FAKE NEWS DETECTION USING MACHINE LEARNING ALGORITHM

BY

MD Alamin Hosen
ID: 181-15-11132

Akhi Mony
ID: 181-15-11130

MD Touhid Hasan
ID: 181-15-10500

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Computer Science and Engineering

Supervised By

Israt Jahan
Lecturer
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

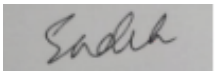
JANUARY 2022

APPROVAL

This Project titled "**Fake news detection using machine learning algorithm**", submitted by MD Alamin Hosen – ID No: 181-15-11132, Akhi Mony – ID No: 181-15-11130 and MD Touhid Hasan – ID No: 181-15-10500 to the Department of Computer Science and Engineering, Daffodil International University, has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering (BSc) and approved as to its style and contents. The presentation has been held on January 4, 2022.

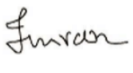
BOARD OF EXAMINERS

Dr. Touhid Bhuiyan (DTB) **Chairman**
Professor and Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University



Md. Sadekur Rahman (SR) **Internal Examiner**
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Afsara Tasneem Misha (ATM) **Internal Examiner**
Lecturer
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University



Shah Md. Imran **External Examiner**
Industry Promotion Expert
LICT Project, ICT Division, Bangladesh

DECLARATION

We hereby declare that, this project has been done by us under the supervision of **Israt Jahan, Lecturer, Department of CSE** Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

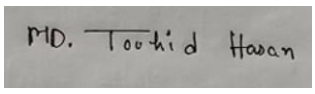
Supervised by:

Israt Jahan
Lecturer
Department of CSE
Daffodil International University

Submitted by:

MD Alamin Hosen
ID: 181-15-11132
Department of CSE
Daffodil International University

Akhi mony
ID: 181-15-11130
Department of CSE
Daffodil International University



MD Touhid Hasan
ID: 181-15-10500
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First we express our heartiest thanks and gratefulness to almighty God for His divine blessing makes us possible to complete the final year project/internship successfully.

We really grateful and wish our profound our indebtedness to **Israt Jahan**, **Lecturer**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of "*Data Mining & Machine Learning*" to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts and correcting them at all stage have made it possible to complete this project.

We would like to express our heartiest gratitude to **Professor Dr. Touhid Bhuiyan, Head,** Department of CSE, for his kind help to finish our project and also to other faculty member and the staff of CSE department of Daffodil International University.

We would like to thank our entire course mate in Daffodil International University, who took part in this discuss while completing the course work.

Finally, we must acknowledge with due respect the constant support and patients of our parents.

ABSTRACT

Recent years have seen an explosion in social media usage, allowing people to connect with others. Since the appearance of platforms such as Facebook and Twitter, such platforms influence how we speak, think, and behave. This problem negatively undermines confidence in content because of the existence of fake news. For instance, false news was a determining factor in influencing the outcome of the 2016 presidential election. Because this information is so harmful, it is essential to make sure we have the necessary tools to detect and resist it. It's difficult to determine what news is false and what is true. We've hardly put in any effort for such a high-quality outcome. This work is for analyzing & delectating the fake news from a fresh collected dataset. We applied Bidirectional Long Short-Term Memory (BiLSTM) to determine if the news is false or real in order to showcase this study. This machine learning technique and approach are being used since there is a lot of study into how people can improve the efficiency and accuracy of their work. A number of foreign websites and newspapers were used for data collection. After creates & running the model, the work achieved 84% model accuracy and 62.0 F1-macro score with training data.

LIST OF FIGURES

LIST OF TABLES

TABLE OF CONTENTS

CHAPTER 1
INTRODUCTION

1.1 Introduction

Social media now gathers and shares news as well as providing a constant supply of fresh information on the internet every day. The function of news is to inform people about current events across the world. Social media as a source of collecting news is growing increasingly popular over conventional media, such as television, radio, and newspapers, due to a number of factors. There are a variety of reasons for this. One is because it is simple to access, which results in reduced costs and less time required to get news. The more outrageous the story, the more likely it is to become viral on social media, which increases the likelihood of its being false news. Social media platforms like Facebook, Twitter, and Instagram open up new facilities with some obstacles like bogus news being shared, and plenty of social bots [1].

Fake news is deception or rumor provided to the people as if it were real information with the goal of influencing people's thinking and molding public opinion into political action, and fake news publications have often found ways to earn money online through their misinformation. Fraudulent news can dupe the public by seeming similar to legitimate websites or utilizing identical URLs to reputable news outlets. These websites may be found across the world, such as in the United States, China, Russia, and other nations. To a large extent, humans are solely responsible for fake news and political inconsistency, both of which serve selfish interests. There are already around 23 million social bots on Twitter, with 140 million on Facebook, and 27 million on Instagram. The spread of political division as a result of fake news is inarguable. The current US presidential election has brought attention to the issue of false news. It has become the home of posting bogus news on social media [2].

