

FAKE NEWS DETECTION USING MACHINE LEARNING APPROACH

BY

MD. ABDUS SALAM MOLLA

ID:173-15-10299

AND

RAUNAK AHMED

ID:183-15-12023

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

Supervised By

Mr. Md. Jueal Mia

Assistant Professor

Department of Computer Science and Engineering
Daffodil International University

Co-Supervised By

Mr. Aniruddha Rakshit

Assistant Professor

Department of CSE

Daffodil International University



DAFFODIL INTERNATIONAL UNIVERSITY

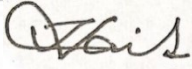
DHAKA, BANGLADESH

12 September, 2022

APPROVAL

This Project titled “Fake News Detection Using Machine Learning Approach”, submitted by Md. Abdus Salam Molla and Raunak Ahmed to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfilment 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 12 September, 2022.

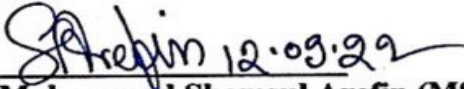
BOARD OF EXAMINERS



13.09.2022

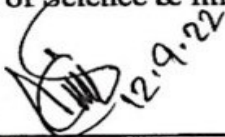
Chairman

Dr. Sheak Rashed Haider Noori
Associate Professor and Associate Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University




Dr. Mohammad Shamsul Arefin (MSA)
Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology

Internal Examiner



Md. Abbas Ali Khan (AAK)
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner



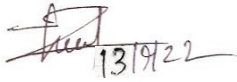
Dr. Md Sazzadur Rahman
Associate Professor
Institute of Information Technology
Jahangirnagar University

External Examiner

DECLARATION

We hereby declare that, this project has been done by us under the supervision of **Mr. Md. Jueal Mia, Assistant Professor, Dept. of CSE**, Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for the award of any degree or diploma.

Supervised by:

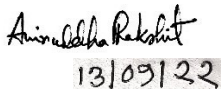


13/09/22

Mr. Md. Jueal Mia

Assistant Professor
Department of Computer Science and Engineering
Daffodil International University

Co-Supervised by:



13/09/22

Mr. Aniruddha Rakshit

Department of Computer Science and Engineering
Daffodil International University

Submitted by:

Salam

Md. Abdus Salam Molla

ID: 173-15-10299
Department of CSE
Daffodil International University

Raunak

Raunak Ahmed

ID: 183-15-12023
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First and foremost, I appreciate God for guiding me down the path to earning an honorable B.Sc. in Computer Science & Engineering from Daffodil International University. Then I'll always be grateful to my parents for their love and support.

I'd like to express my gratitude to **Supervisor Mr. Md. Jueal Mia, Assistant Professor**, Department of CSE Daffodil International University, Dhaka for providing me with the necessary guidance and advice in order to accomplish this fantastic research project on “Fake News Detection Using ML Approach “. His encouragement and guidance gave me the confidence I needed to accomplish my research assignment correctly. He provided me with all of the necessary materials and knowledge to begin this investigation from beginning. I'd want to express my gratitude to my coworkers for their assistance in shaping the dataset and other relevant duties.

I am also grateful to my **Co-Supervisor Mr. Aniruddha Rakshit, Assistant Professor**, Department of CSE Daffodil International University, Dhaka whose advice was a huge help in accomplishing the project's aim.

ABSTRACT

People are getting more exposed to fake news as their use of the web increases. to get notoriety while profiting from clickbait news outlets and social media, The media promotes incorrect information to deceive people. fascinating content about a current topic Though the spread of incorrect information has recently become more serious throughout the world, several existing techniques for categorizing and detecting have recently been created. There hasn't been a lot of research done on misleading news in an English news report. There was coverage of Bengali news. In this article, we look at Bengali forgeries. When classifying news, the South Asian context is considered. More than 200 million people use Bengali as their first language, and it is their way of life. Communication necessitates are fundamental in Bengali. Our main intention towards this research was to initiate an interpretation between the ML and DL assumptions. The machine learning classifiers which were used in this case were Random Forest, SVM, Decision Tree, XGB, Gradient boost classifier and Ada boost classifier. The best accuracy was achieved by the GB which was 89%. And afterwards we have used various well known deep learning approaches to conduct our second stage of experiments. Where we have used RNN, LSTM, Bi-LSTM, GRU, BERT. Then we comprehensively shown the models comparison to product the best evaluation result. There was also a comparative analysis to show the otherward works comparison. Which we think were beneficial in case of understanding the whole purpose behind our work. As Deep Learning methods were initiated our model with the base of RNN has achieved overall 94% accuracy. By which we propose this article in the elaborate discussion of our full procedure.

