



APPLICATION OF DEEP LEARNING ALGORITHMS DETECTING FAKE AND CORRECT TEXTUAL OR VERBAL NEWS

SAMAD DADVANDIPOUR

Institute of Information Engineering, University of Miskolc, Hungary
aitsamad@uni-miskolc.hu

YAHYA LAYTH KHALEEL

MSc student, University of Miskolc, Hungary
Employee at Tikrit University / Iraq
yahya@tu.edu.iq

Abstract. The ongoing spread and expansion of information technology and social media sites has made it easier for people to access different types of news – political, economic, medical, social etc. - through these platforms. This rapid growth in news outlets and the demand for information has blurred the lines between real and fake news, and led to the dissemination of fake news, which is a dangerous state of affairs. The outbreak of the coronavirus pandemic and a rising awareness of the dangers posed all across the globe saw a parallel rise in fake news and rumors, as like as unsubstantiated statements and deceptive ideas. The main aim of this study is supposed to set out to overcome these kind of problems in the future, with application of deep learning algorithms (LSTM, Bi-LSTM, BERT), using a large dataset (39279 rows) to identify fake and correct textual or verbal news. The results of the deep learning application using different algorithms show that the BERT model performed the best, achieving a text classification accuracy of 96.63 %.

Keywords: Fake News, Misinformation, Deep Learning, BERT

1. Introduction

The internet now plays a major role in every aspect of our daily lives, while traditional news outlets, such as television and newspapers, no longer have a monopoly over how we acquire and consume news. This change is further exacerbated by the spread of social media platforms for, as noted by [1], popular social media sites like Facebook and Twitter have only driven this trend. The growth of a large social media community of users and the trustworthy

information offered by leading media outlets have allowed people who generate misinformation to raise their profiles and reputation in the eyes of their followers. They aim to monetise their own content through dramatic headlines and captions, which are intended to be shared. In addition dissemination of misinformation is profitable, thanks to advertising platforms such as Google AdSense, and this encourages people to cash in on sensational statements. Nevertheless, this trend can be extremely harmful and damaging. The majority of the information people access in today's world is not tested and is simply assumed to be true – and therein lies a problem. The main aim of this study is supposed to set out to overcome these kind of problems in the future because it had peaked, as a result of political initiatives and changes which caused lots amount of problems in the last past two years during the pandemic.

1.1. Statement of the problem

Social media transmission of information is speedy, straightforward, and easier to access than traditional news outlets. As a result, it has been growing in popularity in recent times. However, the lack of restrictions on sharing content on social media – and the absence of any meaningful verification processes – has allowed fake news to flourish and grow at a fast rate.

It has never been easy to identify fake news and, as noted by Yang et al. [2], it is unlikely to be an infallible process in the near future, simply because so much user-generated content is being produced in a sophisticated and convincing language, to deceive readers. The spread of false content and misinformation will not merely affect online users, but will also have a significant impact on the general public when it is widely shared.

The rapid spread of fake news relating to COVID-19 in the last past years on social media had made some serious problems for every society.

According to the WHO, the first three months of 2020 saw more than 6,000 people hospitalized globally, as a direct result of fake news about the coronavirus. It has also been calculated that more than 800 people died, as a result of believing false information and fake news about the pandemic during this time [3].

It has been established that deep learning can be successfully used for image classification and object recognition, so the large quantity of natural language data and advances in representing this type of data make deep learning well-suited to texts and speech processing – thereby making it a cutting-edge tool in many NLP assignments [4].

Patwa et al. [5] used machine learning to analyze a dataset made up of 10,700 rows, achieving 93.32 % accuracy. The researchers recommended that subsequent studies should gather more data and use deep learning in place of machine learning –which is what has determined the approach of this study.

1.2. Objectives of the Study

1.2.1. General Objective

The main objective of this research study is to minimize the dissemination of fake news by recommending a model which uses deep learning to identify fake news with a high degree of accuracy.