Approximately 6 in 10 adults in the United States use the internet, and about one in six uses Twitter. [3] claims that in the 2016 US presidential election, false news attracted people's attention and had an impact on people's opinions. It therefore became a global menace. In order to stop the spread of false news, it was important to design ways to detect

it. The work introduced and also applied a machine learning approaches which is Bidirectional Long Short-Term Memory (BiLSTM). And we got the best accuracy on in this model.

1.2 Motivation of research

Classifying news is a key part of detecting disingenuousness. After that, analyze the characteristics of the two news stories' categorized Database. Ahead of that, we're driven mostly by the need to create news headlines and content that news organizations' databases crave. Using all this information, we discovered that news in a variety of areas is in great demand in the world, including sports, politics and business news. All of this is happening mostly in Bangladesh, but there is considerable interest from other countries as well. When we compile all of this information, we're left with nothing more than names and dates. Because it momentarily confirms what news individuals are most likely to find, it has been really beneficial to us. If human harassment becomes a pattern as a result of spreading false information, our social media landscape will unquestionably change for the worse. As a result, we'll have to start recognizing and classifying news based on its qualities from now on. As part of our study, we identified false news and true news to disseminate to the public, with the latter being 84% accuracy.

1.3 Rationale of the study

This study will pave the road for personal and societal progress. Despite the fact that it is a difficult task, we can claim that the identification of false news plays a significant role. This information is gathered from a number of different publications, which we then analyze. We can also learn about news classification techniques from there. Other people's blogs and websites provide us with ideas on how to identify false news reports.

1.4 Research Questions

- How can we collect data from the newspapers, web portals or other authorized sites?
- How can we find a real news from the huge amount of mixed (real+fake) news?
- What are the approaches of finding accurate news?

- What is the approach of preprocessing data?
- What is the accuracy of proposed model?

1.5 Expected Output

For individuals, detecting fake news in social media or newspapers posts is extremely difficult, and inferring misleading information is crucial if they want to stop the rumor. Using our technology, we want to reduce the amount of misinformation that spreads on social media by the author without any hindrances. We're all aware that it's impossible to halt the spread of rumors, but we can restrict the amount of information that gets out there. Our analytics, both qualitative and quantitative, can cover a wide range of topics, including false news, and do so quickly. False information will not spread as widely if we can stop it from spreading. Already, rumors and fake news are wreaking on our society, and this will have long-term consequences for our children and families.

Hopefully this research work will produce a machine learning model which will detect real, fake or partially fake news from the inputted data with high accuracy.

1.6 Report Layout

The report has total 5 Chapters which will be followed given by instructions:

This study is summarized in Chapter 1 of the report. This chapter's primary focus is on introduction and discussion. This chapter does a good job of explaining why people get motivated. This sensible study is also essential since it shows what the research questions will be and what the expected results of the investigation will be, as explained in the previous section.

In Chapter 2, What has been done before this issue was researched? What exactly were the researchers trying to achieve with this study, and how did they go about it? What are the research's issues, and how will it solve them? As a conclusion, the study design has been provided in the final section.

This work's statistical methodologies are discussed in Chapter 3's theoretical explanation of research. These procedures have been illustrated in this chapter, and the final section describes how the model is evaluated through the use of machine learning models.

In Chapter 4, the findings of this study are detailed and discussed. Some research-related images have made it easier to grasp the work's criteria.

This section contains the conclusions from Chapter 5. It's important to the achievement of the entire segment. The idea of a substantial research study was presented. In addition, what are the restrictions on conducting this research, which will be useful to other scholars in the future?

CHAPTER 2
BACKGROUND

2.1 Introduction

Related work refers to some work about the Fake News Detection that the previous researcher will be discussed. This is the part where we discuss various methods, limitations & also accuracies of what the researcher found by their study. Various online journals served as inspiration for this work. Our effort was summarized in a short report that tells a narrative about our goals and how we achieved them. As a result of the problem scope, we will have a complicated issue with our research. The challenges we experience are primarily due to technical problems.

2.2 Related Work

After the U.S. presidential election in 2016, the machine learning research community began to dedicate a lot of time to finding fake news, which gained a lot of attention in the general media [2] has done an extensive study on the identification of false news by leveraging information from the sources. They noted the following characteristics that are utilized in content-based classification: Lexical, Syntactic, Visual, Statistical, Users, Post, and Network. Deep neural network models were developed by [4] to identify false news. The dataset used in the experiments was real-world (Buzzfeed and PolitiFact). Their technique of classifying has been separated into three distinct sections.

The first two are text-based categorization systems; the other two utilize both social context and social network data. The model that performs best has been determined to be a deep neural network. To be certain that a news story was genuine, [5] looked at language characteristics to spot false news. The results of their study evaluated four language aspects. In this case, we see evidence from syntax, emotion, grammar, and readability.

The sequential neural network-based model improved on the prior machine learning based approach, outperforming it by over eight percent. The authors, [6] conducted a study that was content-based. Consider the number of unique words, the average number of words per post, and the average amount of characters each post to determine the average number

of characters submitted each week. They implemented several machine learning methods and they managed to achieve the best results using SVM (Support Vector Machine). They obtained a result of 95.70%.

