A Hybrid LSTM-BERT and Glove-based Deep Learning Approach for the Detection of Fake News

Name: Kajal Saini

Department Of Computer Science

Mail id - jainkajal2112@gmail.com

Guide: Dr. Ruchi Jain

College: LNCTS, Bhopal

Mail id - drruchij@lnct.ac.in

2023 3rd International Conference on Smart Data Intelligence (ICSMDI) | 978-1-6654-6487-1/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICSMDI57622.2023.00077

Abstract-

Since the growth of the internet, there has been an increase in the circulation of false information. The very network that keeps us informed about what's going on in the world also provides the ideal environment for the spread of bad content and fake news. Fighting against this fake news is vital since information is what shapes people's perspectives around the world. People don't just establish their own beliefs, but also make significant judgments based on the information that they gather. Should this information turn out to be wrong, the repercussions might be catastrophic. It is entirely impossible for a person to verify each and every piece of news individually. This article has proposed a hybrid deep learning model based on LSTM and BERT with Glove followed by a feature extraction method using TFIDF vectorizer, implement machine learning methods like naïve Bayes, ensemble learning, and XG-boost, and evaluate the performance using accuracy and loss, the BERT model outperform with accuracy 99% and 3% loss.

Keywords: Fake News Classification, Deep Learning, Bidirectional Encoder representation of transformer (BERT), Long short term memory (LSTM), Glove, Encoder.

I. INTRODUCTION

Today's users have access to an online platform where they can produce and share content. Even while the majority of its users behave responsibly when using the internet, a small percentage of users engage in so-called anti-social conduct[1]. Intentional disinformation dissemination is one of the most prevalent examples of this type of conduct in internet settings. Misinformation is spread by the development and dissemination of deceptive content, such as hoaxes or biased or false news. Such antisocial behavior could be problematic because of its potentially harmful repercussions[2].

People have always made an effort to interact with one another. Simple information was initially communicated verbally between people, but as technology advanced, more modern techniques for communicating information evolved[3]. Unfortunately, not all of the information was accurate. Slanderous falsehoods were spread regularly to deceive the enemy or harm the competitor. Specifically, a lot of false information was spread for political or financial gain. Similarly, to that, in the Internet era, people are constantly bombarded with news about global events in the form of media pieces, and people frequently cannot tell whether the information offered is accurate or not. Due to this, a field of study called fake news has evolved to determine the accuracy of the information and spotting false information[4].

Machine learning algorithms are currently finding widespread use in the problem of identifying fake news. It is essential to have correctly annotated datasets in order to train these models to reliably separate fake news from the text. Numerous datasets in the English language are already accessible and are frequently used to test fake news detectors[5]. First, we need to gather the data in order to be able to recognize bogus news stories in the Slovak internet environment. Along with gathering the data, annotations must also be gathered to determine which entries do or do not include misleading information[6].

However, given the availability of social media, this material may be generated and modified in large quantities by regular people, and its dissemination is careless. Social media sites like Facebook and Twitter have made it possible for all kinds of dubious and misleading "news" items to spread without being properly regulated. These phony tales spread widely via and across various platforms due to social media

978-1-6654-6487-1/23/$31.00 ©2023 IEEE
DOI 10.1109/ICSMDI57622.2023.00077

users' tendency to believe what their peers post and what they view regardless of authenticity [4]. New tools being tested by Google and Facebook now aim to make it easier for people to identify and report fake news websites. Facebook is establishing a new system for users and fact-checkers to flag suspect stories, while Google has begun blocking hoax websites from its ad platform and testing fact-checking badges in Google News [7].By regularly updating the dataset and adding the most recent news, this method may also be simply used to social media sites like Facebook and Twitter. In this composition, the naive Bayes classifier has been specifically used to detect fake news[7]. We tested the disparity in accuracy by using various article lengths to identify fake news. Additionally, We learned about web scraping, which illuminated the process by which we might regularly refresh our data set and so confirm the accuracy of the latest Tweeter and Spam post. This composition differs from other papers on related topics[8].

II. LITERATURE REVIEW

Vasist et al. offers a fresh perspective on the effectiveness of various algorithmic strategies and their adaptability to a multi-domain environment rife with bogus news. The study provides an original and thorough evaluation of classification outcomes when exposed to numerous datasets, including both pandemic-related information and information from other fields, as compared to its effectiveness upon pandemic-related data alone. Selecting an appropriate algorithm for theme-diverse fake news detection is essential. Practitioners will benefit most from our comparative perspective, which we provide as part of our practical contribution[8].