1.2.2. Specific Objectives

- To evaluate number of deep learning models with the collected dataset.
- To recommend criteria to be used to distinguish between fake and real datasets.
- To recommend the best model for deployment on a web server.

2. Related work

Kar et al. [6] presented a method for the recognition of fake news relating to the COVID-19 pandemic early on from Social Media (SM), such as tweets, in numerous Indian and English languages. They also generated an annotated collection of Hindi as well as Bengali tweets for the purpose. In order to identify fake or real tweets, the suggested model was built on Multilingual BERT (mBERT) using embedding, and enhanced with other relevant properties from Twitter. They found that the algorithm detected fake news with an F-score of approximately 89%, which outperformed results from the English dataset. They also discovered that models trained on dataset tweets from different Indian languages performed better.

Shahi and Nandini. [7] viewed 5182 groups of real articles on the COVID-19 pandemic, presenting the first multi-language cross-domain dataset. After acquiring references from Poynter as well as Snopes, they gathered reality articles from 92 different verified websites. The data were extracted between January 4 2020 and May 15 2020. They also worked on manually categorising articles into 11 separate reality news groups based on their content. The data were provided in 40 languages and relevant to 105 countries. Subsequently, they created a classifier and published the findings of the automated fake news detection and classification, which demonstrated the detection of the fake class with an F1 score of 0.76. Madani et al. [8] suggested a categorisation strategy based on NLP, ML and DL that employed new tweet attributes. The approach was used in conjunction with Apache Spark to identify fake news in tweets on the coronavirus epidemic using the phrase "COVID-19". The findings demonstrated that the RF model was superior to other classification models, such as DT and SVM, in terms of accuracy. By applying the approach to a dataset

of 2000 fresh COVID-19 tweets, they were able to recognise 37% of fabricated news. The authors also demonstrated that the attitude of tweets was significant in tweet categorisation; the number of fake tweets was therefore greater than the frequency of actual tweets.

3. Dataset collection

Identifying large datasets pertaining to fake COVID-19 news is considered to be important in addressing the objectives of this study. However, this is a challenging task since the topic is very new and thus there is limited data available. Thus, we carried out two processes in order to address the objectives. Firstly, we collected data manually using the Google Fact Check tool [\[9\]](#) and secondly, we collected and merged multiple COVID-19 news datasets from around the world. These processes are presented below.

3.1. Manual dataset collection

To verify the validity of the information published online, the Google Fact Check tool can be used. Moreover, the API is designed to check the authenticity of data in order to prevent the publishing of fake news and misinformation. In turn, this should minimize confusion. The Fact Check Explorer can be used to verify results obtained from the internet regarding a particular topic or individual.

When using this tool, information is provided in JSON format and contains rating text. This indicates the fact check in the search result (For instance, "True" or "Mostly True").

In the present work, only data containing a rating text value (True) and (False) was employed. Altogether, 5763 rows of data were collected using the Google Fact Check tool.

3.2. Collecting datasets through searches

After keywords relevant to the topic were searched (including "COVID-19", "Coronavirus" and "Corona Pandemic"), four datasets were revealed [\[10\]](#) [\[11\]](#) [\[12\]](#) [\[13\]](#) These datasets contain data pertaining to real and fake COVID-19 news from across the globe. Once the datasets are identified and collected, the datasets can be filtered into unified columns, after which they can be combined into a single dataset.

