IJEMD-CSAI, 1 (2) (2022), 17 – 23
https://doi.org/10.54938/ijemdcsai.2022.01.2.137

International Journal of Emerging Multidiciplinaries: Computer Science & Artificial Intelligence

Research Paper
Journal Homepage: www.ijemd.com
ISSN (print): 0000-0000



Sentiment Analysis on Education Transformation During COVID-19 Using Arabic Tweets in KSA

Nida Aslam[1] , Irfan Ullah Khan[2] , Taif AlKhales[3], Reem AlMakki[4], Shahad AlNajim[5], Shaden Almarshad[6], Rana Saad[7]
[1]College of Computer Science and Information Technology, Imam Abdulrahman bin Faisal University, Dammam, 34221,Saudi Arabia naslam@iau.edu.sa
[2]College of Computer Science and Information Technology, Imam Abdulrahman bin Faisal University, Dammam, 34221,Saudi Arabia iurab@iau.edu.sa
[3]College of Computer Science and Information Technology, Imam Abdulrahman bin Faisal University, Dammam, 34221,Saudi Arabia 2170003781@iau.edu.sa
[4]College of Computer Science and Information Technology, Imam Abdulrahman bin Faisal University, Dammam, 34221,Saudi Arabia 2170002112@iau.edu.sa
[5]College of Computer Science and Information Technology, Imam Abdulrahman bin Faisal University, Dammam, 34221,Saudi Arabia 2170005277@iau.edu.sa
[6]College of Computer Science and Information Technology, Imam Abdulrahman bin Faisal University, Dammam, 34221,Saudi Arabia 2170002319@iau.edu.sa
[7]College of Computer Science and Information Technology, Imam Abdulrahman bin Faisal University, Dammam, 34221,Saudi Arabia 2170000689@iau.edu.sa

ARTICLE INFO

Submitted:
14-09-2022
Revised:
12-10-,2022
Accepted:
18-11-2022
Keywords:
Sentiment analysis, online education, Social media, machine learning.
Corresponding Author:
Nida Aslam naslam@iau.edu.sa

ABSTRACT

The emergence of COVID-19 pandemic has changed the whole world. To prevent the spread of the virus, different precautionary measure and policies have been defined, one of them is distance learning. It has led to the educational transformation from physical education to online learning. Similarly, in KSA online education is adopted since March 2020. Inorder to extract the individual perception about the online education in KSA, twitter data was used. Arabic tweets were collected using twitter API. Furthermore, several machine learning models such as Support Vector Machine (SVM), Naïve Bayes (NB), Random Forest (RF), Logistic regression, K-nearest neighbors was used develop an automated sentiment analysis of online education related tweets in KSA. For feature extraction and selection N-gram and Term Frequency–Inverse Document Frequency (TF-IDF) was used. Logistic Regression achieved the highest accuracy of 69.33% for multiclass and Random Forest achieved accuracy of 80.35%. According to the dataset, most of the individuals have negative opinion about the online learning as the number of negative tweets are higher as compared to positive and neutral class.

1.Introduction

With the rapid growth of using social media platforms, people tend to express their feelings easily. Especially on the Twitter platform recently people use Twitter to catch up on what happened to the world through trends and events. World Health Organization (WHO) has declared COVID-19 as pandemic in March 2020. As per the statistics on 22nd September 2021, 228 countries has infected worldwide. The pandemic has the devasting affect on the society and individual lives. Most countries started quarantine in 2020 to control the disease from spreading among people, and that caused the need for distance learning.

Due to quarantine and self-isolation number of internet users also increased at an exponential rate. Individuals were using social media to get an up-to-date information and share their feelings. The transformation from physical education to online has revolutionized the traditional education process. Twitter is one of the easiest platforms for people to share their feelings, related to virus, government policies such as self-isolation, lockdown, and eLearning as well.

Twitter showed how people thought, reacted toward many things, it can be used a tool to collect data about people's feelings toward many topics since social media contains a rich source of emotions, opinions, and reviews that can be used for sentiment analysis. Sentiment Analysis (SA) is an area related to text mining, that can study subjective information toward a certain subject using text. Furthermore, it is very helpful in social media monitoring because people tend to express their feelings on social media.

