IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 17 December 2023, accepted 7 January 2024, date of publication 15 January 2024, date of current version 1 February 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3354165

RESEARCH ARTICLE

Detection of Turkish Fake News From Tweets With BERT Models

GÜLSÜM KAYABAŞI KORU AND ÇELEBİ ULUYOL
Computer Forensics Department, Gazi University, 06640 Ankara, Turkey
Corresponding author: Gülsüm Kayabaşi Koru (gulsum.kayabasikoru@gazi.edu.tr)

ABSTRACT As the number of people using social networks increases, more people are using social media platforms to meet their news needs. Users think that it is easier to follow the agenda by accessing news, especially on Twitter, rather than newspaper news pages. However, fake news is increasingly appearing on social media, and it is not always possible for people to obtain correct news from partial news pages or short Twitter posts. Understanding whether the news shared on Twitter is true or not is an important problem. Detecting fake tweets is of great importance in Turkish as well as in any language. In this study, fake news obtained from verification platforms on Twitter and real news obtained from the Twitter accounts of mainstream newspapers were labeled and, preprocessed using the Zemberek natural language processing tool developed for the Turkish language, and a dataset named TR_FaRe_News was created. Then, the TR_FaRe_News dataset was explored using ensemble methods and BoW, TF-IDF, and Word2Vec vectorization methods for fake news detection. Then a pre-trained BERT deep learning model was fine-tuned, and variations of the model extended with Bi-LSTM and Convolutional Neural Network (CNN) layers with the frozen and unfrozen parameters methods were explored. The performance evaluation was conducted using seven comparable datasets, namely BuzzFeedNews, GossipCop, ISOT, LIAR, Twitter15, and Twitter16, including an LLM-generated fake news dataset. Analyzing Turkish tweets and using fake news datasets generated by LLM is considered an important contribution. Accuracy values between 90 and 94% were obtained with the BERT and BERTurk + CNN models with 94% accuracy.

INDEX TERMS Fake news, generated news, ensemble learning, deep learning, BERT.

I. INTRODUCTION

Internet journalism is a concept frequently used in our age. Examples, which started with the messages sent by news groups to their subscribers, have led to the emergence of virtual journalism, which we call Internet journalism, with the development of various software and hardware applications over the Internet [1]. With the emergence of virtual journalism, the tendency to receive news through social media has begun.

Unlike traditional media, social media should not be considered just as media; it is an integrated concept that includes many dimensions. Journalists need to reach audiences in the cyberworld beyond known means of communication [2]. Social media statistics show that there are approximately 5.6 billion active Internet users worldwide as of January 2023 [3] and according to the research conducted by eBizMba, news sites are among the 15 most popular web pages in 2022 and the first quarter of 2023 [4].

When the "We Are Social Digital 2023 Global and Türkiye Report" is examined, Twitter ranks 4th in the most preferred social media list after Instagram, WhatsApp and Facebook with a rate of 66.5% [5]. Considering the statistics above, it is clear that people access news via Twitter as much as they prefer to access news from web pages [6].

On the other hand, fake news causes serious harm to society and increasingly appear on news pages and on Twitter. This has triggered several studies on fake news, including the development of fake news detection and real news verification techniques. News verification experts have created platforms that detect fake news in Turkish to verify the news that users share on social networks. The news

The associate editor coordinating the review of this manuscript and approving it for publication was Wentao Fan.

shared on the webpages of these platforms is also shared on Twitter accounts [7]. The main feature of these accounts is that they prove whether the news they verify is fake. The mentioned accounts include teyitorg, dogrulukpayicom, dogrulaorg, gununyalanlari, and malumatfurusorg. This study utilizes these accounts to identify and uncover instances of fake news. Teyitorg and dogrulukpayicom accounts are affiliated with the International Fact-Checking Network (IFCN), an organization that unites verification platforms worldwide to enhance supervision and responsibility [8]. The other accounts (dogrulaorg, gununyalanlari and malumatfurusorg) are non-members. They meet the network's criteria. These accounts engage in the dissemination of fake news and also create posts to increase public awareness and facilitate the identification of fake news.This study utilizes these stories to identify and uncover instances of misinformation. Real news, which is another focal point in this study, was taken from Twitter accounts of the most clicked mainstream newspapers [8]. Detailed information regarding this section is included in the dataset section of the study.

This study considered three aspects. First, within the scope of the study, a dataset named TR_FaRe_News was created, consisting of tweets taken between January 2020 and December 2022 and shared on the Twitter [7] platform. Second, the manually labeled dataset was converted into a vector using word representation methods, and the fake news classification process was carried out with our model created using machine learning algorithms and BERT. Finally, we included fake news datasets generated by large language models like GPT-2 in our scope and compared the model we built with BERT with human-generated datasets like our own dataset.

To summarize the purpose of the study, it was to create a fake news data set consisting of fake and real news in Turkish, introduce it to the literature and make it available to future researchers, and conduct experiments on the fake news dataset created by a large language model such as GPT-2 with human-generated datasets and state-of-the-art models we have created.

The following are the contributions to the literature made by this study:

- Fake news obtained from verification platforms on Twitter and real news obtained from the Twitter accounts of mainstream newspapers are labeled,
- The labeled news items were preprocessed with the Zemberek NLP tool for Turkish,
- The dataset generated after preprocessing is named TR_FaRe_News,
- Then the TR_FaRe_News dataset was explored using ensemble methods and BoW, TF-IDF, and Word2Vec vectorization methods for fake news detection,
- After that, a pre-trained BERT deep learning model was fine-tuned, and variations of the model extended with Bi-LSTM and CNN layers with the frozen and unfrozen parameter methods were tested,
- The performance evaluation was conducted with seven comparable datasets: BuzzFeedNews, GossipCop, ISOT, LIAR, Twitter15, and Twitter16, including even a GPT-2-generated fake news dataset,
- The TR_FaRe_News dataset was built in Turkish language and used for classification of fake news generated by LLM,
- Accuracy values between 90 and 94% were obtained with the BERT and BERTurk + CNN models with 94% accuracy.

The subsequent sections of this work are structured as follows: Section II presents a comprehensive analysis of existing literature on the subject matter. The specific architectural components of the proposed model are outlined in Section III. The experiments are showcased in Section IV. The findings and analysis are outlined in Section V. Section VI outlines the issues and constraints. Section VII provides an overview of potential enhancements for study, future endeavors, and final conclusions.

II. RELATED WORKS

Scientists in the field of natural language processing (NLP) are employing machine learning and deep learning techniques to identify and counteract fake news, a challenging endeavor that necessitates thorough comprehension and effective countermeasures.

Fake news detection is the basis of many tasks, such as news accuracy detection and classification. In the literature, there are many studies on fake news detection. Because collecting these studies under a single heading would cause semantic confusion, they are analyzed under three headings in this article. The initial two research employ supervised machine learning algorithms and ensemble learning techniques, whereas the latter studies utilize deep learning methods. The third heading is studies that detect fake news by creating a Turkish dataset, since the language used in our study was Turkish. The primary objective of this study is to analyze research conducted in languages other than Turkish, categorizing them into two distinct groups. The aim is to highlight the achievements of studies conducted in Turkish, considering the intricacies of the language, and to compare them with our own research. Additionally, this study aims to create a dataset in Turkish that can be utilized by other researchers.

A. FAKE NEWS DETECTION USING MACHINE LEARNING AND ENSEMBLE LEARNING APPROACHES

A multitude of research projects utilize machine-learning algorithms to identify and categorize fake news, and a substantial amount of inquiries have been undertaken, so enhancing the existing body of knowledge on ensemble learning techniques [9].

A study combining linguistic features and knowledge-based approaches achieved 94.4% accuracy, outperforming 89.4% using linguistic features separately. Support Vector

Machine (SVM) and Random Forest (RF) achieved 97% accuracy using LIAR dataset [10].

The study discusses spam detection using n-gram analysis, highlighting its advantages for fake content, with a 90% success rate achieved using the SVM algorithm [11].

The study [12] presents two news datasets for fake news detection, detailing data collection, annotation, validation, and linguistic differences. Comparative analyses show a 73% f1-score for automatic and manual identification.

A study conducted to identify fake news on COVID-19 in both Hindi and English languages produced an impressive accuracy rate of 93.45% in English and 97% in Hindi [9].