4. Research methodologies

4.1. Dataset balancing check

Altogether, there were 39279 rows and two columns (namely ‘News’ and ‘label’) in the dataset. Each piece of news was categorized as real or fake, and this generated a total of 20515 real news items. Contrastingly, there were 18764 fake news items. This is presented in the following figure:

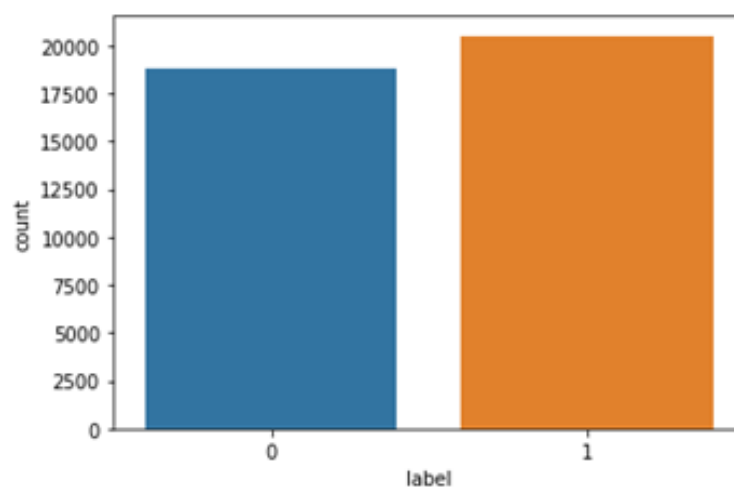


Figure 1. Dataset balancing check

4.2. Natural Language Processing

4.2.1. Tokenization

During the tokenization stage, the textual data is broken down into smaller components called ‘tokens’. As a whole, the dataset is made up of long paragraphs which consist of many lines of words. Analyzing such long paragraphs is quite challenging and thus it is more effective to break the paragraphs down into smaller lines, after which the lines can be further broken down into words.

4.2.2. Normalization

In the dataset, suffixes and prefixes have been added to single words to generate many more words. However, this can make the dataset redundant and will not provide a better or more efficient output. Thus, such words must be converted into their root forms to minimize the number of unique words in the

dataset and to enhance the overall outcomes of the study.

There are two methods that are commonly used in NLP to normalize datasets. These are as follows:

-Stemming

The purpose of the stemming process is to delete any suffixes from words and revert the word to its original root form. However, it may be the case that the root word is non-meaningful or does not exist in the English dictionary.

-Lemmatization

Although the process of lemmatization is similar to stemming, it is far more efficient. During lemmatization, all words produced after removing the suffix are meaningful and existent in the English dictionary. In other words, no incorrect words are formed. Once the lemmatization process has been performed, the resultant word is called a lemma. Lemmatization is much more effective in obtaining the root form of a word than stemming since the former process generates a word that has a genuine meaning. In the present work, lemmatization was performed on the dataset. In the following figure, the differences between stemming and lemmatization are summarized:

