v1: 6 November 2023

Research Article

Sentiment Analysis of Opinions about Online Education in the Kurdistan Region of Iraq during COVID-19

Peer-approved: 6 November 2023

Qeios, Vol. 5 (2023)
ISSN: 2632-3834

Maryam Sami[1], **Hossein Hassani**[1]

1. University of Kurdistan Hewlêr (UKH), Erbil, Iraq

Sentiment analysis is widely used in various areas and has versatile applications. For example, it is used in market research, customer retention strategies, and product analysis, to name a few. Although a few works on the topic exist for the Kurdish language, similar to other fields in Kurdish processing, it is not well-studied, and particularly it suffers from data inadequacy. In this paper, we present research we conducted to analyze the sentiments of learners/educators toward online education during COVID-19 in the Kurdistan Region of Iraq. We collected the data from tweets tweeted in Kurdish (Sorani) up to March 2022. We used four Machine Learning algorithms: Naïve Bayes, SVM, Random Forest, and Logistic Regression, and analyzed their performance on our dataset. We retrieved about 600 tweets, which after preprocessing, yielded 511 items. We conducted five experiments, four of which included testing all algorithms using two scenarios of balanced and unbalanced datasets of positive/negative items, each using 80/20 and 90/10 training/testing data splitting methods. The fifth experiment included four parts, setting a limit for feature selection starting at 500 features and increasing it to 500 at a time until 2000 features, testing for both 80/20 and 90/10 data splitting approaches. The results showed that the best algorithm to build a sentiment analysis model is SVM, with an accuracy of 89% and a maximum feature selection of 1000. The dataset is publicly available for non-commercial use under CC0 1.0 Universal license.

Correspondence: papers@team.qeios.com — Qeios will forward to the authors

1. Introduction

The spread of Coronavirus (COVID-19) in 2020 dramatically affected every aspect of life. As a precaution, governments had to set laws for people's safety. In almost all countries, the authorities enforced quarantine at different levels. That caused the world to descend into chaos, and like the rest of the world, people in the Kurdistan Region of Iraq (KRI) had to adapt to managing their lives in a new way. One of the concerns was education, and COVID-19 forced educational institutions to switch from face-to-face to online education. However, regardless of the readiness level of different regions and countries for such a dramatic shift, many wondered about its efficiency.

Sentiment analysis has been an instrument to study the efficiency of various educational methods for a while [1], but its usage in the analysis of the "enforced" online education to understand the opinion of students emerged rapidly during the COVID-19 pandemic (see [2] [3]). We also attempted to use the technique to analyze the situation in the KRI and developed a dataset from

tweets about online education in the region. We focused on the tweets written in Sorani Kurdish and tested different machine-learning algorithms to assess their accuracy in analyzing the sentiments.

In this paper, we present an overview of our research and the developed dataset developed regarding sentiment analysis of opinions about online education during the COVID-19 pandemic in the KRI. The dataset is available under CC0 1.0 Universal license at https://github.com/KurdishBLARK/SentimentAnalysis/tree/main/OnlineEducation-during-COVID-19. The rest of this paper is organized as follows. Section 2 provides a brief background about the Kurdish language. In Section 3 we present a summary of related work. Section 4 briefly presents the method we followed, Section 5 demonstrates the results of data collection and preparation, Section 6 describes the conducted experiments of applying various algorithms on the dataset, Section 7 illustrates the results and discusses the outcome of the experiments, and finally, Section 8 concludes the paper.

2. A brief Introduction to Kurdish Language

Kurdish is a multi-dialect Indo-Iranian language that is spoken by more than 30 million people in several countries. It uses different scripts, such as Latin, Persian-Arabic, and Cyrillic. [4]. It is considered a less-resourced language from Natural Language Processing perspective [5]. Sorani is one of the Kurdish dialects that is mostly written in an Arabic-based script that first has been modified by Persian writers and then revised further to accommodate the letters that represent the Kurdish phonemes that the original scripts did not support. Table 1 shows the Kurdish alphabet for Persian-Arabic script.