The study successfully detected fake news on multiple languages using conventional machine learning algorithms, with results ranging from 81% for TwitterBR to 95% for btvlifestyle [13].

The study used the TF-IDF method to obtain vector representations of news texts, followed by classification successes using 23 supervised AI algorithms, and evaluation metrics were compared [14].

FakeNewsNet, an extensive compilation of fake news, was launched in [15] with the purpose of facilitating studies on fake news by offering a wide-ranging collection of news articles, social context, and spatio-temporal data. The analysis examines the datasets from BuzzFeed and GossipCop from several angles and emphasizes the benefits of FakeNewsNet in identifying fake news on social media.

This study conducts a comparison of supervised machine-learning algorithms in order to automatically detect fake news. The systems are evaluated based on the features extracted from the news [16].

A study [17] assessed the efficacy of five machine learning and three deep learning models on two distinct datasets by employing deep learning in conjunction with conventional techniques. The claim was that the key to achieving high test accuracy was chunking.

The study employed capsule neural networks to detect fake news, adopting several word embedding models for news of different durations. Static word embeddings are employed for brief news articles, whereas non-static embeddings enable progressive training and updating during the training phase for moderate and extensive news articles [18], [19].

The work presents a machine learning ensemble method to automatically classify news articles, utilizing linguistic characteristics to differentiate between fake and real content. The approach outperforms individual learners in four real-world datasets [20].

The paper suggests utilizing an ensemble learning technique to tackle the issue of imbalanced data in Indonesian fake news datasets. The study showcased that the random forest classifier surpassed the multinomial classifier in ensemble classification, achieving an impressive f-1 score of 0.98. The Naïve Bayes and support vector machine classifiers, which were not ensemble models, were used to evaluate 660 documents. The f-1 scores obtained were 0.43 and 0.74, respectively [21].

The study proposes an intelligent detection system using an Ensemble Voting Classifier for real and fake news classification, utilizing 11 machine learning algorithms and detection techniques like Gradient Boosting and Ada Boosting, achieving 94.5% accuracy [22].

The UNBiased dataset, a new corpus of text, uses advanced linguistic features, word embeddings, ensemble algorithms, and SVMs to accurately classify fake news [23].

B. FAKE NEWS DETECTION USING DEEP LEARNING APPROACHES

The study investigates the efficacy of 19 machine learning methods in identifying fake news across three English datasets. Out of the total of 19 models, 8 were conventional deep learning models, while the remaining 5 were pre-trained sophisticated language models such as BERT. The findings indicate that models based on BERT exhibit superior performance compared to other models, but Naïve Bayes algorithms can reach comparable outcomes [24].

A study using a Convolutional Neural Network (CNN) achieved 85% accuracy in analyzing fake news, highlighting the need for a comprehensive understanding of its characteristics [25].

Another study proposes a system using a deep learning model to convert any word in an information message into an ideal measurement vector. Word vectorization effectively manages high-dimensional data variation, with LSTM model's accuracy reaching 91.73%, surpassing CNN and RNN models [26].

A study using active learning techniques achieved a 97.1% f1-score performance using Multilingual-BERT for solving multilingual fake news detection problems [27].

Researchers developed a deep neural network model to automatically detect truth in Arabic news or claims, achieving 91% accuracy when applied to an Arabic dataset [28].

The AugFake-BERT approach employs a cutting-edge BERT language model to classify data and mitigate underclassing problems by generating synthetic fake data. This approach achieves an impressive accuracy score of 92.45% [29].

A method for automatically detecting fake news integrates both textual and visual characteristics, while maintaining the semantic connections among words. The model attained classification accuracies of 93% and 92% for the PolitiFact and GossipCop datasets, respectively [30].

The study aims to predict fake news items using a NLP-based classifier, comparing results from multiple models and presenting a new design with an attention-like mechanism in a CNN [31].

The researchers used RNN to read news headlines and articles, comparing it with advanced systems, but found a significant issue with overfitting [32].

The Bi-LSTM model demonstrated the highest accuracy in feature extraction and stance classification using deep neural networks, outperforming RNN models and their extensions [33].

A study used deep learning techniques to develop a classifier for predicting fake news stories using RNN models and LSTMs, utilizing the LIAR dataset [34].

The study effectively attained a 98.9% accuracy rate by employing a BERT-based deep learning method to categorize parallel segments of a single-layer deep CNN with different kernel sizes and filters [35].

An investigation on the impact of margin loss CNNs exhibited inferior performance on the LIAR dataset in comparison to the ISOT dataset for the purpose of fake news detection [36].

C. FAKE NEWS DETECTION USING TURKISH LANGUAGES

The study focuses on fake news detection in Turkish using SVM and NB classifiers. The datasets used include term frequency, TF-IDF, n-gram, style markers, slang usage, url, accessible link features, Headline and News content compatibility, and Exaggerated Headline usage. The study obtained a 79% f1-measure using an SVM classifier [37].

The study suggests automatic mechanisms to verify digital content reliability in libraries without manual verification, preventing fake news spread. A dataset was generated by utilizing both real news and fake news. The ExtraTrees classifier was employed, resulting in an impressive accuracy rate of 96.81%. [38].

A study on Twitter identified two popular fake news topics in August 2019. Data was collected using Twitter Scraper and labeled using a labeling platform. The dataset comprises 1287 tweets and is the inaugural attempt at identifying Turkish fake news on social media [39].

The study presents Turkish fake news detection approaches using the SOSYalan Turkish dataset, demonstrating that deep learning models outperform existing literature in both Turkish and English [40].

This study highlights the importance of developing a fake news identification model for COVID-19 in the Turkish language, proposing an advanced deep-language transformer. The study developed a model to identify genuine COVID-19 news in Turkey sourced from social media. There are five conventional machine learning algorithms and deep learning algorithms like LSTM, Bi-LSTM, CNN, and GRU were tested, with BERT and variations improving efficiency and achieving 98.5% accuracy [41].

The study evaluated supervised and unsupervised machine-learning algorithms on Turkish pseudonym datasets, achieving an 86% f1-score for supervised algorithms and 72% for unsupervised algorithms [42].

III. MACHINE LEARNING AND BERT BASED MODELS

A. EXPLORED MACHINE LEARNING MODELS

This study employed four machine learning algorithms, including MultinomialNB (MNB), RF, LR, and SVM, to classify fake news. The study also evaluated the effectiveness of the Voting Classifier (VC) and ensemble systems in classifying fake news. Fig. 1 displays the architectural design of the model.

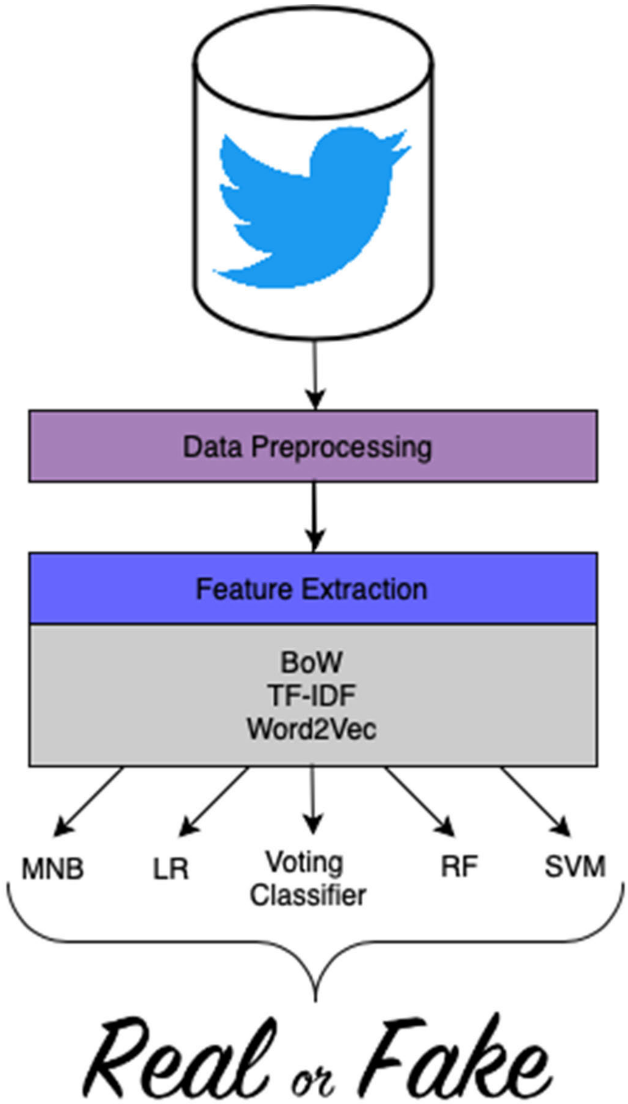

FIGURE 1. Machine learning models.

