Online-Journals.org

iJIM International Journal of Interactive Mobile Technologies

iJIM | eISSN: 1865-7923 | Vol. 18 No. 12 (2024) | OPEN ACCESS

https://doi.org/10.3991/ijim.v18i12.47197

PAPER

Machine Learning Algorithms for Attitude Prediction from Arabic Text: Detecting Student Attitude towards Online Learning

Esra'a Alshdaifat(✉), Ala'a Al-shdaifat, Ayoub Alsarhan

Department of Information Technology, The Hashemite University, Zarqa, Jordan

esraa@hu.edu.jo

ABSTRACT

Due to its flexibility, accessibility and the increasing importance of digital literacy, online learning has gained priority in the last few years. However, several challenges have led students to resist it. Predicting students' attitudes towards online learning could assist educators and educational institutions in addressing these challenges and enhancing its effectiveness. This paper presents a Learning Attitude Prediction Model (LAPM) that can detect students' attitudes from their informal Arabic texts. To generate the desired LAPM, five machine learning algorithms and three different approaches for text representation are employed. In addition, handling stop words is another important issue when dealing with informal Arabic text. Two scenarios are commonly adopted: preserving and eliminating stop words. The best result was obtained when using a support vector machine (SVM) classifier coupled with the term frequency-inverse document frequency (TF-IDF) approach and preserving stop-words, achieving an F1-score of 85.4%. Therefore, an effective LAPM could be developed to predict students' attitudes. Using LAPM, educators and educational institutions can monitor students' attitudes toward online learning and provide personalized support to individual students. Consequently, an enhancement in student satisfaction and an improvement in academic achievement could be achieved.

KEYWORDS

text mining, Arabic text, student attitude, online learning, sentiment analysis and opinion mining

1 INTRODUCTION

In the last few years, online learning, also referred to as "e-learning," has been widely adopted in education. It has not been considered an alternative to traditional learning; however, it has become a fundamental component of the educational process. Many factors have contributed to the growth of online learning in recent years,

Alshdaifat, E., Al-shdaifat, A., Alsarhan, A. (2024). Machine Learning Algorithms for Attitude Prediction from Arabic Text: Detecting Student Attitude towards Online Learning. *International Journal of Interactive Mobile Technologies (iJIM)*, 18(12), pp. 42–56. https://doi.org/10.3991/ijim.v18i12.47197

Article submitted 2023-12-06. Revision uploaded 2024-02-14. Final acceptance 2024-02-17.

such as: (i) the flexibility of online learning, which meets the demands of the current busy life; (ii) the COVID-19 pandemic, which demonstrated the value of online learning in the face of unexpected disruptions in face-to-face education; (iii) technological evolution, starting with the availability of high-speed internet and portable devices and culminating in effective platforms and software applications; and (iv) providing students with the opportunity to achieve their educational and career objectives. In contrast, there are several issues related to online learning that limit its effectiveness, such as: (i) the absence of face-to-face interactions among students and between students and educators, (ii) reduced enthusiasm and engagement in the learning process, and (iii) issues related to time management.

Hence, learners may display different attitudes towards online learning. Identifying these attitudes enables educators and institutions to: (i) improve the overall online education process, (ii) offer tailored support to individual students, (iii) respond to students' concerns, thus creating a successful learning environment and improving student outcomes, and (iv) proceed with online learning as a feasible learning mode. Consequently, developing an attitude detection model has become a necessity for educators and educational institutions. One way to achieve this objective and discover students' attitudes toward online learning is by analyzing their written texts. Several methods can be used to collect students' texts, including surveys and questionnaires, discussion forums, social media, and course evaluation forms. Some researchers attempted to construct a prediction model for detecting student attitudes based on sentiment analysis instead of solely analyzing questionnaire data and obtaining static observations to enhance the learning process [1]. The work on predicting student attitudes can be differentiated based on the data used, the text representation approach adopted, and the machine learning algorithms employed.

In this paper, a Learning Attitude Prediction Model (LAPM) is fully described. LAPM is considered an opinion classification model that can be utilized by educational institutions and educators to enhance the online learning process. The proposed research methodology utilizes several classification algorithms to generate the desired LAPM. No matter which classification algorithm is employed, an essential issue is how to represent the text using a set of features. This issue is addressed by considering three different approaches: (i) the bag of words approach, (ii) the term frequency-inverse document frequency approach, and (iii) the bidirectional encoder representations from transformers (BERT) approach. In addition, due to the unique nature of the informal Arabic language, various pre-processing techniques are being explored.

The contribution of the work presented in this paper can be summarized as follows:

- Collecting a novel Arabic text dataset from students at the Hashemite University in Jordan and utilizing it to generate a potential attitude prediction model. Given the lack of freely available resources for Arabic opinion mining datasets.
- Combining various text representation approaches and machine learning algorithms to create the most effective Learning Attitude Prediction Model.
- Employing various text pre-processing techniques to handle Arabic text in order to determine the most effective one for generating the Learning Attitude Prediction Model.

The rest of this paper is organized as follows: Section 2 provides a review of related work on opinion mining, including a brief description of the text representation approaches and classification algorithms utilized in this study.

Section 3 discusses the nature of the proposed LAPM. Section 4 presents the experimental setup and the evaluation results obtained. Section 5 concludes the work presented in this paper.

2 LITERATURE REVIEW

This section is organized as follows: Sub-section 2.1 presents a review of related work on opinion mining; Sub-section 2.2 explains an explanation of the adopted text representation approaches; and Sub-section 2.3 provides a brief description of the machine learning algorithms utilized.