Gupta et al. The World Wide Web's development and the quick adoption of online platforms cleared the path for previously unheard-of news dissemination. In comparison to the previous five years, users of social media sites are now creating and disseminating more material, some of which has no real-world relevance. Automatically classifying the text is a difficult and time-consuming task. [9].

Michail et al. s constructed on social media by ignoring the textual content of the news or even the messages that spread it and instead considering the profiles of the users who engage in the graph, the accounts of their social connections, and the method the news spread. As a result, the approach is more robust and has a wider range of scenarios in which it can be used. The study's findings demonstrate that the suggested approach outperforms approaches that rely just on textual data and offer a model that can be used to identify similar disinformation operations on various contexts within the same social medium[10].

Rai et al. The spread of false information via social media has only exacerbated this problem in many parts of the world. Given that the goal of spreading false information is to sway public opinion, it has far-reaching consequences. Researchers have showed a lot of interest in this field because it is difficult and expensive to manually confirm the accuracy of news. Several approaches were investigated for detecting fake news: classification based on sentiment, classification based on images, classification based on content, classification based on interpersonal context, and hybrid context. [11].

Sudhakar et al. To test the data, these two types of methodologies and algorithms are used. More than 1.5 million records were examined by this algorithm. [12].

III. PROPOSED METHODOLOGY

A. Data Collection

It was manually taken from of the Grumble text website and consists of 425 sample of spam text. The National University in Singapore Department for Computer Science Research created the NUS SMS Corpus (NSC), a collection made up of roughly 10,000 real texts [13]. 3,375 randomly selected ham messages from the NSC make up this particular subset. Most of the messages come from students at various Singaporean educational institutions, where the majority of individuals live. The volunteers whose names have been suppressed to respect their privacy made the following contributions. The URL is where the initial Ham and or Spam data was gathered. http://www.dt.fee.unicamp.br/tiago/smsspamcollection/.

	target	message
0	ham	Go until jurong point, crazy.. Available only ...
1	ham	Ok lar... Joking wif u oni...
2	spam	Free entry in 2 a wkly comp to win FA Cup fina...
3	ham	U dun say so early hor... U c already then say...
4	ham	Nah I don't think he goes to usf, he lives aro...

Figure 1. Data Frame of NUS Corpus.

B. Preprocessing

The procedure of cleaning and enhancing in this work includes determining the size of messages, trying to count the total number of desired values, cleanup text using regular expression techniques, and removing punctuation. Punctuation Deleter allows you to rapidly remove punctuation characters like!"#$%&'()*+,-./:;=>?@[] '| from a text. employing pad sequencing, stop - word, lemmatization, tokenization, count vectorization, n-grams, & feature extraction using TFIDF[13].

a. Tokenization

Tokenization is a technique used in natural language processing to split up larger blocks of text like paragraphs or phrases into smaller, more manageable tokens. Separating a sentence into its constituent parts is the first stage in Natural Language Processing (NLP), sometimes known as "gathering the data" (words)[14].

b. Lemmatization

For the goal of text normalization, Natural Language Processing (NLP) makes use of a method called lemmatization. Using this method, you can change the style of any word into the key of its etymological root. Lemmatization is the reduction of a word's several inflected forms to its unchanging root form.

C. TF-IDF

Word frequency is a quantitative indicator of a word's relative importance within a group or corpus of content. This technique is often used as a weighted sum in text mining, user modelling, and information retrieval searches. If a word appears more frequently in the document, the TF-IDF value will be higher; however, this will be offset by the larger number of documents in the corpus that include the term. This allows for a correction to be made again for fact that some words tend to be used more frequently than others in the English language. One of the most popular term-weighting methods now in use is the TF-IDF technique. It was found in a 2015 survey that TF-IDF was used by 83% of text-based recommendation systems in digital libraries.[15].

$$W_{i,j} = tf_{i,j} * \log(\frac{N}{df_i}) \tag{1}$$

D. Exploratory data Analysis

Way of evaluating data sets to highlight their most important properties, frequently making use of statistical graphics as well as other methods of data visualization. In contrast to the testing of traditional hypotheses, EDA can be employed with or without a statistical model, and its primary goal is to learn what the data might tell us in ways that go beyond the formal modelling[16].

