |
| Stems | 4548 |

Table 3: Tree portions from each corpus

| Polarity/corpus | Comments | ID |
|-----------------|----------------------------|-----|
| Positive | ”تطبيق في قمة الروعة“ | 2 |
| Negative | ”تطبيق سيئ جدا لا انصح به“ | 455 |
| Neutral | ”الحجز الصحي 14 يوم“ | 300 |

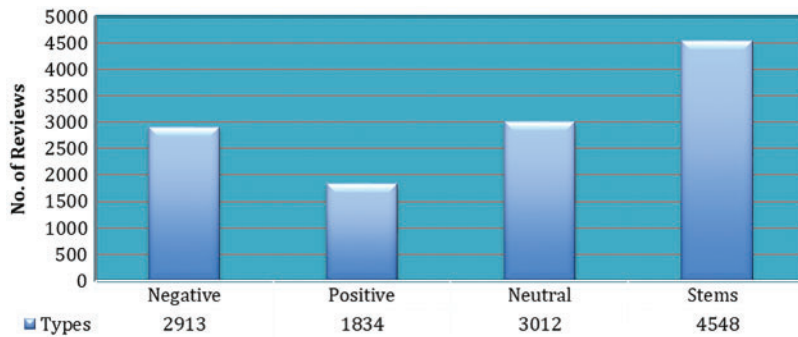


Figure 2: Dataset Statistics

3.2.2 Word Tokenization

The process of splitting a sentence into a list of words is called word tokenization. Each record in the corpus was segmented into small units, which were the words; these words were counted in every record. Counts were registered for each unique word against each record. Every unique word was then grouped depending on its root, i.e., the stem. For example, the tokenization of the Arabic sentence “الروعة تطبيقي في” is as follows: “تطبيقي”, “في”, “رعة”, and “الروعة”.

3.2.3 Dataset Corpus and Noise Cleaning

Any real text scraped or collected from the web includes many undesired aspects of letters or words. Therefore, each record was checked for noise, such as non-letter characters, stop words, and non-Arabic words. By reading plain text sources and distributing a dataset to records, each record was ordered into three components: Record number, record text content, and record label. Class labels are polynomials of three values, as mentioned before. For the previous example and after the noise was removed, the remaining words were as follows: “تطبيقي”, “رعة”, and “الروعة”. Then, the root of these words was used in the word stemming step.

3.2.4 Words Stemming

Word Stemming is the process used to find the root of words where selected words are reduced to their word stems. For example, the root of the words “تطبيقي”, “طبقات”, “أطباق”, “مطبق” and “تطابق” is “طبق”. Stems are used in the feature selection methods. They are the roots of every similar word morphologically. In the Arabic language, roots or stems are mostly three or four letters. The Khoja stemmer algorithm [52] was used for stemming the words in the dataset.

3.3 Feature Selection and Extraction

Feature selection and extraction are a significant part of any sentiment analysis method to filter irrelevant or redundant features in the dataset. After pre-processing, the text must be converted into an understandable form. Hybrid and multiple features selection are trends in the literature of sentiment analysis research as in [53,54], In this research, the following methods were explored for feature selection and extraction.

For each word in the review collection, we calculated a set of linguistic and statistical features using the aforementioned collections and then used machine learning algorithms for term classification

- I) Term Frequency Matrix (TFM) is one of the feature selection methods used in this research. It is the easiest way to map the text into a numerical representation that is used to gather the frequency of each word in the collected reviews. Stems are counted in the overall records, the $n * m$ found matrix consider the n records and number of m words in each record, and the existence of any unique word related to the particular stem is added to the stem count.
- II) Terms-document matrix (TDM). In the binary values version of the TFM, zero values are still zero and larger values become zero.
- III) Terms frequency and inverse document frequency matrix (TF-IDF). A popular method in text mining, extracted from TFM by a specific formula.
- IV) Attributes aggregation (summation and averaging of both TDM and TFM). Four attributes are generated by aggregating TDM and TFM attribute values. The four generated attributes are the TDM Sum, TDM Avg, TFM Sum, and TFM Avg.
- V) TDM K-means Clustering involves clustering TDM by the k-means method, where k is 3, 5, and 8.
- VI) Aggregations and aggregations clustering and TDM K-means clustering involves combining No. V and No. IV with the clustering of No. IV by k-means with $k = 3, 5$ and 8.
- VII) Aggregations clustering and TDM K-means clustering involves combining No. IV with No. V only.
- VIII) Supervised lexicon weights only.
- IX) {TDM, TDM Sum and Avg, TFM Sum and Avg, supervised lexicon weights}
- X) {TDM, supervised lexicon weights}

