Konya Mühendislik Bilimleri Dergisi, c. 11, s. 2, 341-353, 2023
Konya Journal of Engineering Sciences, v. 11, n. 2, 341-353, 2023
ISSN: 2667-8055 (Electronic)
DOI: 10.36306/konjes.1173939

SENTIMENT CLASSIFICATION ON TURKISH TWEETS ABOUT COVID-19 USING LSTM NETWORK

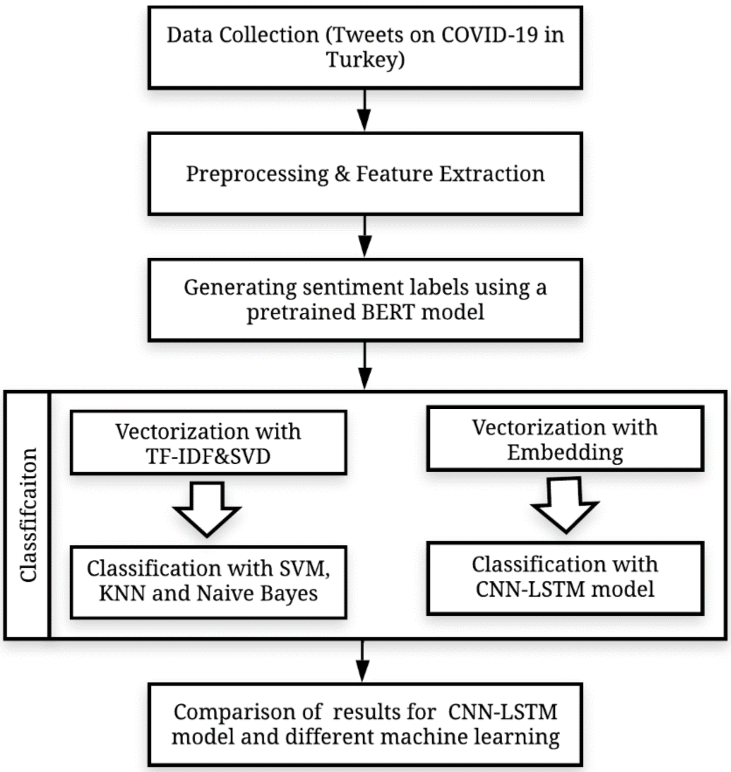
[1*] Mustafa ÇATALTAŞ , [2]Büşra ÜSTÜNEL , [3]Nurdan AKHAN BAYKAN

Konya Technical University, Engineering and Natural Sciences Faculty, Computer Engineering Department, Campus, Konya, TÜRKİYE
[1]mcataltas@ktun.edu.tr, [2]f181213042@ktun.edu.tr, [3]nbaykan@ktun.edu.tr

Highlights

- Collection of Turkish tweets related to COVID-19 pandemic in Turkey and labelling the collected data according to the sentiment classes to form a dataset using a pre-trained BERT model.
- Training a CNN-LSTM deep neural network model to predict sentiment classes for entries in collected dataset.
- Comparison of CNN-LSTM model and different machine learning algoritms' performance on predicting sentiment classes for entries in collected dataset.

Graphical Abstract



The workflow of the study

***Corresponding Author:** Mustafa ÇATALTAŞ, mcataltas@ktun.edu.tr

Konya Mühendislik Bilimleri Dergisi, c. 11, s. 2, 341-353, 2023
Konya Journal of Engineering Sciences, v. 11, n. 2, 341-353, 2023
ISSN: 2667-8055 (Electronic)
DOI: 10.36306/konjes.1173939

SENTIMENT CLASSIFICATION ON TURKISH TWEETS ABOUT COVID-19 USING LSTM NETWORK

[1]Mustafa ÇATALTAŞ , [2]Büşra ÜSTÜNEL , [3]Nurdan AKHAN BAYKAN

Konya Technical University, Engineering and Natural Sciences Faculty, Computer Engineering Department, Campus, Konya, TÜRKİYE
[1]mcataltas@ktun.edu.tr, [2]f181213042@ktun.edu.tr, [3]nbaykan@ktun.edu.tr

(Received: 23.09.2022; Accepted in Revised Form: 17.01.2023)

ABSTRACT: As Covid-19 pandemic affected everyone in various aspects, people have been expressing their opinions on these aspects mostly on social media platforms because of the pandemic. These opinions play a crucial role in understanding the sentiments towards the pandemic. In this study, Turkish tweets on Covid-19 topic were collected from March 2020 to January 2021 and labelled as positive, negative, or neutral in terms of sentiment using BERT which is a pre-trained text classifier model. Using this labelled dataset, a set of experiments were carried out with SVM, Naive Bayes, K-Nearest Neighbors, and CNN-LSTM model machine learning algorithms for binary and multi-class classification tasks. Results of these experiments have shown that CNN-LSTM model outperforms other machine learning algorithms which are used in this study in both binary classification and multi-class classification tasks.

Keywords : *Sentiment Analysis, Turkish Twitter, Classification, LSTM*

1. INTRODUCTION

Since the start of Covid-19 pandemic, the society has faced different situations in various aspects from economics to daily life activities [1]. It caused people to experience unexpected situations like wearing mask in public or staying inside for long periods of time and so they share their opinions on these situations and experiences in some virtual environments. Nowadays, the opinion-sharing process is mostly happening on social media platforms such as Twitter, Instagram and Facebook. Furthermore, the pandemic has increased social media usage even more since conventional ways of communication between people have been less accessible like office environments, coffee-shops etc. This had led to an expected increase in usage of social media. One of these social media platforms, Twitter, is a platform where users express their opinions mostly in the form of text and within the 280 character per tweet limit and Twitter has become a valuable resource for opinion mining of mass groups of people [2]. Twitter has become quite popular in Turkey in 2020 as well.

