2022 18th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob) | 978-1-6654-6975-3/22/$31.00 ©2022 IEEE | DOI: 10.1109/WIMOB55322.2022.9941689

Smart Edge-based Fake News Detection using Pre-trained BERT Model

Yuhang Guo
Electrical Engineering And Computer Science
Khalifa University, EECS department
Abu Dhabi, UAE
yuhang.guo@ku.ac.ae

Hanane Lamaazi
Center for Cyber-Physical System
Khalifa University, EECS department
Abu Dhabi, UAE
hanane.lamaazi@ku.ac.ae

Rabeb Mizouni
Center for Cyber-Physical System
Khalifa University, EECS department
Abu Dhabi, UAE
rabeb.mizouni@ku.ac.ae

***Abstract*—Today, online media applications are an important source of information. People are creating and sharing more information than ever before around the world. Being provided by unreliable sources, some news can be misleading. In fact, the assessment of the correctness of the news can be region related. In other words, news can be true in a specific region while fake in another. Existing proposed solutions for fake news detection developed in centralized platforms are not considering the location from where the news gets announced, but they are focused more on the news content. In this paper, a region-based distributed fake news detection framework is proposed. The framework is deployed in a mobile crowdsensing (MCS) environment where a set of workers responsible for collecting news are selected based on their availability in a specific region. The selected workers share the news to the nearest edge node, where the pre-processing and detection of fake news are executed locally. The detection process uses a pre-trained BERT model where it achieved an accuracy of 91%.**

***Index Terms*—Fake News, BERT, Text Classification, Deep Learning, Fine-Tuning, Edge Computing, Distributed Architecture.**

I. Introduction

The fast development of the Internet of Things (IoT) leads to rapid increase of information shared between individuals. News is one of the information that has an important impact on our daily life. Today, the goal of news are to eliminate uncertainty among audiences, reflect and guide public opinion, serve society, spread knowledge and popularize education, and provide entertainment and guide life. News requires the reporting of true facts, through which people can gain insight into the magnitude of the universe, the trend of social development, and the dynamics of life evolution.

Our capacity to make decisions is largely dependent on the type of information we receive; our worldview is formed based on the information we digest. There is growing evidence that consumers react absurdly to news that later turns out to be false. A recent case is the spread of the new coronavirus, where false reports about the origin, nature, and behavior of the new virus spread across the Internet. The situation has gotten worse as more and more people read these falsehoods online. Identifying such fake news online can be an arduous task.

A set of research studies [1], [2], [3] proposed solutions to detect fake news deployed in centralized platforms. In terms of architecture design, the centralized system compute and store the news in one set of hardware system, which does not need to face the problem of a network partition. It can easily achieve high consistency and high reliability through redundancy of storage and high optimization of hardware and software combination. However, in terms of availability, since the centralized architecture is designed as a single point, can face a set of problems that can interrupt a several important services from performing. Distributed architecture design, which naturally has multiple nodes, can easily achieve high availability by means of primary backup, redundancy, hashing, in addition to the parallel computing and storage.

In this paper, a region-based distributed fake news detection framework is proposed. The framework is deployed in mobile crowdsensing (MCS) environment, where a set of workers responsible for collecting news are selected based on their availability in a specific region. The detection of fake news is processed using BERT model [4], which represents Bidirectional Encoder Representations from Transformers. BERT is trained to perform well in terms of text classification, and it is designed to pre-train deep bidirectional representations of unlabeled text by acting on the joint left and right contexts of all layers conditionally. Therefore, the pre-trained BERT model can be fine-tuned by adding only one output layer, and creating state-of-the-art models for a wide range of tasks (*e.g.,* answering questions and linguistic reasoning) without major modifications to the task-specific architecture. We use BERT and fine-tune it to perform our specific tasks. Moreover, we develop two-layer distributed architecture for fake news detection, Cloud Server - Edge Nodes- Workers. The overall framework is shown in Figure 1.

The main contributions of our work are summarized as follows:

- Overcome the limitation of centralized platforms by adopting a distributed architecture using smart edge nodes;
- Offload the cloud server by assigning the selection of workers to the edge servers based on their availability on a specific region;
- Detect fake news and report only the legitimate information by applying a pre-trained BERT model locally by the edge servers.