Features VIII, IX, and X were experimentally selected to be further explored in the next phase (ML classification). These features produce a weight for each record based on a supervised automatic sentiment lexicon. The supervised lexicon is gathered from the corpus, and each record's unique words are grouped according to the label of the record, so we have three lists of words: A negative list, a positive list, and a neutral list. We eliminate neutral words and clear the remaining two lists from shared words, so any negative list word that exists in the positive list is removed from both and vice versa. The final two lists comprise the supervised lexicon. The lexicon is used to produce a weight for each record by the summation of existing positive list words and the subtraction of existing negative list words, which produces an extracted feature that we call a supervised feature. This feature is then combined with traditional TDM and TFM selected features to increase the accuracy of classification.

4 Classification Algorithms

For classification, the four most common machine learning (ML) classifiers are selected to perform the comparative analysis. These classifiers are (i) decision tree, (ii) support vector machine (SVM), K-nearest neighbor (KNN), and Naïve Bayes (NB).

4.1 Regular Naïve Bayes (NB)

The naïve Bayes classifier is a popular classifier that assumes independence between every pair of features. The following Eq. (1) for naïve Bayes is adopted from [55] as follows:

$$y = P(c_k \prod_{i=1}^n P(x_i|c_k)) \quad (1)$$

where y represents the text final class, which is used to match the labels of the data. c_k relates to the class labels, and x_i is a feature vector of d -dimensional.

4.2 Decision Tree (DT) with the Gain Ratio Equation

Another popular ML classifier is a decision tree where each node used denotes a choice among several alternatives and each leaf node presents a classification (decision). DT starts with a root that branches off into several solutions, similar to a tree. DT has been used extensively in ASA.

4.3 K-Nearest Neighbor with $k = 3$

KNN (K-nearest neighbours) is a supervised learning method used as a classifier in machine learning. It looks for the similarity of a given vector to another vector available in the dataset. Two main parameters need to be set for KNN: (i) the k value and (ii) the distance metric. To calculate the distance, the Euclidean function is used, where $K = 3$. The k set to 3 based on numerical analysis shows the best results obtained when $k = 3$. The KNN compared new vectors with k training examples that are its closest neighbour. KNN is among the popular methods used for ASA, as in [18,21].

4.4 Support Vector Machine

SVM is a supervised machine learning algorithm whose main concept is to assign labels to objects based on the learning process through examples. SVM is used extensively for ASA in the literature, as in [18,40,56]. In this work, SVM with sigmoid kernel was used according to the following equations [57]: where w^* denotes the vectors' weight used to specify the hyperplane with maximum margin, $\varphi(x)$ is used for the predefined functions of input vector x, the optimal coefficients are denoted by a_i^* that are determined during the training process, $K()$ is used for the selected kernel function, y specifies the class labels, and parameter b is the bias.

Among the different kernels used with SVM, we have chosen the sigmoid kernel due to its origin from neural networks. The sigmoid function gives values between -1 and 1; therefore, it can classify the predictions based on a particular limit refer to Eq. (2).

$$k(x, y) = \tanh(\mathbf{ax} \cdot \mathbf{y} + \mathbf{b}) \quad (2)$$

where a is alpha and b is the intercept constant. These parameters can be attuned using the kernel parameters a and b.

5 Result

Generally, the performance of a classification algorithm is measured based on its accuracy, recall, precision and F-measure. Accuracy refers to the ratio of the number of accurately estimated samples to the total number of predicted samples based on Tab. 5, accuracy in Eq. (3). Recall refers to the positive samples that are classified correctly divided by the total number of true positives and false negatives available in the dataset, as presented in Eq. (4). Precision refers to positive samples that are classified correctly divided by the total samples labelled positive by the true-positive and false-positive samples as in Eq. (5). The F-measure is the accuracy of the test by considering both the precision and recall of the test based on Eq. (6).