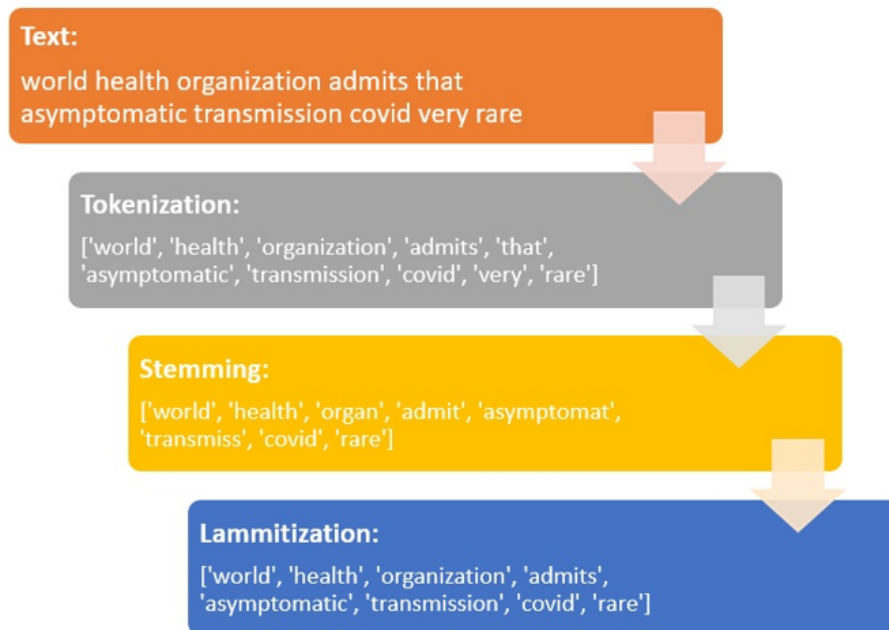


Figure 2. Steps involved in NLP

4.2.3. *Cleaning the dataset*

There are several cleaning steps that must be performed to remove any unnecessary content and to ensure that the dataset is more representable. The stages involved in this process include:

- Remove emojis. Emojis can be defined as small icons that are used to represent symbols, objects or emotions. They are used frequently in communication applications, especially social media. Text processing and understanding can be impeded by the use of such symbols and should thus be removed.
- Delete hashtags: Hashtags can be defined as content labels that help other people interested in a specific topic locate relevant information. A hashtag is usually a simple keyword or phrase (without spaces) that is preceded by a (#) sign.
- Remove punctuation: a list of punctuation items that will be eliminated must be carefully selected based on the usage case.
- Remove stopwords: this ultimately enhances the quality and performance of the classifications system.
- Remove HTML Tags: no value is added to text data through these tags and removing them can facilitate more efficient browsing.
- Remove any non-UTF-8 or ASCII characters.
- Remove URLs.
- Expand words: People typically shorten common phrases in their everyday verbal and written communications. For example, most individuals would contract “you are” to “you’re”. By converting contractions into their natural form, the data will be more accurate.

4.3. **Feature extraction (Vectorization)**

A majority of models and similarity measures use numeric vectors as input. Nonetheless, each document must be converted into a numeric vector before any operations can be carried out on a text. This is a key issue with data mining. However, it is important to make unstructured text documents numerically computable by representing them in numerical format.

In the present work, we thus tested multiple methods on the different models. Word2Vector was employed in LSTM and Bi-LSTM, whereas one hot encoding was employed in BERT.

4.4. **Models**

Following the preprocessing and representing of the dataset in numerical form, the dataset was separated into two sections, the first of which was the ‘train’ section (70% of the dataset) and the second of which was called ‘test’ (30%

of the dataset). Subsequently, the dataset was applied to the deep learning models and then compared all of the results.

4.4.1. LSTM

Long Short-Term Memory (LSTM) networks comprise input and output strata, together with at least a single hidden layer. The neuron population sizes within the input and output layers equate to the number of explanatory covariates, i.e. feature space, and the output space, respectively [14]. The principal trait of LSTM networks is present in the hidden layer(s), which are made up of memory cells. The individual memory cells exhibit a triad of gates which sustain and modify the cell condition, i.e. a forget (ft), input (it) and output (ot) gate, respectively. The stream of memory cell input and output into the remainder of the network are governed by the input and output gates, respectively. The addendum of the forget gate to the memory cell allows the output data with high weights to be passed from a prior to a subsequent neuron. The high activation data dictate the information retained in the memory; if high activation of the input unit were present, memory cell data storage would occur, together with passage of data to the subsequent neuron. The alternative is that input data with high weights remains within the memory cell [15].

The researchers employed the "model.summary()" function in Python to briefly summarize the model. The input data was then applied to train the parameters of the newly-produced LSTM model. Moreover, to determine the cost function and the lowest point, these parameters were used. LSTM consist of three layers in the Neural Network, which are repeated t times (t denotes the number of time steps in the data).

The final layer is dense and fully connected, meaning that the neurons from the previous layer are fed into each neuron in the final layer. Each neuron has a sigmoid function, which means that the output of this unit is consistent and will always fall somewhere between 0 and 1. All of the other dense layers are deeply connected to their preceding layers and thus the layer's neurons are connected to neurons in the preceding layer.