In this study, we propose machine learning based model for exploring the sentiment analysis of people in Kingdom of Saudi Arabia (KSA) related to online education using twitter data. We collected Arabic tweets from KSA using Twitter API and labeled as positive, negative, and neutral. Several machine learning models like Support Vector

© 2022 The Author(s). This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0). https://creativecommons.org/licenses/by/4.0/

Machine (SVM), Naïve Bayes (NB), Random Forest (RF), Logistic regression (LR), K-Nearest Neighbors (KNN) and feature selection technique such as N-gram and Term Frequency–Inverse Document Frequency (TF-IDF) was be used in the proposed study for the automated sentiment analysis of Arabic tweets.

2.Review of Related Literatures

With the advancement in technology today many platforms offer online courses, and people have different opinions. A study was conducted by J. Sultana et al. to capture students' feedback on the effectiveness of online courses platform and the quality of education back in 2018. Their experiments have been done using Kalboard 360 dataset repository. Several conventional machine learning models were used and SVM achieved the highest accuracy of 78.75%. Multi-Layer-Perceptron (MLP) and SVM classifiers have shown similar performance.

Furthermore, Onan presented a sentiment analysis model for Massive open online courses (MOOC). The data was collected by crawling an online website coursetalk.com to get different university courses reviews by students. The feedback collected reached up to 93,000 MOOC reviews. The reviews were initially pre-processed and used to train different ML models, ensemble learning, and deep learning-based approach. The maximum predictive performance was achieved by deep learning architectures specifically Long Short-Term Memory Networks (LSTM) with a classification accuracy of 95.80%.

Similarly, Clarizia et al. aimed to detect students' moods across social media such as Twitter, Facebook related to different topics. They used Latent Dirichlet Allocation (LDA) approach, and mixed Graph of Terms for the descriptive analysis of the dataset. that was obtained by LDA. Senti Word-Net was used for the lexicon and senti analysis of the words.

Furthermore, Al-Twairesh et al. collected a huge dataset of Arabic tweets and especially for Saudi dialect tweets. They collected three datasets the first one was EMO-TWEET which used distant supervision as noisy label through emoticons. The second data set was KEY-TWEET, collected the tweets using sentiment Arabic words. The third dataset was SAUDI-TWEET which was collected from previous datasets whose location was in KSA. The tweets were labeled as negative, positive, mixed, and neutral. They used multi-way sentiment classification. They experimented with different number of classes, Binary classes (Negative and Positive), Three classes (Negative, Positive and Neutral), and Four classes (Negative, Positive, Neutral and Mixed). For classification SVM with linear kernel was used. The best results achieved by the proposed model was 62.27% for the binary.

AL-Rubaiee et al. performed the sentiment analysis of the tweets related to e-Learning. They collected two thousand tweets from students on the DELDE Twitter account then annotated manually into positive, negative and neutral for the sentiment analysis of Arabic tweets in e-learning. The study achieved the highest accuracy of 84.84% using SVM TF-IDF 3-gram.

Moreover, another research was done to analyse the user's experience during pandemic period by Arambepola. The study conducted some statistical data analysis to explore user experience on distance learning for several countries based on their economic stability and internet usage. They used Term Frequency Inverse Document Frequency (TF_IDF) vectorizer module to score the words, then used TextBlob for calculating the subjectivity and polarity. The number of tweets collected were 202,645 during the covid-19 period with Twitter API using hashtags and keywords such as #distancelearning, #virtuallearning. They applied some pre-processing techniques and classified them into three classes such as positive, negative, and neutral. After the annotation some statical analysis was performed and conclude that that if effective internet and other requirement were available then distance education is better.

Consequently, Z. Solangi et al. conducted a study using a statistical approach by distributing a questionnaire to indicate the most important factors to achieving effective implementation of online education in Jubail, Saudi Arabia. They measured the effectiveness by distributing the questionnaire to the students and instructors in Jubail Industrial College (JIC), and Jubail Technical Institute (JTI) to figure out what are the important factors in E-learning for higher education students in 2018. The students at the university highlighted that self-efficacy and compatibility were the most important parts if one wanted to effectively gain knowledge from an E-Learning experience. Secondly, the instructors in the university emphasized that when they conducted online preparations to utilize technical efficiency that it would have a successful effect on their student's outcomes. Another element to maximize their throughput was the facilitating conditions offered to the instructors, like good internet connections, the availability of a quiet environment helped to achieve effective implementation of eLearning for higher studies in Jubail University.