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (4)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (5)$$

$$F\text{-measure} = 2 \times (\text{precision} \times \text{recall}) / (\text{precision} + \text{recall}) \quad (6)$$

where true positive (TP) refers to the number of reviews that are classified correctly and belong to the current class. True negative (TN) refers to the number of reviews that are classified correctly that do not belong to the current class.

False-positive (FP) refers to the number of reviews that are classified mistakenly to belong to the current class. False negative (FN) refers to the number of reviews that are classified mistakenly and do not belong to the current class, as shown in [Tab. 4](#).

Table 4: Correspondence between the actual and predicted classes in the confusion matrix

| Algorithm | | Predicted | |
|-----------|------------|---------------------|---------------------|
| | | Relevant | Irrelevant |
| Actual | Relevant | True Positive (TP) | False Negative (FN) |
| | Irrelevant | False Positive (FP) | True Negative (TN) |

[Tab. 5](#) illustrates the results produced by the classifiers in terms of accuracy, recall, precision, and F-score. [Tab. 6](#) presents a comparison of different classifiers based on execution time.

Table 5: Comparison of different classifiers based on Accuracy, Recall, Precision, and F-Score

| | | DTree | KNN | SVM | NB |
|-----------|------|--------|---------------|--------|--------|
| Accuracy | IIIV | 59.92% | 55.16% | 52.78% | 54.78% |
| | IX | 50.30% | 78.46% | 55.38% | 51.26% |
| | X | 50.30% | 75.14% | 52.69% | 51.26% |
| Recall | IIIV | 59.08% | 52.34% | 48.27% | 51.38% |
| | IX | 51.18% | 78.01% | 50.55% | 57.49% |
| | X | 51.18% | 74.87% | 48.13% | 57.49% |
| Precision | IIIV | 64.05% | 67.69% | 73.83% | 72.38% |
| | IX | 74.48% | 79.94% | 75.82% | 72.64% |
| | X | 74.48% | 76.31% | 73.95% | 72.64% |
| F-Score | IIIV | 61.46% | 59.03% | 58.37% | 60.10% |
| | IX | 60.67% | 78.96% | 60.66% | 64.18% |
| | X | 60.67% | 75.58% | 58.31% | 64.18% |

Table 6: Comparison of different classifiers based on execution time

| | ML FS | DTree | KNN | SVM | NB |
|----------------|----------|-------|-------|-------|------|
| Execution Time | IIIIV | 1 s | 5 s | 9 s | 1 s |
| | IX | 54 s | 12:08 | 30:35 | 23 s |
| | X | 33 s | 11:36 | 32:12 | 46 s |

6 Discussion

Tab. 5 shows the obtained results based on accuracy, recall, precision, and the F-measure. Tab. 6 shows a comparison based on execution time. We can see that KNN using feature IX produced the highest accuracy and obtained 78.46% and 59.92%, 55.38%, and 54.78% for DT, SVM, and NB, respectively. The NB model has the worst accuracy compared to the other methods.

The recall is up to 78.1% for KNN, which is the highest, while the other methods obtained 59.08%, 57.49%, 50.55% for DT, NB, and SVM, respectively. The best precision was 79.94% by KNN, while 75.82%, 74.48%, and 72.64 were obtained by SVM, DTree, and NB, respectively. For the F-measure, is KNN scored 78.69%, whereas NB, DTree, and SVM scored 64.18%, 61.46%, and 60.66%, respectively. The last evaluation criterion was the execution time, where the DT and NB were the fastest methods while SVM was the slowest. Fig. 3 shows the comparison of different classifiers based on accuracy.