One approach that can be used to avoid overfitting a deep learning neural network with training data is regularization. This approach can significantly improve the model's performance when used with new data. Regularization penalties are applied on a per-layer basis. Moreover, dropout layers must be considered. When the gradient is calculated, there is a chance that each unit and its relative connections will be omitted from the calculation. In turn, this can result in a lack of co-adaption of units in the network. In other words, the unit cannot depend solely on the input of another unit because it is possible that this unit will be omitted in training. When testing is carried out, all units are included. They are also weighted according to the probability that they

will be included during training.

This approach to avoiding overfitting can be very effective. The dropout probability will be high if small numbers of training data are used (and vice versa). The following parameters have been established for this model:

- batch size=512
- Number of epochs=400
- Learning rate= 0.000001
- Activation='sigmoid'
- Loss = 'binary_crossentropy'
- Optimizer='adam'
- Dropout layers rate are set at 20%.

4.4.2. Bi-LSTM

The bidirectional LSTM (Bi-LSTM) model constitutes a paradigm in which, as the name implies, training occurs input to output and volte face [16]. Specifically, a Bi-LSTM paradigm initially introduces input data into the feedback layer of an LSTM model, and then performs the training once more using a second LSTM but with the input data sequence in reverse. The parameters have been established as follows:

- batch size=400
- Number of epochs=400
- Learning rate= 0.000001
- Activation='sigmoid'
- Loss = 'binary_crossentropy'
- Optimizer='adam'
- Dropout layers rate are set at 20%.

4.4.3. BERT

The bidirectional encode representations from transformers (BERT) is a transformers model that is pre-trained using a vast amount of English data. Moreover, the BERT operates in a self-supervised manner, meaning that it is only pre-trained on raw texts that have not been labelled by human beings. Thus, it is able to work with vast quantities of publicly-available data. Additionally, the process of generating inputs and labels from the texts is automatic. To be more specific, the model is pre-trained with two key objectives. These are as follows:

1) Masked language modelling (MLM): this is where the model randomly masks 15% of words in the input sentence and processes the whole masked sentence

in order to predict the masked words. This varies from traditional recurrent neural networks (RNNs) because the latter typically view words one after the other in sequence. It also differs from autoregressive models (such as Generative Pre-trained Transformer (GPT)) since future tokens are internally masked in the latter. Nonetheless, the process that takes place in MLM enables the model to learn the sentence bidirectionally.

2) Next sentence prediction (NSP): this is where two masked sentences are concatenated during pretraining. They may relate to sentences that were adjacent in the source texts, or they may not. The model must then determine whether the two sentences were next to each other. This enables the model to learn an inner representation of the English language. This can subsequently be employed to extract features that will be helpful for downstream tasks. For example, by employing the features created by the BERT model as inputs, you can train a standard classifier on a dataset of labelled sentences.

This model is created using the configuration outlined below:

- 1) 24-layer
- 2) 1024 hidden dimension
- 3) 16 attention heads
- 4) 336M parameters.

There is a cased and uncased version of both approaches. In the uncased version converts, words are all converted into lowercase.

The following parameters were established for this model:

- batch size=512
- Number of epochs=400
- Learning rate= 0.0000001
- Activation='sigmoid'
- Loss = 'binary_crossentropy'
- Optimizer='adam'
- Dropout layers rate are set at 20%.

5. Results

There were 3 results for models. In the table below (Table 1), the accuracy of each model presented.

Table 1. The results for all used models

	Accuracy (%)
LSTM	94.56
Bi-LSTM	95.62
BERT	96.63

6. Discussion

The findings indicate that the BERT model was the most accurate in classifying results, the accuracy was found to be (96.63%) when classifying 6025 of the 6242 total real news items and 5363 of the 5542 fake news items.