2.1 Related work on opinion mining

Opinion mining is the process of automating the extraction of sentiments expressed by users from unstructured texts [2]. The predominance of English as a global language results in a significant amount of research and resources available for English opinion mining [3]. On the other hand, the lack of freely available specialized resources and datasets for Arabic opinion mining has motivated researchers to contribute to the development of essential resources and methodologies for this specific linguistic context [4]. Handling Arabic text can be challenging due to the extensive morphological changes that Arabic words can undergo, such as the addition of suffixes, prefixes, and infixes. This results in variations of words that express different aspects of sentiment or meaning [5, 6, 7]. Improving sentiment analysis for the Arabic language is a critical and ongoing research area. Addressing the linguistic complexities, dialectal variations, and resource limitations is essential to developing robust sentiment analysis models tailored for Arabic content [8]. The field of opinion mining in Arabic text (OMA) is particularly timely and relevant, given the widespread use of Arabic in online content [9]. As a result, many researchers focused their work on predicting Arabic sentiments in various application domains [10].

Commencing with the work conducted by Najar and Mesfar [11], a media monitoring system for opinion mining in the political field, particularly in Arabic-speaking countries, was implemented. The system employed a set of local grammars to identify various structures within political opinion phrases. These grammars utilized an opinion lexicon containing opinion-related words (verbs, adjectives, and nouns) with associated semantic markers for polarity and intensity. According to the experiments, the extraction method used in the system produced an F1-score of 83%.

Another attempt to predict opinions is proposed by Elzayady et al. [12]. They developed several models using deep learning techniques (LSTM and CNN) and traditional machine learning algorithms (KNN, Naïve Bayes, and decision trees (DT)). The embraced diversity enables a thorough evaluation of various techniques and their appropriateness for detecting opinions from customer reviews in Arabic texts. The most notable result from this work is the high accuracy achieved by the combined CNN-LSTM model. An average accuracy of 85.83% and 86.88% was achieved for the two datasets under consideration.

Al-Mutawa et al. [13] utilized an ensemble-based deep learning approach that integrated text embeddings and a newly introduced Arabic emoji and emoticon opinion lexicon to forecast user opinions, especially concerning Arabic hotel reviews. The publicly available Arabic HARD dataset was used for testing. It contains hotel reviews along with ratings ranging from one to five stars. The proposed approach resulted in a significant increase in prediction accuracy compared to other studies using the HARD dataset. More specifically, the accuracy increased by 3.21% for the balanced HARD dataset and 2.17% for the unbalanced HARD dataset.

Al-Obaidi and Samawi [14] focused on sentiment analysis for reviews and comments in various Arabic dialects. Their objective was to ascertain whether the sentiments conveyed in these dialects were positive or negative. A corpus of reviews written in two specific Arabic dialects, Jordanian and Saudi, was collected. Two feature extraction (FE) approaches were considered: (i) bag of words (BOW) and (ii) N-grams of words. Three machine learning algorithms were adopted: (i) Naïve-Bayes, (ii) support vector machine (SVM), and (iii) maximum entropy (ME). The reported results indicated that the ME algorithm outperformed the other two machine learning algorithms (Naïve-Bayes and SVM). Additionally, utilizing N-grams (with N = 3) as a FE method enhanced the performance of all three machine learning algorithms.

A study conducted by Guo et al. [15] shows some similarities with the work presented in this paper. An attempt is being made to predict students' attitudes toward blended learning. The students' data from a C++ programming course was used to generate the potential model. Several categorical, numerical, and textual data sets were used. Categorical and numerical features were extracted from student records, while textual data was obtained from student comments and discussions. Textual data was converted into numerical sentiment scores. All the acquired features were inputted into machine learning algorithms. According to the reported experimental results, the SVM classifier was the most effective classifier based on the F1-score evaluation measure, with a score of 75.1%. This research contributed to the ongoing efforts to enhance the quality of blended learning experiences by providing insights into students' sentiments and attitudes. Note that the dataset under consideration was in the English language.

From the foregoing, the work on opinion mining can be categorized based on:

- The considered dataset. More specifically, the language and the source of the data.
- The machine learning algorithms were used to generate the desired model.
- The adopted text representation approach.
- The pre-processing techniques used.

Regarding the work presented in this paper, a novel dataset was used, which can be considered an online course evaluation survey. Several machine learning algorithms, different text representation approaches, and various pre-processing techniques were utilized to effectively handle slang Arabic text and generate the optimal Learning Attitude Prediction Model.

2.2 Overview of text representation approaches

This section provides a general overview of the text representation approaches that were considered in the work presented in this paper.

Commencing with the BOW approach, which is considered a straightforward method for text representation, this approach handles documents or sentences as unordered sets of words. It neglects the order and sequence of words in the document, focusing on the presence or frequency of individual words. In detail, each document is represented as a vector where each feature corresponds to a unique word, and the value in each feature indicates the presence or frequency of that word in the document [16]. BOW is commonly used because of its simplicity and efficiency in the classification process.