Furthermore, Khan et al. has proposed a deep learning model for the sentiment analysis of the tweets related with COVID-19 in KSA. The study has used the Arabic tweets from KSA. Several machine learning and deep learning models were compared, and the study found that LSTM outperformed the other models. Furthermore the study found that most of the tweets were positive and neutral.

Recently, a study has been made to explore the impact of the educational transformation in Indonesia due to COVID-19. The study collected the data through interview. They found that online education has some advantages interms of flexibility, reduced cost and time. Despite of the advantages, it suffers from some limitation like lack of interactivity etc.

In summary, only few studies have been performed related with the sentiment analysis or exploring the impact of online education during COVID-19. All the previous studies were related to the statistical analysis of the dataset.

Nevertheless, there are some studies mentioned above that is related to the machine learning based sentiment analysis of online education. But all those studies were performed before the educational transformation i.e., offline to online due to COVID-19. Similarly, there are some studies on sentiment analysis of twitter data in KSA related to government policies, mental health, awareness etc. Therefore, this study aimed to propose a machine learning model for the sentiment analysis of Arabic tweets in KSA related to online education.

3.Material and Methods

The block diagram of the proposed methodology is presented in the Figure 1.

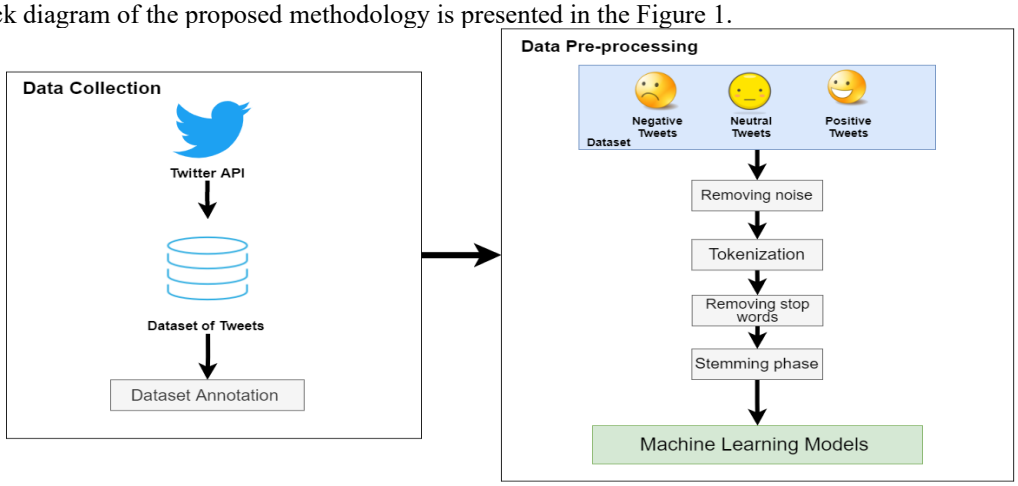


Fig. 1. Proposed methodology

3.Dataset collection and Description

In the current study, the dataset was collected from Twitter platform using Twitter API python script, the tweets were collected within the Kingdom of Saudi Arabia during the COVID-19 pandemic. The dataset consists of tweets with their location mentioned and time of tweet .The date of data collection is during the beginning of the lock down and excluding the holidays by using keywords related to distance learning such as " برنامج", "التعليم عن بعد", "زوم", "بلاك بورد", and it will be labeled into three categories such as positive, negative, and neutral. Figure 3 indicates the most frequent 200 words in the dataset. The size of the word is increased more if it appears more in the document. It indicates the most common words used in the Twitter related to distance learning.

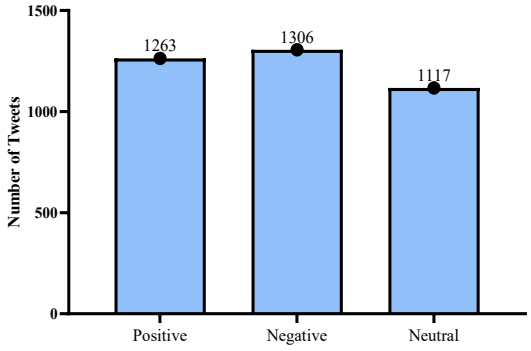


Fig. 2. Distribution of tweets per class



Fig. 3. Word Cloud for the top 200 words in the dataset

Preprocessing

The tweets were written in different dialects and not in classical Arabic. Several preprocessing techniques were applied such as Normalization, Stemming, and Stop-word removal etc.