An ensemble model outperformed a single algorithm-based model, [7] examined. A capsule neural network was used by [8] to identify fake news. There are two distinct categories of news in their dataset. An infographic which has short texts and another that has medium to long words. The keywords “dynamic,” “non-static,” and “multichannel” have been used in word embedding techniques. [9] collect statistics on false news. The data contained in their dataset has several languages (English, Spanish, French, Portuguese, Hindi, Turkish, Italian, Chinese, Croatian, Telugu), thus their dataset offers a great opportunity for those who want many languages. In contrast to well-crafted artificial datasets, fake news datasets are often made up of data from just one language. They did a bench analysis of their dataset and they scored 76% with the BERT-based classifier.

The benchmark dataset for false news presented by [10] included fact-checked news articles from 92 fact-checking websites in many languages for COVID-19. In his paper, [11] outlined an annotation system for social media data collected in several modes. They've developed a semi-automated system for gathering multi-modal annotated social media data that combines the efforts of humans and robots. This framework has also been used to obtain false information on COVID-19.

The transformer model suggested by [12] includes human reasoning and metadata for improved performance in classifying false news inside a given domain dataset. To deal with diverse inputs, they employed several BERT models, each with shared weights. Fake news identification might be made easier with an integrated multi-task model, according to [13]. Considering that either one of the two might have a larger rate of false news, they looked at both subjects and writers. They're looking into how subject labels and other contextual information affect short-form false news.

2.3 Research Summary

Fake news and rumors are the greatest dangers in today's digital culture since they may have a variety of harmful effects on the population. Readers have a hard time telling the difference between false and legitimate news. In order to attract readers' attention to their rumors or worthless news, some news portals, blogs and websites that do not have proper authorization to publish news are constantly publishing a variety of rumors or news with spicy headlines, making it difficult to identify reliable and authorized news sources.

Fake news identification utilizing machine learning and deep learning algorithms is demonstrated in this study. However, we can limit the spread of false information by not allowing it to be shared. We require the help of the relevant state or federal department to impose restrictions on any website. For this work, we want to narrow down the number of news items we use for machine learning and deep learning and block the flow of material that has been classified as fraudulent by the algorithms and by country's central information cell. Our major goal is to limit the dissemination of fake news on social media in order to reduce the spread of disinformation. The most difficult part of this study is finding high-quality data, such as examples of both false and legitimate news items over a wide range of themes. In spite of the difficulty, our new algorithm does an excellent job of spotting bogus news and has shown to be very accurate over time.

2.4 Scope of the Problem

Due to the fact that we must work with a large dataset, and gathering large amounts of data and checking competence with a model are both extremely tough tasks. Fake news identification also faces the challenge of detecting language patterns. Using text categorization system to detect false and legitimate news sources. We set out to find a solution to this problem, but it's possible that users of various social media platforms will be unwilling to share their ideas with us using our model since they aren't committed to doing so. If such is the case, then we must raise public awareness of the problem. Creating social awareness will be a major challenge for us.

2.5 Challenges

There are some challenges we have faced during this work -

- Collecting quality data.
- Data preprocessing.
- Data purification and source authentication.
- Preparation and publication of a data collection for research study.
- Getting accuracy above 80%.
- Making a decision based on the results of tests.

CHAPTER 3
RESEARCH METHODOLOGY

3.1 Introduction

It is in this chapter that we explore our research's theoretical foundations. Some methods and datasets will help us get the task done faster and more accurately. Even this is sufficient to get a clear picture of the work involved. There are a number of instruments and technologies that we'll explore in this chapter. Machine learning and data mining are used to assess the collected data and implement it in a way that is beneficial for everyone.

3.2 Research Subject and Instrumentation

The most important aspect of research is the subject matter and the field in which it is being conducted. Our study is primarily news-based. When it comes to Research, it's all about what you're studying. We also digest the news and offer it to our audience in a unique way. It's not enough to convey a notion of what research is all about; it has to be explained in detail. Consequently, in order to have a comprehensive understanding of the subject, you must be familiar with its several subfields. For the sake of this study, we've covered all of the relevant topics. Researchers employ a variety of tools to help them conduct their studies, including a wide range of instruments.

3.3 Data Collection Procedure

Data is one of the most important aspects of any research project, and without it, it is nearly impossible to conduct a study. The results of research are utilized for a variety of testing reasons. And, armed with this information, we can proceed forward with our investigations. We have collected data from various website. We have collected several kinds of news from different international websites.

3.4 Statistical Analysis

Fake news and rumor are the biggest hazards to society in this digital age since they may have a wide range of harmful effects. It's quite difficult for readers to tell the difference between true and bogus news. There are a number of online news portals, blogs, and sites that do not have the legal authority to publish news, but they continue to post rumors or

useless news with sizzling headlines in order to attract visitors. Table 1 shows the quantity of data collections along with platforms.