Since tweets contains tweet texts, likes, retweets, followers, hashtags, mentions and so on, there is a lot of data in them, and all these types of data could be useful in different contexts. However, to extract valuable information from massive amounts of data Twitter contains, manual techniques may be insufficient. Therefore, different computational techniques should be used for sentiment analysis, topic modelling and clustering using tweet texts, classification of tweets using hashtags and like counts [3].

Sentiment analysis, which deals with sentiments and subjectivity in a text using computational methods, is one of the most popular techniques to extract or identify information from massive data [4]. One of the methods that sentiment analysis is performed is classification using machine learning algorithms. However, machine learning algorithms require labelled data for training. Labeled data refers to data being tagged with labels which identify a property of each instance in the dataset such as sentiment of a tweet.

There are two main types of machine learning algorithms which are unsupervised and supervised. Unsupervised algorithms learn by itself with unlabeled data how the output should be. Whereas supervised algorithms require labelled data for training and uses information learnt from training set to

***Corresponding Author:** Mustafa ÇATALTAŞ, mcataltas@ktun.edu.tr

predict target feature given input features. Regression and classification are two main types of supervised learning. Classification is a supervised task where model is first trained with a train set which consists of data points and labels of these data points [5]. Using training sets, classification algorithms learn to predict labels of new data points. Some supervised machine learning algorithms that are generally used in classification are Support Vector Machines, k-Nearest Neighbors, Decision Trees, Multi-Layer Perceptrons etc.

In this study, Turkish tweets were used to measure Turkish society's opinions to Covid-19 and the preventions using machine learning and deep learning methods. During the pandemic process, it was thought that it could help to understand the psychology of the society according to the positive or negative opinions of people in the country. Moreover, the performances of machine learning and deep learning methods were compared, and most successful model was determined according to the data content.

For this study, sentiments of Turkish Covid-19 tweets in 2020 are analyzed separately by restricting the region to Turkey and language to Turkish. The labels indicate the sentiment of tweets as positive, negative, or neutral. Unfortunately, this information is not available in Twitter, so a pre-trained model named BERT was used to label tweets. The proposed model has utilized a combination of Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) models to perform sentiment analysis. Classification was also made with k-Nearest Neighbors, Naïve Bayes and Support Vector Machine traditional machine learning algorithms and compared with the results of the CNN-LSTM model.

The organization of this paper is as follows; Section I gives a brief introduction of problem and methodology used in the project. Section II contains selected related works, Section III provides detailed information on methodology used in the project, and Section IV includes experimental setups and results from these experiments. Finally, Section V concludes the problem, methods, results and discusses on possible future work of this project.

2. RELATED WORKS

Opinion mining on social media platforms using machine learning and natural language processing techniques has been quite popular topic since 2010s. In previous works on opinion mining on Twitter, Öztürk *et al.* [6] applied both classification and clustering techniques on tweets which are related to flu and allergy. They used a handful search strategy to label the tweets while collecting data from Twitter. In clustering task, they obtained 3 clusters using K-Means algorithm and when compared to original clusters, they achieved 68% accuracy. In classification task, they achieved 97% accuracy using Random Forest algorithm. In [7], multi-label classification problem for textual data in which every entry in the dataset could have multiple target classes is studied. The method which was proposed to solve this problem combined CNN and LSTM to create a hybrid model. With the proposed model, they achieved 99% accuracy and 81% F1-score on Wikipedia comments dataset which contains Wikipedia comments labelled by their toxicity level.

There has been works before that dealt with sentiment analysis of Covid-19 related tweets in English [8-10]. In [9], sentiment analysis on English Covid-19 dataset collected from Twitter from January 2020 to July 2020 is performed by utilizing a logistic regression classifier which is fed with collected tweet texts. Their main goal was to analyze sentiment change of the society through different stages of the pandemic.

There has also been an ongoing effort for sentiment analysis on Turkish social media. In previous works in this area [11-15], various traditional machine learning algorithms were utilized for sentiment classification of microblogging entries in earlier works [11]. In recent years, with the rapid enhancement of deep learning model, sophisticated models like LSTMs, RNNs and Transformers has been deployed for this task [15]. In [12], a comparison of machine learning models and pretrained language models on sentiment analysis task was given. The results showed that pretrained language model scored 98.7% accuracy, while the most successful machine learning model scored 98.4% accuracy.

In [13], to perform sentiment analysis on English Twitter-Covid-19 dataset, an LSTM model was utilized with tweets labelled with sentiments obtained by a pre-trained model. The results of the LSTM model were compared to traditional machine learning models. From these comparisons, it was seen that the LSTM model perform much better than traditional machine learning algorithms. LSTM model gave 0.75 F1-score, while Support Vector Machine, the best performing traditional machine learning algorithm in the study, gave 0.60 F1-score.

3. METHODOLOGY

An overall flow of the methods used in this study is given in Figure 1. Detailed explanations of these methods are given in below subsections.