ئ	ا	ب	پ	ت	ج	چ	ح	خ	د	ر	ڕ	ز	ژ	س	ش	ع	غ	ف	ق	و	گ	ل	ڵ	م	ن	ه	ۆ	وو	ی	ێ

Table 1. Kurdish alphabet for Persian-Arabic script

Because of the commonality of the phonemes of Kurdish and Persian, the Kurdish linguists kept the four added letters to the Arabic alphabet (پ , چ , گ, and ژ) but elided eight Arabic letters that they found unsuitable for Kurdish phonemes (ث, ذ, ص, ض, ط, ظ, ض, and ع) [6].

Although Sorani is usually written in the modified Persian-Arabic script, users of social media sometimes use Latin or simply use English alphabets to express their views. That makes the processing of Kurdish texts that are retrieved from social media a challenging task. We refer back to this issue in Sections 4 and 5 when we describe the research method and the data collection process.

3. Related Work

The research about sentiment analysis has attracted the research community in the past decade, and it seems to keep its status, at least for the near future. Table 2 summarizes the related work and presents the topics, number of entries in the studied datasets, methods with the highest accuracy, and the languages of the contents.

Reference	Subject of dataset	Entries	Best method(s)	Accuracy	Language
[7]	Review tweets	1200	SVM, Ensemble & Maximum Entropy	90%	English
[8]	Tweets about about World Cup 2014	4162	Bayesian Logistic Regression	74.84%	English
[9]	Tweets about Jakarta Governor election	1356	Multinomial Logistic Regression	74%	English
[10]	Book reviews	2000	Naïve Bayes	81.45%	English
[11]	Camera reviews	3106	Naïve Bayes	98.17%	English
	Laptop reviews	1946	Naïve Bayes	90.22%	English
[11]	Mobile reviews	1918	Naïve Bayes	92.85%	English
[11]	Tablet reviews	1894	Naïve Bayes	97.17%	English
[11]	TV reviews	1596	Naïve Bayes	90.16%	English
[11]	Video surveillance	2597	Naïve Bayes	91.13%	English
[12]	Tweets in Jordanian	1800	SVM	88.72%	Arabic
[13]	Self-deriving cars	7156	SVM	59.9%	Arabic
[14]	Apple products	3884	SVM	71.2%	Arabic
[15]	Tweets of various opinions	4242	Logistic Regression	57%	Arabic
[16]	Tweets about depression	4542	Random Forest	82.39%	Arabic
[17]	Tweets about governmental preventive measures to contain COVID-19	58000	1-gram Naïve Bayes	89%	Arabic
[18]	Arabic book reviews	3315	Deep Learning	82%	Arabic
[19]	Tweets about online education	3480	Logistic Regression	89.9%	Arabic
	Tweets about politics	3000	Naïve Bayes	95%	Persian
[20]	Movie reviews	2010	Long Short-Term Neural Network	95.61%	Persian
[21]	Social media comments	15000	Naïve Bayes	66%	Kurdish
[22]	Medical sentiments	6756	N/A	N/A	Kurdish

Table 2. A summary of literature review.

Overall the literature we reviewed suggests the following points:

- Naïve Bayes, Support Vector Machine (SMV), Logistic Regression and Random Forest have better performance and score higher accuracy than other algorithms. Therefore, we use those four algorithms in this research.
- The increase in the training set enhances the accuracy and performance of the sentiment analysis model, particularly, when the datasets are small.
- Term Frequency-Inverse Document Frequency (TF-IDF) and N-grams are common approaches for

feature extraction.

- Tweepy and other applications interacting with Twitter's API work better for tweet retrieval.
- Research on sentiment analysis in Kurdish is extremely limited. We were able to only retrieve two papers regarding sentiment analysis in Kurdish at the time of preparing this paper.

Based on the points mentioned, we devise our method to conduct this research as we describe in Sections 4.

4. Method

We obtain a developer's account from Twitter to collect the required data by setting the proper attributes that filter the data. We preprocess the data and clean it to prepare it for labeling. As we expect not to be able to collect a large amount of data, we select the most suitable algorithms that have shown appropriate performance on a small amount of data (see Section 3). Figure 1 illustrates the main steps of the research method, and this section explains the details of those steps.