TABLE 1: DATA QUANTITY WITH PLATFORMS

Platform	**Data**
CNN	550
Fox News	220
The Guardian	340
American News	670
BVA News	425
The Buston Tribune	410
Fox-nees24.com	362
Total	**2977**

We have collected total 2977 data from several websites. There is a variation of different data. The data was classified in to three categories; True, False & Partially false. Where True data contains 1712, False contains 856 & partially false contains 406 news data. Figure 3.4.1 shows the categories of data with size.

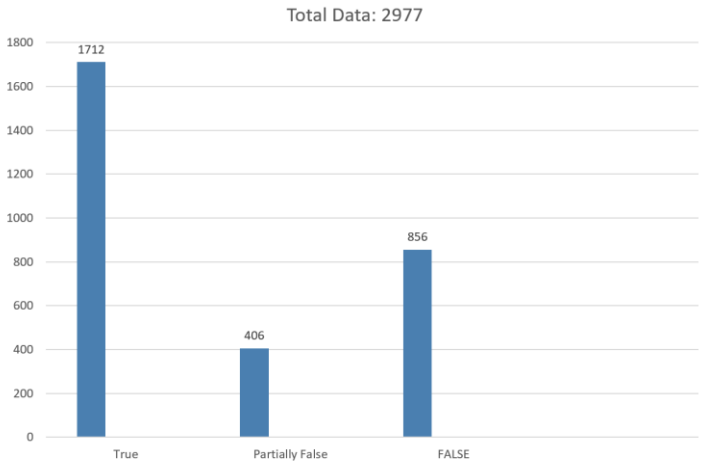


Figure 3.4.1: Data categories with size

3.5 Automatic Detection of Fake News

When we're doing research, the first step we do is pre-process data. The first step in data mining is to transform the data through a process known as data processing. We get news from a variety of sources for this purpose. The data is then preprocessed to fix the inaccuracy. Each news outlet publishes a variety of stories, and these stories form a Beginning Data Set. These data sets are broken down into a variety of numbers. Processes this data on a one-by-one basis.

3.6 Data Pre-Processing

When we're doing research, the first thing we do is pre-process data. The initial step in data mining is to process the data. We get news from a variety of sources for this purpose. In order to fix this, we begin by preparing the data. Each news outlet publishes a variety of stories, and these stories form a Beginning Data Set. These data sets are broken down into a variety of numbers. Processes this data on a one-by-one basis.

Data processing steps

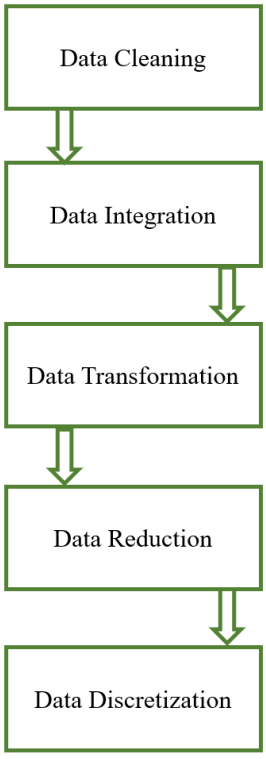


Figure 3.5.1: Proposed data preprocessing method

Removing Puntuation

There are a lot of punctuations, links, numerals, and special characters in news items that in most circumstances have no bearing on whether or not the news is accurate or false. Adding to this, punctuation appears often and has a significant influence on the metrics for punctuation, but has no effect on the categorization of the text.

Capitalization

Although it is possible to use upper and lowercase letters in a sentence, it's preferable to use the same register level in a computational model. When it comes to digits, it doesn't matter what sort of register level you employ. We utilized lowercase registers in this paper.

Lemmatization

Using lemmatization, the number of words with similar meanings may be reduced. Using either lemmatization or stemming, the entire form of a word is reduced to its root form. In contrast to lemmatization, stemming merely removes the beginning and end of a word to get its root. In this paper, we employed stemming to accomplish our goal.

Removing Stop-words

We can cut the number of words in our data even further without affecting the accuracy of our model. A stop-word is a term that appears repeatedly in a piece of writing but has no bearing on the meaning of the piece. Cleansing our data is now complete, and we are ready to begin putting it together into a bag of words [14].

In order to make the data trainable, we need to remove all of the text and strings from the data. For this reason, we transform all of our text into numbers so that it may be used as a learning tool. Our final trainable data is created by vectorizing our text data using the TF-IDF vectorization approach.