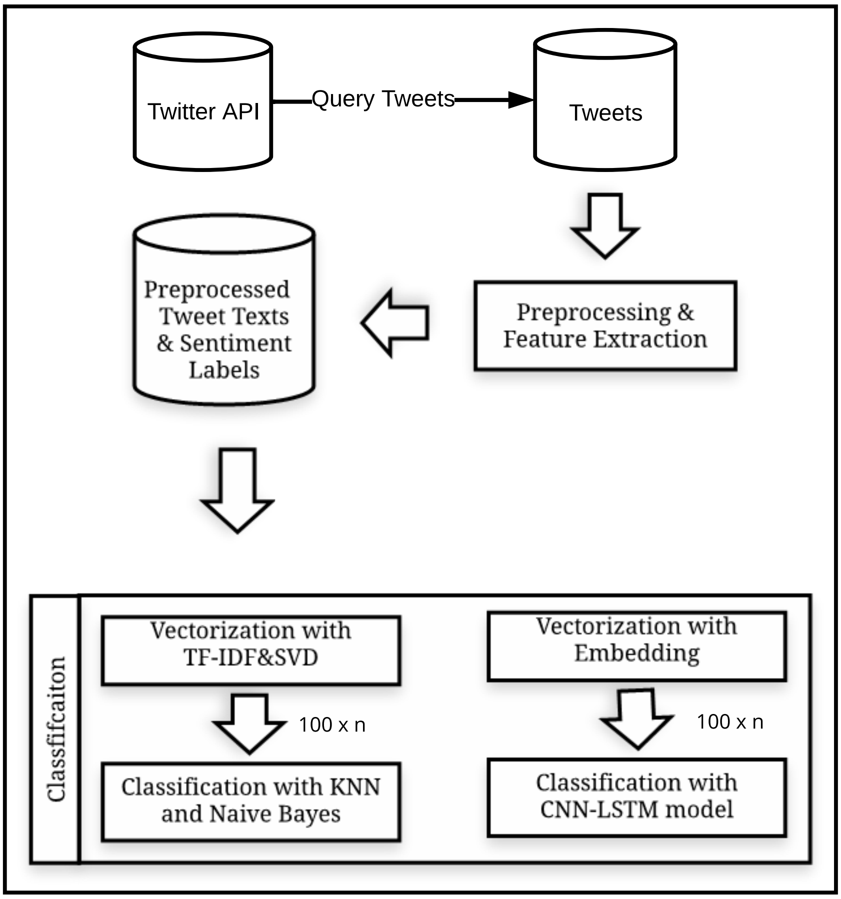


Figure 1. An overview of proposed methods.

3.1. Dataset

To be able to satisfy the data needs, tweets were retrieved from Twitter API. Using the utilities of Twitter API, tweets from March 2020 to January 2021 which contain "Türkiye", and "Corona" terms were retrieved. At the end of this process, 55K tweets on Covid-19 topic were collected.

Since labelled data is required for classification task, the collected data needs to be labelled where labels indicate the sentiment class of each tweet. To achieve this, a pre-trained model BERT (Bidirectional

Encoder Representations from Transformers) on Turkish Sentiment Datasets was used [16]. In Table 1, example tweets are given for each sentiment class from the dataset.

Table 1. Example tweets, their preprocessed versions, and sentiments classes for collected dataset.

Example Tweet	Preprocessed Tweet	Sentiment Class
Corona Türkiye'de yokken gelmeyecek gibi hissediyorduk simdi gitmeyecek gibi hissediyoruz	corona turkiyede yokken gelmeyecek hissediyorduk simdi gitmeyecek hissediyoruz	Negative
31 Mart - Türkiye için corona vaka sayısı açıklandı!	31 mart turkiye corona vaka sayisi aciklandı	Neutral
Bence şu olay biterse Türkiye anlaşıp aynı anda sokaklara inip koşalım, düşünsenize deli gibi koşan insanlar bence muhteşem #corona #EvdeKalTurkiye #COVID19	bence su olay biterse türkiye anlasip ayni anda sokaklara inip kosalim düsünsenize deli kosan insanlar bence muhtesem	Positive

As can be seen from Table 2, the dataset contains mostly negative tweets which are expected because of the topics of these tweets. When the data distributions are examined, it is seen that there is an unbalanced data in terms of the classes. For classification, this may affect the outcome success of experimental studies. For this reason, the results are given not only by accuracy, but also by precision and recall values, and by the confusion matrices. Thus, the classification successes of each class are shown according to classifiers.

Table 2. Distribution of data for each sentiment class

Sentiment Class	Tweet Count
Positive	12534
Negative	38910
Neutral	4072

3.2. Preprocessing

Preprocessing steps were applied to text for normalization. These steps include removal of duplicate tweets, non-alphanumerical character including emojis, punctations, equalizing the number of documents in sarcastic and non-sarcastic classes.

To improve the results in classification, tweets in the dataset were subjected to several preprocessing steps before text vectorization. At first, duplicate tweets were removed, and all tweets were converted to lowercase so that there will not be any ambiguity. Then, tweets were cleaned from any unwanted character in the text by removing URLs, any non-alphanumerical character including emojis, punctations. After these steps, documents that are shorter than 5 token were removed from the dataset just as duplicate documents. In Figure 2, these steps are shown.

