

**“C Becomes The Most Popular Language For Machine Learning”
Fake News Detection**

By

**MD. ABDULLAH AL AMIN
161-15-6851**

AND

**AQUIBUZZAMAN MD. SAYEM
162-15-7728**

AND

**MUHAIMINUL ISLAM JIM
162-15-7923**

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

Supervised By

Saiful Islam

Senior Lecturer

Department of CSE

Daffodil International University

Co-Supervised By

Nusrat Jahan

Lecturer

Department of CSE

Daffodil International University



DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

NOVEMBER 2019

APPROVAL

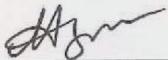
This Project titled “C Becomes The Most Popular Language For ‘Machine Learning’ Fake News Detection” submitted by Md. Abdullah Al-Amin, ID No: 161-15-6851 and Aquibuzzaman Md. Sayem, ID No: 162-15-7728 and Muhaiminul Islam Jim, ID No: 162-15-7923 to the Department of Computer Science and Engineering, 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 and approved as to its style and contents. The presentation has been held on December 6, 2019.

BOARD OF EXAMINERS



Dr. Syed Akhter Hossain
Professor and Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Chairman



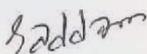
Nazmun Nessa Moon
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner



Dr. Fizar Ahmed
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner



Dr. Md. Saddam Hossain
Assistant Professor
Department of Computer Science and Engineering
United International University

External Examiner

DECLARATION

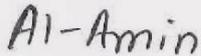
We hereby declare that, this project has been done by us under the supervision of **Saiful Islam, Senior Lecturer, Department of CSE** Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

Supervised by:

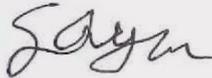


Saiful Islam
Senior Lecturer
Department of CSE
Daffodil International University

Submitted by:



Md. Abdullah Al-Amin
ID: 161-15-6851
Department of CSE
Daffodil International University



Aquibuzzaman Md. Sayem
ID: 162-15-7728
Department of CSE
Daffodil International University



Muhalminul Islam Jim
ID: 162-15-7923
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First we express our heartiest thanks and gratitude to almighty Allah 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 **Saiful Islam, Senior Lecturer**, Department of CSE Daffodil International University, Dhaka. Deep knowledge & keen interest of our supervisor in the field of “*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 draft and correcting them at all stages have made it possible to complete this project.

We would like to express our heartiest gratitude to Prof. **Dr. Syed Akhter Hossain, Head, Department of CSE**, for his kind help to finish our project and also to other faculty members and the staff of CSE department of Daffodil International University.

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

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

ABSTRACT

Now-a-days political people or other people are spreading fake news for their own benefit. Ordinary people easily believe these fake news. As a result, riots are spreading among the people across the country which is risky for a developing country. In this paper, we are working to detect fake news and provide a model for checking fake news. We collected our dataset which has different type of news and labeled them as 0 and 1, which means true and fake respectively. For detecting the fake news we use Long Short-Term Memory, Bidirectional Long-Short Term Memory and Random Forest algorithms in our dataset and compared the results among these model for checking which one gives us a better result. After our experiment we found that Random Forest algorithm had 87.75% accuracy to detect the fake news.

TABLE OF CONTENTS

CONTENTS	PAGE
Board of examiners	ii
Declaration	iii
Acknowledgements	iv
Abstract	v
CHAPTER	PAGE
CHAPTER 1: INTRODUCTION	10-13
1.1 Introduction	1
1.2 Objectives	2
1.3 Motivation	2
1.4 Rationale of the Study	3
1.5 Research Questions	3
1.6 Expected Output	3
1.7 Report Layout	4
CHAPTER 2: BACKGROUND	14-16
2.1 Introduction	5
2.2 Related Works	5
2.3 Research Summary	6
2.4 Scope of the problem	6
2.5 Challenges	7

CHAPTER 3: RESEARCH METHODOLOGY	17-23
3.1 Introduction	8
3.2 Research Subject and Instrumentation	8
3.3 Data Collection Procedure	8
3.4 Statistical Analysis	9
3.5 Implementation Requirements	13
CHAPTER 4: EXPERIMENTAL RESULTS AND DISCUSSION	24-25
4.1 Introduction	14
4.2 Experimental Results	14
4.3 Summary	15
CHAPTER 5: SUMMARY, CONCLUSION, RECOMMENDATION AND IMPLICATION FOR FUTURE RESEARCH	26-27
5.1 Summary of the Study	16
5.2 Conclusions	16
5.3 Recommendations	16
5.4 Implication for Further Study	17
REFERENCES	18
APPENDIX	19