As seen in Fig. 1., the tweet texts in the TR_FaRe_News dataset were first passed through data preprocessing steps. Feature extraction was performed with BoW, TF-IDF and word2Vec and then classified.

The performance of the classification models was assessed using commonly employed measures such as accuracy and f1 values. Calculating these measures necessitates the use of two parameters: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN).

The study presents the formulas for calculating the performance metrics that are used to evaluate the model performance. These formulas are shown in Equations (1) and (2).

$$Accuracy := \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$F1 - Measure := \frac{2xTP}{2xTP + FP + FN} \quad (2)$$

Classification results are shown in Table 1 below.

TABLE 1. Machine learning results.

Method	TR_FaRe_News Dataset			GPT-2 Fake News Dataset	
	Accuracy	f1-score		Accuracy	f1-score
MNB+BoW	%71	%66.2		%63.2	%63
MNB+TF-IDF	%71	%66.2		%62.7	%62.7
MNB+Word2Vec	%85.2	%85.1		%69.6	%69.5
RF+BoW	%73.5	%79.1		%65.3	%66
RF+TF-IDF	%73.4	%79.4		%65.3	%65.2
RF+Word2Vec	%90	%90.9		%76.7	%76.6
LR+BoW	%73.6	%71.3		% 65.7	%65.5
LR+TF-IDF	%72.9	%71.2		%64.7	%64.6
LR+Word2Vec	%88.5	%88.4		%77.1	%77.1
SVM+BoW	%73	%70.4		%65.6	%65
SVM+TF-IDF	%71.9	%69.7		%64.1	%64.1
SVM+Word2Vec	%88	%88		%77	%77
VC+BoW	%73.5	%71.2		%65.8	%65.5
VC+TF-IDF	%72.8	%71		%64.9	%64.8
VC+Word2Vec	**%89**	%88.9		**%77.6**	%77.6

TWhen Table 1 is examined, when we compare the TR_FaRe_News dataset we created with the fake news dataset created by GPT-2, which is a large language model, the success rate of the community systems' voting classifier is 89% with the word2vec development process, the most performant algorithm for the TR_FaRe_News dataset. For the GPT-2 dataset, the best performing algorithm was found to be the voting classifier and achieved 77.6% success.

B. BERT MODELS

The BERT language representation model is considered a very powerful model for language based tasks [43]. It was first introduced in 2018. It is designed to perform pre-training in a bi-directional way in all layers that can handle unlabelled sequential data. The BERT model employs an attention mechanism to acquire the contextual associations among words in the input text of the transformer. The system employs an encoder to analyze the input text and build word embeddings, and a decoder to forecast the outcome [43]. In this way, because the vectors are produced in accordance with the semantic context of the words, the language is better understood and homophones used for different purposes can be distinguished [12].

A pre-trained state-of-the-art BERT model can be adapted or fine-tuned for various tasks. These tasks can include question answering, language recognition, etc. The BERT model is considered state-of-the-art because it consistently achieves the best accuracies in several natural language processing tasks [43]. The process of adapting BERT involved both pre-training and fine-tuning. During the pre-training phase of the BERT model, the model underwent training for various pre-trained tasks utilizing unlabeled data [43]. Features were extracted from a pre-trained model [44]. The fine-tuning method incurs lower costs compared to the pre-training process. The fine-tuning process begins by utilizing the pre-trained parameters. The parameters were updated with labelled data prepared according to the type of study. Except for the output layers, the architectures used for fine-tuning and pre-training were the same. DistilBERTurk and BERTurk were used in this study.

1) DISTILLBERT/DISTILLBERTURK (MODEL 1)

DistilBERTurk [45] has a transformer architecture that is similar to that of BERTurk. Distillation, a process carried out during the pre-training phase, is executed in the fine-tuning step based on the specific task at hand. The number of layers was reduced by half, and the algebraic processes were optimized. By implementing multiple adjustments, DistillBERTurk achieved comparable outcomes despite its 40% smaller size compared to BERTurk [46].

The sentences in the dataset were first tokenized using the DistilBERTurk tokenizer trained to create word embeddings (768 dimensions), converted into tensors, and provided to the model. Subsequently, a basic neural network architecture consisting of Dense and Dropout layers was used for the forward classification task and training with DistilBERTClass. This is the first model we created for BERT models within the scope of this study.

DistilBERT is a variant of BERT that demonstrates commendable performance. Consequently, we incorporated it into our analysis to evaluate its performance in comparison to other BERT models. The settings utilized in the DistilBERTurk configuration are enumerated in Table 2.

TABLE 2. DistillBERT parameters.

Parameters	Explanation	Value Used
activation	Activation function	sigmoid
Epochs	Number of epochs	5
Dim	Word vectors dimension	768
Hidden_dim	Hidden dimension	3072
Vocab_size	Vocabulary size	32000
N_layers	Number of layers	6

DistilBERT is a highly effective variant of BERTurk. Consequently, we utilized it to evaluate its efficacy in comparison to other BERT models. The DistilBERT configuration includes the settings indicated in Table 2.

The results obtained in line with the parameters used were calculated as the accuracy, precision, recall, and f1-score. The truth table of DistillBERTurk is presented in Table 3.

Additionally, the DistillBERTurk model used was tested on BuzzFeed, GosspCop, ISOT, LIAR, Twitter15, Twitter16 and GPT-2 fake news datasets. The fields used in the datasets were updated to suit our model and were used in our experiments. The parameters used in our dataset were also used in the datasets in question.

TABLE 3. DistillBERTurk truth table.

	TR_FaRe_News Dataset		
	Precision	Recall	F1-score
Fake	0.91	0.92	0.91
Real	0.93	0.92	0.93
Accuracy			**0.92**
Macro avg	0.92	0.92	0.92
Weighted avg	0.92	0.92	0.92

A comparison of the results obtained with our DistillBERTurk model with the TR_FaRe_News dataset created within the scope of this study is presented in Table 4.

TABLE 4. A comparison table of the results obtained with our DistillBERTurk model.

Dataset	Accuracy
BuzzFeedNews	0.73
GossipCop	0.85
ISOT	0.97
LIAR	0.61
Twitter15	0.92
Twitter16	0.91
GPT-2	0.90
TR_FaReNews	0.92

2) BERT/BERTURK(MODEL2-6)

The Turkish BERT model (BERTurk) [47]was pre-trained on the Oscar Corpus, Opus Corpora and Wikipedia. The model consists of 12 transducer layers. BERTurk models vary in word sizes of 32K and 128K, and both are available in cased and caseless versions.

The classification task in this study utilized the original BertForSequenceClassification model. Subsequently, the BERT model underwent fine-tuning, followed by the replication of the same experiment. Finally, following the fine-tuned model, additional layers were added for both BERTurk+CNN and BERTurk+Bi-LSTM, such as freezing and unfreezing parameters in the fine-tuned model. Training was then performed for model adaptation, hyperparameterisation, and testing. In addition to the success of BERTurk in word-embedding tasks, the aim of using CNN is to provide a deep learning model that uses neural network layers to define the model and is powerful enough to process large amounts of data using a network of hidden layers [50]. When using Bi-LSTM and BERT together, the aim is to extract the improved features of BERTurk and achieve better learning performance. It is also believed that Bi-LSTM will better capture the global context. The results of CNN and Bi-LSTM on short sentences datasets such as Twitter are very good when analyzed in the literature. The fact that each component of the Bi-LSTM input sequence contains information from both past and present has helped us to produce more meaningful output [51]. The procedure of our main model, the improved BERTurk model, is presented in Figure 3.