978-1-6654-6975-3/22/$31.00 ©2022 IEEE

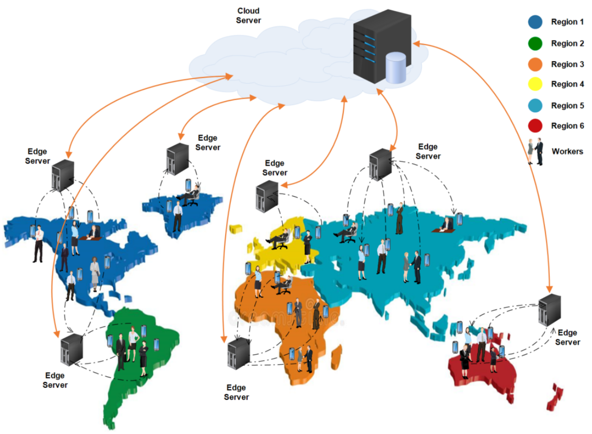

Fig. 1. System Illustration

II. Related work

Fake news is becoming increasingly inundate and proliferating on the Internet. Several researchers are working to address this issue. The core of the fake news detection task lies in how to obtain vector representation of the text and extract vector features. Using pre-trained language models can help us extract the vector representation of the text faster and more accurately.

In [5], the author proposed a model named TraceMiner to reference the embedding of users in the social network and created an LSTM-RNN model to stand for the tracks of messages. Using the TraceMiner approach, they have high classification accuracy and excel at interpreting real-world datasets than traditional methods. Similarly, authors in [6] proposed a comprehensive method with a pre-trained word embedding model, GloVe, for detecting fake news using a combination of CNN and GRU algorithms. The proposed model achieved higher accuracy through the experiments compared to the combination of CNN with LSTM and other existing models. Using a position recognition, the authors in [7] studied a sub-task to detect fake news. Given a news article, the aim was to conclude the relevance of the body and its context. They showed an original idea that combines neural, statistical and external features to give a valid answer to this problem and better detect fake news. Authors in [8] proposed three new models applied to semantic analysis for detecting fake news. Two of them are created, optimized and trained from scratch, and the last one fine-tunes the BERT. The models presented achieved an accuracy of 98% and obtained greater metrics compared to all the other relevant ones.

In [9], the authors adopted three Natural Language Processing models, TF-IDF Vectors with Dense Neural Network (TF-IDF-DNN), Bag of Words Vector with Dense Neural Network (BoW-DNN), and pre-trained word embedding's with Neural Networks (Word2Vec). Among them, TF-IDF-DNN and BoW-DNN outperformed Word2Vec in terms of accuracy. This is due to incapability of the Word2Vec to capture the importance of the semantic level of words when the length of the news is too large. Table I summarizes the above-mentioned approaches and provides more details about the research findings.

All the existing solutions use centralized platforms. However, in this work, we advocate the use of our model on a distributed architecture where a set of Edge servers is responsible for detecting fake news. This detection is fostered by a region-driven selection of workers.

III. Pre-trained BERT Model Overview

BERT [4] is a language processing model built on a neural network that focuses on recognizing word-to-word relationships or sentence-to-sentence relationships, using a semi-supervised learning and language representation model. It is a bi-directional transformer-based model which co-adjusts the left-to-right and right-to-left transformers. In the pre-trained phase, the model uses an unsupervised prediction task consisting of a masked language model (MLM) [10]. Then, the model applies a fine-tuning phase to the model parameters, for the downstream task, to achieve the best fit.

1) Bidirectional transformer: A natural language is a form of communication that has evolved in human society, and a sentence or a word usually needs to be contextualized to reflect its meaning. This means that a computer cannot simply interpret words from above (sequential parsing) or below (inverse parsing), but requires a contextual approach. BERT's deep bidirectional transformer embodies this idea.

2) Masked Language Model: There may be loops when using bidirectional interpretations, leading to misunderstandings about the word's understanding of "itself." BERT uses the Masked Language Model (MLM) [10] to address this misunderstanding. The MLM model randomly masks the words in the input sentence (OpenAI GPT takes a similar approach). For example, the following sentence:
"Claire travelled to Europe with her favorite bag."

- 80% probability of replacing with "[MASK]" token —- "Claire travelled to [MASK] with her favorite [MASK]."
- 10% probability of replacing a single word with a randomly sampled word —- "Claire travelled to orange with her favorite moon."
- 10% probability of no replacement —- "Claire travelled to Europe with her favorite bag."