E. Machine Learning Modeling

ML branch of study concerned with developing and analyzing techniques that "learn," or make use of existing or new data to enhance their performance on a given set of tasks. It's generally considered to fall within the umbrella of AI. Machine learning algorithms can make inferences and judgments without it being explicitly programmed by using a model they construct Based on a Selected Body of Data. Training data refers to information used to teach a model. Medical, email filtering, speech recognition, agricultural, and computer vision are just some of the many applications of machine learning algorithms. In these areas, classical algorithms would be extremely difficult to design, if not impossible to implement[17].

In this work implemented naïve bayes, ensemble models and XGBoost for machine learning models implementation over the used data.

1. Naïve bayes

Classifiers belonging to the naive Bayes family use Bayes' theorem under the premise of feature independence, hence the name (see Bayes classifier). However, when combined with the kernel density estimation, these models, which are among the smallest Bayesian network models, can reach impressive precision. Naive Bayes classification algorithms are extremely scalable due to the fact that they only require a number of parameters that is linear inside the number of variables (characteristics) in a training issue. As opposed to training many other kinds of classifiers via costly iterative approximation, maximum-likelihood it is possible to train simply evaluating a shuttered expression[5].

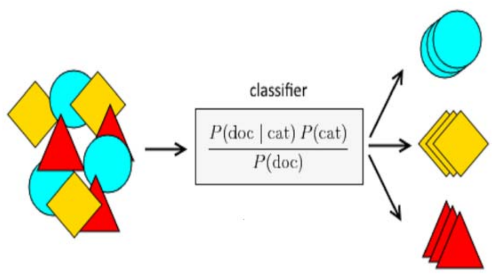

Figure 2. Naïve Bayes[18]

2. XG-Boost

When compared to more unusual data sources like photos and movies, XGBoost excels at working with the tabular data stored in Pandas DataFrames. Many Kaggle tournaments are dominated by XGBoost models. XGBoost models necessitate more expertise and model adjustment to achieve optimal accuracy than methods like Random Forest. When you're done with this guide, you'll understand how to use XGBoost in every step of the modelling process. Tweak XGBoost models until they function admirably While scikit-learn also provides a Gradient Boosted Decision Trees solution, XGBoost offers several technical advantages. The term "Gradient Boosted Decision Trees" is being bandied about, but what exactly does it mean? Here, let's go over a diagram together[19].

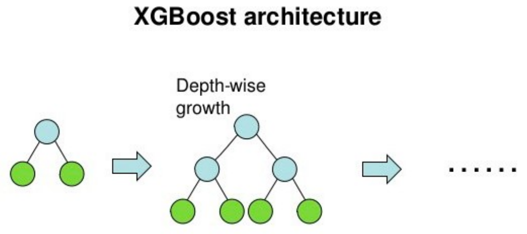

Figure 3. XGBoost Tree[20].

F. Deep Learning

Deep learning is a subfield of machine learning that refers to the overall practice of combining artificial neural networks with representation learning (or deep structured learning). There are three unique methods of education, which are supervised, semi-supervised, and unsupervised learning. The long-short term memory with the glove and the bidirectional encoder representation with the glove were both utilized in this proposed study[9].

1. LSTM

Information can be retained in a long-short-term memory network, which is also known as a complex RNN or sequential network. The RNN's vanishing gradient issue can be resolved by using this method. Recurrent neural networks (RNNs) are utilized for long-term storage. In a manner similar to RNNs, they retain and make use of acquired knowledge while analyzing new information. However, RNNs have the downside of forgetting long-term dependencies when their gradient decreases. Long-term dependency issues are avoided on purpose during the LSTM building process. On a high level, LSTM performs quite similarly to the an RNN cell. The internal workings of the LSTM network are illustrated below. The LSTM is made up of three portions, each of which serves a different purpose, as can be seen in the figure below[21].

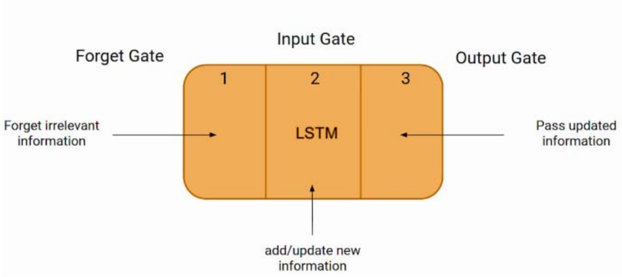

Figure 4. LSTM State Architecture[22].