TABLE OF CONTENTS

CONTENTS	PAGE
Board of examiners	ii
Declaration	iii
Acknowledgments	iv
Abstract	v
CHAPTER	
CHAPTER 1: INTRODUCTION	1-3
1.1 Introduction	1-2
1.2 Motivation	2
1.4 Research questions	2
1.5 Expected output	3
1.6 Report layout	3
CHAPTER 2: BACKGROUND STUDIES	4-8
2.1 Introduction	4
2.2 Related work	4-7
2.3 Research summary	7
2.4 Scope of the problem	7
2.5 Challenges	7-8
CHAPTER 3: RESEARCH METHODOLOGY	9-15
3.1 Introduction	9-10
3.2 Classifier Algorithm	11-12
3.3 Research Subject and Instrumentation	12-13
3.4 Preprocessing	13-14
3.5 Stop Word Remove	15
3.6 Tokenization	15
3.7 Implementation requirements	15

CHAPTER 4: EXPERIMENTAL RESULTS AND DISCUSSION	16-22
4.1 Introduction	16
4.2 Model Performance	16-23
4.3 Summary	23
CHAPTER 5: IMPACT ON SOCIETY, ENVIRONMENT, AND SUSTAINABILITY	24-25
5.1 Impact on Society	24
5.2 Impact on Environment	24
5.3 Ethical Aspects	25
5.4 Sustainability	25
CHAPTER 6: CONCLUSION AND FUTURE WORK	26-27
6.1 Conclusion	26
6.2 Recommendations	26
6.3 Implication for further study	27
REFERENCES	28-30

LIST OF FIGURES

FIGURES	PAGE NO
Figure 1: Method ML Flow Chart	9
Figure 2: LSTM cell structure	10
Figure 3: Data Statistics	14
Figure 4: Data Preprocessing Method	14
Figure 5: ROC Curve model performance	17
Figure 6: Confusion matrixes of different Classifiers	19-20
Figure 7: ROC curves of ML Classifiers	20
Figure 8: Training and Validation Accuracy & Loss Of RNN	21
Figure 9: Predicted Label	21

LIST OF TABLES

TABLES	PAGE NO
Table 1.0: Highest accuracy evaluation	17
Table 1.1: The comparison between classifiers	18
Table 1.2: The comparison between Deep learning model	18
Table 1.3: Comparative Analysis (CA)	22

LIST OF ABBREVIATION

NLP	Natural Language Process
ML	Machine Learning
RF	Random Forest
SVM	Support vector Machine
DT	Decision Tree
XGB	XGBoost
GBC	Gradient Boost Classifier
ABC	Ada Boost Classifier
RNN	Recurrent Neural Network
LSTM	Long-short Term Memory
BI-LSTM	Bidirectional Long-short Term Memory
GRU	Gated Recurrent Unit
BERT	Bidirectional Encoder Representations from Transformers

CHAPTER 1

Introduction

1.1 Introduction

In recent years, false news identification has been a prominent issue among scholars all around the world. Various research reported bogus news detecting methods. Produced artificially false news is referred to as "Neural Fake News" and is produced using deep learning-based methods that were developed in response to threads of neural fake news that the Allen Center for AI had released [1]. To categorize Bengali news as true or false, we used data mining classification methods such as Naive Bayes, Random Forest, and Logistic regression in our system. Natural Language Processing (NLP) is used to mine text from a dataset. To construct a Bengali news dataset, a Bengali newspaper scrapper was created. An online URL checker is being developed to provide real-time categorization of legitimate and fraudulent news [3]. We employ Word Clouds of actual and false news data to visualize the outcome. We end our study with a maximum accuracy of 85 percent using the Random Forest Classifier [5]. In such cases we can see the concept being upheld in many other works which can elaborate with the consistency with such profound techniques [6]. Fake news is no longer just a buzzword; it has become one of the most serious democratic concerns [7]. Fake news may come from anybody or any source. It might be done for selfish or envious reasons, to inflict retribution, for political reasons, to eliminate rival businesses, to promote goods, or to obtain notoriety [8]. Real news takes longer to propagate than fake news. False news rapidly travels among 1000–100,000 people [9] but the truth seldom does so among thousands of people. The rise of the internet and social media has increased the spread of fake news. The impact of false news on American culture is negative. In the case of, just 42% of respondents believe local news is real, while only 27% believe national news is true. The propagation of purposeful false news during the 2016 US election caused tremendous damage. Fake news is becoming increasingly commercialized and globalized. The social media are designed to help people communicate and able to reach each other, but they have also become a means of causing havoc [10]. Fake news is known as one of the most widespread on Facebook, but it may also be found on internet interaction platform. [11], a research published in Science, examined 160,000 tweets tweeted by three million individuals between 2006 and 2017 and discovered that false news went viral more frequently than

accurate stories. A false narrative is likely to be spread 1500 times more than a factual tale, and humans may be more prejudiced towards disseminating fake news by design since it is more engaging and fascinating. According to [12], Open AI just released the GPT-2 language model, which can create language virtually like a human. This has tremendous ramifications for the spread of bogus news. If deep learning-based algorithms can write news like humans, the cost of making false news falls dramatically, and anyone can simply generate and circulate a limitless amount of fake news. As when it comes to the conception of linguistic data researches, we are now highly appreciate the newly condemned released Deep Learning approaches. As the world has grown efficient towards language interpretation and conceptual contradiction, we have adjusted towards the research skepticism to accept such models with open arms. The linguistic process of detection, segmentation and many other works have become much more popular and helpful towards the evolution of the computer science.