3) Embeddings: The word embedding of the BERT model is not a simple word encoding but a combination of embeddings that contain three layers of meaning. The first layer of embedding is the encoding of the word itself, which is performed by initializing BERT with an external input word list containing all natural language words. The second layer of embedding is based on the position information of the word for encoding. In order to reflect the position information of the word in the sentence, BERT will perform position embedding for each word in each sentence. The third layer of embedding is sentence-level encoding. To reflect the independence of the sentence (BERT calls it segment embedding), BERT uses two-sentence splicing to construct the encoding. After the three embedding layers have been completed, BERT will combine the three embeddings to determine the word vector.

TABLE I
SUMMARIZATION OF RESEARCH FINDINGS

Ref	Model	Benchmark	Dataset	Distributed Architecture	Accuracy(%)	Conclusion
[5]	TraceMiner	SVM, XGBoost, TM(DeepWalk), TM(LINE)	A real-world Twitter message trace	No	93.80	TraceMiner(infer embeddings; LSTM-RNN); optimization to guarantee the correctness

[6]	CNN+GRU	CNN+LSTM	”BanFakeNews” dataset,
			”Fake News” dataset
				No	98.94	Ensemble approach (CNN+GRU) with GloVe;
						Different Languages;
						Comparison Experiments

[7]	Combine Neural, Statistical and External Features	word2vec+ external features;
		skip-thought;
		TF-IDF
			FNC-1 dataset	No	89.50	Stance detection;
						combines the neural, statistical and external features; classifying the headline-body news pair

[8]	BERT	LSTM, CNN	TI-CNN dataset; Fake News Corpus	No	98.00	Created, optimized, and trained LSTM and CNN from scratch, fine-tuned BERT for textual analysis
[9]	TF-IDF-DNN, BoW-DNN, Word2Vec	DNN, CNN, RNN	FNC-1 dataset	No	94.21	Stance Detection; detecting news article and headline pair; regularization techniques

Figure 2 shows the BERT input representation. The input embeddings are the sum of the token embeddings, the segmentation embeddings, and the position embeddings.

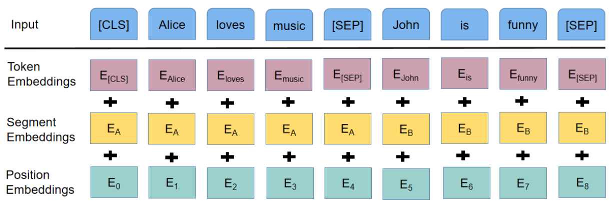

Fig. 2. BERT Input Representation

4) Fine-Tuning BERT: The parameters of a language model are determined at the time of training, but it is unscientific to utilize the same set of parameters for different downstream tasks, and it would be time-consuming to retrain. BERT is a pre-trained language model that adopts fine-tuning to slight adjust the trained model for different downstream NLP tasks to achieve the best model matching results.

IV. Proposed Framework

The BERT model is deployed into a distributed architecture. Centralized platforms have a set of shortcomings that make detecting fake news very hard and cannot stop the spread of such anomalous information. A centralized platform is a single unit where all the news is stored. Processing a huge amount of news coming from different sources, domains and regions is very complex and requires a high computation performance which is expensive. Using a distributed architecture can help on reducing the computation cost by using parallel computation and also can reduce the computational delay and react in real-time with the end-user [11]. Also, the different entities used in distributed architecture can cooperate between each other to break off the spread of fake news between individuals and over different regions. As known, fake news can have a critical impact on individuals' and countries' decisions which can lead to serious consequences. However, detecting fake news over a region can reduce or even stop its propagation over other regions. Our proposed solutions are a smart edge-based fake news detection deployed in MCS environment where a set of workers are selected based on their availability in a specific region specified by the task. This work is based on a pre-built distributed architecture [12] where the edge nodes are responsible for selecting the best group of workers based on their outcomes.

Crowdsensing relies specifically on recruiting workers to collect information from physical environment. Deploying workers to collect news according to their availability in a specific region is an important parameter to be considered to detect the fake news. The proposed framework is region-related, where every requested task has different region requirements. In the proposed framework, a sensing task t_j is defined as $t_j = (L_j^T, R_j^T)$, where L_j^T is the task location, R_j^T is the region from where the news have to be reported.