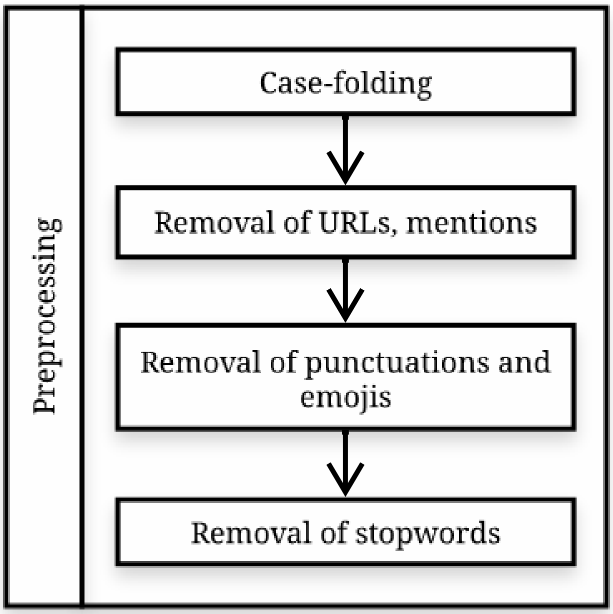


Figure 2. Preprocessing steps of study

3.3. Feature Extraction

To extract features from text, the methods used are TF-IDF [17] which creates a vector for each document in the dataset. TF-IDF is a scoring system for determining the relative importance of a word to the entire document in a corpus [17]. It utilizes two terms: Terms Frequency (TF) which represents number of occurrences of a word in a document and Inverse Document Frequency (IDF) which represents the ratio of total number of documents that contains the term. Equations (1)-(3) show the detailed calculation steps of TF-IDF.

$$tfidf(t,d,D) = tf(t,d) \; x \; idf(t,D) \tag{1}$$

$$tf(t,d) = 1 + \log f(t,d) \tag{2}$$

$$idf(t,D) = \log(\frac{N}{n(t)}) \tag{3}$$

In here, tf is term frequency, idf is inverse document frequency d representing current document and t representing current term. Equation 3 shows how inverse document frequency is calculated, with N being number of documents in the dataset and $n(t)$ representing the number of documents which contain term t.

As shown in [18], dimension reduction to TF-IDF can be achieved using Singular Value Decomposition (SVD) which is a dimensionality reduction method that works well with sparse matrices. Equation 4 shows how SVD is calculated. In this equation, TF-IDF matrix which is represented by C is decomposed into factorization of three matrices. For a C matrix with the dimension mxn, these matrices are U, Σ and V. U is an nxn unitary matrix whose columns are the orthogonal eigenvectors of CC^T, Σ is a

mxn diagonal matrix with singular values of $C$and V is an nxn unitary matrix whose columns are the orthogonal eigenvectors of $C^T C$.

$$C = U\Sigma V^T \tag{4}$$

Truncated Singular Value Decomposition is based on SVD which is also called Latent Semantic Analysis (LSA) in information retrieval [19]. Truncated SVD makes it possible to restrict the *number* of features in the training set by applying low-rank approximation to TF-IDF matrix. In Equation 5, k represents the number of features. In truncated SVD, only highest k singular values are kept in Σ, others are set to 0. When C_k is computed, all rows are comprised of 0s except first k rows as shown in Equation 5.

$$C_k = U\Sigma_k V^T \tag{5}$$

For LSTM model, input text is first tokenized by splitting the input text by spaces. In the result of tokenization, *since* number of tokens in different input texts differs, a different sequence length is seen in resulting feature set. To make all of them have the same length, sequences are padded until they reach the length of 100. Therefore, final sequence lengths for all sequences have become 100.

3.4. Classification

3.4.1. Traditional Machine Learning Algorithms

Text vectors generated in previous phase are first fed into traditional machine learning models which are Support Vector Machine (SVM), k-Nearest Neighbors (k-NN) and Naïve Bayes (NB). k-NN is one of earliest and *simplest* machine learning algorithms [20]. Despite this, it is an effective baseline method for classification task. The algorithm of k-NN is quite simple. To assign a class to a given data point, it calculates determined distance between the data point and all other points in dataset and finds closest k data points. Dominant class with these k points' classes is selected as the class of the given data point. The number of neighbors indicated by parameter (k) is selected as 10 and Euclidean distance was used as distance function for this study. Naïve Bayes (NB) is one of the basic probabilistic models that is utilized in text classification, especially with binary text classification [21]. Naïve Bayes uses conditional probability to calculate most probable target class in classification tasks. Support Vector Machines are known as powerful method for finding decision boundaries between classes which is also called hyperplanes [22]. Although SVM was originally designed to have a linear classifier, it can be used in higher dimensional problems by using SVM with kernel functions such as sigmoid kernel function or hyperbolic tangent kernel function.

3.4.2. CNN-LSTM Model

In the processing of variable-length sequences, the Recurrent Neural Network (RNN) has been widely used [23]. However, because most RNNs are multi-layer feed-forward neural networks, a huge amount of historical data provided by long sequences will result in vanishing gradient and information loss. Long Short-Term Memory (LSTM) [23] is an enhanced version of RNN where information loss and vanishing gradient problems are resolved by adding memory cell state and three control gates to RNN design. "Cell State" can be explained as a communication line and network memory that carries meaningful information across cells to make predictions. In this way, the short-term memory problem is solved, and old data can be moved along the network chain. The information that the Cell State must carry throughout this journey is determined through the gates. These gates can determine what information is necessary or unnecessary which is decided using a sigmoid function.

Convolutional neural networks (CNN) are a type of artificial neural network (ANN) that is used to evaluate visual data [24]. CNNs are based on a shared-weight architecture of convolution kernels or filters that slide along input features to produce translation-equivariant outputs called feature maps. They're employed in a variety of applications, including image and video recognition, recommender systems. As stated in [25], adding convolution layers before LSTM layer could be beneficial. Therefore, we end up with model that utilizes convolution layer before LSTM layer. A representation of the model used in this study can be seen in Figure 3 and parameters of the model are given in Table 3.