doi.org/10.32388/SU5BR9.3

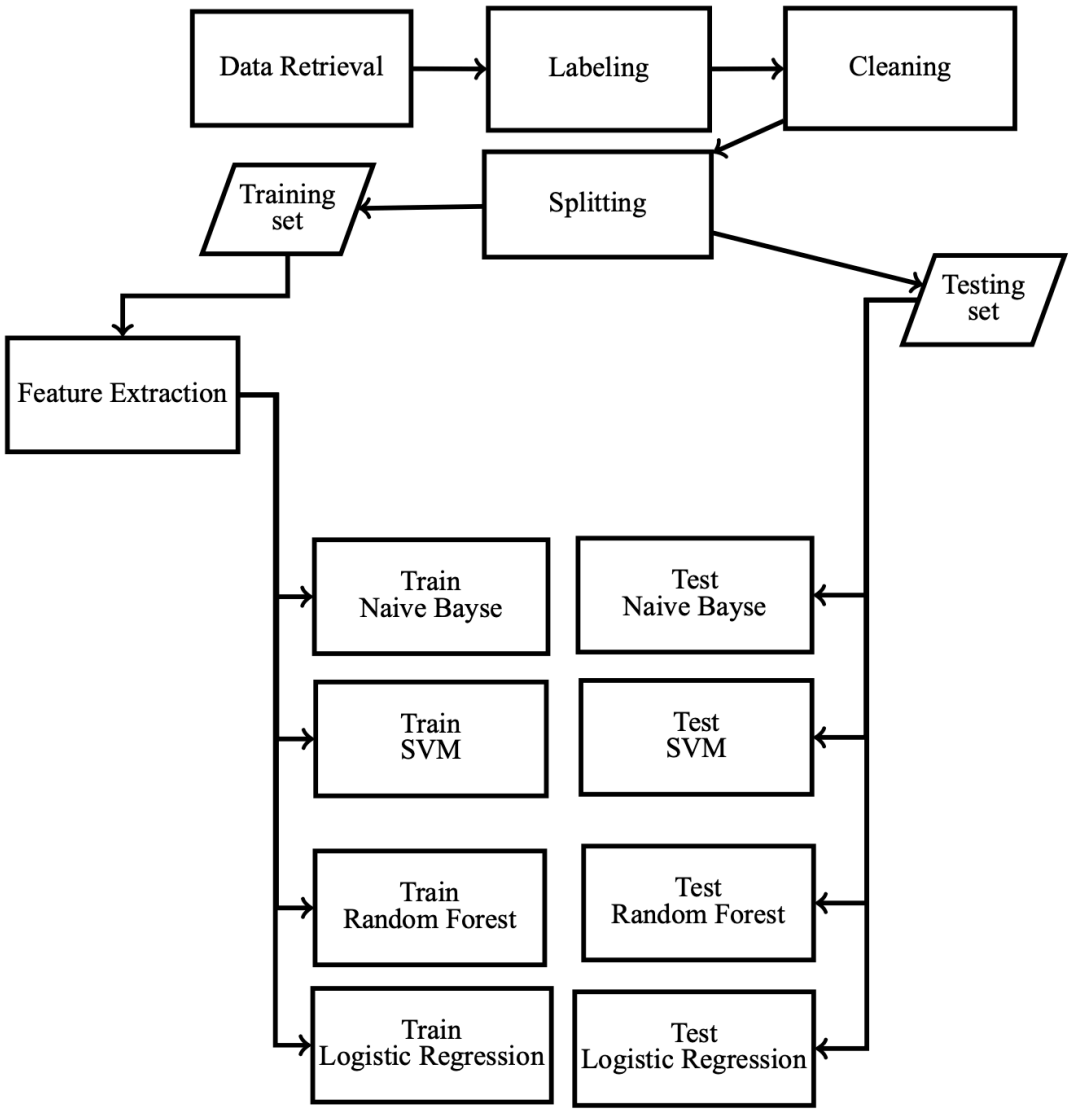

Figure 1. The proposed sentiment analysis process.

Tweets usually include informal language containing slang words, abbreviations, and grammatical mistakes. Preprocessing steps must be applied to the data to transform the text into a format that a machine learning model can analyze. In the preprocessing phase, we do the following:

- Removing all irrelevant tweets, such as those made in a language other than Kurdish, tweets about ads, news, and topics unrelated to how the the online education process is received in Kurdistan.
- Removing punctuation, emojis, symbols, URLs, stop words, numbers
- Removing Arabic diacritics
- Removing duplicates.
- Transliterating texts in Latin into Persian/Arabic script.