The region-driven selection is processed in three-stages:

- **The first stage** consists of worker selection where workers are eligible to send the news if they are located in the same region as specified in the task requirement.
- **The second stage** consists of news processing. The edge server receives the collected news provided by the workers and start processing the news locally using the specific algorithm to detect the fake news.
- **The third stage** consists of a could server. The news reported to it is only true news and the fake news has been removed.

The proposed architecture of fake news detection using a pre-trained BERT model, is illustrated in Figure 3. First, the edge server receives the task requirements consisting of the task ID and sensing region **Phase 1**. Then, the edge server

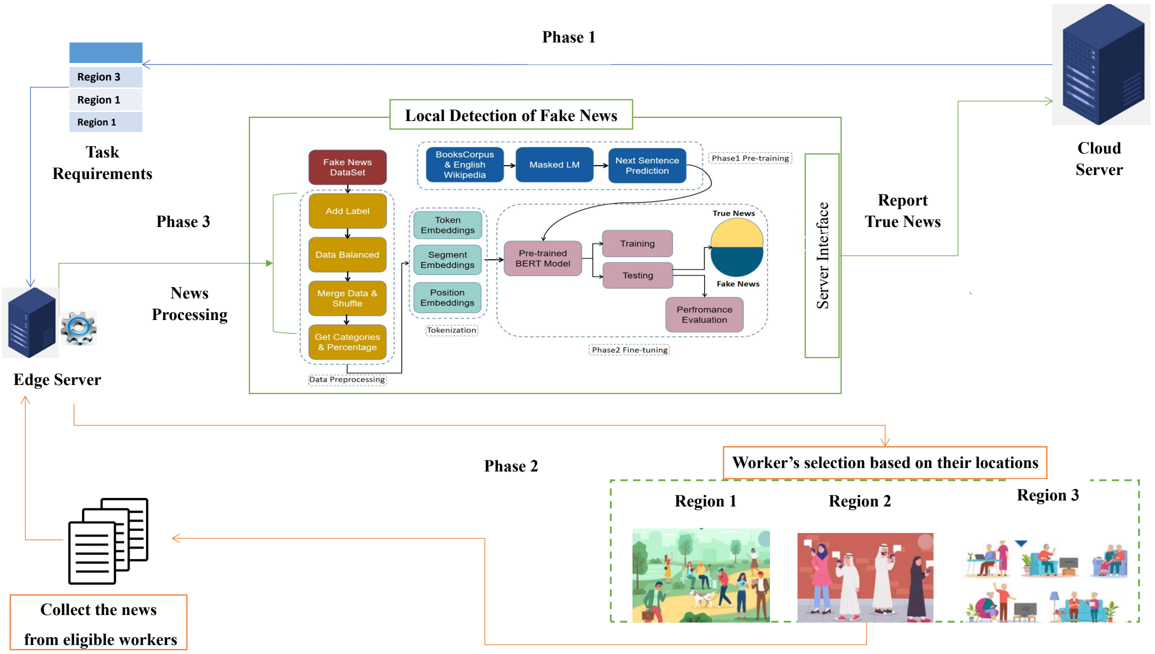

Fig. 3. The Proposed Architecture of Fake News Detection using Pre-trained BERT Model

selects the workers located in the same region as the requested task to collect the news. The workers select a set of news in different domains and report it to the edge server **Phase 2**. Once the news are collected, the edge server start the detection of fake and real news by applying the BERT model. The model is based on four steps needed for an accurate detection, namely: Data pre-processing, Tokenization, Pre-training and Fine-Tuning. Finally the fake news are removed and only the true ones are reported to the cloud server **Phase 3**.

A. Distributed Pre-trained BERT Model

The training of the BERT model requires two phases, the first being pre-training and the other being fine-tuning. The first is unsupervised pre-training on a large corpus in the pre-training phase, followed by supervised fine-tuning of the downstream NLP tasks in the fine-tuning phase.

1) Pre-training Phase: General BERT is pre-trained from scratch using a large corpus (Wikipedia and BooksCorpus) and fine-tuned on the unsupervised corpus to the downstream supervised fake news detection task.