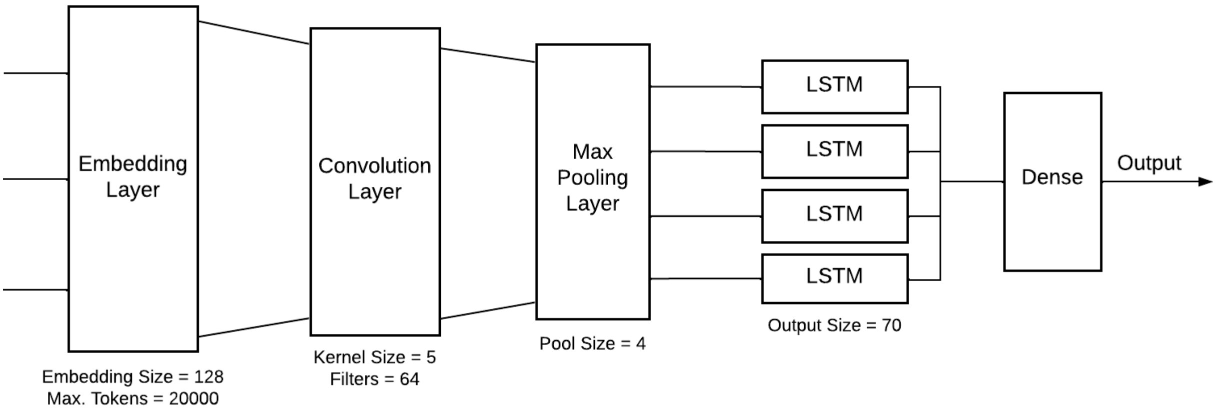


Figure 3. CNN-LSTM network architecture

Table 3. Parameters of CNN-LSTM model

Parameter	Value
Optimizer	Adam
Learning rate	0.001
Loss Function	Categorical Cross Entropy
Dropout	0.25
Batch Size	30
Number of Epochs	2

4. EXPERIMENTS AND RESULTS

4.1. Experimental Setup

All experiments in this study were conducted in Google Colaboratory which provides its users with Tesla K80 GPU and support Python programming language. For deep learning and machine learning part of the study, Scikit-learn [26], TensorFlow [27] and Keras [28] libraries were used. For experiments to be repeatable, random state for all random operations was selected as 42. For all experiments, randomly selected 25% of the dataset was reserved as test set and remaining data in the dataset was used for training which is 75% of the whole dataset. Since the dataset contains sufficient data for classification algorithms, cross-validation techniques were not applied.

The experimental setup includes two different classification tasks, a binary classification task where tweets are classified as positive or negative and a multi-class classification task where tweets are classified as neutral, positive, or negative. The reason for this, as can be seen from Table 1, neutral class contains much less data than other classes which causes imbalance in the dataset. To eliminate this issue, tweets which are labeled as neutral were discarded for binary classification task.

For traditional machine learning algorithms, tweet texts were vectorized through two different ways which TF-IDF with 2000 maximum features limitation and TF-IDF + SVD with 100 maximum features limitation. For CNN-LSTM model, padded sequences were used to vectorize tweet texts with parameters maximum features as 20000 and sequence length as 100 tokens. These parameters are used for both 2-class and 3-class classification tasks. While in Table 4, scores for 3-class classification are given, Table 5 gives 2-class scores.

4.2. Metrics

Metrics used to evaluate result of classification task in this study are accuracy, precision, recall and F1-score [29]. These parameters are calculated using confusion matrices. Confusion matrices are also used to evaluate the results of machine learning algorithms. Creation of confusion matrix for binary classification is given in Figure 4 and same logic can be applied to confusion matrices for multi-class classification.

		Predicted Class	
		1	**0**
True Class	**1**	True Positive (TP)	False Negative (FN)
	0	False Positive (FP)	True Negative (TN)

Figure 4. Confusion matrix for binary classification

The accuracy gives a baseline for comparing the rate at which classifications are correct. From the data, it can be observed that the deep methods perform best. The equation used to calculate accuracy is given in (4).

$$Accuracy = \frac{Correct\ prediction}{total\ number\ of\ predictions} \tag{4}$$

The other metric is F1-score which considers the precision and recall of classifiers and gives a better sense of the classification performance [29]. The F1-score equation is given in (5).

$$F1 - score = 2\ x \frac{1}{\frac{1}{precision} + \frac{1}{recall}} \tag{5}$$

Recall and precision are other classification metrics that constitutes F1-score [29]. Their equations are given in (6) and (7) respectively.

$$Recall = \frac{True\ Positive}{True\ Positive\ +\ False\ negative} \tag{6}$$

$$Precision = \frac{True\ Positive}{True\ Positive\ +\ False\ Positive} \tag{7}$$

4.3. Classification Results

As can be seen from Table 4 and Table 5, TF-IDF alone and TF-IDF with SVD methods performed similar results for most of the cases, however the feature vector length is much higher when TF-IDF used alone. This increases the computation of TF-IDF and thus TF-IDF with SVD becomes more suitable for computation time.

The difference in results of classification models for 2-class and 3-class classification tasks can be seen from Table 4 and Table 5. Classification model performs better in 2-class classification task, which is expected since the probability of guessing the correct label. Furthermore, the eliminated class -neutral- constitutes a small portion of the dataset which results in affecting the classification models adversely.

One observation that is depicted from confusion matrices in Figure 5 and Figure 6 is that SVM can guess wrong labels if the sample is not from the dominant class. Another observation is CNN-LSTM model used in this study outperform traditional machine learning algorithms in both scenarios.