4.1. Labeling

We ask three native Kurdish experts to evaluate the data and manually label them. The three experts delete spam and unrelated tweets, transliterate those written in Latin to Persian/ Arabic Script, label them as Negative or Positive by majority voting, and tag them with either of the two classes. This process dictates that this research is a supervised machine-learning model. The preprocessing phase results in two datasets due to

doi.org/10.32388/SU5BR9.3

splitting the preprocessed dataset into training and testing datasets.

4.2. Feature Extraction

We use the Term frequency-inverse document frequency (TF-IDF) for the feature extraction technique. Using TF-IDF, we find the most significant words in a document. We use the method here to find the words that have the most essential roles in expressing the sentiments. A document's TF-IDF is composed of two important parts: the Term frequency (TF) and the Inverse Document Frequency (IDF). TF indicates how often a term appears in a document; IDF measures how often a term appears in all documents. In this research, we use TF-IDF as Formulae 1, 2, and 3 show.

$$TF(i,j) = \frac{count\ word\ i\ inside\ tweet\ j}{count\ all\ words\ in\ tweet\ j} \qquad (1)$$

$$IDF = log\frac{1+M}{1+DF_i} \qquad (2)$$

Here, M is the total tweets count, and DF_i is the number of tweets that have word i. Additional 1 resolves divide by zero situations.

$$TF\_IDF_{ij} = TF_{ij} \times IDF \qquad (3)$$

4.3. Evaluation

We evaluate the performance of the chosen algorithms by constructing a confusion matrix for each method, obtaining *precision* and *recall* measures, and calculating *F-Score*. Formula 4 is used to calculate *precision*, Formula 5 gives *recall*, and Formula 6 calculated *F-Score*. *F-Score* has various versions of which we use *F1-Score*.

$$precsision = TP\frac{TP}{TP+FP} \qquad (4)$$

$$recall = TP\frac{TP}{TP+FN} \qquad (5)$$

$$F1 = 2\frac{precisio \times recall}{precision + recall} \qquad (6)$$

Here, TP stands for True Positive, FP for False Positive, and FN for False Negative.

5. Data Collection and Preparation

We obtained a developer's account from Twitter and prepared a program to collect the data. We set the date to March 2022, the location to the KRI, and used Persian-Arabic and Latin for Kurdish terms along with English terms for the search parameters. Table 3 shows examples of terms we used for searching.

Persian-Arabic	Latin	English
ئیلیکترۆنی خوێندنی	Xwendny online	online_education
ئۆنالین خوێندنی	Xwendny electrony	OnlineLearning
کۆرونا خوێندنی	Xwendny kati corona	UniversityOnlineCourse
	Xwendny zamani corona	DistanceLearning
	Xwendn ba shewazy nwey electrony	

Table 3. Examples of search terms.

Using this technique, we collected 720 tweets, including spam, unrelated news, replies, retweets, and ads. After preprocessing, the resulting dataset consisted of 511 tweets in Total, 368 (5852 words) items being negative, and 143 (3535 words) labeled Positive. Table 4 shows samples of the labeled tweets.

Table 4.3: sample from the collected labeled tweets.

Negative	Positive
خیندنی ئۆنالین ومعاشی ئۆفالین	کۆرسی نوسینی ،کەیتڤ هەتم بۆت!! زۆر ئیکساتدم
خوێندنی ئۆنالین حاڵم خۆشە #onlinelearning	فێرخوازانی زانکۆ خۆتان بۆ خوێندنی #ـۆنائلین ئامادەکەن سبەی بڕیار لەسەر خوێندنی ئەمسائی زانکۆکان دەدرێت ! #StayAtHome
براەڕێک هەیە ئەڵێی پشتگیری لە خوێندنی ئۆنالین و بڕیارێکی هەموو فاشلی زانکۆ نەکات معاشەکەی ئەبڕن مەبەست یشم نییە @shkoagha !	خوێندنی #ئۆنالین دەستپێدەکاتەوە

Table 4. Examples of labeled tweets.

The resulting data of this stage still included noises, such as emojis, URLs, numbers, Arabic diacritics, and English words. We used the KLPT toolbox [23] to clean the data further and to stem it. Table 5 shows examples of this activity.