3.7 Flow Chart

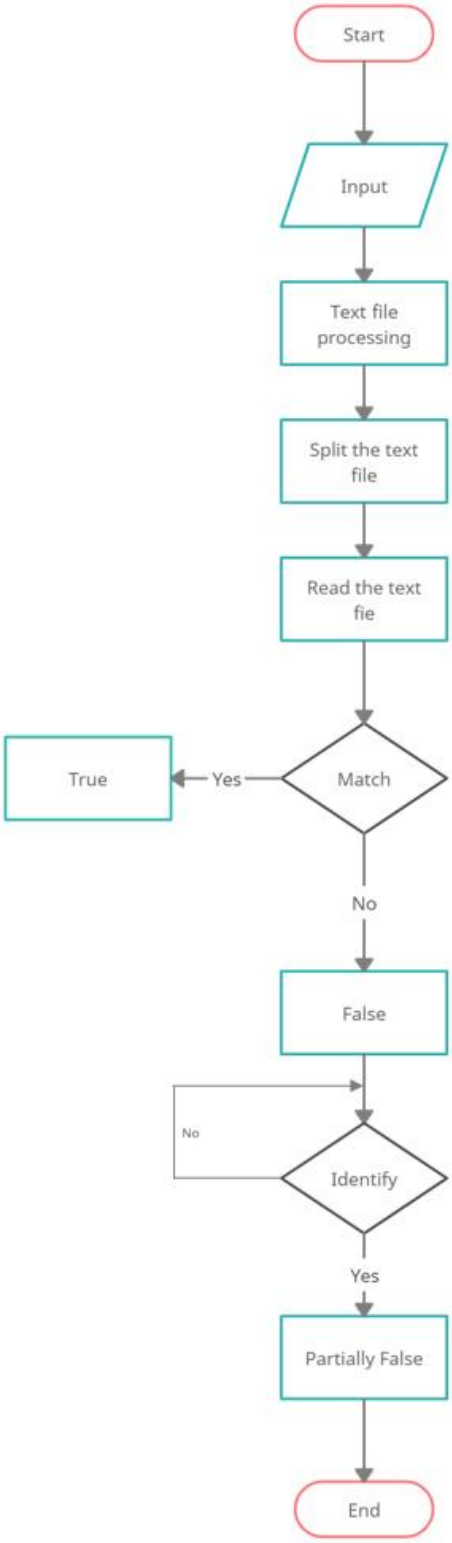


Figure 3.7.1: Proposed flowchart of this work

3.8 Implementation Requirements

There are some requirements to implemented this project.

Hardware/Software Requirements

- Operating System (Windows 7 or above)
- Hard Disk (minimum 250 GB)
- Ram (minimum2 GB)
- Web Browser (preferably chrome)
- Testing tools

Developing Tools

- Python
- Tensor Flow
- Colab
- Notepad++
- Orange

CHAPTER 4
EXPERIMENTAL RESULTS AND DISCUSSION

4.1 Introduction

Specifically, the descriptive analysis of the data utilized in the research, as well as the experimental outcomes of our project, are the primary emphasis of this chapter 4.

4.2 Dataset

4.2.1 Raw data

CNN, Fox News, The Guardian, American News, BVA News, The Buston Tribune, and Fox-news24.com are the sources of our raw data. Our researcher collects our data manually. The news is saved in a text document file when it has been gathered. There are a lot of unnecessary elements in these files, such as an html tag name and a lot of unnecessary data. Since then, it has become clear that the data must be cleaned. That signifies that the raw data has been pre-processed in order to be used in the model.

4.2.2 Cleaning Data

Our data pre-processing task is made easier by the usage of a script file. It is this python script that removes all html tag names. Make sure there are no superfluous gaps in the text. Take off all of the new lines from each news item and arrange it in a single line. Pre-define each news category by assigning an integer number.

We can really receive all of our categorized news in separate files using this method, but the data that is returned is already pre-processed and classified.

4.2.3 Excluded Words Removal

Python code is used to categorize a news item into a certain category. Our system is now ready to develop a model after combining all of the news into a single file. This necessitates some preparatory work. We compile a list of English words that has nothing to do with the news category. Using the term Excluded words, we dubbed the list Excluded words list. Our input file is being checked to see whether any of the terms we've omitted are there. If it's there, it must be eliminated.

4.2.4 Feature Selection and Extraction

Feature selection and extraction are the primary components of this approach's classifying step. When it comes to categorization, it really makes the final decision. We acquire enormous volumes of data. An approach is required in order to make sense of this data. They cannot be processed manually. In this case, feature extraction becomes relevant.

4.3 Building Model and Fit Dataset for Classifier

We split our dataset in two to begin the modeling process.

1. Training Dataset
2. Testing Dataset

80% of the data will be used for training, and 20% will be used for testing. And this is what we anticipate to see in our model, as well We import the sklearn package because we are working with Bi-LSTM classifiers. This classifier is capable of generating an integer that reflects the predicted news category. The figure 4.3.1 illustrates the data ratio of this research work.