Wikipedia and BooksCorpus were used to train the original BERT (Wikipedia totaling 13GB of text with about 300 million words; BooksCorpus totaling 15GB of text with about 400 million words), both of which are general-purpose domain corpora.

2) Fine-tuning Phase: We imported the pre-trained BERT model from library transformers and adopted the following steps to perform the fine-tuning:

- **Dataset Division.** We split our dataset as the training set, testing set, and validation set according to the ratio 7:2:1.
- **Parameters Selection.** Batch size was set as value 32, and the max length of the token was set as 15. We plotted histograms of word counts and labeled the text because almost all of the text had about 15 words, and for computational reasons, we truncated all of the text to 15 with little corruption.
- **Hyperparameters' Definition.** We selected AdamW as the optimizer, set the learning rate to 1e-5, chose the cross_entropy as the loss function, and computed the weights of the classes.
- **Results Comparison and Evaluation.** We run the code three times by setting epochs 10, 20 and 40, respectively. Then, we used three performance metrics to evaluate the model.

V. Simulation Results

A. Dataset

1) Dataset Description: A specialized dataset in the field of News was downloaded from the FakeNewsNet [13] website, which covers fake news from Asia, Europe, Africa and many other regions.

- **Region:** we selected eight regions for the following fake news detection, namely Asia, Europe, Africa, US, Middle-east, Australia, Western and World-news, respectively.
- **Files:** we reported true and fake news into two different files, 'True.csv' and 'Fake.csv'.
- **Columns:** we organized all the data as four columns, namely 'title', 'text', 'region', and 'date'.

2) Dataset Pre-processing: We performed the dataset pre-processing in the following steps:

- **Add Label.** We added the label 'True' or 'Fake' for True and Fake news files, respectively.

- **Merge and Shuffle Data.** We merged and shuffled the two files for the following detection.
- **Data Balance.** We got the percentage of the True and Fake News. From the Figure 4, we can see that Fake news accounts for 52.3% and True news takes up 47.7%. Therefore, the dataset is well-balanced.
- **Get Correlation.** We got the correlation of text in every region.

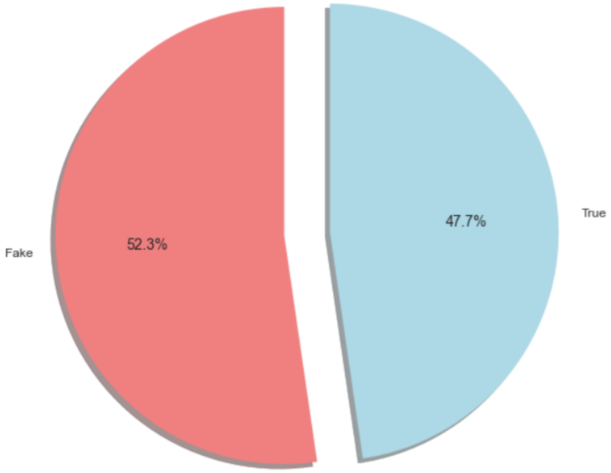

Fig. 4. Well-balanced Data

B. Evaluation Parameters

The main performance metrics are [14] [15]:

- Precision rate:

$$Precision = \frac{TP}{TP + FP} * 100\% \tag{1}$$

where TP denotes true cases, FP denotes pseudo-positive cases, and precision rate indicates the proportion of samples predicted to be true positive cases in the sample of positive cases.

- Recall rate:

$$Recall = \frac{TP}{TP + FN} * 100\% \tag{2}$$

where TP denotes true cases; FN denotes pseudo-counter examples, and recall rate indicates the proportion of samples predicted as positive cases to all positive samples.

- F1-score:

$$F1 - score = \frac{2 * Precision * Recall}{Precision + Recall} \tag{3}$$

C. Results Analysis

Figure 5 shows the increasing accuracy with increasing epochs. At 10 epochs, the model achieved 88% of accuracy and this value is improved over the increase of the number of epochs where it reach the final accuracy of 91%.