Before preprocessing	After preprocessing
تووخوا تێچوو و پارەی خوێندنی دوور چیە لەکاتی پێندەمێک؟ گەمەی بە عەسابم دەکا	تووخوا چ پارە خوێن دوور چیە لەکات ندەم گەمە عەساب دەکا
بۆچی فێربوونی دور هەڵناوەشینەوە و کانسلی ناکەن؟	بۆ فێر دور هەڵناوەشین کانسل کەن
ترسناکە !!! #فێربوونی دوور	ترسناکە فێر دوور

Table 5. Examples of preprocessed tweets.

6. Experiments

We assessed four methods for their accuracy in the sentiments classification: Naïve Bayes, Support Vector Machine (SVM), Random Forest, and Logistic

Regression. We chose those algorithms because the literature has reported their performance on small datasets is reasonable.

Because we could not find a large amount of data, we conducted the experiments by splitting the dataset in two ways: first, 80% training to 20% testing and second, 90% training and 10% testing.

We applied five scenarios. In the first scenario, the data is divided into a ratio of 80% training set and 20% testing set with random selection for the sets and with no specification to a maximum features selection (all features extracted are used).

The second scenario is conducted by splitting the data training and testing set ratio to 90% to 10% instead of 80% to 20% with no specification to a maximum feature selection (all features found are used).

In the third scenario, we reduced the Negative data to balance out the positive data and have a balanced dataset where there are 231 data labeled Positive and 231 labeled Negative forming a dataset of 462 preprocessed tweets. Dividing the dataset into 80% training and 20% testing. For this experiment, all features found are used.

The fourth scenario is conducted with a balanced dataset and dividing the dataset into a 90% training set and a 10% testing set. For this experiment, there is no restriction to the number of features selected (all features found are used).

In the fifth experiment, we manipulated the maximum number of features involved. We set the number of features for the TF-IDF vectorizer to different values. Before that, we conducted a series of experiments to find an optimal number of maximum features that could help us enhance the accuracy of our unbalanced dataset. The initial experiments did not count for restricting the number of features selected because the tools used for assigning weights to the terms arrange the list of features from best (Heavyweights) to worst (Lightweights) but do not restrict the numbers it obtains from the model to include all terms with weights assigned.

Our dataset includes about 2250 features. Therefore, to search for the optimal number of features that include terms related to emotions and exclude irrelevant terms, the number of maximum features was selected randomly for each part of the experiment starting from 500 items and increasing 500 at a time up to 2000 items. The 500-at-a-time increase is because increasing 100 or 200 at a time showed no noticeable change.

7. Results and Discussion

Figures 2 and 3 show the performance of the algorithms for 80-20 and 90-10 data splitting, respectively.

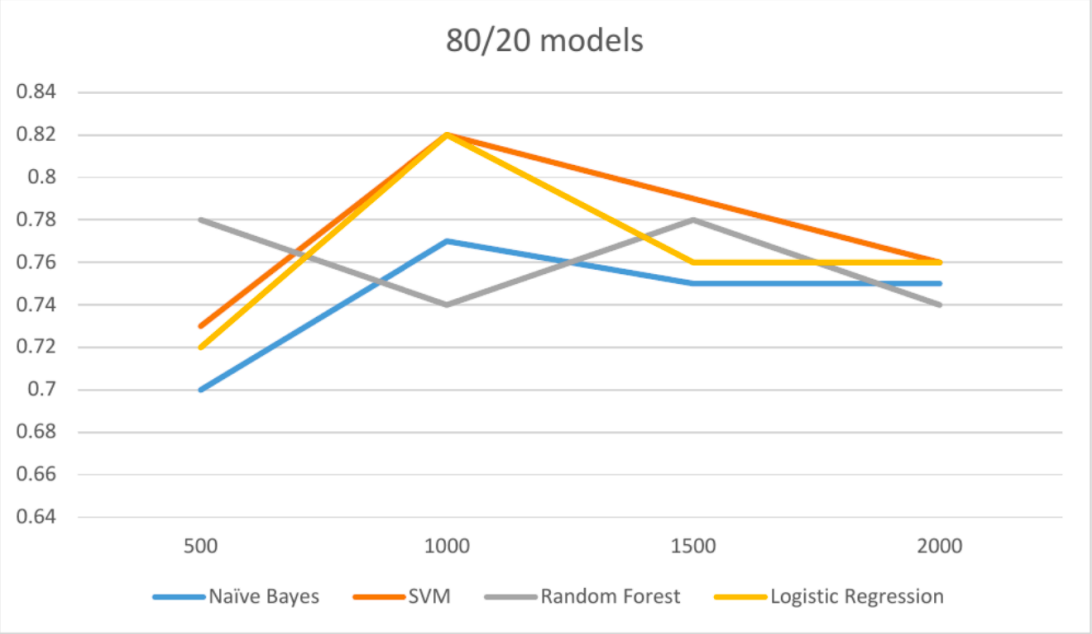

Figure 2. Classifiers evaluations for 80/20 models through features incrementation.