Normalization

Normalization is a pre-processing technique for cleaning data to uniform the text. In the collected data all the URLs, mentions, links, media, emojis and special characters was removed. Also, the repeated letters are handled

in this phase. Such as "اليووووم" into "اليوم". Also, the multiple forms of the Arabic characters like letter Alef "إ, آ, ا" will be normalized into "ا". We used sub method provided from Regular Expression re package. It has several types which we can use them together like Tatweel and diacritics (Tashkeel) removal.

- Tashkeel will be removed using "strip_tashkeel" method provided from the PyArabic package. For example, the word "اليَومُ" become "اليوم"
- Tatweel will be removed using "strip_diacritics" and "strip_tatweel" method provided from the PyArabic package. For example, the word "الـيــوم" become " اليوم".

Stop Words Removal

Stop-words are words that are meaningless and add nothing to the sentence, so we need to remove them from the text. Such as " هاتان , انتم , هذه" and will be adding the dialectal Arabic stop word like "هذي, هذول, انتي".The tokenized text will be compared with the dictionary of Arabic stop-words and if it is matched it will be removed from the text. Excluding the negation stop words because they add to the meaning of the sentence.

We used open method and pass the file path to read the stop-word dictionary. We used the dictionary as a list by splitting it. Then the tokens were compared with the words in the dictionary and stop words were removed and leave only the words that are not matched with the words in the dictionary.

Stemming

Stemming is pre-processing technique that removes suffixes, infixes, and prefixes to extract the word's root. Here is an example of stemming the word "اكلت "its root is "اكل ". Stemming for the Arabic language is challenging as compared to the other languages ISRI Stemmer was used, from nltk.stem.isri.

Feature Extraction and Selection

In the current study, N-gram and Term Frequency–Inverse Document Frequency (TF-IDF) technique was applied for feature extraction and selection. N-gram model is a sequence of n-words that are normally written after each other to give a well-known meaning. N could be Unigram (one-word level), Bigram (two-word level), Trigram (three-word level). It is popular among the other models because it is provided the best performance in NLP comparing with other models.

Term-frequency times Inverse Document-Frequency (TF_IDF), is a commonly used feature extraction and selection technique in text classification. Term-frequency is the number of times the word appears in a document. And the Inverse Document-Frequency is a weight term scheme that gives the tokens that appears more frequently in the different documents a lower impact and give the tokens that occur in small fractions a higher impact. Equation 1 and 2 represents the mathematical equation for TF-IDF.

$$TF - IDF(t,d) = TF(t,d) * IDF(t) \qquad (1)$$

$$IDF(t) = log(\frac{n}{DF(t)}) + 1 \qquad (2)$$

Where, n is the total number of documents, and DF(t) is the number of documents that contain the word t.

Description of the classifiers

Several machine learning models were used such as Support Vector Machine (SVM), Naïve Bayes (NB), Random Forest (RF), Logistic Regression (LR) and K-Nearest Neighbors (KNN) to classify the tweets. Each model was experimented with TF-IDF and different values of N-grams to find the best performance combination.

Naïve Bayes (NB)

Naïve Bayes is probabilistic machine learning model that uses Bayes theorem, and works effectively in text categorization. NB assumes that features are independent of each other that means the presence of some features doesn't affect one feature. Equation 3 represents Naïve Bayes.

$$NB(c|d) = P(c) * \sum_{i=1}^{m} P(f|c)^{ni(d)} \qquad (3)$$

Logistic Regression (LR)

Logistic regression is a machine learning model used for classification and regression. It can be used for the binary and multi-class problem. Equation (4) represents logistic regression:

$$F(x) = \frac{1}{1+e^{(b0+b1x)}} \qquad (4)$$

We trained the LR model for binary and three class labels. For the binary class the N-gram range was (1,2), and C was set to 10. However, for the multi-class, N-gram range was (1,3) and C was set to 1. For both the classes we used l2 norm.