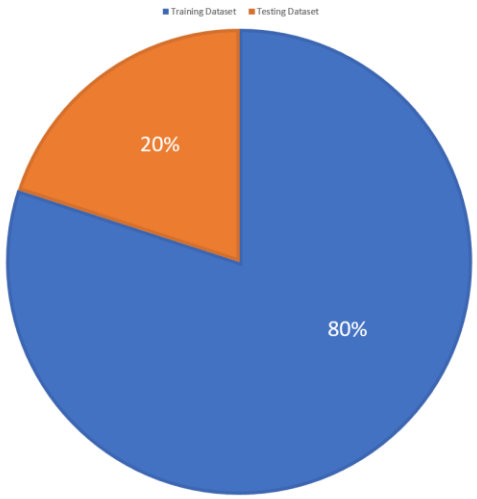


Figure 4.3.1: Data ratio

This figure shows the percentage of data ratio for the training and testing.

4.4 Bi-LSTM Model

Compared to Long Short-Term Memory, Bi-Directional long short-term memory (BiLSTM) is superior in sequence categorization (LSTM). The BiLSTM is made up of two LSTMS, one for forward input and the other for backward input [15]. Using concealed states, it is possible to exchange data in both directions. On every time step, two LSTMs' outputs are combined to form one. The BiLSTM approach helps to reduce the limitations of standard RNN. The context is better understood because to BiLSTM's high level of accuracy.

4.5 Experimental Results

Our BiLSTM model was built with the help of the Keras library. Using a glove embedding of 100d, a model may be created This research use a sequential model as the basis for its analysis. A variety of techniques are employed, including embedding, dropout layers, and a layer with 256 neurons that is completely linked. We have a multiclass dataset.

Because of this, the output layer is applied using soft-max activation. As a result of this design, no data from the test set or the training dataset are ever used. The model is trained with 20 epochs and 128 batch sizes of training data. We found the accuracy of this model is 84%, and the F1-macro score is 62.0.

4.6 Descriptive Analysis

The suggested model's output reveals the identify of the news item that was shown. The news is real, fake, or half false, depending on who you ask. True is assumed to be 2, false is considered to be 0, and partially false is considered to be 1 for determining the identification of the news because the RNN model cannot operate with text. In order to train our model, we employed 20 epochs and obtained an accuracy of 84% and a f1 -macro average score of 62 %. The accuracy and f1 macro average score of the model are shown in the following table 2, which also contains information on the classification report.

TABLE 2: CLASSIFICATION REPORT OF THIS MODEL

Class	Precision	Recall	F1-macro score
True	0.82	0.97	0.89
False	0.92	0.91	0.91
Partially False	0.50	0.03	0.06

The following figure also shows the accuracy and f1 macro average score of the model.

Evaluating Model ...

	precision	recall	f1-score	support
0	0.92	0.91	0.91	174
1	0.50	0.03	0.06	68
2	0.82	0.97	0.89	353
accuracy			0.84	595
macro avg	0.74	0.64	0.62	595
weighted avg	0.81	0.84	0.80	595

Figure 4.6.1: Accuracy and f1 macro average score of the model

TABLE 3: ACCURACY REPORT

Methodology	Accuracy	F1-macro average
Bi-LSTM	84%	62%

The following diagram demonstrates the graphical representation of our suggested model's accuracy vs evaluation accuracy (Figure 4.3) and loss vs evaluation loss (Figure 4.4), which are both obtained by our proposed model. It is evident from those images that our proposed model is gaining knowledge from its previous.