doi.org/10.32388/SU5BR9.3

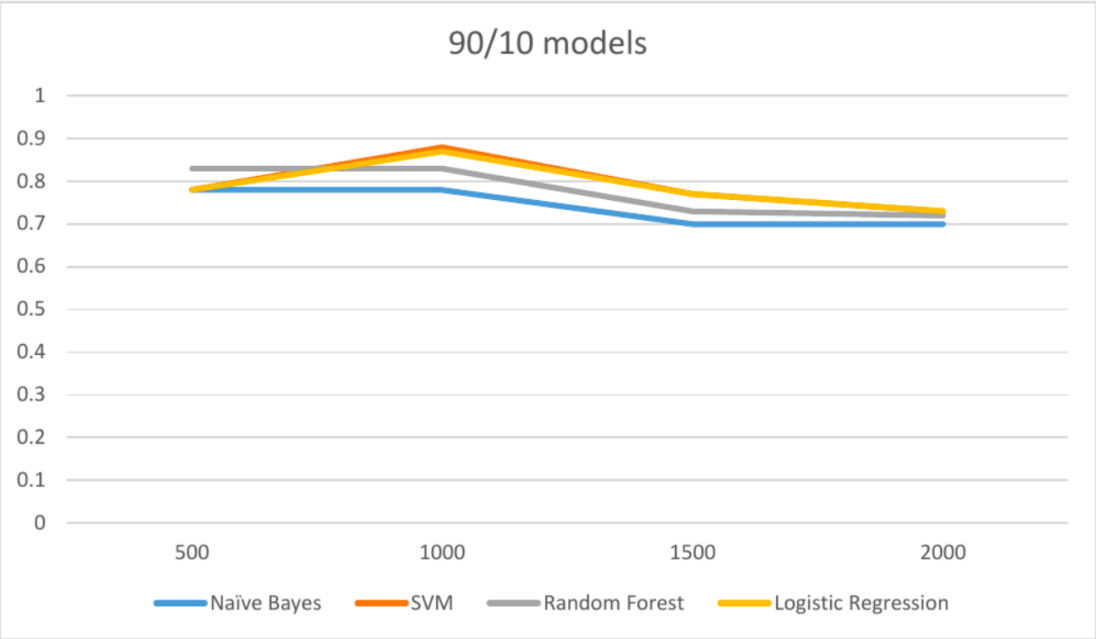

Figure 3. Classifiers evaluations for 90/10 models through features incrementation.

Conducting the experiments resulted in the following outcomes:

- Collected data shows more negative feedback (emotions) about online education in KRI, as more data was labeled negative, which indicates that students are not taking online education positively.
- Although the dominant approach for dividing the dataset is 80/20, 90/10 works better for smaller datasets. This is in accordance with the literature regarding the classification approaches. As the results show for the 80/20 models, the accuracy was about 65% for SVM and 62% for Logistic Regression, which increases in the 90/10 method to 80% for both algorithms.
- Scenario 3 and 4 models recorded higher accuracy scores because of the balanced dataset. The negative labeled data equals the positive, which, even with random selection for the training and testing sets, the sets should be balanced or semi-balanced. The accuracy increased from 74% for both the SVM and logistic regression using the original unbalanced sets to 80% and 83% using the balanced datasets for the 80/20 models. Furthermore, for the 90/10 models, the accuracy for both SVM and logistic regression went from 85% for the unbalanced sets to 90%.
- While the Naïve Bayes model is expected to score the highest accuracy, it scored low in most scenarios because the algorithm assumes that each word is independent. While the Random Forest models scored lower in all scenarios because of the small amount of data available, the random forest would not have many entries for the categories of the decision trees.
- SVM and Logistic regression had great accuracy and similar performances. That is due to both algorithms separating training data linearly.
- For the fifth experiment, as is shown in figure 2 and figure 3, the optimal maximum number of features is 1000 for most algorithms regardless if the training set was set to 80% and testing 20% or the training set was set to 90% and the testing 10%; the accuracy increased. The 500 maximum features part of the experiment scored a relatively lower accuracy because 500 is considered a small number that does not cover all terms used to express emotions. Moreover, for our dataset of size 600, having 2000 features is considered overwhelming, and it included many useless terms due to the small dataset.
- Conducting all these experiments with the dataset gathered is to find the best Classification model. Having a 90% training to 10% testing is rare but due to having a small dataset, increasing the test dataset reduces the variance of the random process in