K-Nearest Neighbors (KNN)

K-Nearest Neighbors (KNN) is one of the widely used model for supervised and unsupervised learning. KNN relies on labeled data to be capable to give an appropriate output when new unlabeled data given to it. KNN assumes that points that are close to each other are similar, and if there is a point that is unknown then the distance between it and other points will be calculated and it will be classified based on the nearest neighbors. The distance can be calculated using Euclidean distance equation see equation(5) below.

$$d(x,x') = \sqrt{(x_1 - x_1')^2 + \cdots + (x_n - x_n')^2} \qquad (5)$$

The Figure 4 and 5 shows the effect of the k value on the accuracy for binary and multi class. Several ranges will be used to investigate the impact of different value of K on the accuracy of the classifier.

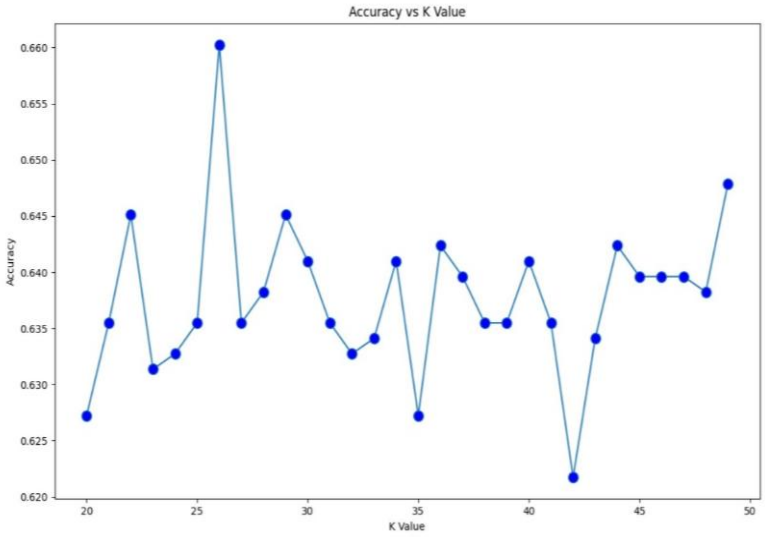


Fig. 4. Different values of K in KNN for Multiclass

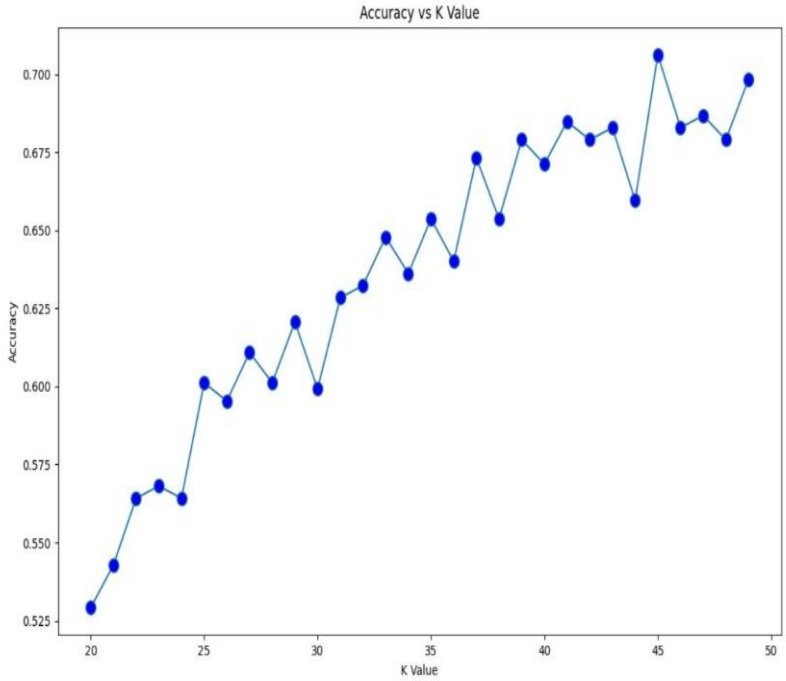


Fig. 5. Different values of K in KNN for Binary class

Random Forest (RF)

Random Forests is an ensemble-based classification and regression machine learning model. It is a combination of tree predictors that predicts the class label by generating randomly a forest, and the forest is a collection of multiple Decision Trees, each tree has the value of a random vector sampled independently, distributed equally among all tree. Table 1 shows the optimal parameters for Random Forest using grid search optimization technique.