LIST OF FIGURES

FIGURES	PAGE
Figure 3.4.1: Diagram of a LSTM model	19
Figure 3.4.2: Diagram of a BLSTM model	19
Figure 3.4.3: Diagram of a Random Forest model	20
Figure 3.4.1.1: Unprocessed Data	21
Figure 3.4.2.1: Dataset Summary	21
Figure 3.4.3.1: Train-Test data summary	22
Figure 3.4.3.2: Train History	22
Figure 4.2.1: Accuracy Comparison	25

LIST OF TABLES

TABLES	PAGE
Table 4.2.1: Accuracy Metrics	14

CHAPTER 1

INTRODUCTION

1.1 Introduction

Judging everything happening around us, we can easily state that the outspread of fake news is greatly hampering the peace of society. We can see that people are getting bullied, attacked and being shamed because of fake news spreading about them or certain events surrounding them.

If we look closely we can get some idea about why these fake news are being spread between people. Rolling out made up news isn't new in our society. If we indulge ourselves in history we can see that this messed up technique is used from long before any news media were invented. People would spread political propaganda to make get biased opinion from the people. Also people would tell lies about others in order to defame them. So this problem isn't new. But now many more reasons may have been added.

These days political propaganda is also spread to push biased thoughts in people's minds. But another thing added to the reason is the change of medium of news. Back in the old days the only way to get news was through the newspaper or the radio. Even televisions when it was available. But now because of the advancements of technology we get news through online media. Like from Facebook, twitter and online mediums of the same newspapers or television channels. In this field there are a whole lot of competition so what these mediums do is they try to make catchy headlines to attract people to click on their news. But in the name of making catchy headlines they sometimes spread fake news. Just to get more visitors on their pages. Which seems like a big reason for the spread of fake news.

There can be many different ways to detect these fake news. Print media has less fake news. So just from knowing which media news came from can give us a probability of the news being true or fake.

But in this paper we look at the problem from another perspective. Here we tried to detect whether the news is true or fake from a machine learning and NLP perspective. We will use different algorithms here on the data set and observe how each of them perform and why they perform in that way. We will overlook some things such as source of the news, author, and medium. We will only observe from total NLP perspective. We will take the headline and body of a news and try to detect whether a news is fake by taking the words in account.

1.2 Objectives

- ❑ Determine which algorithm is best for this classification
- ❑ Be able to tell which news is fake

1.3 Motivation

The present era is the age of technology. In this technological era, we can access everything by staying in our home. This technology has made our world smaller and brings us closer than ever.

Nowadays online newspapers are very popular. Through it, we can know what is happening around the world instantly. But some people misuse this technology. Some sources spread false news by using online news-paper and social media like Facebook or Twitter. It creates confusion among the mass people and sometimes it becomes very dangerous because people can't identify what is real.

In the last few years, Bangladesh also faced incidents of fake news. In Cox's Bazar's Ramu_[4] (12th September 2012), suddenly a piece of news spread that the holy Quran and the Prophet of Islam were humiliated by a boy named Uttam Barua in Facebook. This news spread quickly all over the country. In that time nineteen Buddhist temples were destroyed or damaged by some groups from the local Muslim community and many Buddhist houses were attacked, vandalized, and set on fire. Another incident happened on 5th May 2013. On that day thousands of supporters of Quami Madrasa-based organization Hefajot e Islam gather at Shapla Square to blockade Dhaka as several bloggers were accused of conspiring against Islam. At night thousands of

RAB, police and BGB evacuated the place in a joint operation. The next day, the whole Motijheel area felt like a war-torn region. On that day, it was claimed that about two and a half thousand were killed and various pictures were spread on Facebook. But the government was informed that 28 people were killed that night. For these types of incidents, riots between people of the country are rife which is not desirable for an independent country.

In this regard, we are intending to discover a system that will identify the false news.

