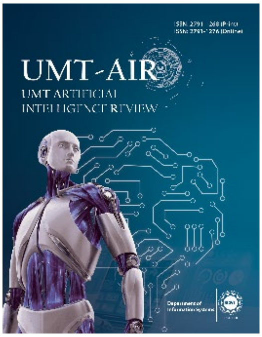


UMT Artificial Intelligence Review (UMT-AIR)
Volume 1 Issue 1, Spring 2021
ISSN(P): 2791-1276 ISSN(E): 2791-1268
Journal DOI: https://doi.org/10.32350/UMT-AIR
Issue DOI: https://doi.org/10.32350/UMT-AIR/0101
Homepage: https://journals.umt.edu.pk/index.php/UMT-AIR

Journal QR Code:

Article: **Rumor Identification on Twitter Data for 2020 US Presidential Elections with BERT Model**

Author(s): Abdul Rahim

Affiliation: Addo.ai

Article QR:

Citation: R. Abdul, "Rumor identification on twitter data for 2020 US presidential elections with BERT model," *UMT Artificial Intelligence Review,* vol. 1, pp. 44–54, 2021. https://doi.org/10.32350/UMT-AIR/0101/03

A publication of the
Dr Hasan Murad School of Management
University of Management and Technology, Lahore, Pakistan

Rumor Identification in Twitter Data for 2020 US Presidential Election using BERT Model

Abdul Rahim[1]*

ABSTRACT: Social media platforms provide rich resources to their users to connect, share and search for the information of their interest. These platforms are even more significant for governmental issues and political campaigns. As information spreads within seconds, it is incredibly challenging to control and monitor the authenticity of the information. Many attempts have been made in this regard. This paper briefly overviews some significant efforts and discusses the patterns of rumors and fake news using the latest machine learning techniques. For this purpose, we extracted the tweets, specifically with the hash tag Donald Trump, during the high time of the 2020 US presidential election in order to test their authenticity. Similar data was extracted from the FactCheck websites Snopes.com, factcheck.org, and politifact.org. We applied the already established BERT model to train the data and tested one million tweets. We found the model as reliably accurate and proposed that once all the truthful information is saved and pretrained in the model, it can auto-identify the validation of the information shared. Also, once established, such models help find the behavior of rumors and patterns in American politics.

KEYWORDS:
BERT Model, rumor detection, social media, US elections

I. Introduction

Social media and microblogging platforms are great examples of the latest technologies becoming part of our daily lives. These communication advancements are used in personal and professional domains, although they are used primarily in journalism and for extending political influences. Nevertheless, such communication platforms allow their users to experience more information flow quickly, easily, and cost-free.

These are undeniable advantages, although they have given rise to unnecessary competitiveness, which, in turn, has resulted in inappropriate use of these exciting new technologies. Indeed, their dark side is becoming more prominent

*Corresponding Author: abdulraheem4622@gmail.com with time due to the spread of false propaganda, fake news, and tampered information, creating a battleground for hate speech [1], [2]. All of this has resulted in irreversible damage in various forms, that is, damage to mental health and repute and social shaming [3], [4]. Such negative and unethical use of social platforms remains in the limelight for quite some time, and this fact challenges the reliability and survival of these modern communication platforms. It also creates prospects for machine learning techniques to provide solutions.

With the rise of online and fact-checking platforms and machine learning techniques, control over the spread of incorrect information is being improved, though the need for a reliable autodetection system is still there [5], [6]. This study focuses on rumors and disinformation propagated during elections, suggesting how these can be tackled using machine learning techniques. Specifically, we were inspired by the studies on rumors regarding the 2016 US presidential election [7], [8], [9]. In this work, we mainly focus on the 2020 presidential election.

The current research aims to review the previously utilized techniques and then evaluate the BERT model to classify the rumors based on a small set of information. Efforts are being made to control incorrect information flow; indeed, remarkable efforts have been made by Twitter in this regard [10]. However, fake news is still a norm in the glamour world when it comes to journalism and politics. Keeping in view the said consideration, this study investigates the circulation of rumors related to politicians during elections as it provides us with selected patterns for rumor propagation.