2. BERT

That's why there's BERT, or Bidirectional Encoder Representations with Transformers. Pre-training unlabeled text bidirectional LSTM representations by concurrently conditioning on left and right context is the goal. Pre-trained BERT models can thus be refined with a single additional output layer to yield state-of-the-art models for a wide range of NLP applications. That sounds a bit too complicated to begin with. But let's break it down because it does a good job of summarizing what BERT accomplishes[23]. Bidirectional Encoder Representations of Transformers is an acronym that is simple to understand. Each word in this sentence has a specific meaning, which we will explore in this article one at a time. For the time being, the most important thing to remember is that BERT is built on the Transformer design.

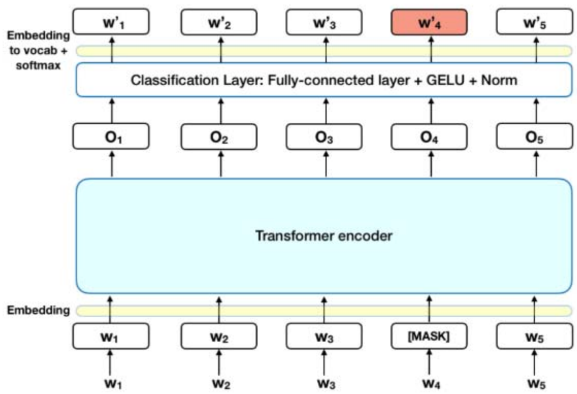

Figure 5. BERT Internal Architecture[24].

IV. Results and Discussion

Firstly start form data collection about ham and spam contain 425 samples and contain 10000 text messages for preprocessing an data cleaning using regular expression, remove punctuation, apply stop words for text, lemmatization, tokenization, perform vectorization with n-grams with 1-grams and 2-grams and apply pad sequencing and finally TIIDF and perform exploratory data analysis over processed data, after processing apply Glove with 100 embedding dimension and convert data into NumPy array, to implementation machine learning model used naïve bayes, ensemble learning and XGBoost and for implementation of deep learning model used LSTM and BERT with Glove model with following hyper parameters, deep learning model based on TensorFlow library, table 1 shows the hyper parameters used like layers LSTM, BERT, bidirectional, global average pooling , batch normalization, to activate the layers uses sigmoid and ReLU, optimizer used as RMSprop, loss were calculated by binary cross entropy the learning rate is 0.001 and epoch for LSTM is 7 and for BERT is 3, there were attention layers also be present. The work was implemented using python language with google colab with i7 and 8 GB of ram[19].

TABLE 1. HYPER PARAMETERS USED FOR LSTM AND BERT

Model	Sequential
Neural Network	LSTM, BERT
Directions	Bi
Pooling	Globalmaxpooling
Normalization	Batch Normalization
Layer	Dropout, Dense
Activation	Sigmoid, ReLU
Optimizer	RMSprop
Loss	Binary cross entropy
Metrices	Accuracy
Word Representation	Glove
Learning rate	0.001
Split	80:20
Epochs	7,3

A. Exploratory Data Analysis

Figure 6 illustrates the frequency of ham and spam, with ham targets containing and over 400 samples with spam text containing 800 or so samples. Both the blue and grey colour display values of 4285 for the blue and 747 for the grey. The word clouds for Ham & Spam texts are shown in Figures 7 and 8.

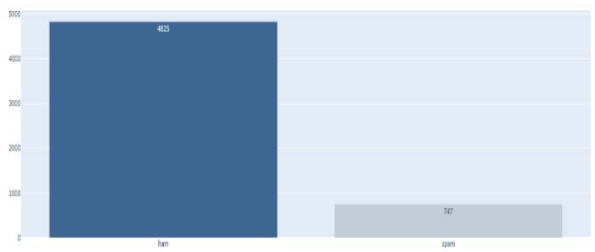

Figure 6. Dataset Distribution by Target.

Figure 7. Word Cloud of Ham Text[25].

Figure 8. Word Cloud of Spam Messages[25].

B. Results Analysis

The table 2 shows the performance over machine learning model with accuracy and roc score in which XG-Boost out perform with 98.34 % accuracy and naïve bayes perform average accuracy like 97 and pipeline process perform worst in comparison to all models.