1.4 Rationale of the Study

Spreading fake news can have serious consequences in our society. It can cause clashes between people of different ideologies and religions. May cause political disruption, create chaos and destroy the peace of the society we live in. A country's people and economy may get hampered. So, being able to detect if a news is fake or not will have serious impact. There are many studies on this area so we already know it is possible. We will try to find which algorithms will be more efficient in this process and also see which algorithm gives more accuracy. 100% accuracy hasn't been found in this area. So to detect fake news we must know how we can get the highest accuracy. Our study will compare different algorithms to find the best one up for the task. This will help the future researchers by pointing them to the right algorithm for solving this problem with more accuracy.

1.5 Research Questions

- Is it possible to detect fake news?
- Which algorithms to use?

1.6 Expected output

We already know how important it is to know if a news is real or fake. With this study we will have the capability to state whether a news is fake or not. To know this we will use LSTM, BLSTM and Random Forest algorithm. And know which one will get us our desired outcome with less error.

1.7 Report Layout

In the first chapter, we give an introduction to the problem. How this problem came and what are the intentions of doing it. And small overview of what we are doing in this research and why. What we want to achieve. The second chapter is more of a literature review and what we want to do and why we think it's important. The next chapter gives is on what we did and why we did it. It gives a clear idea about what we took in account to do the research. And how we managed to be successful. The fourth chapter talks about the result and describes. In the last chapter we put a closure on the research.

CHAPTER 2

BACKGROUND

2.1 Introduction

We wanted to make a system using natural language processing to detect the false news. To implement this system we used some natural language processing model and algorithm, tools and studied some related works. All of these will be described in detail here.

2.2 Related Works

Fake news detection is not a new area. The problem has been around for quite a long time and also there has been a lot of work to solve this problem. To have a better understanding and figure out what we should work with, we had to study some of the relevant works.

We saw the use of LSTM (Long Short-Term Memories) done by Samir Bajaj^[1] where this algorithm had a good performance and we think this algorithm can set a base for our study.

Kurt Miller and Alex Oswalt^[2] tried a different approach using Bidirectional Long Short-Time Memory (BLSTM) to detect the fake news. They implemented their model using Keras deep learning framework and Tensorflow for weight calculation. In their work, they found that Recurrent Neural Network models are most efficient when LSTM units are used in place of standard RNN units. After the research, they designed a model that could detect fake news with an accuracy of 57% using Bidirectional LSTM and Multilayer Perceptron.

Ray Oshikawa, Jing Qian and William Yang Wang^[5], they surveyed on different level algorithms Support Vector Machine (SVM), Naive Bayes Classifier (NBC), Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM). After their research they saw, in the field of Natural Language Processing, LSTM is better to use for

automated learning methods. It is because they can analyze long-term and content-transitional information.

William Yang Wang^[6], he designed a hybrid CNN model to integrate meta-data with text. He used Logistic Regression classifier (LR), a Support Vector Machine classifier (SVM), Convolutional Neural Network (CNN) and Bidirectional Long Short-Term Memory (Bi-LSTM) randomly to detect the fake news. After his research, he compared various models. In his result, SVM and LR performed well but Bi-LSTM didn't perform well.

Mykhailo Granik and Volodymyr Mesyura^[7] used Naive Bayes Classifier algorithm to detect fake news. They used this algorithm in Facebook news posts dataset and achieved approximately 74% accuracy to get the fake news. They think this algorithm can be improved for getting better results.

2.3 Research Summary

We decided to look into the problem from the perspective of NLP and Machine Learning. There has been many different works on our topic but we aim to add more to the field. In work done before we can see the use of many different algorithms. Such as LSTM^{[1][2][3]}, BLSTM^[8], CNN, RNN, Random Forest. Different models can get us different outcomes. After knowing about some models we decided to go with LSTM, BLSTM and Random Forest. Details about these models will be discussed and an efficient model between these will be established.

2.4 Scope of Problem

This issue becomes serious because people are now more online faced. They are on many social sites and they are very active on those. And people seem to believe what they see there. Parties with ill intentions grab this weak point. But these drawbacks can be resolved with a system that has good accuracy. Natural Language processing along with Machine Learning cracks this issue. Here we will be analyzing news headlines and body written in English and determine if news are fake or not.

2.5 Challenges

Determining the authenticity of news isn't an easy task. News will have different punctuations to express different sentiments. These needs to be processed. Another thing is that sometimes the same sentence can express different emotions. There can be satires in between news, which we won't be able to determine easily.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Introduction

An overview of our research and the procedures we went through will be in this part of the report. It gives us detailed information of the methods we used to and how they were implemented. It will also discuss about the sources we have drawn our data from and how we plan to use those to classify news. Tools used in this research are also discussed in this section.