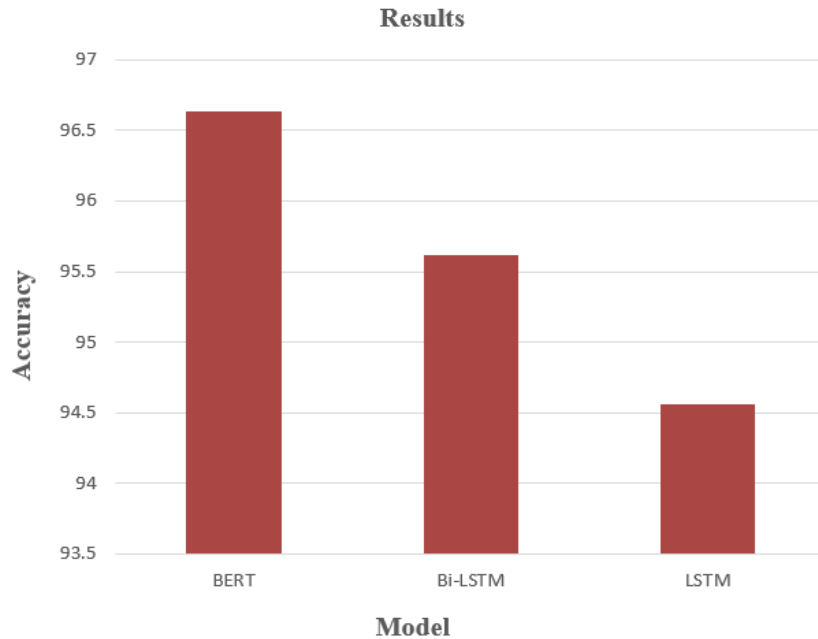


Figure 3. Accuracy of the models.

On the other hand, BERT model outperformed in the precision, it was 96% for fake and real news. Moreover, in terms of recall, the BERT model outperformed the other models. This value was found to be 97% for classifying real and fake news. The BERT was also found to be more efficient than the other models in terms of the f1-score measure. The BERT had an f1 score of 97% and 96% for real and fake news, respectively.

The models were found to perform very well during data training. However, the affinity was significant between training loss and validation loss, as well as between training accuracy and validation accuracy.

A total of 400 epochs is required for all models. Moreover, a learning rate of (lr=0.000001) was employed with the (LSTM and Bi-LSTM) models and (lr=0.0000001) with (BERT). For all the models, the sigmoid activation function was applied. In other words, the input function was converted into a value between 0 and 1. Inputs greater than 1.0 were converted to the value of 1.0.

Likewise, values less than 0.0 are converted to 0.0.