TABLE 2. PERFORMANCE EVALUATION OF MACHINE LEARNING MODELS

Model	Accuracy	ROC
Naïve Bayes	97.48	97.42
Pipeline	95.97	---
XGBOOST	98.34	---

TABLE 3. PERFORMANCE EVALUATION OF DEEP LEARNING MODELS

Model	Accuracy%	Loss%	Val Accuracy%	Val Loss%
Glove Bi LSTM	98.18	6.79	98.06	12.82
Glove BERT	99.21	3.16	98.57	4.99

In the table 3 shows the performance of deep learning models like LSTM and BERT in which BERT has accuracy of 99% and loss of 3% and LSTM has 98%

accuracy with 6% loss and has validation 98% and loss is 12% for LSTM model and 98% validation accuracy for BERT model with 4% validation loss.

V. Conclusion

Since machine learning is opened a brand-new front in the struggle against fake news, it is imperative that this new front be fully exploited and that its advantages be taken advantage of. This paper has demonstrated that the front has a chance of success. The application of machine learning to the detection of fake news is still in the early stages of development. Every new model that is developed and every new method that is suggested brings us one step closer to an internet free of bogus news.

While using machine learning model with naïve bayes, ensemble learning, and XGBoost in which these models perform average so try to implement hybrid deep learning model based on LSTM and BERT with Glove models and extract important features using TF-IDF vectorizer in which BERT outperform has accuracy like 99% and loss is 3% and LSTM has 98% with 6% loss and has validation 98% and loss is 12% for LSTM model and 98% validation accuracy for BERT model with 4% validation loss.

VI. REFERENCES

[1] M. Sarnovsky, V. Maslej-Kresnakova, and N. Hrabovska, "Annotated dataset for the fake news classification in Slovak language," *ICETA 2020 - 18th IEEE Int. Conf. Emerg. eLearning Technol. Appl. Proc.*, pp. 574–579, 2020, doi: 10.1109/ICETA51985.2020.9379254.

[2] S. M. and K. P. Kaliyamurthie, "a Novel Learning Vector Quantization With Hamming Distance Measure for Effective Prediction of Fake News," *ECS Trans.*, vol. 107, no. 1, pp. 19423–19433, 2022, doi: 10.1149/10701.19423ecst.

[3] P. Ksieniewicz, P. Zyblewski, M. Choraś, R. Kozik, A. Giełczyk, and M. Woźniak, "Fake News Detection from Data Streams," *Proc. Int. Jt. Conf. Neural Networks*, 2020, doi: 10.1109/IJCNN48605.2020.9207498.

[4] J. Jing, H. Wu, J. Sun, X. Fang, and H. Zhang, "Multimodal fake news detection via progressive fusion networks," *Inf. Process. Manag.*, vol. 60, no. 1, p. 103120, 2023, doi: 10.1016/j.ipm.2022.103120.

[5] R. R. Mandical, N. Mamatha, N. Shivakumar, R. Monica, and A. N. Krishna, "Identification of Fake News Using Machine Learning," *Proc. CONECCT 2020 - 6th IEEE Int. Conf. Electron. Comput. Commun. Technol.*, 2020, doi: 10.1109/CONECCT50063.2020.9198610.

[6] A. Jain, A. Kasbe, Maulana Azad National Institute of Technology, IEEE Asia-Pacific Region. MP Subsection, IEEE Asia-Pacific Region. Bombay Section, and Institute of Electrical and Electronics Engineers, "2018 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS) : February 24-25, 2018, Maulana Azad National Institute of Technology, Bhopal," *2018 IEEE Int. Students' Conf. Electr. Electron. Comput. Sci.*, pp. 1–5, 2018.

[7] P. Faustini and T. Ferreira Covoes, "Fake news detection using one-class classification," *Proc. - 2019 Brazilian Conf. Intell. Syst. BRACIS 2019*, pp. 592–597, 2019, doi: 10.1109/BRACIS.2019.00109.

[8] P. N. Vasist and M. P. Sebastian, "Tackling the infodemic during a pandemic: A comparative study on algorithms to deal with thematically heterogeneous fake news," *Int. J. Inf. Manag. Data Insights*, vol. 2, no. 2, p. 100133, 2022, doi: 10.1016/j.jjimei.2022.100133.

[9] V. Gupta, R. S. Mathur, T. Bansal, and A. Goyal, "Fake News Detection using Machine Learning," *2022 Int. Conf. Mach. Learn. Big Data, Cloud Parallel Comput. COM-IT-CON 2022*, pp. 84–89, 2022, doi: 10.1109/COM-IT-CON54601.2022.9850560.