3.2 Research Subject and Instrumentation

3.2.1 Research Subject

The goal of this research is to perceive an efficient method that would classify news with better accuracy and less error rate. To achieve it, we decided to analyze the text and find out whether the news is fake or not. There were other options to consider to get to the very same result. Like, news sources. We can see that fake news is mostly spread through online media. So we could use news source to classify the news. But we cannot be 100% confident that print media won't spread fake news. Rather we analyze the text of the headline and body of a news and determine patterns to recognize fake news.

3.2.2 Instruments

We used a dataset that contained vast data, 4049 rows of data to be exact. These were labeled as 1 or 0 which meant fake or true respectively. Models like LSTM, BLSTM and Random Forest were used to analyze all these data and classify the news.

3.3 Data Collection Procedure

Data collection is one of the hardest challenges when the project is related to machine learning. If the data isn't of good quality, only that can bring the whole research to its knees. So we collected our data very carefully. We collected data from multiple

sources including but not limited to Kaggle, Socrata etc. Then we had checked some of the data randomly for being sure of the quality of our data.

3.4 Methodology and Data Analysis

We built multiple models to analyze the results and single out the best model for fake news detection. Models we've tried are:

1. Long Short Term Memory (LSTM)
2. Bidirectional Long Short Term Memory (BLSTM)
3. Random Forest

Both LSTM and BLSTM are variations of RNN (Recurrent Neural Network).

RNNs are a bit different from feed forward networks. Feed forward networks can't properly address sequence of information. That makes it hard to use those networks in some applications where the data is sequential. For example, speech recognition, handwriting recognition etc. For doing these tasks efficiently RNNs come into play. RNNs can use their internal memory to process sequences. They are networks with loops in them, allowing information to persist.

LSTM is a bit different from regular neural networks as it have feedback connections. And this property makes it a good choice for analyzing continued data like speech or video. It's widely used for video recognition, speech recognition and handwriting recognition. A regular LSTM has four parts; a cell, an input gate, an output gate and a forget gate.

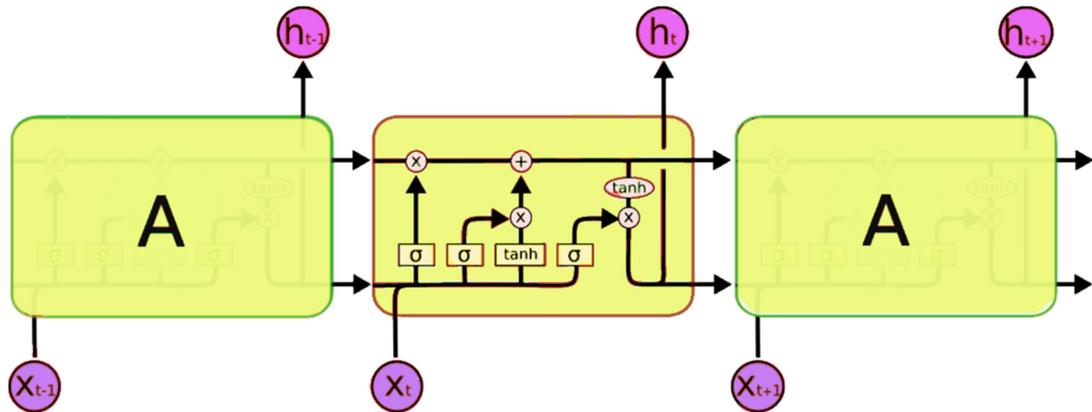


Figure 3.4.1: Diagram of a n LSTM model

BLSTM is a variant of LSTM that has two opposite directed layers. BLSTMs make more information available to the network which results in better learning in some cases.