Table 4. Results for 3-class classification (negative-neutral-positive)

Classifier	Vectorizer	Accuracy	Precision*	Recall*	F1-score*
k-NN	TF-IDF	0.74	0.71	0.74	0.71
k-NN	TF-IDF + SVD	0.74	0.71	0.74	0.71
Naive Bayes	TF-IDF	0.46	0.63	0.46	0.48
Naive Bayes	TF-IDF + SVD	0.46	0.64	0.46	0.48
SVM (linear kernel)	TF-IDF	0.73	0.65	0.73	0.65
SVM (linear kernel)	TF-IDF + SVD	0.72	0.65	0.72	0.64
LSTM-CNN model	Padded Sequences	**0.76**	**0.82**	**0.77**	**0.77**

*Weighted averages were used when overall precision, recall and F1-score.

Table 5. Results for 2-class classification (negative-positive)

Classifier	Vectorizer	Accuracy	Precision*	Recall*	F1-score*
k-NN	TF-IDF	0.79	0.77	0.79	0.77
k-NN	TF-IDF + SVD	0.79	0.77	0.79	0.77
Naive Bayes	TF-IDF	0.54	0.72	0.54	0.56
Naive Bayes	TF-IDF + SVD	0.54	0.72	0.54	0.57
SVM (linear kernel)	TF-IDF	0.77	0.77	0.77	0.71
SVM (linear kernel)	TF-IDF + SVD	0.77	0.77	0.77	0.71
LSTM-CNN model	Padded Sequences	**0.84**	**0.84**	**0.84**	**0.84**

*Weighted averages were used when overall precision, recall and F1-score.

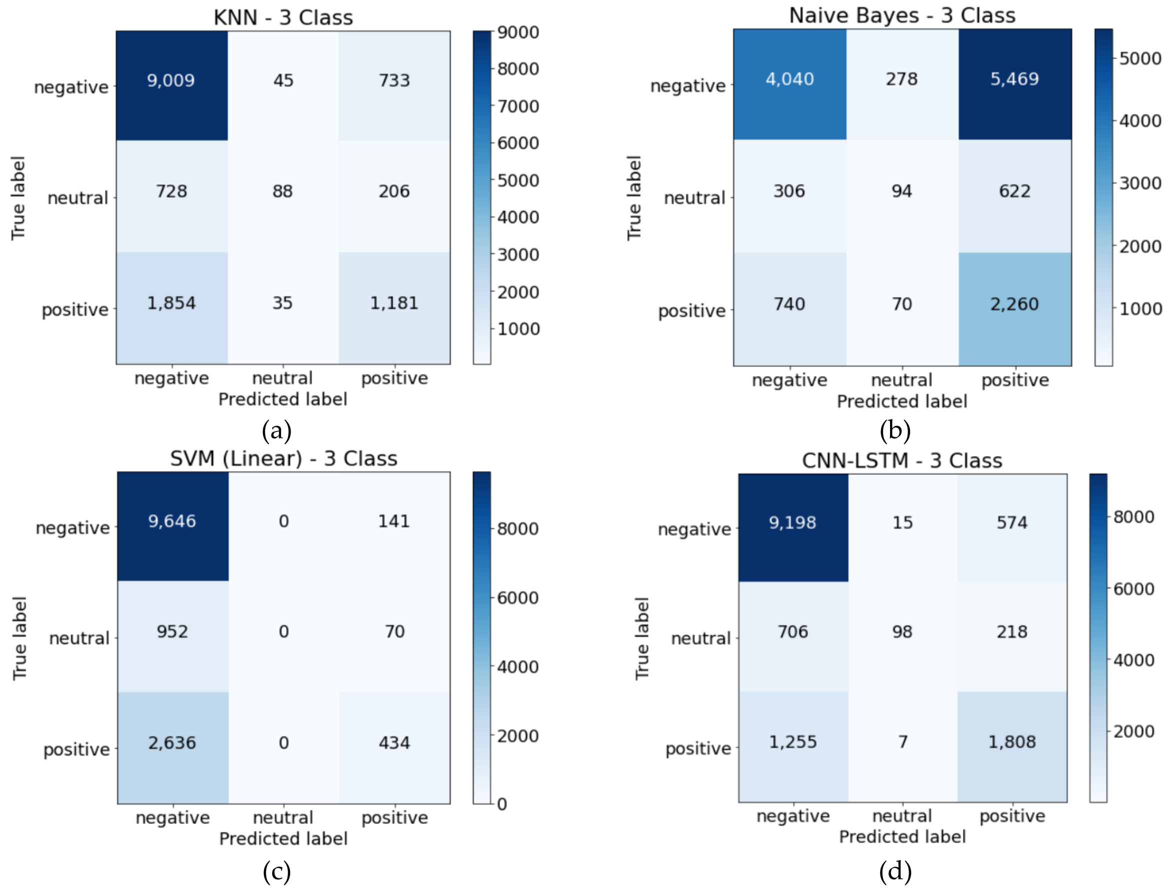


Figure 5. Confusion matrices for 3-class classification results of (a) k-NN, (b) Naïve Bayes, (c) SVM trained with vectors created by TF-IDF + SVD and (d) CNN-LSTM model trained with vectors created by padded sequences.

5. CONCLUSION

Sentiment analysis has become a significant field of research in information retrieval in the field of NLP. Sentiment analysis techniques aim to extract or identify subjective information from textual data with rule-based or machine learning techniques. In recent years, the increase in the popularity of this field comes from advancements in deep learning field which includes more powerful architectures and pre-trained language models.