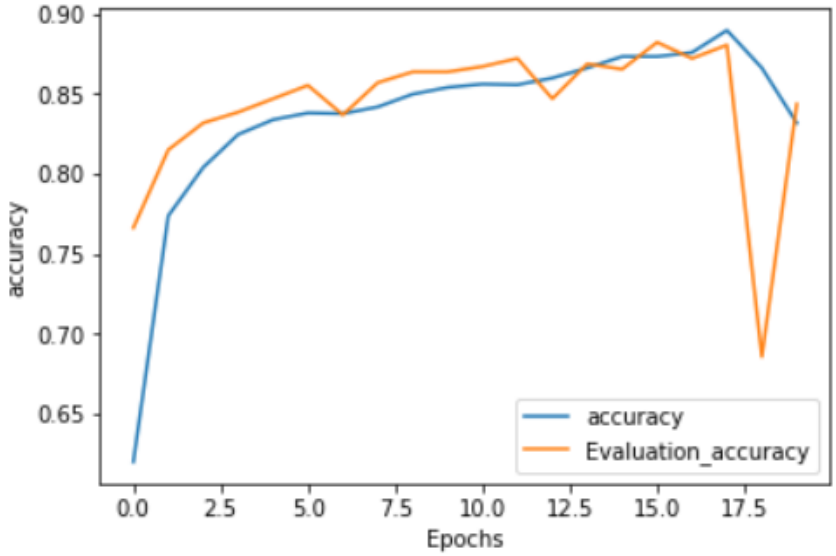


Figure 4.6.2: Accuracy VS Evaluation Accuracy

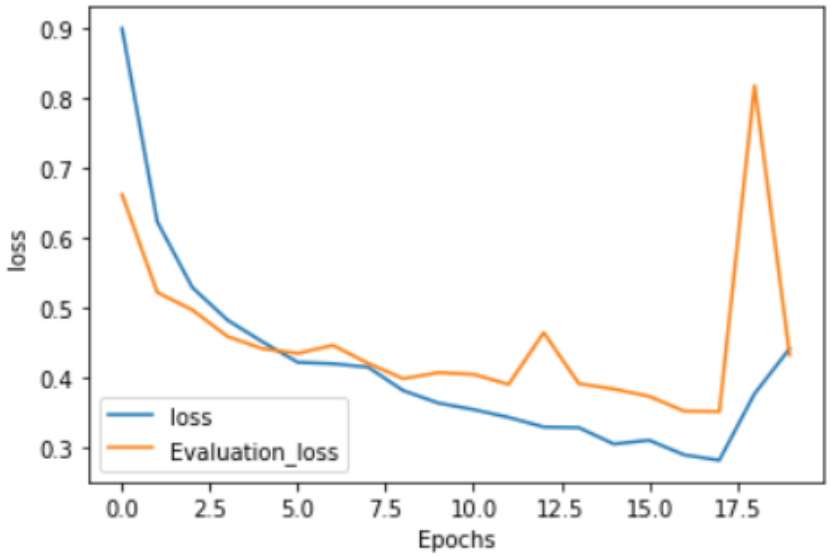


Figure 4.6.3: Loss VS Evaluation Loss

CHAPTER 5
SUMMARY

5.1 Summary of the Study

It's difficult to determine what news is false and what is true. We've hardly put in any effort for such a high-quality outcome. This machine learning technique and approach are being used since there is a lot of study into how people can improve the efficiency and accuracy of their work. In order to determine how much false news and how much actual news there is, we use machine learning techniques to do so. That's how we were able to get these impressive findings and accuracy in our study. For our research, we concentrated mostly on social and political news. Don't let yourself be swayed by misinformation. Classified the datasets and learnt from our own approaches to identify true and fraudulent news.

5.2 Conclusions

A BiLSTM model for identifying fake news was suggested in this study. Filter techniques were utilized to remove unnecessary attributes from our dataset in order to speed up the training process. In order to make the features machine understandable, count vectorizer and word tokenizer are utilized. Therefore, to reduce the time it takes to train our proposed model, features, stopwards, Stemming, and other data preparation activities are employed. This model has an overall accuracy rate of 84% and a macro score of 62.0, which is satisfactory. There are a few issues that need to be addressed. We can improve accuracy and performance by combining BiLSTM with CNN. In order to improve our model's performance, we must spend a lot of effort and money on high-quality hardware components. In the future, we want to use a larger, more diverse dataset to test our suggested approach and overcome our current constraints.

5.3 Recommendations

There are some recommendations for this research work

- To improve the quality of this research, it is necessary to make the data collection process more efficient.

- Collect news that has been placed on news portals or social media in order to reach the desired outcome.
- Machine learning techniques should be used to calculate the result of your constructing methods.
- To acquire the most accurate results, the dataset should be balanced.

5.4 Implication for Further Study

Fake news detection relies heavily on the ability to classify the news. Take a look at the attributes of these two categorized IDs from the news reports. Before that, our main goal is to provide the news headlines that certain people are seeking for more and more, and we do this by developing a database for them. Last but not least, the ability to detect low-quality or even malevolent individuals who are disseminating bogus news is critical for intervention efforts.

- The datasets would be classified to increase productivity.
- This study can be done with high precision if the proper methodologies are used.
- Identified the algorithms that are most conducive to obtaining an accurate outcome.

REFERENCES

[1] Shao, C., Ciampaglia, G. L., Varol, O., Flammini, A., & Menczer, F. (2017). The spread of fake news by social bots. arXiv preprint arXiv:1707.07592, 96, 104.

[2] Granik, M., & Mesyura, V. (2017, May). Fake news detection using naive Bayes classifier. In 2017 IEEE first Ukraine conference on electrical and computer engineering (UKRCON) (pp. 900-903). IEEE.

[3] Dwoskin, E., Dewey, C. and Timberg, C., 2016. "Why Facebook And Google Are Struggling to Purge Fake News." Washington Post, 1.

[4] Abedalla, A., Al-Sadi, A., & Abdullah, M. (2019, October). A closer look at fake news detection: A deep learning perspective. In Proceedings of the 2019 3rd International Conference on Advances in Artificial Intelligence (pp. 24-28).

[5] Dey, A., Rafi, R. Z., Parash, S. H., Arko, S. K., & Chakrabarty, A. (2018, June). Fake news pattern recognition using linguistic analysis. In 2018 Joint 7th International Conference on Informatics, Electronics & Vision (ICIEV) and 2018 2nd International Conference on Imaging, Vision & Pattern Recognition (icIVPR) (pp. 305-309). IEEE.

[6] Hiramath, C. K., & Deshpande, G. C. (2019, July). Fake News Detection Using Deep Learning Techniques. In 2019 1st International Conference on Advances in Information Technology (ICAIT) (pp. 411-415). IEEE.