The rest of the paper is organized as follows. In Section 2, we briefly review the existing studies related to rumors in general by applying machine learning models, particularly in US elections. We proceed with the paper by presenting Section 3, including data preparation, the modeling aspect, and the results derived. Later, in Section 4, we discuss the results and the limitations and prospects of the current work.

II. LITERATURE REVIEW

Many studies have been carried out to identify and classify rumors and fake news by employing machine learning methods; however, the recent trend has shifted towards advanced deep learning and hybrid approaches [10], [11], [12]. New tools and technologies are emerging with advancements in the machine learning domain. We utilized complex techniques by eschewing the details to achieve the goals more meaningfully. This section discusses some prominent and latest studies explicitly aimed at rumor detection using advanced machine learning approaches. Extensive work has been carried out by leveraging BERT and its variants as the foundation model. For instance, in the study of [13], the authors proposed a combination of Convolutional Neural Network (CNN) with BERT. Adding the CNN layer was to enhance the word's semantic representation with varying sentence lengths. In doing so, the said authors achieved results with 98.9% accuracy. Similarly, Harrag et al. [14] carried out another study that employed the BERT model with GPT2 to recognize and classify the information (in this case, tweets) as either human-generated or machine-generated. They specifically targeted the Arabic language tweets and compared their predictions with hybrid models, such as RNN, LSTM, GRU, and their variants. They reported 98% accuracy for their data.

Indeed, much work has been carried out with considerable accuracy by employing the BERT model, though lack of resources, absence of context, and unavailability of standard corpora for fake or propaganda news are the challenges faced in this domain [13]. Da San Martino et al [13] proposed a dataset of annotated news articles with 14 fake approaches to address this issue. This study proposed a BERT-based model adopted by Patil, Singh, and Agarwal [14] for SemEval 2020-Task 11. The proposed approach consists of two aspects: identification of propaganda and classification of the techniques used to disseminate it (among 14 classes), such as exaggeration, minimization, name-calling, bandwagon, and others.

For the 2016 US election, extensive studies were carried out on the role of social media. Since these platforms are exploited at their best by politicians and influencers to promote their respective election campaigns. Notable work was conducted by Jin et al. [7] for the 2016 US presidential election. For instance, in their study, Jin et al. analyzed the false and fabricated information dispensed via eight million tweets posted by campaigners of vital presidential candidates, including Hillary Clinton and Donald Trump. Conclusions were drawn by matching them with confirmed news articles and classifying them with TF-IDF and BM25, Word2Vec and Doc2Vec, and lexicon matching approaches.

Moreover, Boynton, Shafique, and Srinivasan [10] carried out the analysis of suspended accounts related to the 2016 US presidential election by dividing the accounts into those that belonged to suspended and regular communities, respectively (Trump-IRA, Gay rights, BlackLivesMatter, and others) and strived to find the behavioral characteristics. They found no similarities among them. Their work highlighted the heterogeneous behavior exhibited by suspended accounts.

As there exist quite exciting and distinctive perspectives about the use of social media to disseminate fake news, rumors, disinformation, and trolls, they have given rise to a variety of machine learning models. No doubt, all the existing attempts are remarkable. They provide an opportunity for the research community to review the literature on misinformation from a much broader perspective and develop a sophisticated corpus and model better to understand the nature of the problem and its challenges. In the same vein, in this work, we use the BERT model to analyze the data retrieved from social media and related to the 2020 US presidential election.

III. DATA AND METHODS

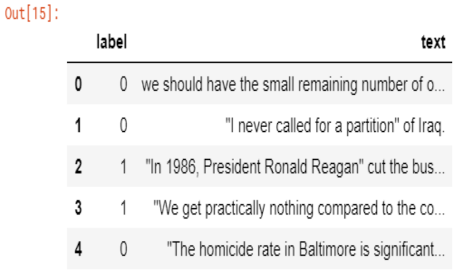
In this section, we discuss the proposed model. Fig. 1 depicts the layout of the methodology used for this experimental study.We applied the Bidirectional Encoder Representations from Transformers

(BERT) technique to classify election rumors intelligently. We divided the methodology into two components.