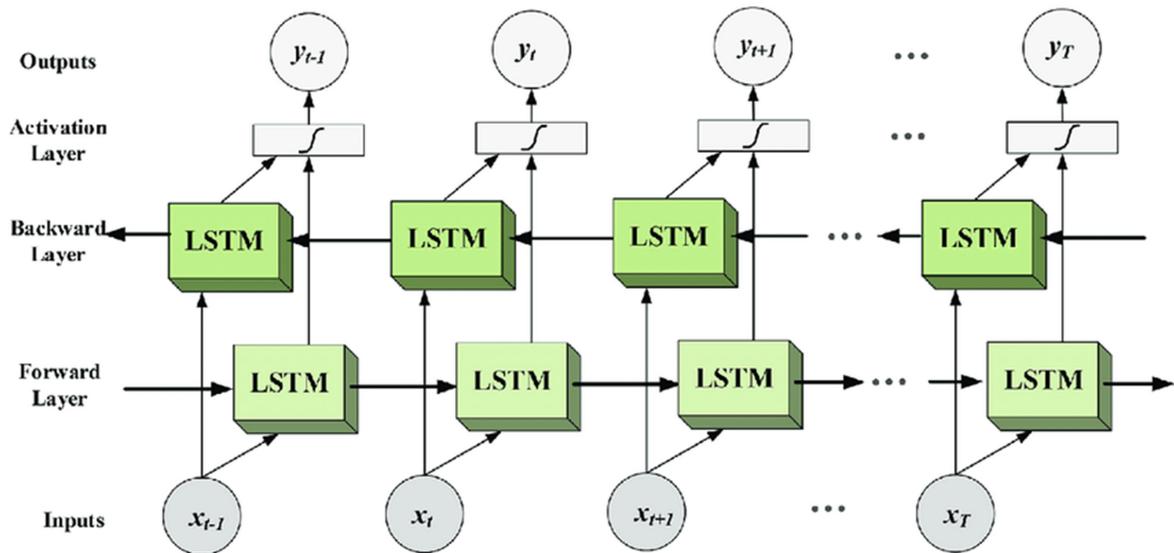


Figure 3.4.2: Diagram of a BLSTM model

Random forest on the other hand works by creating numerous decision trees at the time of training.

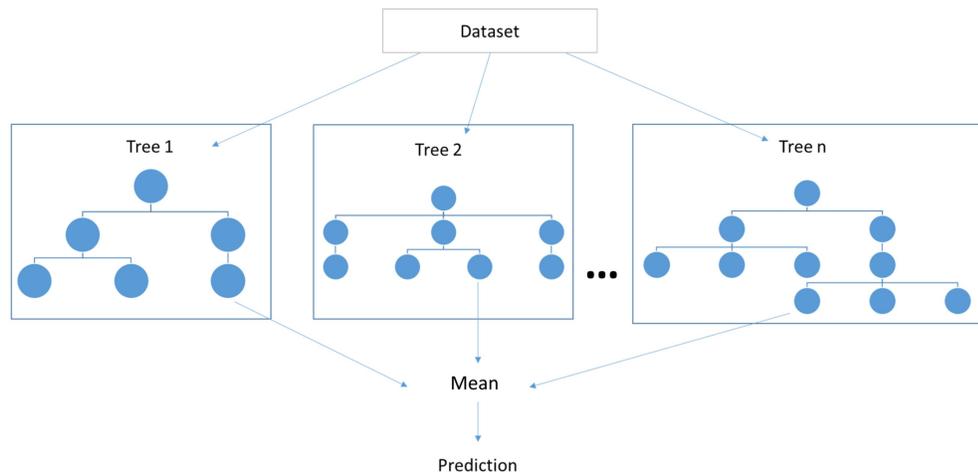


Figure 3.4.3: Diagram of a Random Forest model

We used these three models for our case. And roughly work for all the models can be classified in three parts.

3.4.1 Pre-Processing

At the pre-processing stage we processed our data so that a model can be trained easily using the clean set of data. So we removed all the special characters, tokenized the inputs. Created a new column, changed the file format to CSV etc. After doing these our dataset was ready for use.

	URLs	Headline	Body	Label
3408	http://beforeitsnews.com/entertainment/2017/10...	10 Awful Moments Caught on Video!	10 Awful Moments Caught on Video!\n% of reader...	0
2990	https://www.reuters.com/article/us-britain-eu-...	In Brexit poker, clock narrows transition options	BRUSSELS (Reuters) - Nerves are fraying in the...	1
2165	http://www.bbc.com/news/entertainment-arts-415...	Dolly Parton to read CBeebies bedtime story	Image caption Dolly has just released her firs...	1
1465	https://www.reuters.com/article/us-usa-healthc...	Trump says he's likely to sign healthcare orde...	WASHINGTON (Reuters) - President Donald Trump ...	1
3878	http://beforeitsnews.com/u-s-politics/2017/09/...	The NFL kneels its fans	The NFL kneels its fans\n(Before It's News)\nC...	0