doi.org/10.32388/SU5BR9.3

selecting the training and testing sets. As shown throughout all experiments, the SVM and Logistic regression models scored the highest accuracy.

- The fifth experiment concentrates on trials to enhance the accuracy of the original unbalanced dataset by finding the best maximum number of extracted features that exclude unreverent features with small weights assigned to them according to the TF-IDF preprocess technique. According to that experiment, the optimal number that separates relevant from irrelevant features is 1000 forthe 90% training and 10% testing datasets. SVM outperformed other algorithms with an accuracy of 89%, an f-score of 91% for the negative, and 84% for the positive.

8. Conclusion

Emotions are embedded in social media posts, making it the perfect place for data scraping for SA research. With COVID-19 and the quarantine, schools in KRI switched to remote teaching using online classrooms.

This research uses machine learning models for sentiment analysis on Kurdish tweets about online education in the Kurdistan Region of Iraq (KRI). It presents sentiment analysis models for Kurdish (Sorani). We collected 721 tweets in total, of which we developed a dataset of Sorani tweets about online education during COVID-19, resulting in 512 tweets. We developed a language model and trained four algorithms, Naïve Bayes, SVM, Random Forest, and Logistic Regression, to classify the sentiments into positive and negative.

We conducted five experiments the results are as follows:

- For the 80/20 data splitting model, Logistic regression and the SVM algorithms scored the highest accuracy of 74%.
- For the 90/10 data splitting model, Logistic regression and the SVM algorithms scored the highest accuracy of 85%.
- Balancing the dataset (positive and negative labeled data are equal), for the 80/20 model, the Naïve Bayes algorithm scored the highest accuracy of 86%
- Using the balanced dataset for 90/20 model, Logistic regression and SVM scored the highest accuracy of 90%.
- Experimenting with the number of features selected for both 80/20 and 90/10 models showed that the optimal number for maximum features that enhances the accuracy of the unbalanced dataset is 1000.

In the future, we are interested in using different Kurdish dialects, investigating the impact of Hyperparameter tuning using optimization techniques on increasing the performance of the proposed models, and increasing the size of the dataset by gathering more data from other social network platforms regarding the subject of this work using, and studying the impact of using different feature extraction techniques on the models' classification performance.

Acknowledgments

We would like to extend our gratitude to those who helped us in checking the relevance of the tweets' contents to the subject and assisted us in labeling the data.

References

1. ^*Kechaou, Z., Ammar, M.B., Alimi, A.M.: Improving e-le arning with sentiment analysis of users' opinions. In: 2 011 IEEE global engineering education conference (ED UCON), pp. 1032–1038. IEEE (2011)*
2. ^*Mujahid, M., Lee, E., Rustam, F., Washington, P.B., Ulla h, S., Reshi, A.A., Ashraf, I.: Sentiment analysis and topi c modeling on tweets about online education during c ovid-19. Applied Sciences 11(18), 8438 (2021)*
3. ^*Waheeb, S.A., Khan, N.A., Shang, X.: Topic modeling a nd sentiment analysis of online education in the covid -19 era using social networks based datasets. Electroni cs 11(5), 715 (2022)*
4. ^*Hassani, H., Medjedovic, D.: Automatic Kurdish dialec ts identification. Computer Science & Information Tec hnology 6(2), 61–78 (2016)*
5. ^*Hassani, H.: BLARK for Multi-dialect Languages: To wards the Kurdish BLARK. Language Resources and E valuation 52(2), 625–644 (2018)*
6. ^*Idrees, S., Hassani, H.: Exploiting script similarities to compensate for the large amount of data in training te sseract lstm: Towards kurdish ocr. Applied Sciences 11 (20), 9752 (2021)*
7. ^*Neethu, M., Rajasree, R.: Sentiment Analysis in Twitt er using Machine Learning Techniques. In: 2013 fourth international conference on computing, communicati ons and networking technologies (ICCCNT), pp. 1–5. IE EE (2013)*
8. ^*Barnaghi, P., Ghaffari, P., Breslin, J.G.: Opinion Mining and Sentiment Polarity on Twitter and Correlation be tween Events and Sentiment. In: 2016 IEEE second int ernational conference on big data computing service a nd applications (BigDataService), pp. 52–57. IEEE (201 6)*