1. Data Preparation
2. Model Preparation

a. Data Preparation

We collected the data from Twitter Kaggle and factcheck websites. We extracted the data generated from 1st October 2020 - to 25th December 2020. This period contained the peak time for the US 2020 election campaign, and leaders and their followers circulated a flood of information and disinformation to attract voters.



Out[15]:

	label	text
0	0	we should have the small remaining number of o...
1	0	"I never called for a partition" of Iraq.
2	1	"In 1986, President Ronald Reagan" cut the bus...
3	1	"We get practically nothing compared to the co...
4	0	"The homicide rate in Baltimore is significant...

Figure .1. Proposed Methodology

The collected data was intended to be used for testing and predicting the efficacy of the proposed model. Initially, we intentionally eschewed the followers' data and quotes of the same tweet to focus on the uniqueness of false particulars, facts, and figures communicated via these tweets.

We randomly collected around 1394 rumors and non-rumors from FactCheck websites, namely snopes.com, factcheck.org, and politifact.org. We ensured that data was focused on the news circulating during the said time period and specifically targeting the US 2020 election. Also, it is imperative to note that these factcheck websites use their own standards to label the data: partial truth, entirely false, undetermined, and others. However, for this study, we specifically targeted rumors and non-rumors and assumed that partial truth also falls in the rumor category. We updated the annotation of the extracted data to be normalized using the Prodigy annotation tool (as shown in Fig. 2),

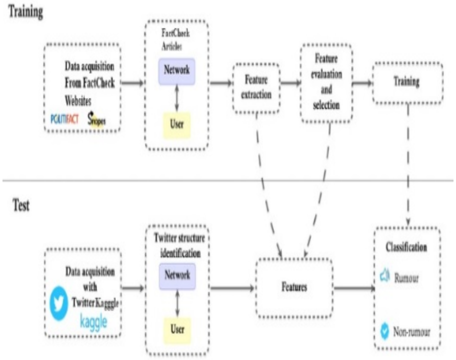

Training
Data acquisition From FactCheck Websites
POLITIFACT
Snopes
FactCheck Articles
Network
User
Feature extraction
Feature evaluation and selection
Training
Test
Data acquisition with TwitterKaggle
kaggle
Twitter structure identification
Network
User
Features
Classification
Rumour
Non-rumour

Figure.2. Sample Annotation of Data using Prodigy

making it easier for us to know if the data requires modification in annotations. Moreover, we manually cross-validated the results based on the information provided by FactCheck and people voter websites.

This practice enabled us to prepare reliable data corresponding to the desired objectives and the identification and reliable classification of rumors.

b. Model Preparation

We used a pre-trained BERT model as a sentence encoder. A significant benefit of using this model is that it accurately extracts the context of the sentence. It also removes directional constraints by applying the Masked Language Model (MLM). This feature has made BERT an outstanding model compared to other embedding models, such as TF-IDF, Word2Vec, Lexicon Matching, and Doc2Vec [11].

After annotating the rumor data, the next step was fine-tuning the model with the training set. This process allowed the implementation of the desired strategy on the pre-trained BERT model. We tokenized the data set of 1394 rumors to make word embedding using a pre-trained BERT model. The maximum length of the sentence in the dataset was 55 (as shown in Fig. 3), and we set it to 128 for training purposes. Consequently, it required significantly less effort to develop the calibrated model. We fine-tuned the top fully connected layer with the word embedding vector. Also, we opted for a broad-match strategy that used a minimum number of keywords to predict as many tweets that can be classified as rumors.