Table 1.Optimal parameters for Random Forest

Parameter	Value
Criterion	Entropy
Estimators	100
Max_depth	500
Max_Features	Log2
Min_sample_split	100
Norm	l1

Support Vector Machine (SVM)

Support vector machine (SVM) is a supervised machine learning algorithm that does several tasks like regression, classification, and outlier detection. The goal of SVM is to create a hyperplane that separates n-dimensional space into classes, so the data on one side of the hyperplane shows a category and the other side shows another category. Data that are closest to the hyperplane are called Support Vectors and plays a role in the position of the hyperplane. The distance between data points and hyperplane called margin.

SVM has two types, simple SVM that can be used for linear regression and classification problems, and kernel SVM that can be useful in non-linear data because many features can be added. SVM can be useful in high dimensional spaces and its work in a good way when the separation of the margin is clear between classes and SVM shows that it is not a good choice for large data sets and does not work well if there is a noise in data. Table 2 shows the optimal parameters for SVM using grid search optimization.

Table 2.Optimal parameters for Support Vector Machine

Parameter	Value
C	1
Gamma	1
Kernel	RBF (multi class), Linear (Binary Class)

4.Experimental Setup and Results

The proposed models were implemented using python programming language. We used several libraries such as the Sklearn, and GridSearchCV. The dataset was partitioned into 20% for the testing and 80% for the training. The cross validation was set to 10-folds. And for the feature selection we used TF-IDF vectorizer and the count vectorizer for the n-gram ranges. For finding the optimal parameters we used a gird search method. Python GridSearchCV library that searches for the optimal hyperparameters of a model. Initially experiment was conducted without TF-IDF for both binary and multi-class. Table 3 shows the performance comparison of different classification techniques in combination with different value of N-gram for the binary class i.e., positive and negative label. For the binary class RF achieved the highest accuracy of 80.35%. The performance of NB and LR are similar. However, KNN achieved the worst performance among the other classifiers with an accuracy of 70.62%.

Table 3.Classification result for binary class

Classifier	N-gram	Precision	Recall	F1_Score	Accuracy
NB	(1,3)	79.02	79.24	78.96	78.98
LR	(1,2)	79.20	79.42	79.15	79.18
RF	(1,2)	**80.20**	**80.22**	**80.21**	**80.35**
KNN	(1,3)	76.06	72.44	69.95	70.62
SVM	(1, 2)	78.78	78.99	78.75	78.79

Similarly, Table 4 represent the classification performance of the classifiers for the multi-class tweets i.e., positive, negative, and neutral. For the three-class problem, multinomial NB and LR model was used. Finally, after classifying the last dataset for the three labels (Negative, Positive, Neutral). Multinomial Logistic Regression outperformed the other model with the accuracy of 69.33%.

Table 4.Classification result for Multi class

Classifier	Ngram Range	Precision	Recall	F1_Score	Accuracy
MNB	(1,2)	67.22	66.81	66.61	67.12
LR	(1,3)	**69.51**	**69.08**	**68.88**	**69.33**
RF	(1,2)	64.37	63.94	63.72	64.09
KNN	(1,3)	65.92	65.64	65.49	66.02
SVM	(1,2)	66.99	65.37	65.07	66.05

Furthermore, we investigated the effect of the TF-IDF on the classifiers Results. Table 5 indicates that TF-IDF with N-grams improves the classifier's accuracy for the multi-class. However, for the binary class the performance of the proposed model was same for both experiments i.e., with and without TF-IDF.

Table 5. Effect of TF_IDF on the classification performance for the multi-class tweets

Classifier	TF-IDF	N-gram	Prec.	Recall	F1-Score	Accuracy
MNB	✓	(1,2)	67.22	66.81	66.61	67.12
MNB	✗	—	65.12	64.81	64.56	65.06
LR	✓	(1,3)	69.51	69.08	68.88	69.33
LR	✗	—	67.97	67.78	67.70	67.81
RF	✓	(1,1)	64.37	63.94	63.72	64.09
RF	✗	—	63.07	62.65	62.33	62.86
KNN	✓	(1,3)	65.92	65.64	65.49	66.02
KNN	✗	—	49.26	49.16	49.17	49.24
SVM	✓	(1,2)	66.99	65.37	65.07	66.05