With respect to the TF-IDF approach, it is a widely used method that takes into account both the frequency of a word in the document or sentence and its significance in the entire corpus [17]. In particular, its calculations involve two main parts: (i) term frequency (TF) and (ii) inverse document frequency (IDF). The TF part measures how often a specific term or phrase appears within a given document. It is calculated by counting the occurrences of the term within the document. The IDF part estimates the importance of a term in the entire corpus of documents. It is calculated by dividing the total number of documents by the number of documents that contain the term [18]. In TF-IDF, the TF and IDF values are multiplied together to calculate a score or weight for each term. Terms that are associated with high scores appear frequently in a few documents, but not in all documents. Low scores are associated with common terms that appear in most documents. Thus, the main goal of the TF-IDF approach is to detect terms that are essential or significant within a document. Terms with high TF-IDF scores are typically deemed significant within the context of that particular document.

Regarding BERT, it is a language representation model that serves as a baseline for various natural language processing tasks [19]. BERT can be utilized in two main ways: FE and fine-tuning (FT). With respect to the FE approach, the structure and parameters of the BERT model remain unchanged. Features are extracted from the BERT model and then inputted into machine learning algorithms to create models, including classification models. The contextualized word representations generated by the BERT are called BERT embeddings, where the semantic meaning of words in relation to a given sentence is considered [19]. Regarding FT, it involves adding extra layers to the original BERT architecture and training the model on downstream tasks. In this approach, the model's parameters are adjusted to fit the specific classification task [19]. Several BERT models have been pre-trained to support the Arabic language. These include a multilingual model developed by Devlin's team [19], Arabert by Antoun et al. [20], and MARBERT by Abdul-Mageed et al. [21]. The models vary in terms of the size of the training data used and the approach adopted. AraBERT is a language model specifically designed for the Arabic language. It differs from Multilingual BERT, which is designed for multiple languages, in that it focuses exclusively on Arabic text. Regarding the work presented in this paper, AraBERT is utilized [20].

2.3 Overview of machine learning algorithms

Text classification is considered one of the machine learning tasks. More specifically, machine learning algorithms are used to construct a model that can automatically assign class labels to text documents based on their content. An overview of the machine learning algorithms used in the work presented in this paper is provided in this section.

Support vector machine. Support vector machine is a supervised machine learning algorithm originally developed to handle binary classification tasks, where the dataset contains only two classes [22]. In the context of SVM, data points are represented as tuples, where "x" represents the feature values and "y" represents the class label. In a multi-dimensional feature space, SVM attempts to define a hyperplane that effectively separates data points into distinct classes. The best hyperplane is the one that ensures the maximum margin, which is the distance between the hyperplane and the nearest data points from each class [23].

Decision tree. Decision trees are widely used in machine learning due to their simplicity, effectiveness in modeling and classification tasks, and versatility in handling features [24]. DT classifiers are built based on a tree-like structure. In this structure, class labels are represented by the leaves of the tree. Features are represented by internal nodes. Accordingly, the branches represent the combination of features that lead to the class labels. To construct the DT model, recursive binary partitioning of the feature space is applied. At each node, the dataset is divided into two subsets based on the value of a feature. This process continues recursively until a stopping criterion is encountered, such as reaching a maximum depth or having no more features or data points.

K-nearest neighbor. K-nearest neighbor is a simple and easy-to-implement machine learning algorithm. The key concept of KNN is that similar data points in a feature space are close to each other. The algorithm commences with determining: (i) the value of k, which represents the number of nearest neighbors to consider. It is important to note that the value of k significantly impacts the algorithm's performance. (ii) The distance measure to assess the similarity between examples. Euclidean distance and Manhattan distance are commonly used for this purpose [24]. Then, for classifying a new unseen example, KNN computes the distance between the new example and all examples stored in the training dataset. After that, it identifies the k closest examples based on the chosen distance metric. Finally, the main class label is assigned to the new example.

Logistic regression. Logistic regression is a binary machine learning algorithm where a given example is classified into one of two class labels. More specifically, it estimates the conditional probability $P(y|x)$, which represents the probability of an example x belonging to a particular class label y, by extracting a set of features from the input data and combining them linearly. The linear combination involves multiplying each feature by a weight and summing them up [25].

Bagging. Bagging is an ensemble learning technique that constructs a group of models and combines their predictions to make a collective decision [24]. For constructing the ensemble, a bootstrap sampling technique is used to generate several variations of the dataset by randomly selecting examples with replacements. Multiple classifiers are then learned, each based on a different data variation, resulting in diverse models. For classifying a new example, the individual predictions of the models are combined using a majority voting approach. Here, the class label that receives the majority of votes is assigned to the example.

3 THE ADOPTED EXPERIMENTAL METHODOLOGY

This section presents the development of the desired learning attitude prediction model. As noted earlier, the LAPM is a form of opinion mining. The generation of such a model involves several phases, starting with collecting the dataset and concluding with evaluating the resulting models. Figure 1 summarizes the methodology adopted to construct the potential LAPM. From the figure, the adopted methodology consists of five main phases, which are detailed as follows:

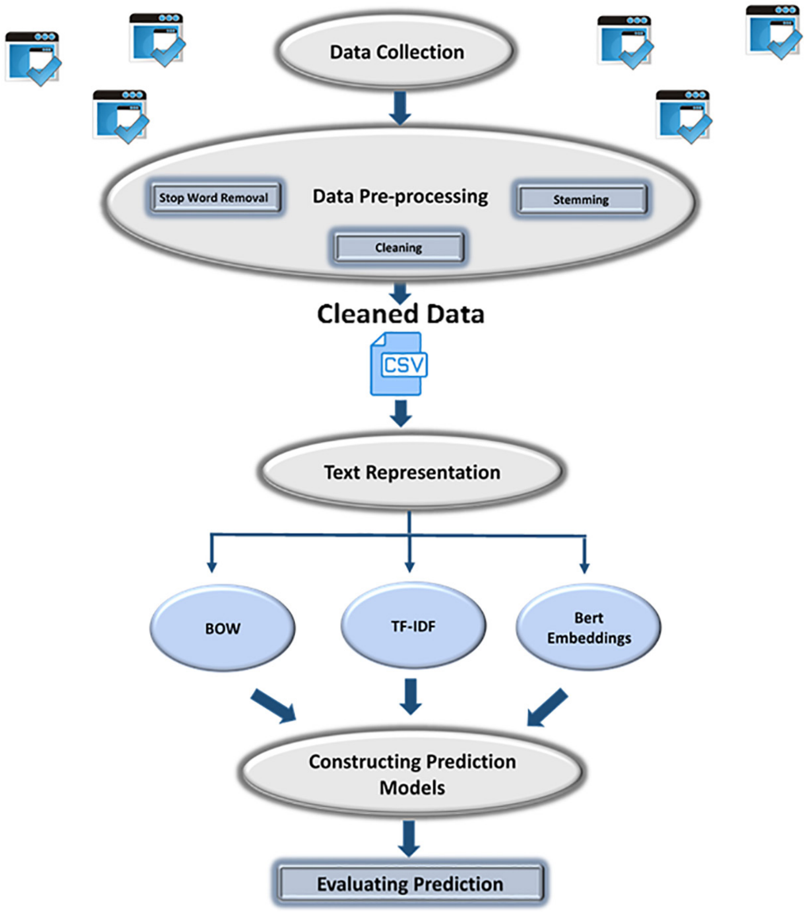

Fig. 1. The adopted methodology to construct the potential LAPM

Phase 1: Data collection. The dataset under consideration was gathered through a survey that asked students to share their opinions on online learning in Arabic. The collected surveys, which contain slang and standard Arabic text, were categorized into positive and negative attitudes. The dataset consists of 523 samples, with 364 associated with positive labels and 159 associated with negative labels.

Phase 2: Data pre-processing. This phase consists of three main steps: (i) text cleaning, (ii) stop-word elimination, and (iii) stemming. Commencing with text cleaning which involves removing noisy data, punctuation marks, and special characters. The next step is removing stop-words. Note that stop-words are frequent words in a language that have little to no meaning in a given context. However, due to the nature of the Arabic language, especially in slang Arabic, stop words might be considered an important part of Arabic sentences. Consequently, preserving and eliminating stop words were considered in the conducted experiments. The last step of this phase is stemming, where words are reduced to their stem or basic form. With respect to the work presented in this paper, two well-known Arabic stemmers were utilized: (i) a light stemmer and (ii) a snowball stemmer.

Phase 3: Feature extraction. In this phase, the sentences are represented by quantitative features that can be input into machine learning algorithms. In this paper, three different approaches were used for data representation that consider both statistical and contextual text representations: (i) BOW, (ii) TF-IDF, and (iii) BERT embedding.

Phase 4: LAPM construction. In this phase, five machine learning algorithms are used to generate the desired LAPM, including: (i) the KNN algorithm; (ii) the SVM algorithm; (iii) the LR algorithm; (iv) the DT algorithm; and (v) the bagging ensemble algorithm. Note here that several values for the key parameters that were associated with the considered classification algorithms were examined. Based on the conducted experiments, $k = 3$ was considered for the KNN classifier, "linear kernel" was adopted for the SVM classifier, and *DT* classifiers were considered as base classifiers to generate the Bagging Ensemble classifier.

Phase 5: Model evaluation. In this phase, the most widely used metrics for evaluating opinion mining models are used to assess the proposed LAPM. These metrics include:

- Accuracy: the proportion of instances that were correctly predicted by the classification model out of all instances [24].

$$accuracy = \frac{TP + TN}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}}$$

- Precision: the proportion of true positive predictions to all positive predictions [24].

$$precision = \frac{TP}{TP + FP}$$

- Recall: the proportion of true positive predictions to all actual positives [24].

$$recall = \frac{TP}{TP + FN}$$

- F1-score: it combines the precision and recall scores of a model [24].

$$F1 = \frac{2 * \mathrm{Precision} * \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}}$$

Where TP, TN, FP, and FN refer to the true positive, true negative, false positive, and false negative samples, respectively. More specifically:

True Positive (TP): refers to the number of positive examples that the model correctly predicted as positive.

True Negative (TN): refers to the number of negative examples that the model correctly predicted as negative.

False Positive (FP): refers to the number of negative examples that the model incorrectly predicted as positive.

False Negative (FN): refers to the number of positive examples that the model incorrectly predicted as negative.

With respect to dividing the considered dataset into training and testing sets, the ten-fold cross validation (TCV) method was adopted for all the experiments reported in this paper. The reasons behind that are: (i) to consider every single example in model construction and evaluation processes; (ii) to obtain precise results that reflect the actual model effectiveness. More specifically, adopting TCV, the considered dataset is divided into 10 disjoint, equal-sized partitions. To construct a classifier, nine partitions are utilized as a training set, and one partition is preserved. The preserved partition is then used to test the constructed classifier. The previous process is repeated 10 times; consequently, each partition is used exactly once as a test set. Then, the average score of the used evaluation measure is calculated. As a result, a single, precise estimation will be obtained.