[10] D. Michail, N. Kanakaris, and I. Varlamis, "Detection of fake news campaigns using graph convolutional networks," *Int. J. Inf. Manag. Data Insights*, vol. 2, no. 2, p. 100104, 2022, doi: 10.1016/j.jjimei.2022.100104.

[11] N. Rai, D. Kumar, N. Kaushik, C. Raj, and A. Ali, "Fake News Classification using transformer based enhanced LSTM and BERT," *Int. J. Cogn. Comput. Eng.*, vol. 3, no. March, pp. 98–105, 2022, doi: 10.1016/j.ijcce.2022.03.003.

[12] M. Sudhakar and K. P. Kaliyamurthie, "Effective prediction of fake news using two machine learning algorithms," *Meas. Sensors*, vol. 24, no. August, p. 100495, 2022, doi: 10.1016/j.measen.2022.100495.

[13] V. Balakrishnan, W. Z. Ng, M. C. Soo, G. J. Han, and C. J. Lee, "Infodemic and fake news – A comprehensive overview of its global magnitude during the COVID-19 pandemic in 2021: A scoping review," *Int. J. Disaster Risk Reduct.*, vol. 78, no. June, p. 103144, 2022, doi: 10.1016/j.ijdrr.2022.103144.

[14] S. I. Manzoor, J. Singla, and Nikita, "Fake news detection using machine learning approaches: A systematic review," *Proc. Int. Conf. Trends Electron. Informatics, ICOEI 2019*, no. Icoei, pp. 230–234, 2019, doi: 10.1109/ICOEI.2019.8862770.

[15] A. Khalil, M. Jarrah, M. Aldwairi, and M. Jaradat, "AFND: Arabic fake news dataset for the detection and classification of articles credibility," *Data Br.*, vol. 42, p. 108141, 2022, doi: 10.1016/j.dib.2022.108141.

[16] J. A. Nasir, O. S. Khan, and I. Varlamis, "Fake news detection: A hybrid CNN-RNN based deep learning approach," *Int. J. Inf. Manag. Data Insights*, vol. 1, no. 1, 2021, doi: 10.1016/j.jjimei.2020.100007.

[17] M. Villagracia Octaviano, "Fake News Detection Using Machine Learning," *ACM Int. Conf. Proceeding Ser.*, pp. 177–180, 2021, doi: 10.1145/3485768.3485774.

[18] "Mathematic behind Naive Bayes algorithm and its application | by Rana singh | Medium." https://ranasinghiitkgp.medium.com/mathematic-behind-naive-bayes-and-its-application-9ec8cc4f0a91 (accessed Feb. 20, 2023).

[19] T. Chen and C. Guestrin, "XGBoost: A scalable tree

boosting system," *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, vol. 13-17-Augu, pp. 785–794, 2016, doi: 10.1145/2939672.2939785.

[20] "Introduction to XGBoost Algorithm | by Nadeem | Medium." https://medium.com/analytics-vidhya/introduction-to-xgboost-algorithm-d2e7fad76b04 (accessed Feb. 20, 2023).

[21] J. Li and M. Lei, "ScienceDirect A Brief Brief Survey Survey for for Fake Fake News News Detection Detection via via Deep Deep Learning Learning Models Models," *Procedia Comput. Sci.*, vol. 214, pp. 1339–1344, 2022, doi: 10.1016/j.procs.2022.11.314.

[22] "LSTM | Introduction to LSTM | Long Short Term Memor." https://www.analyticsvidhya.com/blog/2021/03/introduction-to-long-short-term-memory-lstm/ (accessed Feb. 20, 2023).

[23] Z. Fu, "Computer Network Intrusion Anomaly Detection with Recurrent Neural Network," *Mob. Inf. Syst.*, vol. 2022, 2022, doi: 10.1155/2022/6576023.

[24] "Explainability Of BERT Through Attention | by Mrinal Anand | Analytics Vidhya | Medium." https://medium.com/analytics-vidhya/explainability-of-bert-through-attention-7dbbab8a7062 (accessed Feb. 20, 2023).

[25] "Spam word cloud and ham word cloud | Download Scientific Diagram." https://www.researchgate.net/figure/Spam-word-cloud-and-ham-word-cloud_fig6_364167472 (accessed Feb. 20, 2023).