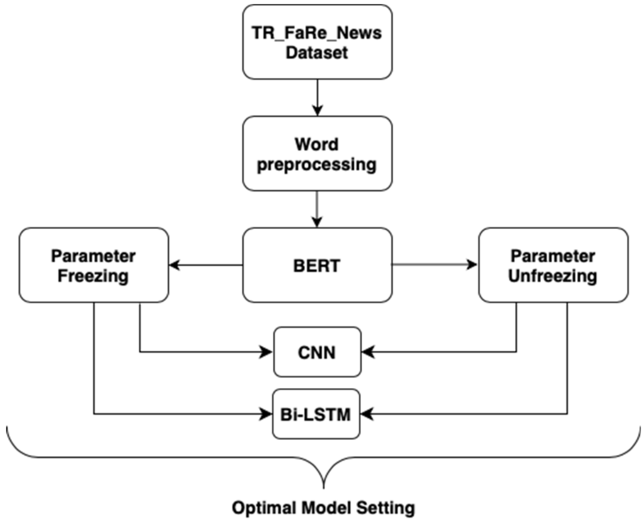

FIGURE 2. Optimal model search architecture.

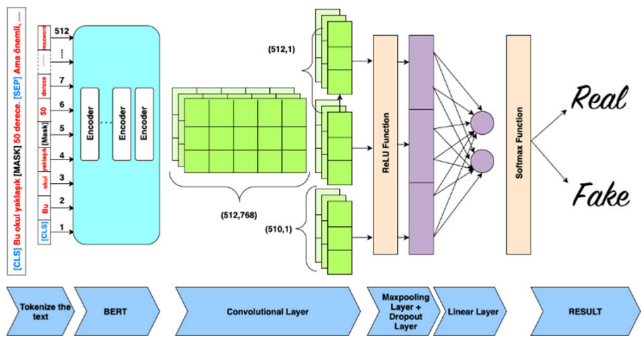

FIGURE 3. BERTurk+CNN architecture.

Five models were created using BERTurk. The aim was to obtain an optimal model by using these models. The process followed to reach the optimal model is illustrated in Fig. 2.

In this study, six different models were created for BERTurk and DistillBERTurk. The initial model was elucidated in the preceding section. The second model was defined as the fine-tuned BERTurk model. The third model is defined as a BERTurk fine-tuned model with CNN layers and frozen parameters. The fourth model was defined as the BERTurk fine-tuned model with CNN layers without frozen parameters. The fifth model is defined as the BERTurk fine-tuned model with BiLSTM layers with frozen parameters. Finally, the sixth model was defined as a BERTurk fine-tuned model with BiLSTM layers without frozen parameters.

There are two techniques available when using BERTurk. these techniques are MLM and NSP techniques. Optimization during training ensures that the loss when using these two techniques is minimized. Since the features extracted in the first stage are generic, parameter freezing was performed to access meaningful features more easily. The use of frozen parameters in large language models and in our own Turkish dataset shows the originality of the study.

BERTurk fine-tuning (Model 2)

The learning rate utilized for Model 2, also known as BERT fine-tuning, was set to 2e-5. The model underwent four rounds of fine-tuning.

BERTurk fine-tuning + CNN (Model 3 and Model 4)

For Models 3 and 4, two CNN layers of kernel size (1,768) and (2,768) were added after the BERTurk fine-tuned Model. The activation procedure is succeeded by a maximum pooling layer, where the kernel size is set to the previous output size and the step size is determined by the prior height of this output. Ultimately, a linear layer is employed, followed by the application of a softmax function. Equation (3) defines the softmax activation function.

$$\sigma(\mathrm{z})_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \tag{3}$$

Because the news in our dataset has two classes (fake and real), *K was set to 2.*Variable *z*represents the input. The class that yields the greatest value when sent through the softmax activation function can be regarded as the outcome of the classification process.In addition, the learning rate of Models 3 and 4 is *2e-5*and the number of rounds is *four,*as in the fine-tuning of BERTurk. The architecture of the BERTurk+CNN model is illustrated in Fig. 3.

BERTurk fine-tuning + Bi-LSTM (Model 5 and Model 6)

After the BERTurk model for Model 5 and Model 6, 2 Bi-LSTM layers were applied to Model 5 and 1 Bi-LSTM layer was applied to Model 6. Afterwards, a linear layer was utilized using a softmax activation function. In addition, the learning rates for Models 5 and 6 were *5e-5,*the number of rounds used for Model 5 was *10,*and the number of rounds used for Model 6 was *6.*The structure of the BERTurk+Bi-LSTM model is depicted in Fig. 4.

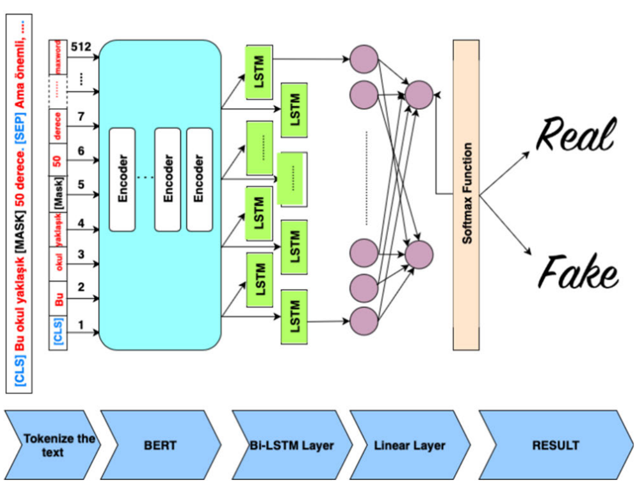

FIGURE 4. **BERT+BiLSTM architecture.**

Various measurements were used to test the prediction results of classifiers on the fake news dataset. First, test accuracy was calculated as the primary measurement. In this study, the calculation shown in (4) was used for test accuracy.

$$\frac{correctclassifiednewsnumber}{totalnewsnumber} \tag{4}$$

The second metric used is the ROC AUC score. The ROC AUC represents the extent of the area enclosed by the ROC curve. The ROC AUC score ranges from 0 to 1, and a number close to 1 indicates excellent performance in predicting classifications. The final metric was the F1 score. The F1 score, similar to the AUC value, varies between 0 and 1, and is calculated by remembering the precision results. The accuracy value is calculated by dividing the number of real positive outcomes by the total number of positive results. Recall, also known as sensitivity, is calculated by dividing the number of real positive results by the total number of samples that should be classified as positive. A higher score corresponds to superior achievement. The calculation of the f1-score is demonstrated by Equation (5).

$$f1 - score = 2\mathrm{x}\frac{precision \cdot recall}{precision + recall} \tag{5}$$

Five models were developed using BERTurk to assess and compare the effectiveness of fake news classification.

After training the created models on our TR_FaRe_News dataset consisting of Turkish news tweets, we obtained various results when we tested them with a validation dataset created with tweet sentences that were not in our dataset, which we call live data. The validation dataset consists of parody accounts on Twitter and current tweets of mainstream Twitter accounts. All model results with GPT-2 Fake News Dataset are presented in Table 5 and all model results with TR_FaRe_News Dataset are presented in Table 6.

TABLE 5. **Model results with GPT-2 fake news dataset.**

Model	GPT-2 Fake News Dataset					
	Train acc	Training loss	Val acc	Val loss	ROC AUC	F1 Score
Model 1	0.99	0.29	0.90	0.53	0.90	0.91
Model 2	0.99	0.003	0.94	0.018	0.93	0.93
Model 3	0.97	0.021	0.94	0.023	0.94	0.94
Model 4	0.98	0.021	0.93	0.023	0.93	0.92
Model 5	0.98	0.021	0.95	0.023	**0.95**	0.94
Model 6	0.98	0.020	0.95	0.023	**0.95**	0.94

TABLE 6. **Model results with TR_FaRe_news dataset.**

Model	TR_FaRe_News Dataset					
	Train acc	Training loss	Val acc	Val loss	ROC AUC	F1 Score
Model 1	0.97	1.11	0.91	1.12	0.90	0.90
Model 2	0.99	0.003	0.94	0.025	**0.94**	0.94
Model 3	0.98	0.0205	0.94	0.0233	**0.94**	0.94
Model 4	0.97	0.0215	0.91	0.0253	0.91	0.91
Model 5	0.97	0.0216	0.91	0.0254	0.91	0.91
Model 6	0.97	0.0212	0.91	0.0247	0.91	0.91

A series of operations were conducted to make it easier for us to access the results in Table 6. The ReLU activation function was utilized due to its low computational cost, while softmax activation created 2 classes and 2 nodes. Max pooling and dropout were employed to reduce large dimensions and improve learning rate. Max pooling was employed to minimize the time-consuming process of selecting the largest value in a matrix to reduce its size. Dropout was employed to prevent data memorizing, while AdamW optimization function was utilized for optimization, enhancing learning rate and creating a timer.