4 EXPERIMENTS AND EVALUATION

In this section, the obtained experimental results are discussed. As noted earlier, three textual representation approaches and five classification algorithms are considered to generate the desired model. In addition, two stemmers are utilized in relation to BOW and TF-IDF approaches, and two scenarios are adopted for handling stop words, including eliminating or retaining them. Thus, the obtained results are organized into the following subsections: Sub-section 4.1 presents the results obtained using the BOW text representation approach; Sub-section 4.2 presents the results obtained using the TF-IDF text representation approach; Sub-section 4.3 presents the results obtained using the BERT text representation approach; and Sub-section 4.4 presents a comparison between the different text representation approaches with respect to the five classification algorithms. Note that the main objectives of Sub-sections 4.1 and 4.2 are to investigate: (i) the effect of stop-word removal on classification effectiveness; (ii) the impact of using different stemmers on classification effectiveness; and (iii) the influence of employing various classification algorithms on classification effectiveness.

4.1 Experimental results obtained from using BOW text representation approach

This section displays the results obtained from using the BOW text representation approach combined with two different stop-word handling strategies, two stemming methods, and five classification algorithms. Table 1 presents the reported results. For simplicity and because the F1-score represents both precision and recall, the results will be discussed based on the F1-scores. Note that the precision and recall results are presented in the Appendix section. According to the table, the performance of stemming approaches varies depending on the classification algorithm used. Regardless of the preprocessing technique adopted, the logistic regression classification algorithm generated the best results. The best-reported result was obtained by using the Snowball stemmer and LR classification algorithm.

Table 1. Results obtained from BOW text representation approach

Technique	SW Removal and Light Stemmer		SW Removal and Snowball Stemmer		SW Preservation and Light Stemmer		SW Preservation and Snowball Stemmer	
	Acc.	F1	Acc.	F1	Acc.	F1	Acc.	F1
SVM	0.742	0.824	0.717	0.800	0.769	0.838	0.757	0.828
KNN	0.707	0.818	0.692	0.801	0.700	0.804	0.721	0.818
DT	0.703	0.795	0.683	0.777	0.717	0.802	0.719	0.798
LR	0.736	0.829	0.730	0.818	0.770	**0.846**	0.780	**0.852**
Bagging	0.723	0.819	0.725	0.814	0.753	0.834	0.738	0.823

In order to observe the effect of removing and preserving stop words, the results are plotted in Figures 2 and 3. From the figures, it can be clearly observed that keeping stop words results in better classification effectiveness, regardless of the adopted stemming technique for most classification algorithms. Note that handling informal Arabic texts has its uniqueness, where stop-words can significantly impact the meaning of the entire sentence.

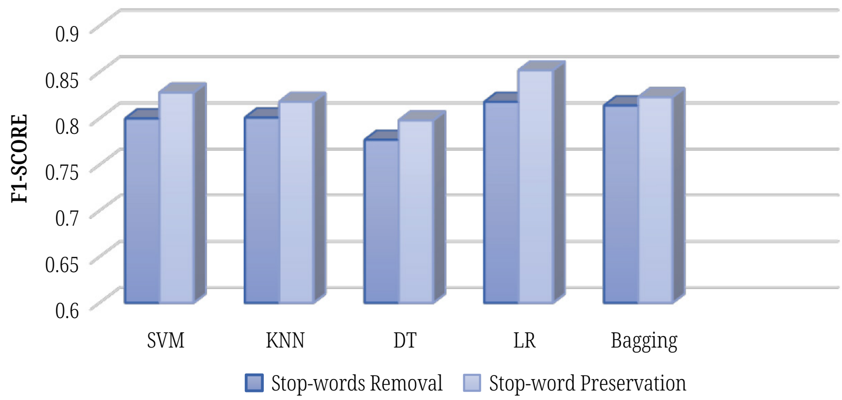

Fig. 2. Stop-words strategies comparison for BOW coupled with snowball stemmer

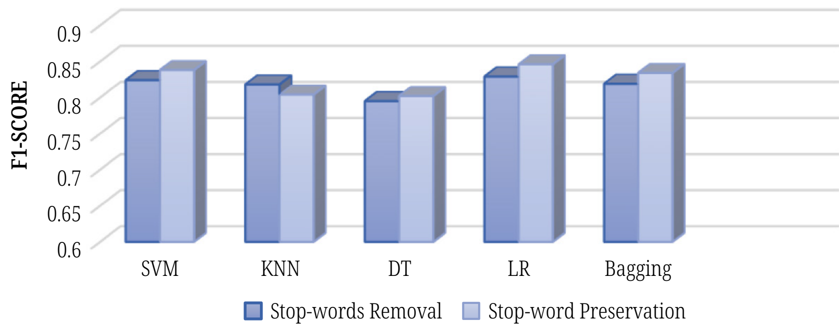

Fig. 3. Stop-words strategies comparison for BOW coupled with light stemmer

4.2 Experimental results obtained from using TF-IDF text representation approach

This section presents the results obtained from using the TF-IDF text representation approach. Table 2 displays the reported results when TF-IDF is combined with two different stop word handling strategies, two stemming methods, and five classification algorithms. As noted in the previous section, the results will be discussed based on the F1 scores. The performance of stemming approaches varies according to the classification algorithm used, as shown in the table. The best-reported result was obtained by using a light stemmer and SVM classification algorithm. Again, to observe the effect of removing and preserving stop words on classification performance, the results are plotted in Figures 4 and 5. It can be seen from the figures that retaining stop words leads to better classification accuracy, irrespective of the stemming technique used for all classification algorithms. It is interesting to mention that the reported observations from using the BOW and TFIDF text representation approaches are consistent.