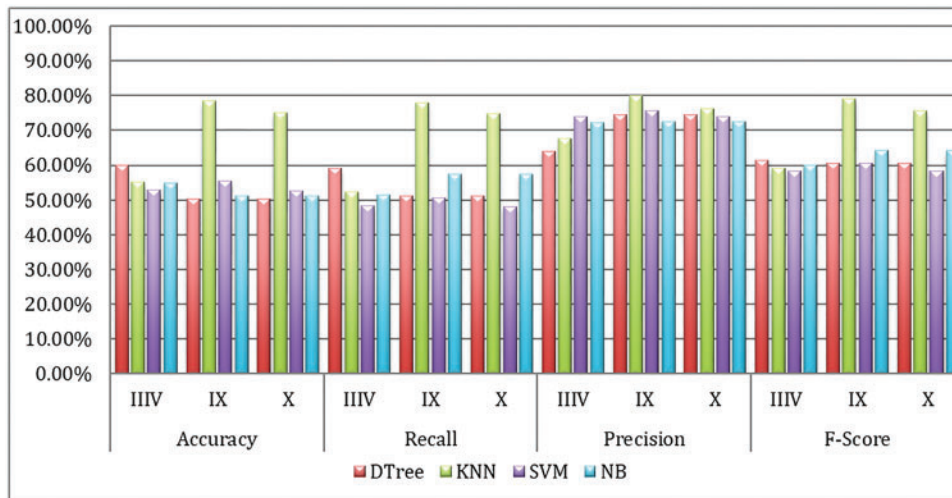


Figure 3: Comparison of different classifiers based on accuracy for ASA

7 Conclusion and Future Work

This paper introduced a proposed hybrid feature selection method used for Arabic sentiment analysis to extract users' opinions of Saudi governmental applications for COVID-19. A new Arabic dataset was developed that includes 7759 reviews collected from Google Play and the app store. There were many challenges for the collected dataset because it was in the Arabic language. Therefore, extensive pre-processing steps were utilized to prepare the data for use for ML classifiers. This was

done by analysing the gathered reviews to label them as positive, negative, and neutral sentiments and performing necessary data cleaning. Four well-known classifier methods were implemented (DTree, SVM, KNN, and NB). The results showed that KNN outperformed the other methods with accuracy of 78.46% compared to 59.92%, 55.38%, and 54.78% for DT, SVM, and NB, respectively. The NB model had the worst accuracy compared to the other methods.

Acknowledgement: Researchers would like to thank the Deanship of Scientific Research, Qassim University for the support of this project.

Funding Statement: The authors gratefully acknowledge Qassim University, represented by the Deanship of Scientific Research, on the financial support for this research under the number (10278-coc-2020-1-3-I) during the academic year 1441 AH/2020 AD.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] R. E. Saady, E. S. Nasr, A. E. M. El-Ghazaly and M. H. Gheith, "Use of Arabic sentiment analysis for mobile applications' requirements evolution: Trends and challenges," in *Advances in Intelligent Systems and Computing*, Cham, Switzerland: Springer, pp. 477–487, 2018.
- [2] L. H. Baniata and S.-B. Park, "Sentence representation network for Arabic sentiment analysis," in *The 43rd Annual Meeting and Winter Conf.*, Gangwon-do, South Korea, pp. 470–472, 2017.
- [3] S. Sari and M. Kalender, "Sentiment analysis and opinion mining using deep learning for the reviews on Google Play," *Innovations in Smart Cities Applications*, vol. 183, no. 1, pp. 1–12, 2021.
- [4] A. Chader, L. Hamdad and A. Belkhiri, "Sentiment analysis in google play store: Algerian reviews case," in *Int. Symp. on Modelling and Implementation of Complex Systems*, Batna, Algeria, Springer, pp. 107–121, 2021.
- [5] N. Genc-Nayebi and A. Abran, "A systematic literature review: Opinion mining studies from mobile app store user reviews," *Journal of Systems and Software*, vol. 125, no. 1, pp. 207–219, 2017.
- [6] A. Ghallab, A. Mohsen and Y. Ali, "Arabic sentiment analysis: A systematic literature review," *Applied Computational Intelligence and Soft Computing*, vol. 2020, no. 1, pp. 1–21, 2020.
- [7] M. Korayem, D. Crandall and M. Abdul-Mageed, "Subjectivity and sentiment analysis of Arabic: A survey," in *Communications in Computer and Information Science*, Cairo, Egypt: Springer, pp. 128–139, 2012.
- [8] A. AlOwisheq, S. AlHumoud, N. AlTwaresh and T. AlBuhairi, "Arabic sentiment analysis resources: A survey," in *Lecture Notes in Computer Science*, Toronto, Canada, Springer International Publishing, pp. 267–278, 2016.
- [9] A. Assiri, A. Emam and H. Aldossari, "Arabic sentiment analysis: A survey," *International Journal of Advanced Computer Science and Applications*, vol. 6, no. 12, pp. 75–85, 2015.
- [10] A. Hamdi, K. Shaban and A. Zainal, "A review on challenging issues in Arabic sentiment analysis," *Journal of Computer Science*, vol. 12, no. 9, pp. 471–481, 2016.
- [11] Y. Zahidi, Y. Elyounoussi and Y. Al-Amrani, "Arabic sentiment analysis problems and challenges," in *Proc. - 10th Int. Conf. on Virtual Campus, JICV 2020*, Tetouan, Morocco, IEEE, pp. 1–4, 2020.
- [12] Y. Zahidi, Y. Elyounoussi and Y. Al-Amrani, "Different valuable tools for Arabic sentiment analysis: A comparative evaluation," *International Journal of Electrical and Computer Engineering*, vol. 11, no. 1, pp. 735–762, 2021.
- [13] G. S. Kaseb and M. F. Ahmed, "Arabic sentiment analysis approaches: An analytical survey," *International Journal of Scientific & Engineering Research*, vol. 7, no. 10, pp. 712–723, 2016.