The ROC curve of the classification process with the TR_FaRe_News dataset for Turkish is shown in Figure 5.

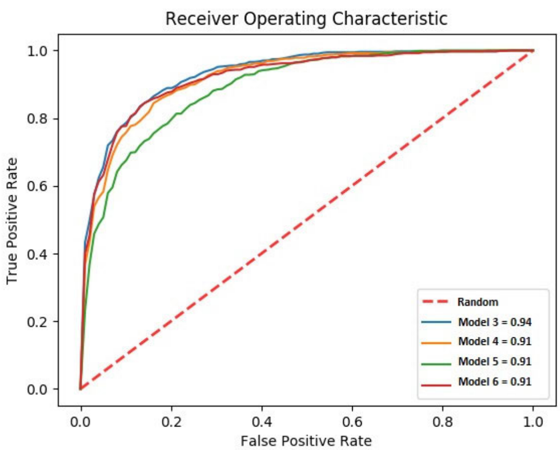

FIGURE 5. ROC curve.

Furthermore, ROC charts were generated for each individual model. Binary classification is a commonly employed technique, particularly when constructing ROC curves in academic research. The primary metric of interest in the ROC curve, which is a properly scaled graphical representation, is the Area Under the Curve (AUC). AUC, or the area under the receiver operating characteristic (ROC) curve, quantifies the accuracy and effectiveness of the model. The value must fall within the range of 0 and 1. A model is considered more successful as its AUC value approaches 1.

IV. EXPERIMENTS

A. DATASETS

For this study, a dataset called TR_FaRe_News (Turkish Fake and Real News) consisting of Turkish fake and real news was created. Tweets taken using the Tweepy module developed for Twitter were manually labelled so that tweets taken from mainstream newspapers were real, and false news shared by the fact-checking platform was labelled fake news. After the labelling process, similar tweets were found in the data and sorted. Before extraction, cosine similarity was calculated for the news in the dataset. Tweets that were very similar were eliminated. Here, if the news tweet in the real news dataset is considered A and the news tweet in the fake news dataset is considered B, the mathematical process is calculated using formula (6).

$$\cos(\theta) = \frac{\mathbf{AxB}}{\|\mathbf{A}\| \ \mathbf{x} \ \|\mathbf{B}\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \sqrt{\sum_{i=1}^{n} B_i^2}} \tag{6}$$

After sorting similar tweets, tweets containing words with high word counts that would negatively affect the classification in the fake and real news datasets were also edited. The word frequencies for the fake and real news datasets are depicted in Fig. 6 and Fig. 7, correspondingly.

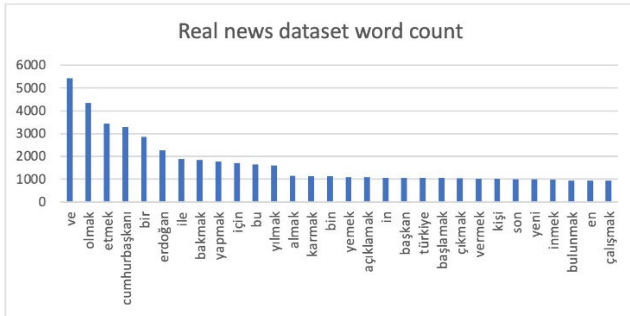

FIGURE 6. Real news dataset word count.

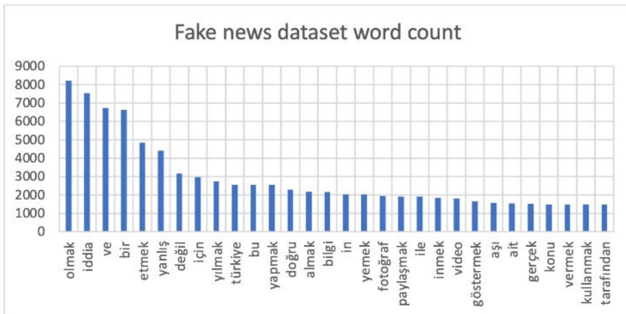

FIGURE 7. Fake news dataset word count.

The manually created TR_FaRe_News dataset was used for the Turkish fake news classification. In addition, BuzzFeedNews, GossipCop, ISOT, LIAR, Twitter15, and Twitter16, which are used to detect fake news in English, were used for verification. Datasets created using large-language models were also used. A fake news dataset created by GPT-2 was also included in this study. The dataset for the experiments is divided into three parts: train, validation (val), and test. The dataset statistics are listed in Table 7.

B. DATA PREPROCESSING

Data preprocessing involves editing the dataset before converting it into numbers that can be understood by the computer. Since Turkish is an agglutinative language, it presents difficulties in terms of natural language processing [48]. To overcome these difficulties, Zemberek was used to analyze texts and access word roots [49]. The StemmingAndLemmatization package of Zemberek, an open source NLP toolkit

TABLE 7. Data istatistics.

Datasets	Split	Real	Fake	Total
BuzzFeedNews	train	73	73	146
	val	8	8	16
	test	19	19	38
GossipCop	train	4,326	4,271	8,597
	val	480	474	954
	test	1,081	1,068	2,149
ISOT	train	16,336	14,323	30,659
	val	1,815	1,591	3,406
	test	4,084	3,580	7,664
LIAR	train	1,995	1,676	3,671
	val	221	186	407
	test	498	419	917
Twitter15	train	244	268	512
	val	27	29	56
	test	61	67	128
Twitter16	train	138	139	277
	val	15	15	30
	test	34	34	68
GPT-2 Fake news dataset	train	5,000	5,000	10,000
	val	2,500	2,500	5,000
	test	1,250	1,250	2,500
TR_FaReNews	train	6,869	6,867	13,736
	val	763	763	1,526
	test	1,717	1,716	3,433

implemented in Java, was utilized for processing the Turkish language. What to do here is:

- The stemming process attempts to find the word root by cutting the suffixes and prefixes in a word,
- Lemmatization is based on the morphological analysis of words. Therefore, the algorithm requires a detailed dictionary to obtain the word root.

To effectively perform text classification and create a classification model, a dataset must be preprocessed. Studies have shown that the results obtained during the classification of data from preprocessed datasets are better than those obtained from non-preprocessed data [48]. The objective of the preprocessing procedures is to enhance performance by reducing the size of the vector space and dimensions. The TR_FaRe_News dataset was utilized for our investigation, and the open-source Zemberek Library [49] was chosen for natural language processing (NLP). The code developments in the library were implemented using Java. The total number of news tweets was calculated according to the operations performed on the news tweets taken from Twitter. The news tweet statistics for the TR_FaRe_News dataset are listed in Table 8.

TABLE 8. News tweet statistics for the TR_FaRe_news dataset.

Process	Mainstream news tweets (AACanli, anadoluajansi, dhainternet, Hurriyet ve ihacomtr)	Fact-checking platform news tweets (dogrulaorg, dogrulukpayicom, gununyalanlari, malumatfurus, teyitorg)	Total
Number of tweets not processed	27,298	38,170	65,468
Deleting duplicate records	27,286	35,209	62,495
Character clearing	20,927	23,819	44,746
Zemberek NLP	26,302	23,507	49,809
Deleting last duplicate records	25,624	21,194	46,818

Following the preparation procedures, our data corpus had a total of 46,818 tweet sentences, but after applying the similarity theorems explained in the previous section and deleting duplicate records again, 18,695 tweet sentences were included in our dataset. The steps we performed for preprocessing are as follows.

- Repeated tweets have been deleted,
    - Special characters, emojis and punctuation marks in tweets are written with words in the tweet sentences. This prevents NLP. Therefore, these have been deleted.
    - Erroneous words or missing letters entered by the users (even by mistake) were normalized or corrected by appliying them to the data corpus.
    - Twitter mentions (starting with @) and tweet hashtags (starting with #) were deleted.
    - Stop words that had no meaning in the sentences or have no effect on the meaning of the sentence were deleted.
    - Because the lowercase mode of the BERTurk model was used, all words were converted to lowercase letters, which have been used in all other models.

The outputs obtained from the applied process are listed in Table 9.