Table 2. Results obtained from TF-IDF text representation approach

Technique	SW Removal and Light Stemmer		SW Removal and Snowball Stemmer		SW Preservation and Light Stemmer		SW Preservation and Snowball Stemmer	
	Acc.	F1	Acc.	F1	Acc.	F1	Acc.	F1
SVM	0.730	0.833	0.725	0.826	0.772	**0.854**	0.771	**0.850**
KNN	0.711	0.808	0.711	0.807	0.723	0.814	0.727	0.815
DT	0.679	0.779	0.667	0.767	0.728	0.810	0.736	0.812
LR	0.700	0.821	0.705	0.824	0.702	0.822	0.713	0.827
Bagging	0.734	0.829	0.719	0.817	0.765	0.840	0.771	0.842

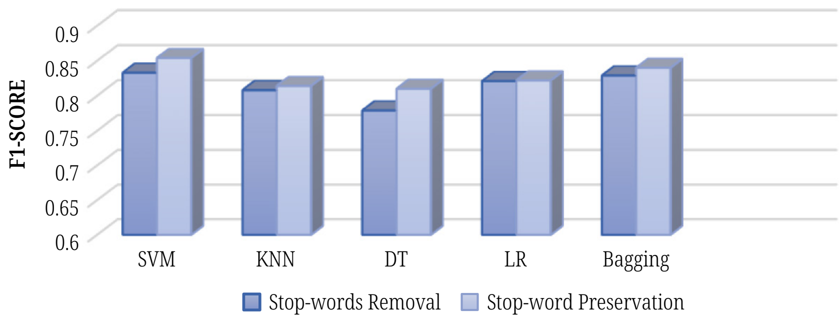

Fig. 4. Stop-words strategies comparison for TF-IDF coupled with light stemmer

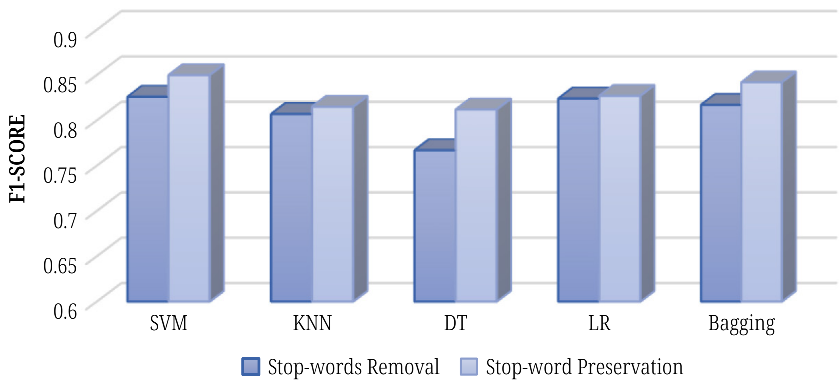

Fig. 5. Stop-words strategies comparison for TF-IDF coupled with snowball stemmer

4.3 Experimental results obtained from BERT text representation approach

In this section, the results obtained from using BERT text representation are presented. Recall that many BERT models were specifically designed and pre-trained to handle the Arabic language. Arabert is the most widely used one [26]. Arabert was utilized in the work presented in this paper. It is worth mentioning that no preprocessing, such as stop-word removal or stemming, was performed before feeding data into BERT. The reason for this is the need for all sentence information to create the text representation. Table 3 presents the results obtained from the BERT text representation approach.

Table 3. Results obtained from BERT text representation approach

Classification Algorithm	BERT Text Representation Approach Results Evaluation Metrics			
	Acc.	F1	Prec.	Recall
SVM	0.686	0.781	0.761	0.805
KNN	0.686	0.797	0.725	0.887
DT	0.620	0.722	0.734	0.715
LR	0.721	0.810	0.768	0.860
Bagging	0.721	**0.825**	0.733	0.942

From the table, it can be noted that the effectiveness of using various classification algorithms vary significantly. The lowest F1-score was generated when using the DT algorithm (0.722), while the highest F1-score was obtained when using the Bagging classification algorithm (0.825).

4.4 Text representation approaches comparison

In this section, a comparison of the text classification approaches under consideration is presented. Figure 6 presents a comparison between BOW, TF-IDF, and BERT. From the figure, it is observed that TF-IDF produced the best results for three out of the five classification algorithms considered. While BOW produced the best results for two classification algorithms, the best overall result was obtained when TF-IDF was coupled with a support vector machine.

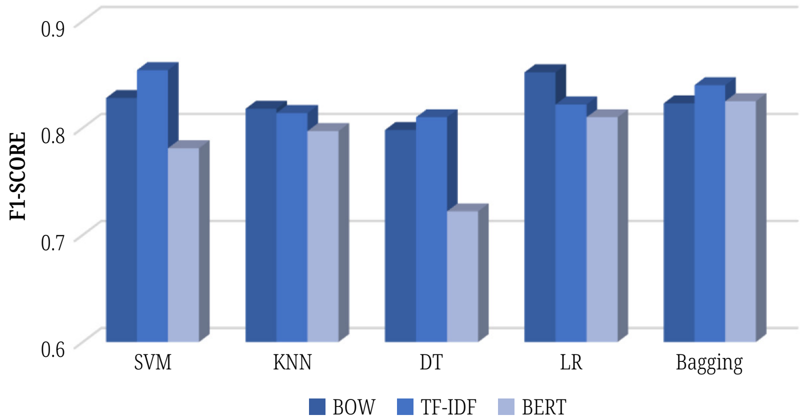

Fig. 6. A comparison between: (i) BOW, (ii) TF-IDF, and (iii) BERT text representation approaches