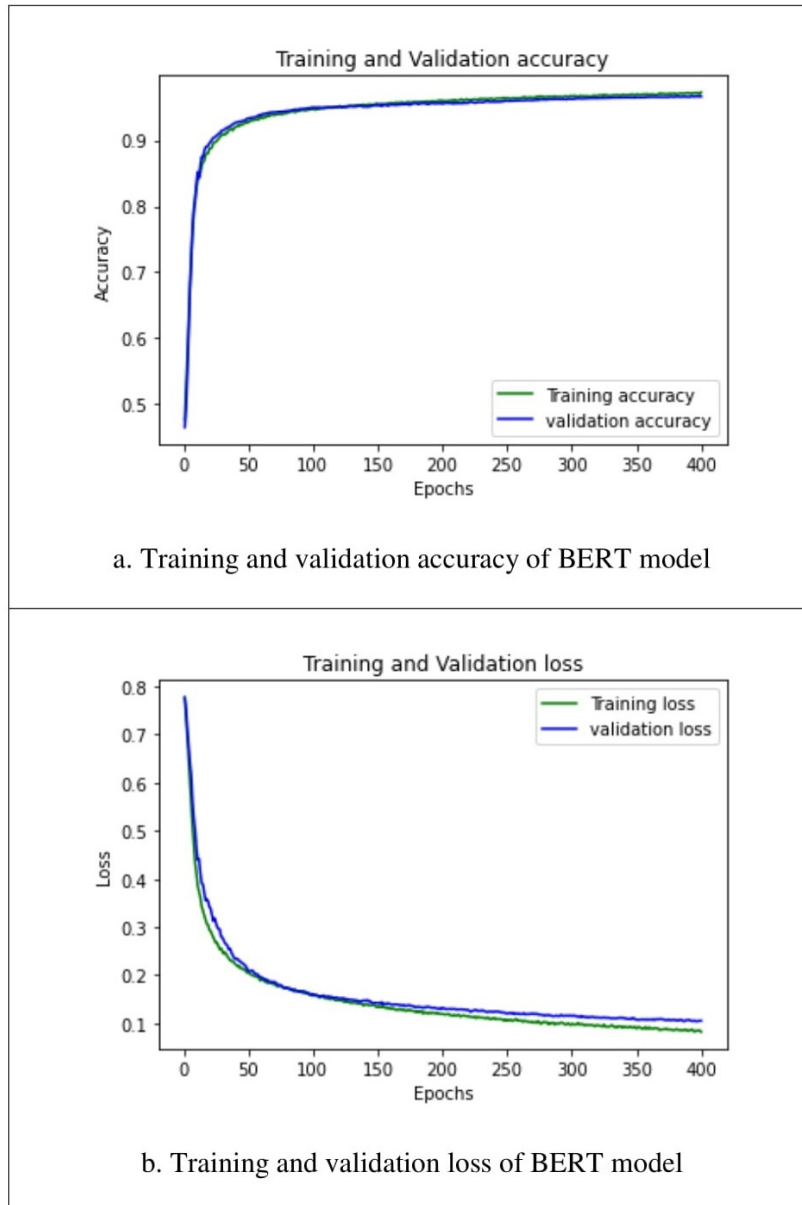


Figure 4. The BERT model performance.

7. Conclusion

The key objective of this work was to use deep learning algorithms to examine fake and correct textual or verbal news pertaining to the COVID-19 pandemic. The main aim of this study is supposed to set out to overcome these kind of problems in the future. At first, the task was challenging due to the lack of available information and data. However, several different sources were used to collect data. Firstly, a manual search for data was performed using the Google Fact Check tool. Other searches were also performed by integrating multiple datasets. The final dataset contained 39279 news items, 20515 of which were real news and 18764 of which were fake news.

Firstly, natural language processing techniques (i.e., data cleaning) were applied to process the data in the dataset. Subsequently, the data were converted into vectors in order to identify features. Deep learning models (LSTM, Bi-LSTM and BERT) were then employed. The key objective was to compare the results and determine which model performed best.

The findings of the present work showed that the BERT model performed the best, achieving a text classification accuracy of 96.63%.

8. Future Work

Combatting fake news is a common problem in the modern world and thus further research into the topic is required. Nonetheless, this work serves as a valuable contribution to studies in this field. In this work, the performance and effectiveness of deep learning methods in detecting COVID-19 fake news were examined. However, this can be further expanded by employing different models and widening the dataset to include data and news items in multiple languages. It may also be beneficial to further develop the study and incorporate it into an online server, after which a browser extension or a mobile application can be created in order to detect fake news.

Future studies should also examine texts embedded in images or videos and published on social media platforms to identify ways to examine whether they are real or fake.

Lastly, transferring data from models trained with one dataset to another is another fascinating topic associated with deep learning that warrants further examination. This process may be able to overcome the barriers impacting unsupervised learning and semi-supervised learning tasks, as well as small datasets.

9. Acknowledgement

The described article was carried out as part of the 2020-1.1.2- PIACI-KFI-2020-00165 "ERPA - Development of Robotic Process Automation solution for heavily overloaded customer services" project implemented with the support provided from the National Research, Development and Innovation Fund of Hungary, financed under the 2020–1.1.2-PIACI KFI funding scheme.