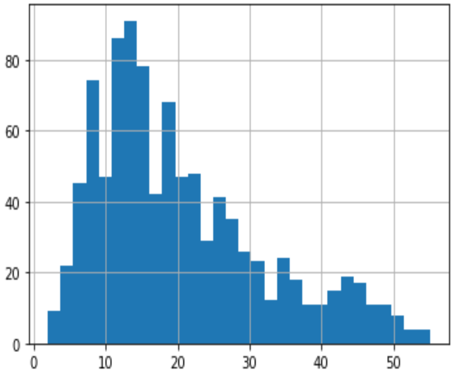


Figure .3. Length of Sentence

Conventionally, two approaches are used to set the hyperparameter values: default optimal and manual. In this study, we opted for the default option as we neither changed the in-built functionality nor created the hybrid of models. For model training, we used Google Colab, an open-source platform for data science training.

IV. RESULTS AND DISCUSSION

After training the model, we presented the model with test data containing the "hashtag_donaldtrump" data of 971073 tweets. The main limitation of the study was inadequate computing resources. The majority of open-source platforms provide a maximum of 16 GBs of RAM for computation. However, nearly 1 million tweets comprised the test data set. To tackle this issue, we made a virtual machine with the following specifications:

1. Operating System: Windows 10 Pro
2. Processor: Intel Xeon X5670 2.9 GHz (8 vCPU cores)
3. RAM: 128 GB
4. Hard Drive: 100 GB

Table I. Models Scores

	PRECISION	RECALL	F1-SCORE	SUPPORT
0	0.88	0.62	0.73	106
1	0.70	0.91	0.79	104
Accuracy		0.77	0.76	210
Macro avg.	0.79	0.77	0.76	210
Weight avg.	0.79	0.77	0.76	210

Table II. Models Scores

	Col_0	**Col_1**
Row_0	66	40
Row_1	9	95

With these specs, we could run the model provided that if we required more RAM for computing data, we used local hardware and saved the file. Moreover, if we needed GPU for computing, we used Kaggle hardware resources. It took around 3 hours to predict the results. We made pandas data frames for these predictions and saved them into CSV.

The model displayed a precision score of 77% accuracy (as given in Table 1). The respective confusion matrix is also shown in Table 2.

The training loss of the model decayed comparatively fast and at 100 epochs with few inconsistencies. We adjust the loss to around 0.4. Fig 4 shows a cross-entropy loss, reduced rapidly by significantly affecting the learning of the data.

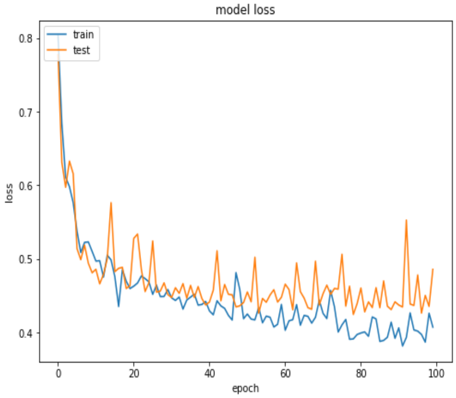


Figure .4. Model Training and Testing

Therefore, the proposed model achieved significantly accurate results with minimal data loss and minor information loss. Further, we compare the train and test data. This comparison contains a smaller set of train data and extensively tested data with the same context. It can be established that bidirectional, pre-trained, word embedding BERT leads to faster training of model and lower cross-entropy loss.

V. CONCLUSION

This work presented a BERT-based model for identifying rumors, specifically those spread by politicians, regarding the US presidential election 2020. The current study intended to utilize the recent advancements in deep learning models by training the proposed model with data from authentic resources and excluding rumors. The tuned model was set to 64 and 100 epochs batch size. The cross-entropy technique was used as a cost function for optimizing the model. The model was able to achieve a precision score of 77%. Later, we used the trained model to detect rumors in almost one million tweets worldwide. Although the initial objective was testing the tweets related to the recent US presidential election, the model can easily be applied to track political rumors on the go, once it is trained with sufficient data from validated sources. In the future, we intend to extend this work by incorporating more diverse tweets in context of political campaigns.