Figure 3.4.1.1: Unprocessed data

3.4.2 Feature Extraction

The preprocessing stage made this step a bit easier. We looked for the most frequent words used differently for both genuine and fake news. We used TF-IDF for extracting features from text corpus. This made the job of feature extraction very simple.

```
↳ count      4009.000000
   mean        60.104016
   std         22.660900
   min         6.000000
   25%         48.000000
   50%         60.000000
   75%         69.000000
   max        653.000000
   Name: Length, dtype: float64
```

Figure 3.4.2.1: Dataset Summary

3.4.3 Training

At this stage we split our dataset into two parts. One for training and another for validation. After that, we used it to train multiple models and validate those. For making training faster we took help of GPU acceleration in Google Colab. That made it possible tinkering every once in a while and rerun without much hassle.

```
print('Train Data Feature (Mixed):', len(x_train_data))
print('Train Data Label (Mixed):', len(y_train_data))
print('Test Data Feature (Mixed):', len(x_test_data))
print('Test Data Label (Mixed):', len(y_test_data))
```

```
↳ Train Data Feature (Mixed): 2996
Train Data Label (Mixed): 2996
Test Data Feature (Mixed): 748
Test Data Label (Mixed): 748
```

Figure 3.4.3.1: Train-Test data summary

Here in the figure below we can see that the loss is gradually decreasing in each epoch.

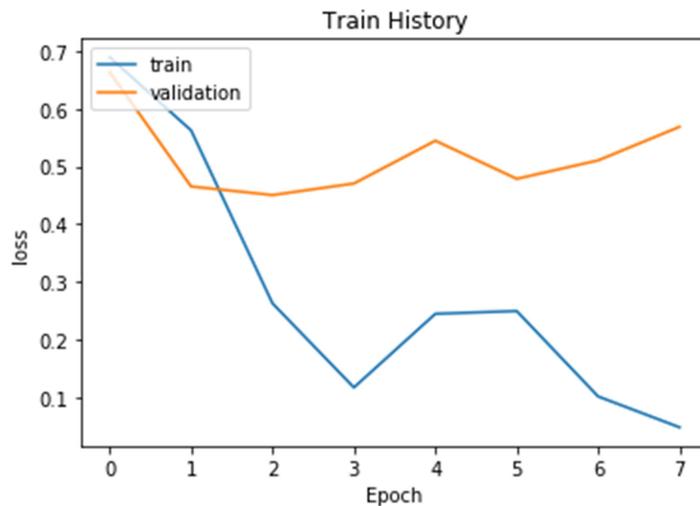


Figure 3.4.3.2: Train History

3.5 Implementation Requirements

We used Google Colab as the platform for our implementation parts. Python3 was the language that we used. Multiple libraries were used. A few of those are Tensorflow, NLTK, Matplotlib, Keras etc.

CHAPTER 4

EXPERIMENTAL RESULTS AND DISCUSSION

4.1 Introduction

To detect the fake news and get the higher accuracy, we used Long Short-Term Memory (LSTM), Bidirectional Long Short-Term Memory (Bi-LSTM) and Random Forest Algorithm in our datasets which contains four columns. Each column has URL of the article from the web that publishes the news, headline of that news article, news body and the label of that news respectively. After our experiment we got a final result which algorithm performed well to detect the fake news, why it is performed well and why other algorithms didn't perform well all discussion are described in detail here.

4.2 Experimental Results

After training we found different models to perform slightly different. The accuracy of random forest was the best while LSTM performed a little worse. The accuracy table is shown below.