- [14] S. O. Alhumoud and A. A. Al Wazrah, "Arabic sentiment analysis using recurrent neural networks: A review," *Artificial Intelligence Review*, vol. 55, no. 1, pp. 707–748, 2022.
- [15] H. Mulki, H. Haddad and I. Babaoglu, "Modern trends in Arabic sentiment analysis: A survey," *Revue Traitement Automatique des Langues*, vol. 58, no. 3, pp. 15–39, 2018.
- [16] M. A. Ibrahim and N. Salim, "Opinion analysis for twitter and Arabic tweets: A systematic literature review," *Journal of Theoretical and Applied Information Technology*, vol. 56, no. 3, pp. 338–348, 2013.
- [17] B. Ihnaini and M. Mahmuddin, "Lexicon-based sentiment analysis of Arabic tweets: A survey," *Journal of Engineering and Applied Sciences*, vol. 13, no. 17, pp. 1–14, 2018.
- [18] A. S. A. AL-Jumaili and H. K. Tayyeh, "A hybrid method of linguistic and statistical features for Arabic sentiment analysis," *Baghdad Science Journal*, vol. 17, no. 1, pp. 385–390, 2020.
- [19] K. Abu Kwaik, S. Chatzikiyriakidis, S. Dobnik, M. Saad and R. Johansson, "An Arabic tweets sentiment analysis dataset using distant supervision and self training," in *Proc. of the 4th Workshop on Open-Source Arabic Corpora and Processing Tools, with a Shared Task on Offensive Language Detection*, Marseille, France, pp. 1–8, 2020.
- [20] M. Nabil, M. Aly and A. F. Atiya, "ASTD: Arabic sentiment tweets dataset," in *Conf. Proc.-EMNLP 2015: Conf. on Empirical Methods in Natural Language Processing*, Lisbon, Portugal, Association for Computational Linguistics, pp. 2515–2519, 2015.
- [21] A. K. Al-Tamimi, A. Shatnawi and E. Bani-Issa, "Arabic sentiment analysis of YouTube comments," in *2017 IEEE Jordan Conf. on Applied Electrical Engineering and Computing Technologies, AEECT 2017*, Aqaba, Jordan, IEEE, pp. 1–6, 2017.
- [22] A. J. S. Al Mukhaiti, S. Siddiqui and K. Shaalan, "Dataset built for arabic sentiment analysis," in *Advances in Intelligent Systems and Computing*, Cairo, Egypt: Springer, pp. 406–416, 2018.
- [23] A. Oussous, F. Z. Benjelloun, A. A. Lahcen and S. Belfkih, "ASA: A framework for Arabic sentiment analysis," *Journal of Information Science*, vol. 46, no. 4, pp. 1–16, 2020.
- [24] W. A. Al-Harbi and A. Emam, "Effect of Saudi dialect preprocessing on Arabic sentiment analysis," *International Journal of Advanced Computer Technology*, vol. 4, no. 6, pp. 1–14, 2015.
- [25] J. O. Atoum and M. Nouman, "Sentiment analysis of Arabic Jordanian dialect tweets," *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 2, pp. 256–262, 2019.
- [26] S. Medhaffar, F. Bougares, Y. Estève and L. Hadrich-Belguith, "Sentiment analysis of Tunisian dialects: Linguistic resources and experiments," in *Proc. of The Third Arabic Natural Language Processing Workshop*, Valencia, Spain, Association for Computational Linguistics, pp. 55–61, 2017.
- [27] N. El-Naggar, Y. El-Sonbaty and M. Abou El-Nasr, "Sentiment analysis of modern standard Arabic and Egyptian dialectal Arabic tweets," in *Proc. of Computing Conf. 2017*, London, UK, pp. 880–887, 2018.
- [28] A. Alnawas and N. Arici, "Sentiment analysis of Iraqi Arabic dialect on Facebook based on distributed representations of documents," *ACM Transactions on Asian and Low-Resource Language Information Processing*, vol. 18, no. 3, pp. 1–17, 2019.
- [29] W. M. S. Yafooz, E. A. Hizam and W. A. Alromema, "Arabic sentiment analysis on chewing Khat leaves using machine learning and ensemble methods," *Engineering Technology & Applied Science Research*, vol. 11, no. 2, pp. 6845–6848, 2021.
- [30] S. Venkatakrishnan, A. Kaushik and J. K. Verma, "Sentiment analysis on Google Play store data using deep learning," *Applications of Machine Learning*, vol. 1, no. 1, pp. 15–30, 2020.
- [31] S. Sadiq, M. Umer, S. Ullah, S. Mirjalili, V. Rupapara *et al.*, "Rupapara et al, Discrepancy detection between actual user reviews and numeric ratings of Google App store using deep learning," *Expert Systems with Applications*, vol. 181, no. 3, pp. 1–11, 2021.
- [32] S. Siddiqui, A. A. Monem and K. Shaalan, "Evaluation and enrichment of arabic sentiment analysis," *Studies in Computational Intelligence*, vol. 740, pp. 17–34, 2018.
- [33] J. Messias, J. P. Diniz, E. Soares, M. Ferreira, M. Araujo *et al.*, "An evaluation of sentiment analysis for mobile devices," *Social Network Analysis and Mining*, vol. 7, no. 1, pp. 2–18, 2017.
- [34] A. Shoukry and A. Rafea, "Sentence-level Arabic sentiment analysis," in *Proc. of the 2012 Int. Conf. on Collaboration Technologies and Systems, CTS 2012*, Denver, CO, USA, IEEE, pp. 546–550, 2012.