9. ^Ramadhan, W., Novianty, S.A., Setianingsih, S.C.: Sentiment Analysis using Multinomial Logistic Regression. In: 2017 International Conference on Control, Electronics, Renewable Energy and Communications (ICCREC), pp. 46–49. IEEE (2017)
10. ^Baid, P., Gupta, A., Chaplot, N.: Sentiment Analysis of Movie Reviews using Machine Learning Techniques. International Journal of Computer Applications 179(7), 45–49 (2017)
11. a, b, c, d, eJagdale, R.S., Shirsat, V.S., Deshmukh, S.N.: Sentiment Analysis on Product Reviews using Machine Learning Techniques. In: Cognitive Informatics and Soft Computing: Proceeding of CISC 2017, pp. 639–647. Springer (2019)
12. ^Alomari, K.M., ElSherif, H.M., Shaalan, K.: Arabic Tweets Sentimental Analysis using Machine Learning. In: Advances in Artificial Intelligence: From Theory to Practice: 30th International Conference on Industrial Engineering and Other Applications of Applied Intelligent Systems, IEA/AIE 2017, Arras, France, June 27-30, 2017, Proceedings, Part I 30, pp. 602– 610. Springer (2017)
13. ^Ahmad, M., Aftab, S.: Analyzing the Performance of SVM for Polarity Detection with Different Datasets. International Journal of Modern Education and Computer Science 9(10), 29 (2017)
14. ^Ahmad, M., Aftab, S., Ali, I.: Sentiment Analysis of Tweets using SVM. Int. J. Comput. Appl 177(5), 25–29 (2017)
15. ^Ahuja, R., Chug, A., Kohli, S., Gupta, S., Ahuja, P.: The Impact of Features Extraction on the Sentiment Analysis. Procedia Computer Science 152, 341–348 (2019)
16. ^Almouzini, S., Alageel, A., et al.: Detecting Arabic Depressed Users from Twitter Data. Procedia Computer Science 163, 257–265 (2019)
17. ^Alhajji, M., Al Khalifah, A., Aljubran, M., Alkhalifah, M.: Sentiment Analysis of Tweets in Saudi Arabia regarding Governmental Preventive Measures to contain COVID-19 (2020)
18. ^Al-Bayati, A.Q., Al-Araji, A.S., Ameen, S.H.: Arabic sentiment analysis (asa) using deep learning approach. Journal of Engineering 26(6), 85–93 (2020)
19. ^Vaziripour, E., Giraud-Carrier, C., Zappala, D.: Analyzing the political sentiment of tweets in farsi. In: Tenth International AAAI Conference on Web and Social Media (2016)
20. ^Dashtipour, K., Gogate, M., Adeel, A., Larijani, H., Hussain, A.: Sentiment analysis of persian movie reviews using deep learning. Entropy 23(5), 596 (2021)
21. ^Abdulla, S., Hama, M.H.: Sentiment analyses for kurdish social network texts using naive bayes classifier. Journal of University of Human Development 1(4), 393–397 (2015)
22. ^Saeed, A.M., Hussein, S.R., Ali, C.M., Rashid, T.A.: Medical dataset classification for kurdish short text over social media. Data in Brief 42, 108,089 (2022)
23. ^Ahmadi, S.: KLPT – Kurdish language processing toolkit. In: Proceedings of Second Workshop for NLP Open Source Software (NLP-OSS), pp. 72–84. Association for Computational Linguistics, Online (2020). DOI 10.18653/v1/2020.nlposs-1.11. URL https://aclanthology.org/2020.nlposs-1.11

Declarations

Funding: No specific funding was received for this work.
Potential competing interests: No potential competing interests to declare.