TABLE 9. An example of preprocessing an original tweet.

Process	Output
Original tweet	#@milliyet Rusya'da savaşa gitmemek için denek olark tecavüzcüler kullanıldı \n\n https://url "
Character clearing	#@milliyet Rusya'da savaşa gitmemek için denek olark tecavüzcüler kullanıldı n n https://url
Misspelling correction	#@milliyet Rusya da savaşa gitmemek için denek olarak tecavüzcüler kullanıldı n n https://url
Deleting Mention and Hashtags	Rusya da savaşa gitmemek için denek olarak tecavüzcüler kullanıldı n n https://url
Deleting URLs	Rusya da savaşa gitmemek için denek olarak tecavüzcüler kullanıldı n n
Semantic root extraction and conversion to lowercase	rusya da savaşa gitmemek için denek olarak tecavüzcüler kullanıldı
Deleting stop words	rusya savaş gitmek denemek olmak tecavüz kullanmak
Output	rusya ilaç gitmek denemek olmak tecavüz kullanmak

C. PARAMETER SETTINGS

The parameters of our first model, DistillBERT, are presented in Section III. An further five models were developed via BERTurk in order to assess and contrast the efficacy of false news classification. The hyperparameter tables for these models are listed in Table 10.

TABLE 10. Hyperparameters used for the BERTurk model.

Model	Learning rate	Epoch	Number of layers added	Optimization algorithm
Model 2	2e-5	4	1	AdamW
Model 3	2e-5	4	2	AdamW
Model 4	2e-5	4	2	AdamW
Model 5	5e-5	10	2	AdamW
Model 6	5e-5	6	1	AdamW

D. EXPERIMENTAL RESULTS

The common table of the results we obtained when we tested the 6 models we created with BERTurk and DistillBERTurk with BuzzFeedNews, GossipCop, ISOT, LIAR, Twitter15, Twitter16, GPT-2 and TR_FaReNews, the dataset we created, is presented in Table 11 A comparison of the results obtained with our BERTurk model with the TR_FaRe_News dataset created within the scope of the thesis is presented in Table 12.

As seen in Table 11, if we compare our BERTurk model with the dataset TR_FaRe_News created within the scope of

TABLE 11. Experimental results.

Datasets	Models					
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
BuzzFeed News	0.73	0.89	0.86	0.82	0.82	0.82
GossipCop	0.85	0.80	0.79	0.77	0.78	0.78
ISOT	0.97	0.98	0.96	0.93	0.93	0.93
LIAR	0.61	0.62	0.65	0.65	0.63	0.64
Twitter15	0.92	0.91	0.93	0.93	0.93	0.93
Twitter16	0.91	0.95	0.95	0.95	0.95	0.95
GPT-2	0.90	0.95	0.94	0.93	0.95	0.95
TR_FaRe News	0.92	0.93	**0.94**	0.91	0.91	0.91

TABLE 12. BERTurk comparison table with all datasets.

Dataset	Acuracy
BuzzFeedNews	0.86
GossipCop	0.75
ISOT	**0.94**
LIAR	0.60
Twitter15	0.89
Twitter16	**0.97**
GPT-2 Fake News Dataset	**0.95**
TR_FaReNews	**0.94**

this thesis, it is seen that it achieves a better accuracy rate in the ISOT dataset than the TR_FaRe_News dataset with a slight difference. We can say that the main reason here is that our dataset consists of short tweets on Twitter and the language analysis difficulties of the Turkish languages. However, if we look at the Section II, we see that good results are obtained with the TR_FaRe_News dataset.

V. RESULTS AND DISCUSSIONS

The word clouds of fake and real news in the TR_FaRe_News dataset created within the scope of the study were extracted. Word clouds were created for the first 50 most common words in the dataset. The word clouds are shown in Figure 8.

In order to assess the effectiveness of the BERT models developed for Turkish fake news detection and classification using deep learning, a comparison was conducted among Turkish studies focused on fake news detection. The comparison is displayed in the Table 13.

When we analyze the table, we observe that among supervised machine learning algorithms, the SVM models perform better for Turkish fake news detection. When [37] and [38] were examined, it was observed that the performance of

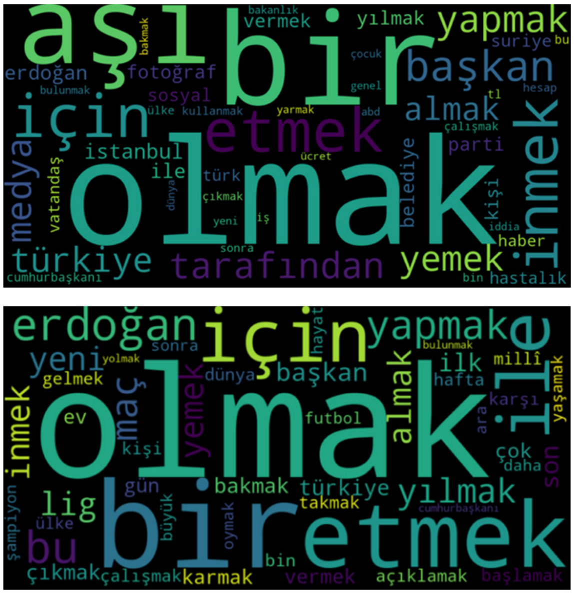

FIGURE 8. Real and fake news dataset word clouds.

the model we created was higher when word2vec word embedding was performed. Additionally, when the ensemble learning perspective we offer in our machine learning model was added, this performance increased even more with the voting classifier, and a performance of 89% was achieved when trained with the TR_FaRe_News dataset and 77.6% when trained with the GPT-2 fake news dataset.

This study, which has not yet been shared with the literature but creates a specific dataset from their studies, includes tweets based on three topics [39]. Because the authors gave f1 score results when evaluating performance, it would not be appropriate to make a comparison in terms of accuracy rate. However, when we compare the f1 score results, it is seen that the score results of the models presented in the study in question are between 57% and 89%, and these results are below those of the BERT models in our study. The f1 score of all five BERT models in our study was over 90%.

Furthermore, the SOSYalan dataset, a notable study in the literature for identifying fake news in Turkish, yielded successful outcomes, but when evaluated in terms of dataset size, it is estimated that the classification success is likely to be high because it is a small dataset [40]. The BERT models were not included in the study.

In terms of the data in the dataset, the study created by selecting tweets on the subject of Covid-19 appears to be very similar to our study. Since our study does not include a single subject, it has a good rate among the literature studies, although there is a slight decrease in performance. In addition, their study did not include a comparison table with important datasets in the literature [41] and [42] has been explained in related works.

Experiments were also conducted for the fake news dataset produced by GPT-2, a large language model we used in our

TABLE 13. The comparison of Turkish studies on fake news detection.

Models	Dataset	Task	Performance
SVM, NB [38]	Specified dataset	Fake news detection with TF-IDF, n-gram, style markers, use of slang/profanity, url, accessible link feature in the news, headline and news content compatibility, and exaggerated headline use features.	f1-score %71-79
NB, KNN, ExtraTrees, SVM, LR, RF, DT [39]	TR_FN	Fake news detection with root count, raw number, word-per-syllable error, news readability, source, news category and news address features.	%89-96
SVM [40]	Specified dataset	Word embedding techniques such as TF-IDF and word2vec are employed for the purpose of detecting fake news.	F1-score %57-89
CNN, RNN-LSTM [41]	BuzzFeed, ISOT, SOSYalan	Fake news detection by word2vec and vector representation per word	%87.14-92.48
LSTM, BiLSTM, GRU, Bi-GRU, BERT, RoBERTa, DistilBERT, BERTurk [42]	Twitter, fact-checking platform Twitter accounts, COVID-19 dataset	Fake news detection using Information Gain, Gain Ratio, Correlation Based Features features	%89-98.5
K-NN, SVM, RF, K-means, NMF, LDA [43]	Specified dataset	Fake news detection with TF-IDF, Word2Vec, Doc2Vec representation methods	F1-score %72-86
MNB, RF, LR, SVM, Voting Classifier, DistillBERTurk, BERTurk	GPT-2 dataset	Fake news detection with BoW, TF-IDF, word2vec representation methods and fine-tuned BERT	%62.7-95

TABLE 13. *(Continued.)* **The comparison of Turkish studies on fake news detection.**

MNB, RF, LR, SVM, Voting Classifier, DistillBERTurk, BERTurk	TR_FaRe News	Fake news detection with BoW, TF-IDF, word2vec representation methods and fine-tuned BERT	%71-94

study. As a result of the experiments, an accuracy between 62.7% and 95% was obtained. This usage does not appear in any study in the Turkish literature. In addition, the performance achieved for our dataset, TR_FaReNews, was between 71% and 94%. This performance is considered to be successful when considering the dataset we prepared.