Table 4.2.1: Accuracy Metrics

Model	Accuracy
LSTM	85%
Bidirectional	87%
Random Forest	87.75%

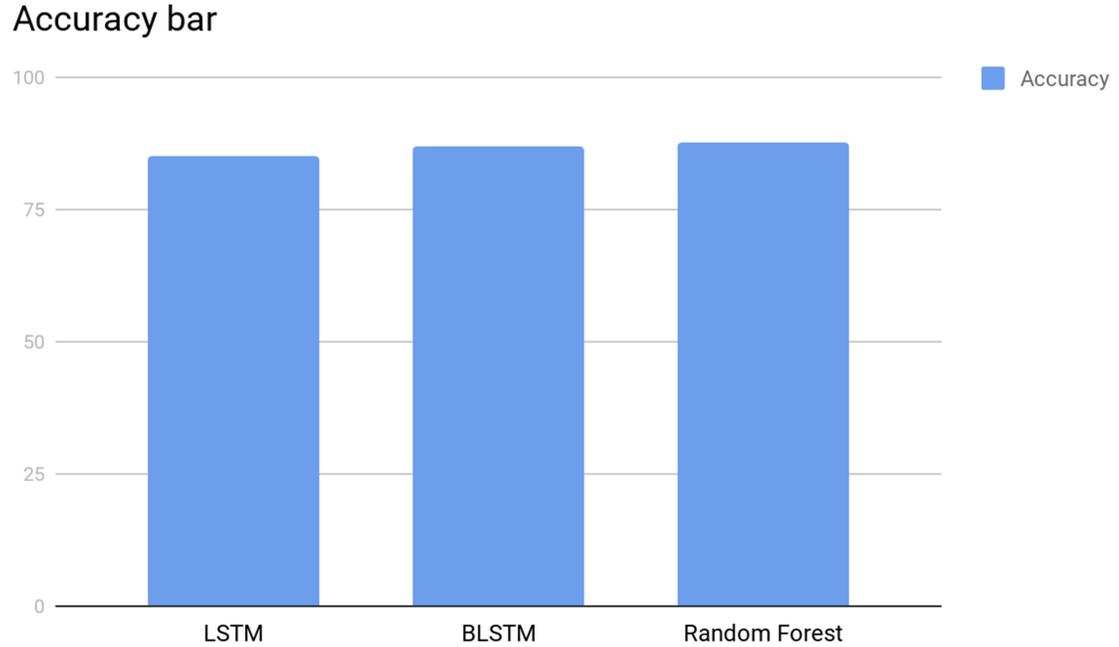


Figure 4.2.1: Accuracy Comparison

The above bar showcases the accuracy of different models for detecting fake news. In the above figure we can see BLSTM performs better than LSTM while Random forest does the best job at detecting fake news with an accuracy of 87.75%.

4.3 Summary

Different models were implemented to find out which one is more efficient to our cause. The dataset we worked on was rich, which helped the classification models to achieve better accuracy. Between the three models we implemented “Random Forest” had 87.75% accuracy which was the most.

CHAPTER 5

SUMMARY, CONCLUSION, RECOMMENDATION AND IMPLICATION FOR FUTURE RESEARCH

5.1 Summary of the Study

The solution we found for this problem is classifying the text data to determine if a news is fake or not. Text classification has been a hot topic for researchers for a long time. So there has been a lot of work in this field. Our research study finds the best model between LSTM, BLSTM and Random Forest. We found “Random Forest” to be the most efficient for text classification to detect fake news. We used a large data set to train the models which increased the probability of identifying fake news better. These data were processed before they were fed through the models which also benefited to make the models more precise.

5.2 Conclusions

This is quite a popular topic and there have been many different research studies. Our study will help future researchers to select models without dilemma. We used different models to detect fake news and we found “Random Forest” to be the best among the models we implemented. We achieved 87.75% using this model.

5.3 Recommendations

We are working with a lot of data for training our model. So there will be a lot of sentences that contain words, digits, punctuation and other symbols. To train our model we have worked on sentences with digits, special symbols in our input data. So if we have these in our dataset, there will be no problem to train our model. But we need to be careful that the dataset we use for fake news detection is accurate. The more accurate our dataset, the more accurate our model will be able to provide fake news detects. If there is something wrong with our dataset, then the data training in our model will also go wrong. Then no matter how good the models or algorithms we use, it will be useless to give better accuracy. So we need to look at the accuracy of our dataset.

5.4 Implication for Further Study

In our research paper we used Long Short-Term Memory, Bidirectional Long Short-Term Memory and Random Forest algorithm to implement our system. But there are some others algorithms such as Convolutional Neural Network, Recurrent Neural Network Support Vector Machine (SVM) and Naive Bayes Classifier (NBC) etc. which can also be used to detect fake news. In future, we will work with these algorithms and again we will compare the results to check which algorithm perform well to detect the fake news. And we will use a more enriched dataset to train our models for detecting fake news. In our system we only used English language to train our models. In future, we will also work with other languages like Bengali, Hindi, German, Spanish etc. So, we can detect the fake news in other languages. In future, we will implement a live system that would detect fake news in real time. So that we can detect the fake news from online news-paper, Facebook posts and other online mediums in real time.