- [35] R. M. Duwairi and I. Qarqaz, "A framework for Arabic sentiment analysis using supervised classification," *International Journal of Data Mining, Modelling and Management*, vol. 8, no. 4, pp. 369–381, 2016.
- [36] M. Alassaf and A. M. Qamar, "Improving sentiment analysis of Arabic tweets by One-way ANOVA," *Journal of King Saud University—Computer and Information Sciences*, vol. 1, no. 0, pp. 1–11, 2021.
- [37] M. A. Alanezi and N. M. Hewahi, "Tweets sentiment analysis during COVID-19 pandemic," in *2020 Int. Conf. on Data Analytics for Business and Industry: Way Towards a Sustainable Economy, ICDABI 2020*, Sakheer, Bahrain, IEEE, pp. 1–6, 2020.
- [38] S. Alosaimi, M. Alharthi, K. Alghamdi, T. Alsubait and T. Alqurashi, "Sentiment analysis of arabic reviews for Saudi hotels using unsupervised machine learning," *Journal of Computer Science*, vol. 16, no. 9, pp. 1258–1267, 2020.
- [39] M. E. M. Abo, N. Idris, R. Mahmud, A. Qazi, I. E. T. Hashem *et al.*, "A multi-criteria approach for arabic dialect sentiment analysis for online reviews: Exploiting optimal machine learning algorithm selection," *Sustainability (Switzerland)*, vol. 13, no. 18, pp. 10018, 2021.
- [40] S. Alhumoud, "Arabic sentiment analysis using deep learning for COVID-19 twitter data," *International Journal of Computer Science and Network Security*, vol. 20, no. 9, pp. 132–138, 2020.
- [41] Y. Al-amrani, M. Lazaar and K. Eddine, "Sentiment analysis using hybrid method of support vector machine and decision tree," *Journal of Theoretical and Applied Information Technology*, vol. 96, no. 7, pp. 1886–1895, 2018.
- [42] M. Sivakumar and U. S. Reddy, "A short review for mobile applications of sentiment analysis on various domains," in *Advances in Intelligent Systems and Computing*, Las Vegas, NV, USA: Springer, pp. 723–726, 2015.
- [43] L. B. Ilmawan and E. Winarko, "Aplikasi mobile untuk analisis sentimen pada Google Play," *Indonesian Journal of Computing and Cybernetics Systems*, vol. 9, no. 1, pp. 53–64, 2015.
- [44] D. Griol, J. M. Molina, A. Sanchis and Z. Callejas, "An approach to sentiment analysis for mobile speech applications," in *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Seville, Spain: Springer-Verlag, pp. 96–107, 2016.
- [45] Y. Zhang, W. Ren, T. Zhu and E. Faith, "MoSa: A modeling and sentiment analysis system for mobile application big data," *Symmetry*, vol. 11, no. 115, pp. 1–16, 2019.
- [46] S. Bhanap and S. Kawthekar, "Prediction of high recommendation mobile brands using sentiment analysis," in *Advances in Intelligent Systems and Computing*, Nanded, India: Springer, pp. 187–192, 2021.