Turkish is an agglutinative language, so it is difficult and different from English. Turkish has some challenges to make sentiment analysis due to its structural difficulties and therefore there are a limited number of Turkish language studies in the literature. In this study, it was aimed to classify Turkish Twitter posts according to their sentiments analysis and to contribute to the literature.

For the study, a Turkish Twitter dataset on Covid-19 which is retrieved from Twitter API and labelled by a pre-trained BERT model was used. After preprocessing applied to collected tweets, they were vectorized through TF-IDF, TF-IDF with SVD and padded sequences. Then k-NN, NB and SVM traditional machine learning (ML) algorithms and CNN-LSTM network were applied for sentiment analysis for binary and multi-class classification tasks. The results show that there is no significant difference between ML algorithms for both classifications, binary and multi-class. Even the most successful model -kNN- had classification accuracy as 0.74 in the 3-class and 0.79 in the 2-class, while the CNN-LSTM model had accuracy as 0.76 and 0.85 in the 3-class and 2-class classification, respectively. The results of experiments conducted have shown that CNN-LSTM model performs better at both classification tasks. The performance comparison of methods can be helped to new future similar studies on Turkish Tweets.

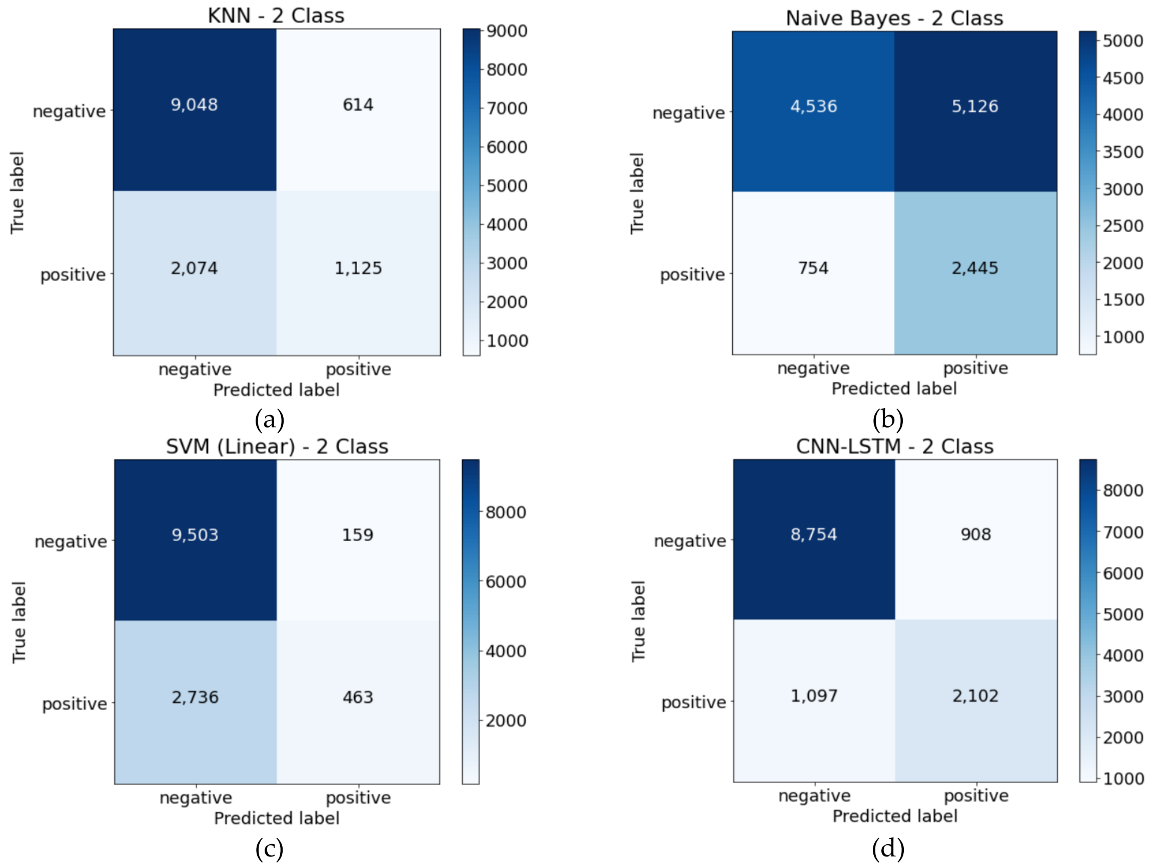


Figure 6. Confusion matrices for 2-class classification results of (a) k-NN, (b) Naïve Bayes, (c) SVM trained with vectors created by TF-IDF + SVD and (d) CNN-LSTM model trained with vectors created by padded sequences.

Declaration of Ethical Standards

Authors declare to comply with all ethical guidelines, including authorship, citation, data reporting, and original research.

Credit Authorship Contribution Statement

Mustafa Çataltaş: Experiments, writing and editing. **Büşra Üstünel:** Data collection, software, and experiments. **Nurdan Akhan Baykan:** Writing, reviewing, supervision.

Declaration of Competing Interest

The authors declared that they have no conflict of interest.

Funding / Acknowledgements

The authors declare that they have not received any funding or research grants during the review, research or assembly of the article.

Data Availability

The datasets collected during the current study are not publicly available but are available from the corresponding author on reasonable request.

REFERENCES

[1] H. Abid, J. Mohd, and V. Raju, "Effects of COVID 19 pandemic in daily life," Current Medicine Research and Practice, vol. 10, no. 2, pp. 78-79, 2020.