[7] Wani, A., Joshi, I., Khandve, S., Wagh, V., & Joshi, R. (2021, February). Evaluating deep learning approaches for covid19 fake news detection. In International Workshop on Combating On line Ho st ile Posts in Regional Languages dur ing Emerge ncy Si tuation (pp. 153-163). Springer, Cham.

[8] Hadi Goldani, M., Momtazi, S., & Safabakhsh, R. (2020). Detecting Fake News with Capsule Neural Networks. arXiv e-prints, arXiv-2002.

[9] Shahi, Gautam Kishore, and Durgesh Nandini. "FakeCovid--A Multilingual Cross-domain Fact Check News Dataset for COVID-19." arXiv preprint arXiv:2006.11343 (2020).

[10] G. K. Shahi, D. Nandini, FakeCovid – a multilingual cross-domain fact check news dataset for covid-19, in: Workshop Proceedings of the 14th International AAAI Conference on Web and Social Media, 2020. URL: http://workshop-proceedings.icwsm.org/pdf/2020_14.pdf.

[11] G. K. Shahi, Amused: An annotation framework of multi-modal social media data, arXiv preprint arXiv:2010.00502 (2020).

[12] Blackledge, C., & Atapour-Abarghouei, A. (2021). Transforming Fake News: Robust Generalisable News Classification Using Transformers. arXiv preprint arXiv:2109.09796.

[13] Jing, Q., Yao, D., Fan, X., Wang, B., Tan, H., Bu, X., & Bi, J. (2021, July). TRANSFAKE: Multi-task Transformer for Multimodal Enhanced Fake News Detection. In 2021 International Joint Conference on Neural Networks (IJCNN) (pp. 1-8). IEEE.

[14] Wilbur, W. J., & Sirotkin, K. (1992). The automatic identification of stop words. Journal of information science, 18(1), 45-55.

[15] Braşoveanu, A. M., & Andonie, R. (2019, June). Semantic fake news detection: a machine learning perspective. In International Work-Conference on Artificial Neural Networks (pp. 656-667). Springer, Cham.

Report Checking Alamin Group's

ORIGINALITY REPORT

20%	18%	8%	15%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Daffodil International University Student Paper	6%
2	dspace.daffodilvarsity.edu.bd:8080 Internet Source	5%
3	Submitted to Bournemouth University Student Paper	1%
4	ceur-ws.org Internet Source	1%
5	Submitted to University of Hertfordshire Student Paper	1%
6	Submitted to The University of Memphis Student Paper	1%
7	Daniel Röchert, Gautam Kishore Shahi, German Neubaum, Björn Ross, Stefan Stieglitz. "The Networked Context of COVID-19 Misinformation: Informational Homogeneity on YouTube at the Beginning of the Pandemic", Online Social Networks and Media, 2021 Publication	1%

8	www.wcse.org Internet Source	<1 %
9	publications.waset.org Internet Source	<1 %
10	libres.uncg.edu Internet Source	<1 %
11	thesai.org Internet Source	<1 %
12	Submitted to American University in Cairo Student Paper	<1 %
13	Submitted to Coventry University Student Paper	<1 %
14	aclanthology.org Internet Source	<1 %
15	pdxscholar.library.pdx.edu Internet Source	<1 %
16	dokumen.pub Internet Source	<1 %
17	"Combating Online Hostile Posts in Regional Languages during Emergency Situation", Springer Science and Business Media LLC, 2021 Publication	<1 %
18	Submitted to Georgia Gwinnett College Student Paper	<1 %

19	Rohit Kumar Kaliyar, Kedar Fitwe, Rajarajeswari P, Anurag Goswami. "Classification of Hoax/Non-Hoax News Articles on Social Media using an Effective Deep Neural Network", 2021 5th International Conference on Computing Methodologies and Communication (ICCMC), 2021 Publication	<1 %
20	onlinelibrary.wiley.com Internet Source	<1 %
21	Somya Ranjan Sahoo, B.B. Gupta. "Multiple features based approach for automatic fake news detection on social networks using deep learning", Applied Soft Computing, 2020 Publication	<1 %
22	Mandeep Singh, Mohammed Wasim Bhatt, Harpreet Singh Bedi, Umang Mishra. "Performance of bernoulli's naive bayes classifier in the detection of fake news", Materials Today: Proceedings, 2020 Publication	<1 %
23	edepot.wur.nl Internet Source	<1 %
24	gautamshahi.github.io Internet Source	<1 %
25	repositorio.ufscar.br Internet Source	<1 %

26	"Communicating Science in Times of Crisis", Wiley, 2021 Publication	<1 %
27	"Intelligent Computing and Innovation on Data Science", Springer Science and Business Media LLC, 2020 Publication	<1 %
28	"Web Information Systems Engineering – WISE 2018", Springer Science and Business Media LLC, 2018 Publication	<1 %
29	"Inventive Communication and Computational Technologies", Springer Science and Business Media LLC, 2021 Publication	<1 %

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	Off		