Refrences

1. T. Davidson, D. Warmsley, M. Macy, and I. Weber, "Automated hate speech detection and the problem of offensive language," in *Pro. Int. AAAI Conf. Web Soc. Media*, vol. 11, no. 1, Mar. 2017, pp. 512-515.
2. A. Seyam, A. Bou Nassif, M. Abu Talib, Q. Nasir, and B. Al Blooshi, "Deep Learning Models to Detect Online False Information: a Systematic Literature Review," in *7th Ann. Int. Conf. Arab Women Comput. Conj. 2nd Forum Women Res.*, Aug. 2021, Sharjah, UAE, pp. 1-5.
3. A. Matamoros-Fernández and J. Farkas, "Racism, hate speech, and social media: A systematic review and critique," *Telev. New Media,* vol. 22, no. 2, pp. 205-224, Jan. 2021. https://doi.org/10.1177/1527476420982230
4. T. Enarsson and S. Lindgren, "Free speech or hate speech? A legal analysis of the discourse about Roma on Twitter," *Inf. Commun. Technol. Law,* vol. 28, no. 1, pp. 1-18, Jul. 2019. https://doi.org/10.1080/13600834.2018.1494415
5. R. K. Kaliyar, A. Goswami, and P. Narang, "FakeBERT: Fake news detection in social media with a BERT-based deep learning approach," *Multimed. Tools Appl.,* vol. 80, no. 8, pp. 11765-11788, Jan. 2021. https://doi.org/10.1007/s11042-020-10183-2
6. M. Mozafari, R. Farahbakhsh, and N. Crespi, "Hate speech detection and racial bias mitigation in social media based on BERT model," *PloS one,* vol. 15, no. 8 , pp. e0237861, Aug. 2020. https://doi.org/10. 1371/journal.pone.0237861
7. Z. Jin, J. Cao, H. Guo, Y. Zhang, Y. Wang, and J. Luo, "Detection and analysis of 2016 us presidential election related rumors on twitter," in *Int. Conf. Soc. Comput. Beh.-Cul. Model. Predic. Behav. Represent. Model. Simula.*, June. 2017, pp.

14-24. https://doi.org/10.1007/978-3-319-60240-0_2

8. H. T. Le, G. Boynton, Y. Mejova, Z. Shafiq, and P. Srinivasan, "Revisiting the american voter on twitter," in *Pro. 2017 CHI Conf. Hum. Fac. Comput. Sys.*, May. 2017, pp. 4507-4519. https://doi.org/10.1145/3025453.3025543
9. A. Khatua, A. Khatua, and E. Cambria, "Predicting political sentiments of voters from Twitter in multi-party contexts," *Appl. Soft Comput.,*vol. 97, pp. 106-743, Dec. 2020. https://doi.org/10.1016/j.asoc.2020.106743
10. N. Aslam, I. Ullah Khan, F. S. Alotaibi, L. A. Aldaej, and A. K. Aldubaikil, "Fake detect: A deep learning ensemble model for fake news detection," *complexity,* vol. 2021, pp. 1-8, Apr. 2021. https://doi.org/10.1155/2021/5557784
11. Z. Khanam, B. Alwasel, H. Sirafi, and M. Rashid, "Fake news detection using machine learning approaches," in *IOP Conf. Ser.: Mater. Sci. Eng.*, Jaipur, India, 2021. pp. 12-40.
12. G. Da San Martino, A. Barrón-Cedeno, H. Wachsmuth, R. Petrov, and P. Nakov, "SemEval-2020 task 11: Detection of propaganda techniques in news articles," in *Proc. 14th Workshop Seman. Evalu.*, Dec. 2020, pp. 1377-1414.
13. R. Patil, S. Singh, and S. Agarwal, "Bpgc at semeval-2020 task 11: Propaganda detection in news articles with multi-granularity knowledge sharing and linguistic features-based ensemble learning," , *Arxiv*, 2006 [Online]. https://doi.org/10.48550/arXiv.2006.00593
14. Harrag, A., Rezk, H. "Indirect P&O type-2 fuzzy-based adaptive step MPPT for proton exchange membrane fuel cell." *Neural Comput. Applic.,* vol. 33, no.15, pp. 9649–9662, Feb. 2021. https://doi.org/10.1007/s00521-021-05729-w