[2] B. Gokulakrishnan, P. Priyanthan, T. Ragavan, N. Prasath, and A. Perera, "Opinion mining and sentiment analysis on a twitter data stream," in International conference on advances in ICT for emerging regions (ICTer2012), 2012: IEEE, pp. 182-188.

[3] B. Özyurt and M. A. Akçayol, "Fikir Madenciliği ve Duygu Analizi, Yaklaşimlar, Yöntemler Üzerine Bir Araştirma," Selçuk Üniversitesi Mühendislik, Bilim ve Teknoloji Dergisi, vol. 6, no. 4, pp. 668-693, 2018.

[4] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," Ain Shams Engineering Journal, vol. 5, no. 4, pp. 1093-1113, 2014.

[5] A. Singh, N. Thakur, and A. Sharma, "A review of supervised machine learning algorithms," in 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), 2016: IEEE, pp. 1310-1315.

[6] A. Öztürk, Ü. Durak, and F. Badilli, "Twitter Verilerinden Doğal Dil Işleme Ve Makine Öğrenmesi Ile Hastalik Tespiti," Konya Mühendislik Bilimleri Dergisi, vol. 8, no. 4, pp. 839-852, 2020.

[7] H. Çetiner, "Multi-Label Text Analysis With A CNN And LSTM Based Hybrid Deep Learning Model," Adıyaman Üniversitesi Mühendislik Bilimleri Dergisi, vol. 9, no. 17, pp. 15-16, 2022.

[8] A. H. Alamoodi et al., "Sentiment analysis and its applications in fighting COVID-19 and infectious diseases: A systematic review," Expert systems with applications, vol. 167, p. 114155, 2021.

[9] C. R. Machuca, C. Gallardo, and R. M. Toasa, "Twitter sentiment analysis on coronavirus: Machine learning approach," in Journal of Physics: Conference Series, 2021, vol. 1828, no. 1: IOP Publishing, p. 012104.

[10] L. Nemes and A. Kiss, "Social media sentiment analysis based on COVID-19," Journal of Information and Telecommunication, vol. 5, no. 1, pp. 1-15, 2021.

[11] Ö. Çoban, B. Özyer, and G. T. Özyer, "Sentiment analysis for Turkish Twitter feeds," in 2015 23nd Signal Processing and Communications Applications Conference (SIU), 2015: IEEE, pp. 2388-2391.

[12] Z. A. Guven, "Comparison of BERT models and machine learning methods for sentiment analysis on Turkish tweets," in 2021 6th International Conference on Computer Science and Engineering (UBMK), 2021: IEEE, pp. 98-101.

[13] Y. E. Karaca and S. Aslan, "Sentiment Analysis of Covid-19 Tweets by using LSTM Learning Model," Computer Science, no. Special, pp. 366-374, 2021.

[14] N. Öztürk and S. Ayvaz, "Sentiment analysis on Twitter: A text mining approach to the Syrian refugee crisis," Telematics and Informatics, vol. 35, no. 1, pp. 136-147, 2018.

[15] H. A. Shehu et al., "Deep sentiment analysis: a case study on stemmed Turkish twitter data," IEEE Access, vol. 9, pp. 56836-56854, 2021.

[16] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," CoRR, vol. abs/1810.04805, 2018.

[17] J. Ramos, "Using tf-idf to determine word relevance in document queries," in Proceedings of the first instructional conference on machine learning, 2003, vol. 242, no. 1: Citeseer, pp. 29-48.

[18] A. I. Kadhim, Y.-N. Cheah, I. A. Hieder, and R. A. Ali, "Improving TF-IDF with singular value decomposition (SVD) for feature extraction on Twitter," in 3rd international engineering conference on developments in civil and computer engineering applications, 2017.

[19] D. M. Christopher, R. Prabhakar, and S. Hinrich, "Introduction to information retrieval," ed: Cambridge University Press, 2008.

[20] T. Cover and P. Hart, "Nearest neighbor pattern classification," IEEE transactions on information theory, vol. 13, no. 1, pp. 21-27, 1967.

[21] L. I. Kuncheva, "On the optimality of Naïve Bayes with dependent binary features," Pattern Recognition Letters, vol. 27, no. 7, pp. 830-837, 2006.

[22] V. Vapnik, "Pattern recognition using generalized portrait method," Automation and remote control, vol. 24, pp. 774-780, 1963.

[23] A. Sherstinsky, "Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network," Physica D: Nonlinear Phenomena, vol. 404, p. 132306, 2020.

[24] M. V. Valueva, N. Nagornov, P. A. Lyakhov, G. V. Valuev, and N. I. Chervyakov, "Application of the residue number system to reduce hardware costs of the convolutional neural network implementation," Mathematics and computers in simulation, vol. 177, pp. 232-243, 2020.

[25] T. N. Sainath, O. Vinyals, A. Senior, and H. Sak, "Convolutional, long short-term memory, fully connected deep neural networks," in 2015 IEEE international conference on acoustics, speech and signal processing (ICASSP), 2015: IEEE, pp. 4580-4584.

[26] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," the Journal of machine Learning research, vol. 12, pp. 2825-2830, 2011.

[27] M. Abadi et al., "Tensorflow: Large-scale machine learning on heterogeneous distributed systems," arXiv preprint arXiv:1603.04467, 2016.

[28] F. Chollet, "keras," ed, 2015.

[29] D. M. Powers, "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation," arXiv preprint arXiv:2010.16061, 2020.