- [47] A. Abdelwahab, F. Alqasemi and H. Abdelkader, "Enhancing the performance of sentiment analysis supervised learning using sentiments keywords based technique," in *Second Int. Conf. on Computer Science, Guangzhou, China, Information Technology and Applications*, pp. 107–116, 2017.
- [48] K. M. Alalayah, I. M. Alwayle, F. A. Alqasemi and N. A. Al-Majmar, "Feature selection optimization for highlighting opinions using supervised and unsupervised learning on Arabic language," *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 10, no. 2, pp. 636–642, 2021.
- [49] E.-S. M. El-Kenawy, A. Ibrahim, S. Mirjalili, M. M. Eid and S. E. Hussein, "Novel feature selection and voting classifier algorithms for COVID-19 classification in CT images," *IEEE Access*, vol. 8, pp. 179317–179335, 2020.
- [50] E.-S. M. El-kenawy, S. Mirjalili, A. Ibrahim, M. Alrahmawy, M. El-Said *et al.*, "Advanced meta-heuristics, convolutional neural networks, and feature selectors for efficient COVID-19 X-ray chest image classification," *IEEE Access*, vol. 9, pp. 36019–36037, 2021.
- [51] E. Cambria, "Affective computing and sentiment analysis," *IEEE Intelligent Systems*, vol. 31, no. 2, pp. 102–107, 2016.
- [52] S. Khoja. and R. Garside, "Stemming Arabic text," *Computing Department, Lancaster University, Lancaster, UK*, vol. 1, no. 1, pp. 1–20, 1999.
- [53] M. S. F. Alharbi and E.-S. M. El-kenawy, "Optimize machine learning programming algorithms for sentiment analysis in social media," *International Journal of Computer Applications*, vol. 174, no. 25, pp. 0975–8887, 2021.

- [54] M. Al-Sarem, F. Saeed, Z. G. Al-Mekhlafi, B. A. Mohammed, M. Hadwan *et al.*, “An Improved multiple features and machine learning-based approach for detecting Clickbait news on social networks,” *Applied Sciences*, vol. 11, no. 20, pp. 1–15, 2021.
- [55] S. Raschka, “Naive Bayes and text classification I—introduction and theory,” *ArXiv*, vol. 1410, no. 5329, pp. 1–20, 2014.
- [56] A. Elhawil, Y. Trabelsi and M. Mahfoud, “Comparison between the NB and SVM methods for multiclass Arabic sentiment analysis,” in *IEEE 1st Int. Maghreb Meeting of the Conf. on Sciences and Techniques of Automatic Control and Computer Engineering MI-STA*, Tripoli, Libya, IEEE, pp. 913–917, 2021.
- [57] J. K. Alwan, A. J. Hussain, D. H. Abd, A. T. Sadiq, M. Khalaf *et al.*, “Political Arabic articles orientation using rough set theory with sentiment lexicon,” *IEEE Access*, vol. 9, no. 1, pp. 24475–24484, 2021.