VI. CHALLENGES AND LIMITATIONS

The challenges and limitations we faced in the study are listed below;

- In the Twitter API's free version, only tweet sentences from the last seven days were available, making it difficult to obtain data.
- We retrieved the data using Twitter usernames, so we had all tweets from these usernames in the last seven days. Considering that we extracted tweets over a two-year period (from January 2020 to December 2022), a maximum of 3250 tweets each time the app was run.
- Considering the way Twitter is used in terms of the number of characters, it has been observed that there are also posts consisting of very few words. One of the factors that reduce the success of text classification processes is the low number of words. As a result of the data preprocessing steps, tweets that did not contain any words were also encountered.
- Turkish exhibits a markedly distinct structure compared to English. Therefore, the Zemberek Natural Language Processing (NLP) tool was employed to examine word affixes, since there is a scarcity of available sources.
- The large size of the datasets generated by a large language model such as GPT-2, has led to temporal problems in utilizing the entire dataset.

VII. CONCLUSION

This study utilized news accuracy platforms to detect and classify Twitter posts' success, creating classification models and testing their effectiveness using real-time tweet data.

The studies conducted a literature review on detecting fake news on Twitter and discussing its connection to disinformation, deception, and misinformation. The study detailed the datasets used in fake news detection applications, tests, and research, examining their role in the experiments conducted within the study's scope. In this context;

- First, The study created Turkish datasets using 18695 tweets and two tags, with 13736 for training, 3433 for testing, and 1526 for verification in the Twitter environment. The Turkish language dataset has been used for detecting fake news, providing a comprehensive source for future studies and contributing to the literature on fake news detection.
- Then, the data was collected and labeled, then preprocessed using the Zemberek library for Natural Language Processing (NLP) processes, bringing the roots to a semantic level. This is suitable for models utilizing machine learning and deep learning.

Machine learning and deep learning studies on TR_FaRe_News dataset detected fake news tweets with six models, including DistillBERTurk, achieving accuracy values of 90-94%. The models tested on different datasets exhibit various differences. The model's creation for the Turkish language and its current results indicate a difference in usage patterns among BERT models developed for the Turkish language.

This study analyzed Twitter tweet news, focusing on short texts with low word frequency and frequent topics. Despite these disadvantages, high accuracy rates were achieved due to considerations like special expressions, abbreviations, spelling errors, and NLP losses.

The study created a Turkish dataset called TR_FaRe_News using Twitter data from fact-checking platforms and mainstream news agencies. The dataset was divided into 10 parts for reliability and classification performance was evaluated using supervised machine-learning algorithms and a voting classifier. The BERTurk + CNN model achieved 94% accuracy.

The study utilized GPT-2's fake news dataset, a large language model, in experiments, resulting in impressive performance, unlike any previous Turkish study using such news.

The Turkish fake news detection field is limited by a lack of references and an accessible dataset in literature. Our study yielded higher results than other Turkish fake news detection classification studies, with similar results observed in English fake news detection studies. Our study utilized the GPT-2 fake news dataset, which is not commonly used in literature for detecting fake news through the creation of a model. The language models developed in this study are deemed to be both original and successful.

We plan to utilize the TR_FaRe_News dataset developed within our study's scope in various models in the future. The study will also involve experiments using GRU models and various other generative models. The models we developed will be tested with larger language models beyond GPT-2. Our study is anticipated to serve as a valuable resource for future research in Turkey. The TR_FaRe_News dataset, utilizing classification models and findings, will significantly contribute to the literature in Turkish Fake News Detection.

This dataset serves as a crucial foundation for academic research.

DATA AVAILABILITY STATEMENT

The datasets utilized in the studies, with the exception of the TR_FaRe_News dataset, are readily accessible to the public and can be obtained by following the links provided below:

- BuzzFeedNews, GossipCop: https://github.com/ KaiD-MML/FakeNewsNet
- ISOT: https://onlineacademiccommunity.uvic.ca/isot/2022/11/27/fake-news-detection-datasets/
- LIAR: https://paperswithcode.com/dataset/liar
- Twitter15, Twitter16: https://www.kaggle.com/datasets/lhyimp/twitter1516
- GPT-2 Fake News Dataset: https://openaipublic.blob.core.windows.net/gpt-2/output-dataset/v1/
- TR_FaRe_News: To acquire it, simply write an email to the corresponding author of the article.

REFERENCES

[1] I. A. Taşkıran, "Dijital Yerli Gazetelerin Sosyal Medya Stratejileri ve Sosyal Medyanın Haber Okunurluğuna Etkisi," *Akdeniz Üniversitesi Iletişim Fakültesi Dergisi*, vol. 30, pp. 218–240, Apr. 2018.

[2] C. Beckett, *SuperMedia: Saving Journalism so it Can Save World*. Hoboken, NJ, USA: Wiley, 2011.

[3] *25 İCİN 2023+ Sosyal Medya İstatİstiklerİ, GerçEkler Ve Trendler*. Accessed: 2023. [Online]. Available: https://www.websiterating.com/tr/research/internet-statistics-facts/#chapter-1

[4] (2023). *Top 15 Most Popular News Websites*. Accessed: Jun. 26, 2023. [Online]. Available: http://www.ebizmba.com/articles/

[5] *We Are Social Dijital 2023 Global ve Türkiye Raporu Yayınlandı*. Accessed: 2023. [Online]. Available: https://omgiletisim.com/we-are-social-dijital-2023-global-ve-turkiye-raporu-yayinlandi/

[6] *En Popüler Web Siteleri Sıralaması*. Accessed: 2023. [Online]. Available: https://www.similarweb.com/tr/top-websites/turkey/news-and-media/

[7] Twitter. *İşte Olup Bitenler*. Accessed: 2023. [Online]. Available: https://twitter.com/?lang=tr

[8] H. A. Ünver, Edam, TürkıYe'de doğruluk kontrolü ve doğrulama kuruluşlari," Centre Econ. Foreign Policy Stud., Tech. Rep., 2020.

[9] N. Seddari, A. Derhab, M. Belaoued, W. Halboob, J. Al-Muhtadi, and A. Bouras, "A hybrid linguistic and knowledge-based analysis approach for fake news detection on social media," *IEEE Access*, vol. 10, pp. 62097–62109, 2022.

[10] D. Choudhury and T. Acharjee, "A novel approach to fake news detection in social networks using genetic algorithm applying machine learning classifiers," *Multimedia Tools Appl.*, vol. 82, no. 6, pp. 9029–9045, Mar. 2023.

[11] A. Hadeer, "Detecting opinion spam and fake news using n-gram analysis and semantic similarity," Ph.D. dissertation, 2017.

[12] V. Pérez-Rosas, B. Kleinberg, A. Lefevre, and R. Mihalcea, "Automatic detection of fake news," 2017, *arXiv:1708.07104*.

[13] P. H. A. Faustini and T. F. Covões, "Fake news detection in multiple platforms and languages," *Expert Syst. Appl.*, vol. 158, Nov. 2020, Art. no. 113503.

[14] F. A. Ozbay and B. Alatas, "Fake news detection within online social media using supervised artificial intelligence algorithms," *Phys. A, Stat. Mech. Appl.*, vol. 540, Feb. 2020, Art. no. 123174.

[15] K. Shu, D. Mahudeswaran, S. Wang, D. Lee, and H. Liu, "FakeNewsNet: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media," *Big Data*, vol. 8, no. 3, pp. 171–188, Jun. 2020.

[16] J. C. S. Reis, A. Correia, F. Murai, A. Veloso, and F. Benevenuto, "Supervised learning for fake news detection," *IEEE Intell. Syst.*, vol. 34, no. 2, pp. 76–81, Mar. 2019.

[17] T. Jiang, J. P. Li, A. U. Haq, A. Saboor, and A. Ali, "A novel stacking approach for accurate detection of fake news," *IEEE Access*, vol. 9, pp. 22626–22639, 2021.