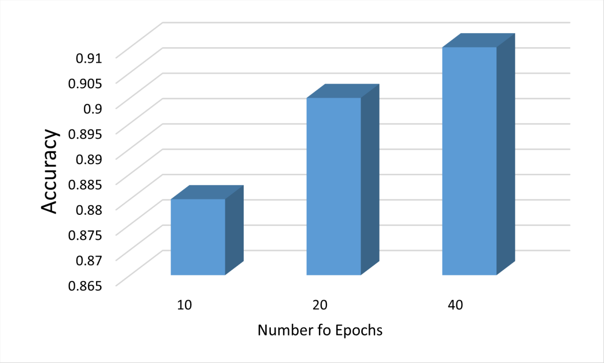

Fig. 5. Model Accuracy

BERT model achieved high precision, recall, f1-score values for each category (fake, true). Figure 6 shows that the model provide a detection precision of 89%, 93% for the false and true classes respectively. However, The fake news achieve 92% for recall while it is 89% for the true news. Also, both fake news and true news have similar f1-score up to 91%.

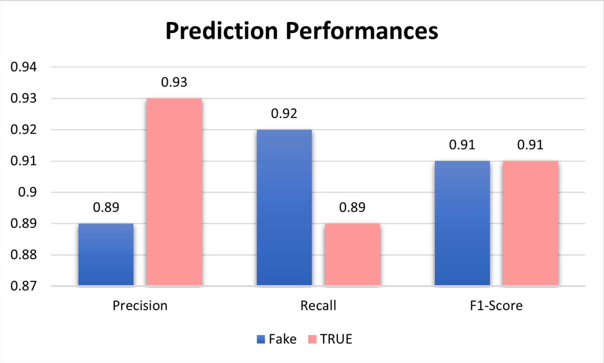

Fig. 6. Prediction Performances of BERT Model (Epochs=40)

From figure 7, Macro-Avg and Weighted-Avg are used to calculate precision, recall, and f1-score for each category and return the average, regardless of the proportion of each category in the dataset. Both the macro-avg and weighted-avg have 91% for precision, recall and f1-score.

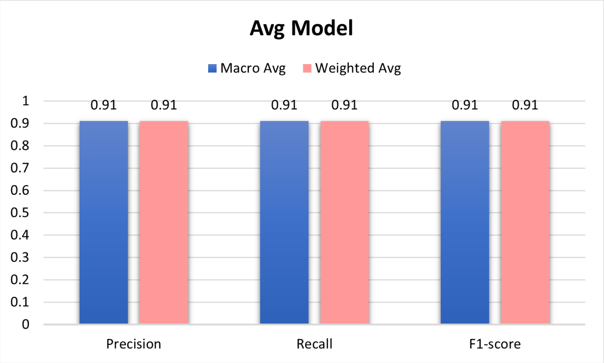

Fig. 7. Prediction Performances of BERT Model (Epochs=40)

Figure 8 presnt the loss function in the training and validation phases for BERT model. It is clear that both loss and validation loss are decreasing to reach lower value of up to 20% with number of epochs equal to 60.

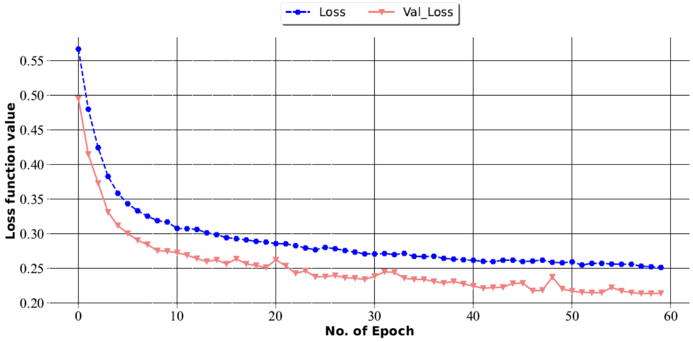

Fig. 8. Training Loss vs Validation Loss.

Figure 9 presents the classification performances where the proposed model is detecting more than 40% of true news while it considers only 5% of true news as fake ones. Similarly, the model accuratly detect more than 45% of fake news while only 3% are wrong detection. To conclude, the model is accuratly detect true news from the fake ones.