References

- [1] BONDIELLI, A. and MARCELLONI, F.: A survey on fake news and rumour detection techniques. *Information Sciences*, **497**, (2019), 38–55, URL <https://dx.doi.org/10.1016/j.ins.2019.05.035>.
- [2] YANG, Y., ZHENG, L., ZHANG, J., CUI, Q., LI, Z., and YU, P. S.: Tt-cnn: Convolutional neural networks for fake news detection. *arXiv preprint arXiv:1806.00749*, URL <https://dx.doi.org/10.48550/ARXIV.1806.00749>.
- [3] AGARWAL, I., RANA, D., DAIVIK, N. N. S. S., and SURYA TEJA, C.: Fake news on covid-19 healthcare. 2021, URL <https://dx.doi.org/10.21227/k0qs-bz38>.
- [4] HASSAN, A. and MAHMOOD, A.: Efficient deep learning model for text classification based on recurrent and convolutional layers. In *2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)*, 2017, pp. 1108–1113, URL <https://dx.doi.org/10.1109/ICMLA.2017.00009>.
- [5] PATWA, P. E. A.: Fighting an infodemic: Covid-19 fake news dataset. In *CON-STRRAINT 2021*, Springer, 2021, pp. 21–29, URL <https://dx.doi.org/10.48550/arXiv.2011.03327>.
- [6] KAR, D., BHARDWAJ, M., SAMANTA, S., and AZAD, A. P.: No rumours please! a multi-indic-lingual approach for covid fake-tweet detection. In *2021 Grace Hopper Celebration India (GHCI)*, 2021, pp. 1–5, URL <https://dx.doi.org/10.1109/GHCI50508.2021.9514012>.
- [7] SHAHI, G. K. and NANDINI, D.: Fakecovid- a multilingual cross-domain fact check news dataset for covid-19. URL <https://doi.org/10.36190/2020.14>.
- [8] Using artificial intelligence techniques for detecting covid-19 epidemic fake news in moroccan tweets. *Results in Physics*, **25**, (2021), 104266, URL <https://dx.doi.org/10.1016/j.rinp.2021.104266>.
- [9] GOOGLE: Google Fact Check Tools. 2022, URL <https://toolbox.google.com/factcheck/explorer/search/SebastianKurz;hl=en>.
- [10] KERNELER, K.: COVID-19 Fake News Dataset. 2021, URL <https://kaggle.com/kerneler/starter-covid-19-fake-news-dataset-19713c8a-1>.
- [11] AGARWAL, I. and RANA, D.: COVIFN : Fake News on COVID19. 2021, URL <https://dx.doi.org/10.21227/vna0-wk05>.
- [12] BANIK, S.: COVID Fake News Dataset. 2021, URL <https://www.kaggle.com/thesumitbanik/covid-fake-news-dataset>.

-
- [13] GARRETT, K.: COVID-19 Open Research Dataset (CORD-19). 2020, URL <https://data.world/kgarrett/covid-19-open-research-dataset>.
- [14] FISCHER, T. and KRAUSS, C.: Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research*, **270**(2), (2018), 654–669, URL <https://dx.doi.org/10.1016/j.ejor.2017.11.054>.
- [15] SHAHID, F., ZAMEER, A., and MUNEEB, M.: Predictions for covid-19 with deep learning models of lstm, gru and bi-lstm. *Chaos, Solitons & Fractals*, **140**, (2020), 110212, URL <https://dx.doi.org/10.1016/j.chaos.2020.110212>.
- [16] SIAMI-NAMINI, S., TAVAKOLI, N., and NAMIN, A. S.: A comparative analysis of forecasting financial time series using arima, lstm, and bilstm. *arXiv preprint arXiv:1911.09512*, URL <https://doi.org/10.48550/arxiv.1911.09512>.