SVM	✗	—	57.00	37.21	26.91	38.53

5.Conclusion

In conclusion, the dataset collected using twitter API, contains most of the tweets were negative towards the distance learning in Saudi Arabia. We developed the model to predict people's perception towards the distance learning that can help to improve the online learning. Several values of N were used for N-gram in combination with TF-IDF. The optimal accuracy for the multiclass was achieved using Logistic Regression with an accuracy equal to 69.33% and the optimal accuracy for the Binary class was using Random Forest with an accuracy of 80.35%. The binary class (Negative and Positive) gives a better performance in our case, the reason could be the diversity of the Neutral class because it contains a lot of unrelated text. For future work, we could expand the dataset and improve the classifiers performance by investigating other classifiers. In addition, training the model using huge dataset.

6.References

"COVID-19 Statictics." https://www.worldometers.info/coronavirus/ (accessed September, 22nd, 2022).

J. Sultana, N. Sultana, K. Yadav, and F. Alfayez, "Prediction of Sentiment Analysis on Educational Data based on Deep Learning Approach," 2018, doi: 10.1109/NCG.2018.8593108.

A. Onan, "Sentiment analysis on massive open online course evaluations: A text mining and deep learning approach," *Comput. Appl. Eng. Educ.*, 2020, doi: 10.1002/cae.22253.

F. Clarizia, F. Colace, M. De Santo, M. Lombardi, F. Pascale, and A. Pietrosanto, "E-learning and sentiment analysis: A case study," in *ACM International Conference Proceeding Series*, Jan. 2018, pp. 111–118, doi: 10.1145/3178158.3178181.

N. Al-Twairesh, H. Al-Khalifa, A. Al-Salman, and Y. Al-Ohali, "AraSenTi-Tweet: A Corpus for Arabic Sentiment Analysis of Saudi Tweets," *Procedia Comput. Sci.*, vol. 117, pp. 63–72, 2017, doi: 10.1016/j.procs.2017.10.094.

H. Al-Rubaiee, R. Qiu, K. Alomar, and D. Li, "Sentiment Analysis of Arabic Tweets in e-Learning," *J. Comput. Sci.*, vol. 12, no. 11, pp. 553–563, 2016, doi: 10.3844/jcssp.2016.553.563.

N. Arambepola, "Analysing the Tweets about Distance Learning during COVID-19 Pandemic using Sentiment Analysis," 2020. .

Z. A. Solangi, F. Al Shahrani, and S. M. Pandhiani, "Factors affecting successful implementation of elearning: Study of colleges and institutes sector RCJ Saudi Arabia," *Int. J. Emerg. Technol. Learn.*, 2018, doi: 10.3991/ijet.v13i06.8537.

Khan, Irfan Ullah, Nida Aslam, Sara Chrouf, Israa Atef, Ikram Merah, Latifah AlMulhim, and Raghad AlShuaifan. "Computational Intelligence-Based Model for Exploring Individual Perception on SARS-CoV-2 Vaccine in Saudi Arabia." Computational Intelligence and Neuroscience 2022 (2022).

R. Firmansyah *et al.*, "Educational Transformation : An Evaluation of Online Learning Due to COVID-19," vol. 16, no. 07, pp. 61–76.

N. Alami, M. Meknassi, S. A. Ouatik, and N. Ennahnahi, "Impact of stemming on Arabic text summarization," in *Colloquium in Information Science and Technology, CIST*, Jul. 2016, vol. 0, pp. 338–343, doi: 10.1109/CIST.2016.7805067.

A. Ayedh, G. TAN, K. Alwesabi, and H. Rajeh, "The Effect of Preprocessing on Arabic Document Categorization," *Algorithms*, vol. 9, no. 2, Apr. 2016, doi: 10.3390/a9020027.

C. Maas, A.L., Daly, R.E., Pham, P.T., Huang, D., Ng, A.Y., Potts, "Learning word vectors for sentiment analysis.," in *49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, 2011, pp. 142–150.

K. S. and T. I. S. Rahman, K. Biplob, M. Rahman, "An Investigation and Evaluation of N-Gram, TF-IDF and Ensemble Methods in Sentiment Classification," *Cyber Secur. Comput. Sci.*, pp. 391–402, 2020.

M. K. Han, *Data Mining: Concepts and Techniques*. Morgan Kauffman, 2012.

N. C. and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. e University Press, 2013.