5 CONCLUSION AND FUTURE WORK

Online learning offers promising benefits for the field of education. In order to obtain these conjectured benefits, educators and educational institutions require continuous monitoring of students' attitudes towards this type of learning to enhance student satisfaction and improve academic achievement. Thus, automatic detection of students' attitudes regarding online learning is needed to improve the overall online learning process. In this paper, a LAPM for Arabic texts is proposed to achieve this goal. The proposed model is considered an opinion-mining model that can predict students' opinions by analyzing their discussions. To generate the desired model, a novel

Arabic text dataset was collected from students at the Hashemite University in Jordan. Five machine learning algorithms, namely SVM, DT, KNN, LR, and Bagging, were utilized. A key point when generating such a model is how to represent the text using a set of features. Three different approaches were considered to address this point: (i) the Bag of Words approach, (ii) the TF-IDF approach, and (iii) the BERT approach. Several pre-processing techniques were examined to obtain the most effective model.

According to the reported experimental results, an effective LAPM could be developed to predict students' attitudes. In addition, the effectiveness of the model is highly affected by the classification algorithm used, text representation approach, and pre-processing methods. The best reported experimental result, in terms of the F1-score evaluation measure, was obtained by using a SVM classifier coupled with TF-IDF, employing a light stemmer, and preserving stop-words, achieving an accuracy of 85.4%. It is important to mention that neglecting stop words resulted in a degradation in the model's effectiveness. The reason behind that is the nature of the dataset being considered. In other words, handling informal Arabic texts has its uniqueness, where stop-words can significantly impact the meaning of the entire sentence. With respect to comparing the three considered text representation approaches, TF-IDF produced the best overall result. Although BERT is a well-known approach for generating text representation, handling a special informal Arabic dataset that includes students' vocabularies could be considered the reason behind the superiority of the TF-IDF approach. In addition, it is interesting to mention here that our classification problem is a binary problem, and the best overall result was produced by SVM, which was originally designed for binary classification problems.

Future research in the field of opinion mining is promising, especially in exploring the combination of various available methods for text pre-processing and representation. Moreover, the authors aim to extract a large Arabic dataset from social networks to support data scientists and promote research in Arabic text mining.

6 REFERENCES

[1] C. Guo, X. Yan, and Y. Li, "Prediction of student attitude towards blended learning based on sentiment analysis," in *Proceedings of the 2020 9th International Conference on Educational and Information Technology*, 2020, pp. 228–233. https://doi.org/10.1145/3383923.3383930

[2] F. Hemmatian and M. K. Sohrabi, "A survey on classification techniques for opinion mining and sentiment analysis," *Artificial Intelligence Review*, pp. 1–51, 2019.

[3] M. Hammad and M. Al-Awadi, "Sentiment analysis for arabic reviews in social networks using machine learning," in *Information Technology: New Generations. Advances in Intelligent Systems and Computing*, Latifi, S. Ed., Springer, Cham, 2016, vol. 448, pp. 131–139. https://doi.org/10.1007/978-3-319-32467-8_13

[4] M. Al-Ayyoub, A. A. Khamaiseh, Y. Jararweh, and M. N. Al-Kabi, "A comprehensive survey of arabic sentiment analysis," *Inf. Process. Manag*, vol. 56, no. 2, pp. 320–342, 2019. https://doi.org/10.1016/j.ipm.2018.07.006

[5] N. M. El-Makky *et al.*, *Sentiment Analysis of Colloquial Arabic Tweets*, 2014.

[6] G. Alwakid, T. Osman, and T. Hughes-Roberts, "Challenges in sentiment analysis for arabic social networks," *Procedia Computer Science*, vol. 117, pp. 89–100, 2017. https://doi.org/10.1016/j.procs.2017.10.097

[7] K. M. Alomari, H. M. ElSherif, and K. Shaalan, "Arabic Tweets sentimental analysis using machine learning," in *Advances in Artificial Intelligence: From Theory to Practice,* Lecture Notes in Computer Science, S. Benferhat, K. Tabia, and M. Ali, Eds., IEA/AIE 2017, Springer, Cham, 2017, vol. 10350. https://doi.org/10.1007/978-3-319-60042-0_66

[8] V. K. Vijayan, K. R. Bindu, and L. Parameswaran, "A comprehensive study of text classification algorithms," in *International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, 2017, pp. 1109–1113. https://doi.org/10.1109/ICACCI.2017.8125990

[9] G. Badaro *et al.*, "A survey of opinion mining in Arabic: A comprehensive system perspective covering challenges and advances in tools, resources, models, applications and visualizations," *ACM Transactions on Asian Language Information*, vol. 384, 2019.