[18] M. H. Goldani, S. Momtazi, and R. Safabakhsh, "Detecting fake news with capsule neural networks," *Appl. Soft Comput.*, vol. 101, Mar. 2021, Art. no. 106991.

[19] S. Hakak, M. Alazab, S. Khan, T. R. Gadekallu, P. K. R. Maddikunta, and W. Z. Khan, "An ensemble machine learning approach through effective feature extraction to classify fake news," *Future Gener. Comput. Syst.*, vol. 117, pp. 47–58, Jun. 2021.

[20] I. Ahmad, M. Yousaf, S. Yousaf, and M. O. Ahmad, "Fake news detection using machine learning ensemble methods," *Complexity*, vol. 2020, pp. 1–11, Oct. 2020.

[21] H. S. Al-Ash, M. F. Putri, P. Mursanto, and A. Bustamam, "Ensemble learning approach on Indonesian fake news classification," in *Proc. 3rd Int. Conf. Informat. Comput. Sci. (ICICoS)*, Oct. 2019, pp. 1–6.

[22] A. Mahabub, "A robust technique of fake news detection using ensemble voting classifier and comparison with other classifiers," *Social Netw. Appl. Sci.*, vol. 2, no. 4, p. 525, Apr. 2020.

[23] G. Gravanis, A. Vakali, K. Diamantaras, and P. Karadais, "Behind the cues: A benchmarking study for fake news detection," *Expert Syst. Appl.*, vol. 128, pp. 201–213, Aug. 2019.

[24] J. Y. Khan, M. T. I. Khondaker, S. Afroz, G. Uddin, and A. Iqbal, "A benchmark study of machine learning models for online fake news detection," *Mach. Learn. With Appl.*, vol. 4, Jun. 2021, Art. no. 100032.

[25] A. Saha, K. Tawhid, M. B. Miah, A. A. Somudro, and D. Nandi, "A comparative analysis on fake news detection methods," in *Proc. 2nd Int. Conf. Comput. Advancements*, Mar. 2022, pp. 392–399.

[26] K. K. Kumar, S. H. Rao, G. Srikar, and M. B. Chandra, "A novel approach for detection of fake news using long short term memory (LSTM)," *Int. J.*, vol. 10, no. 5, 2021.

[27] T. Pu, "An active learning framework for trustworthy classification over COVID-19 multilingual fake news dataset," Doctoral dissertation, Fordham Univ., Bronx, NY, USA, 2023.

[28] F. Harrag and M. K. Djahli, "Arabic fake news detection: A fact checking based deep learning approach," *ACM Trans. Asian Low-Resource Lang. Inf. Process.*, vol. 21, no. 4, pp. 1–34, Jul. 2022.

[29] A. J. Keya, M. A. H. Wadud, M. F. Mridha, M. Alatiyyah, and M. A. Hamid, "AugFake-BERT: Handling imbalance through augmentation of fake news using BERT to enhance the performance of fake news classification," *Appl. Sci.*, vol. 12, no. 17, p. 8398, Aug. 2022.

[30] B. Palani, S. Elango, and V. Viswanathan K, "CB-fake: A multimodal deep learning framework for automatic fake news detection using capsule neural network and BERT," *Multimedia Tools Appl.*, vol. 81, no. 4, pp. 5587–5620, Feb. 2022.

[31] B. Samir, "The pope has a new baby!" in *Fake News Detection Using Deep Learning*, 2017.

[32] A. Ågren, and C. Ågren, "Combating fake news with stance detection using recurrent neural networks," M.S. thesis, Univ. Gothenburg, Gothenburg, Sweden, 2018.

[33] G. Rajendran, B. Chitturi, and P. Poornachandran, "Stance-in-depth deep neural approach to stance classification," *Proc. Comput. Sci.*, vol. 132, pp. 1646–1653, Jan. 2018.

[34] S. Girgis, E. Amer, and M. Gadallah, "Deep learning algorithms for detecting fake news in online text," in *Proc. 13th Int. Conf. Comput. Eng. Syst. (ICCES)*, Dec. 2018, pp. 93–97.

[35] R. K. Kaliyar, A. Goswami, and P. Narang, "FakeBERT: Fake news detection in social media with a BERT-based deep learning approach," *Multimedia Tools Appl.*, vol. 80, no. 8, pp. 11765–11788, Mar. 2021.

[36] M. H. Goldani, R. Safabakhsh, and S. Momtazi, "Convolutional neural network with margin loss for fake news detection," *Inf. Process. Manage.*, vol. 58, no. 1, Jan. 2021, Art. no. 102418.

[37] U. Mertoglu, H. Sever, and B. Genc, "Savunmada Yenilikçi bir Dijital Dönüşüm Alanı: Sahte Haber Tespit Modeli," *Savtek*, vol. 9, pp. 771–778, Jan. 2018.

[38] U. Mertoğlu and B. Genç, "Automated fake news detection in the age of digital libraries," *Inf. Technol. Libraries*, vol. 39, no. 4, pp. 1–19, Dec. 2020.

[39] S. G. Taskin, E. U. Kucuksille, and K. Topal, "Detection of Turkish fake news in Twitter with machine learning algorithms," *Arabian J. Sci. Eng.*, vol. 47, no. 2, pp. 2359–2379, Feb. 2022.

[40] G. Güler and S. Gündüz, "Deep learning based fake news detection on social media," *Int. J. Inf. Secur. Sci.*, vol. 12, no. 2, pp. 1–21, Jun. 2023.

[41] M. Bozuyla and A. Özçift, "Developing a fake news identification model with advanced deep languagetransformers for Turkish COVID-19 misinformation data," *Turkish J. Electr. Eng. Comput. Sci.*, vol. 30, no. 3, pp. 908–926, Mar. 2022.

[42] S. G. Taskin, E. U. Kucuksille, and K. Topal, "Twitter üzerinde Türkçe sahte haber tespiti," *Balıkesir Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, vol. 23, no. 1, pp. 151–172, 2021.

[43] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," 2018, *arXiv:1810.04805*.

[44] A. S. Alammary, "BERT models for Arabic text classification: A systematic review," *Appl. Sci.*, vol. 12, no. 11, p. 5720, Jun. 2022.

[45] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, "DistilBERT, a distilled version of BERT: Smaller, faster, cheaper and lighter," 2019, *arXiv:1910.01108*.

[46] A, Wani, I, Joshi, S. Khandve, V. Wagh, and Joshi, "Evaluating deep learning approaches for COVID19 fake news detection," in *Proc. Int. Workshop Combating Line Hostile Posts in Regional Languages During Emergency Situation*, Cham, Switzerland: Springer, 2021, pp. 153–163.

[47] (2023). *BERTurk*. Accessed: Jun. 16, 2023. [Online]. Available: https://huggingface.co/dbmdz/bert-base-turkish-128k-uncased

[48] K. Oflazer, and M. Saraçlar, "Turkish and its challenges for language and speech processing," in *Turkish Natural Language Processing*, 2018, pp. 1–19.

[49] A. A. Akın, and M. D. Akın, "Zemberek, an open source NLP framework for Turkic languages," *Structure*, vol. 10, pp. 1–5, Jun. 2007.

[50] P. Sharma and R. Sahu, "Fake news detection using deep learning based approach," in *Proc. Int. Conf. Circuit Power Comput. Technol. (ICCPCT)*, Aug. 2023, pp. 651–656.

[51] A. R. Merryton and M. G. Augasta, "A novel framework for fake news detection using double layer BI-LSTM," in *Proc. 5th Int. Conf. Smart Syst. Inventive Technol. (ICSSIT)*, Jan. 2023, pp. 1689–1696.

GÜLSÜM KAYABAŞI KORU received the B.Sc. and M.Sc. degrees in computer engineering. She is currently pursuing the Ph.D. degree in computer forensics with Gazi University, Ankara, Turkey. She is also a Computer Engineer with the Ministry of National Defense. Her current research interests include software project management, information systems security, social network analysis, and predictive machine learning.

ÇELEBİ ULUYOL received the B.Sc. and M.Sc. degrees in electronics and computer education and the Ph.D. degree in educational technologies from Gazi University, Ankara, Turkey. He is currently a Professor with the Department of Forensic Informatics, Gazi University. His current research interests include education, computer education and instructional technology, computer sciences, and software.