References

- [1] “The Pope has a new baby. Fake news detection using deep learning”, SemanticScholar.org, available at: <<<https://www.semanticscholar.org/paper/%E2%80%9C-The-Pope-Has-a-New-Baby-!-%E2%80%9D-Fake-News-Detection-Bajaj/19edb6aa318d70cd727b3cdb006a782556ba657a>>>, last accessed on: 04-09-2019
- [2] “Fake News Headline Classification using Neural Networks with Attention”, SemanticScholar, available at: <<<https://www.semanticscholar.org/paper/Fake-News-Headline-Classification-using-Neural-with-Miller/75da7a21c87f198af775646280785c6d7ecbcf59>>>, last accessed on: 12-09-2019
- [3] Álvaro Ibrain Rodríguez, Lara Lloret Iglesias, “Fake news detection using deep learning”, Cornell University Computer Science Department, V2, Page:3, October 2019
- [4] BDNews24, available at <<<https://bdnews24.com/bangladesh/2012/09/29/buddhist-temples-homes-burned-looted-in-ramu>>>, last accessed on: 30-10-2019
- [5] Ray Oshikawa, Jing Qian, William Yang Wang, “A Survey on Natural Language Processing for Fake News Detection”, Cornell University Computer Science Department, 2 November 2018
- [6] “ "Liar, Liar Pants on Fire": A New Benchmark Dataset for Fake News Detection”, Arxiv, available at: <<<https://arxiv.org/abs/1705.00648>>> last accessed on: 24 September 2019
- [7] Mykhailo Granik, Volodymyr Mesyura, “Fake news detection using naive Bayes classifier”, 2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON), June 2017
- [8] V. Frinken, A. Fischer, H. Bunke, "A Novel Word Spotting Algorithm Using Bidirectional Long Short-Term Memory Neural Networks", *4th Workshop on Artificial Neural Networks in Pattern Recognition*, 2010.

Appendix

NLP - Natural Language Processing

LSTM - Long Short Term Memory

BLSTM - Bidirectional Long Short Term Memory

RNN - Recurrent Neural Network

CNN - Convolutional Neural Network

SVM - Support Vector Machine

NBC - Naive Bayes Classifier

Keras - Deep Learning Library

Tensorflow - A Machine Learning Platform

LR- Logistic Regression

TF-IDF - term frequency–inverse document frequency

GPU - Graphics Processing Unit

NLTK - Natural Language Toolkit

Turnitin Originality Report

Processed on: 04-Nov-2019 14:31 +06
 ID: 1206600850
 Word Count: 4531
 Submitted: 1

Similarity Index

22%

Similarity by Source

Internet Sources: N/A
 Publications: N/A
 Student Papers: 22%

FND By Sayeem Aquib

8% match (student papers from 26-Aug-2019)
[Submitted to Daffodil International University on 2019-08-26](#)

3% match (student papers from 02-Apr-2019)
[Submitted to Daffodil International University on 2019-04-02](#)

1% match (student papers from 03-Apr-2019)
[Submitted to Daffodil International University on 2019-04-03](#)

1% match (student papers from 15-Jul-2017)
[Submitted to Institut Teknologi Brunei on 2017-07-15](#)

1% match (student papers from 02-Apr-2019)
[Submitted to Daffodil International University on 2019-04-02](#)

1% match (student papers from 07-Apr-2018)
[Submitted to Daffodil International University on 2018-04-07](#)

1% match (student papers from 04-Sep-2019)
[Submitted to University of Sheffield on 2019-09-04](#)

< 1% match (student papers from 16-Jul-2019)
[Submitted to Syiah Kuala University on 2019-07-16](#)

< 1% match (student papers from 04-Mar-2018)
[Submitted to National College of Ireland on 2018-03-04](#)

< 1% match (student papers from 01-Mar-2018)
[Submitted to University of Lincoln on 2018-03-01](#)

< 1% match (student papers from 02-Apr-2019)
[Submitted to Daffodil International University on 2019-04-02](#)

< 1% match (student papers from 21-Apr-2018)
[Submitted to University College London on 2018-04-21](#)

< 1% match (student papers from 26-Apr-2019)
[Submitted to Middlesex University on 2019-04-26](#)

< 1% match (student papers from 28-Aug-2018)
[Submitted to University of Oxford on 2018-08-28](#)