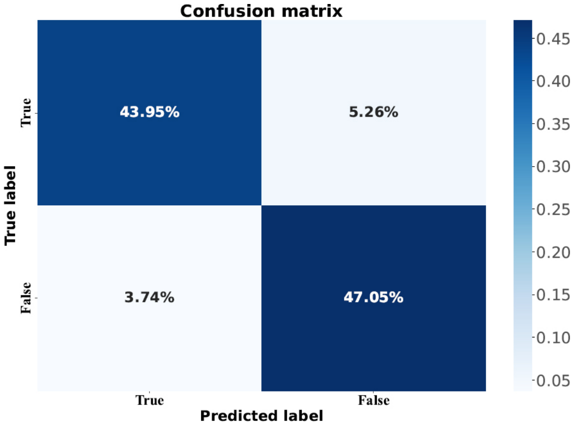

Fig. 9. The Confusion Matrix of BERT Model

VI. Conclusion and Future Work

In this paper, we proposed distributed pre-trained BERT for region-based fake news detection and it performed reasonably well. The framework is deployed in mobile crowdsensing environment where a set of workers responsible for collecting news are selected based on their availability in a specific region. The selected workers share the news to the nearest edge node where the pre-processing and detection of fake news is executed locally. The detection process uses a Pre-trained BERT model where it achieved an accuracy of 91%. The result shows that applying a pre-trained BERT model under distributed architecture is promising to detect region-based fake news.

In the future work, we will expand our work in more domains and regions, and further optimize our distributed architecture in mobile crowdsensing environment.

References

[1] Z. Jin, J. Cao, Y. Zhang, and J. Luo, "News verification by exploiting conflicting social viewpoints in microblogs," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 30, 2016.

[2] Z. Jin, J. Cao, Y. Zhang, J. Zhou, and Q. Tian, "Novel visual and statistical image features for microblogs news verification," *IEEE transactions on multimedia*, vol. 19, no. 3, pp. 598–608, 2016.

[3] Z. Jin, J. Cao, H. Guo, Y. Zhang, and J. Luo, "Multimodal fusion with recurrent neural networks for rumor detection on microblogs," in *Proceedings of the 25th ACM international conference on Multimedia*, pp. 795–816, 2017.

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

[5] L. Wu and H. Liu, "Tracing fake-news footprints: Characterizing social media messages by how they propagate," in *Proceedings of the eleventh ACM international conference on Web Search and Data Mining*, pp. 637–645, 2018.

[6] A. J. Keya, S. Afridi, A. S. Maria, S. S. Pinki, J. Ghosh, and M. F. Mridha, "Fake news detection based on deep learning," in *2021 International Conference on Science Contemporary Technologies (ICSCT)*, pp. 1–6, 2021.

[7] G. Bhatt, A. Sharma, S. Sharma, A. Nagpal, B. Raman, and A. Mittal, "Combining neural, statistical and external features for fake news stance identification," in *Companion Proceedings of the The Web Conference 2018*, WWW '18, (Republic and Canton of Geneva, CHE), p. 1353–1357, International World Wide Web Conferences Steering Committee, 2018.

[8] Á. I. Rodríguez and L. L. Iglesias, "Fake news detection using deep learning," *arXiv preprint arXiv:1910.03496*, 2019.

[9] A. Thota, P. Tilak, S. Ahluwalia, and N. Lohia, "Fake news detection: a deep learning approach," *SMU Data Science Review*, vol. 1, no. 3, p. 10, 2018.

[10] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[11] H. Lamaazi, R. Mizouni, S. Singh, and H. Otrok, "A mobile edge-based crowdsensing framework for heterogeneous iot," *IEEE Access*, vol. 8, pp. 207524–207536, 2020.

[12] H. Lamaazi, R. Mizouni, H. Otrok, S. Singh, and E. Damiani, "Smart-3dm: Data-driven decision making using smart edge computing in hetero-crowdsensing environment," *Future Generation Computer Systems*, vol. 131, pp. 151–165, 2022.

[13] K. Shu, D. Mahudeswaran, S. Wang, D. Lee, and H. Liu, "Fakenewsnet: A data repository with news content, social context and spatialtemporal information for studying fake news on social media," *arXiv preprint arXiv:1809.01286*, 2018.

[14] W. Choukri, H. Lamaazi, and N. Benamar, "Rpl rank attack detection using deep learning," in *2020 International Conference on Innovation and Intelligence for Informatics, Computing and Technologies (3ICT)*, pp. 1–6, IEEE, 2020.

[15] W. Choukri, H. Lamaazi, and N. Benamar, "Abnormal network traffic detection using deep learning models in iot environment," in *2021 3rd IEEE Middle East and North Africa COMMunications Conference (MENACOMM)*, pp. 98–103, IEEE, 2021.