[10] S. R. El-Beltagy, T. Khalil, A. Halaby, and M. Hammad, in *Combining Lexical Features and a Supervised Learning Approach for Arabic Sentiment Analysis*, A. Gelbukh, Ed., Computational Linguistics and Intelligent Text Processing. CICLing 2016. Lecture Notes in Computer Science, Springer, Cham, 2018, vol. 9624. https://doi.org/10.1007/978-3-319-75487-1_24

[11] D. Najar and S. Mesfar, "Opinion mining and sentiment analysis for Arabic on-line texts: Application on the political domain," *International Journal of Speech Technology*, vol. 20, pp. 575–585, 2017. https://doi.org/10.1007/s10772-017-9422-4

[12] H. Elzayady, K. M. Badran, and G. I. Salama, "Arabic opinion mining using combined CNN – LSTM models," *Int. J. Intell. Syst. Appl.*, vol. 12, no. 4, pp. 25–36, 2020. https://doi.org/10.5815/ijisa.2020.04.03

[13] R. Al-Mutawa and A. Y. Al-Aama, "User opinion prediction for Arabic hotel reviews using lexicons and artificial intelligence techniques," *Applied Sciences*, vol. 13, no. 10, p. 5985, 2023. https://doi.org/10.3390/app13105985

[14] A. Y. Al-Obaidi and V. W. Samawi, "Opinion mining: Analysis of comments written in Arabic colloquial," in *Proceedings of the World Congress on Engineering and Computer Science 2016 (WCECS 2016)*, 2016.

[15] C. Guo, X. Yan, and Y. Li, "Prediction of student attitude towards blended learning based on sentiment analysis," in *Proceedings of the 2020 9th International Conference on Educational and Information Technology*, 2020, pp. 228–233. https://doi.org/10.1145/3383923.3383930

[16] Z. S. Harris, "Distributional structure," *Word*, vol. 10, nos. 2–3, pp. 146–162, 1954. https://doi.org/10.1080/00437956.1954.11659520

[17] J. Leskovec, A. Rajaraman, and J. D. Ullman, *Mining of Massive Datasets*, 3rd ed. Cambridge, England: Cambridge University Press, 2020. https://doi.org/10.1017/9781108684163

[18] T. Okeefe and I. Koprinska, "Feature selection and weighting methods in sentiment analysis," in *Proceedings of the 14th Australasian Document Computing Symposium*, Sydney, Australia, 2009.

[19] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "Pre-training of deep bidirectional transformers for language understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Minneapolis, Minnesota: Association for Computational Linguistics, 2019, vol. 1, pp. 4171–4186.

[20] W. Antoun, F. Baly, and H. Hajj, "AraBERT: Transformer-based model for Arabic language understanding," in *LREC 2020 Workshop Language Resources and Evaluation Conference*, 2020.

[21] M. Abdul-Mageed, A. Elmadany, and E. M. Nagoudi, "ARBERT & MARBERT: Deep Bidirectional Transformers for Arabic," in *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, 2021, pp. 7088–7105. https://doi.org/10.18653/v1/2021.acl-long.551

[22] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining and Knowledge*, vol. 2, pp. 121–167, 1998. https://doi.org/10.1023/A:1009715923555

[23] D. M. J. Tax and R. P. Duin, "Using two-class classifiers for multiclass classification," in *2002 International Conference on Pattern Recognition*, Quebec City, QC, Canada, 2002, vol. 2, pp. 124–127. https://doi.org/10.1109/ICPR.2002.1048253

[24] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*. Oxford, England: Morgan Kaufmann, 2012.

[25] D. Jurafsky and J. H. Martin, *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Upper Saddle River: Pearson Prentice Hall, 2008.

[26] A. S. Alammary, "BERT models for Arabic text classification: A systematic review," *Applied Sciences*, vol. 12, no. 11, 2022. https://doi.org/10.3390/app12115720

7 APPENDIX

This section presents the results obtained from BOW and TFIDF text representation approaches in terms of precision and recall. The results are shown in Tables A1 and A2, respectively.

Table A1. Results obtained from BOW text representation approach in terms of precision and recall

Technique	SW Removal and Light Stemmer		SW Removal and Snowball Stemmer		SW Preservation and Light Stemmer		SW Preservation and Snowball Stemmer	
	Prec.	Rec.	Prec.	Rec.	Prec.	Rec.	Prec.	Rec.
SVM	0.786	0.868	0.786	0.816	0.818	0.860	0.816	0.841
KNN	0.721	0.948	0.727	0.893	0.733	0.896	0.749	0.902
DT	0.765	0.830	0.758	0.800	0.780	0.827	0.796	0.805
LR	0.757	0.917	0.775	0.868	0.793	0.907	0.802	0.910
Bagging	0.751	0.904	0.768	0.868	0.784	0.893	0.781	0.871

Table A2. Results obtained from TFIDF text representation approach in terms of precision and recall

Technique	SW Removal and Light Stemmer		SW Removal and Snowball Stemmer		SW Preservation and Light Stemmer		SW Preservation and Snowball Stemmer	
	Prec.	Rec.	Prec.	Rec.	Prec.	Rec.	Prec.	Rec.
SVM	0.733	0.967	0.738	0.940	0.772	0.956	0.781	0.934
KNN	0.754	0.871	0.755	0.869	0.763	0.874	0.771	0.866
DT	0.746	0.816	0.741	0.799	0.789	0.832	0.804	0.825
LR	0.701	0.992	0.706	0.989	0.702	0.992	0.712	0.986
Bagging	0.750	0.928	0.745	0.907	0.795	0.893	0.804	0.888

8 AUTHORS

Esra'a Alshdaifat, Ala'a Al-shdaifat, and **Ayoub Alsarhan** are with the Department of Information Technology, Faculty of Prince Al-Hussein Bin Abdallah II For Information Technology, The Hashemite University, P.O. Box 330127, Zarqa 13133, Jordan.