We offer related works in Section 2, datasets in Section 3, methods in Section 4, and data visualization in Section 5. The complete result analysis is reported in section 6, and the online interface of our system is demonstrated in section 7. The final section of the report also includes our conclusion and future plans for this project.

1.1 Motivation

Scientists have always been interested by the potential to construct and work on languages in order to understand and predict human behavior when it comes to Narrative study. This necessity can only be comprehended via the use of words. That is why I choose this study genre to focus on for future progress. Sentiment determination towards various words is a very natural and continuous method in terms of sentiment detection presently. By utilizing machine learning to distinguish between negative and positive evaluations, we can foresee a liberal approach for those terms. That is our primary motivation for performing this job.

1.2 Research Questions

- What precisely is virtual communities??
- What is Fake News Detection?
- What is NLP?
- Why NLP is important

1.3 Expected Output

The model, which comprises of several separate algorithms that must be trained and then assessed utilizing linguistic Bengali text data acquired from many social networks, was projected to be able to detect which attitudes are unfavorable and which are good. When working with text data, algorithms are frequently utilized. As we all know, dealing with text data in Bengali may be difficult for actual machine learning models; in this instance, the dataset was properly preprocessed so that the output does not vary depending on the criteria, which are the trash values.

1.4 Report Layout

Six sections make the report. Each chapter details many facets of it - the "**Fake News Detection**". Every chapter has different parts described in detail.

Chapter 1: Introduction

The inspiration is clarified and the proposition objective and introduction are presented.

Chapter 2: Background Studies

The applicable work is talked about and significant popular techniques are introduced corresponding related work.

Chapter 3: Research Methodology

Presents the information assortment, information pre-handling, and the element determination methodology.

Chapter 4: Design Specification

the philosophies for assessment grouping are clarified and the result discussed.

Chapter 5: Implementation

The 3-assessment plan, the precision assessment, and the investigation are introduced.

Chapter 6: Conclusion and Future Scope

The end is drawn and my commitments are portrayed.

CHAPTER 2

Background Studies

2.1 Introduction

In recent times, fake news has gained widespread attention. Creating media, stories, or pranks with the intent to deceive viewers in order to profit from their agreement or viewpoint on political, social, religious, or other issues is known as fake news. These articles may be financially rewarding for online news sources since they are typically created to either sway people's attitudes, forward a political goal, or trap people. False news periodically disseminated by websites that seem trustworthy or have the same identities or web links of very well news sources can fool people. In this work, we first built a longitudinal data and then built a false news detection approach using natural language analysis. Five machine learning classifiers were compared. On top of that, we applied the Extra Tree technique to identify key characteristics. The voting classifier outperforms the other methods when evaluated on the dataset using a cross-validation approach. We also ran a variety of analyses on the data patterns and the nature of the characteristics. The experimental data, as well as performance measurements and graphical and tabular representations of the results, are presented. It contains the conclusion and future work, which highlights the contributions of the intended study and suggests some potential next steps.

2.2 Related Work

By incorporating textual elements with auxiliary data, including user social interactions on social media, Shu et al. [12] improved accuracy with several models. The biopsychosocial theories and how to spot fake information online were also covered by the writers. Several methods analyze how false news is spread in comparison to legitimate news [13]. More precisely, the method examines the differences in how a factual article and a fraudulent one spread over a network. Theoretically, it is possible to categorize an article as authentic or fraudulent based on the responses it receives. Numerous research have mostly concentrated on identifying and categorizing bogus news on social networking sites like Facebook and Twitter [14]. In their research, Ahmed et al. [15] extracted lingual information from various blog posts, such like n-grams, and trained a diverse range of Machine learning techniques, which would include logistic

regression (LR), K-nearest neighbor (KNN), linear support vector machine (LSVM), stochastic gradient descent (SGD) of SVM as well as logistic regression attain greater accurateness (92 percent). It is also clear that modern businesses must contend with the difficulty of having their staff duped into divulging their credentials in addition to the possibility of malware being injected into its information management systems. According to some academics, a certain subset of material creators is increasingly employing clickbait to further its phishing goals [16]. Data are frequently seen as a precious commodity in the corporate world today, and as a result, it is crucial that organizations have the essential protections in place to prevent unwanted people from accessing confidential material. Nevertheless, the presence of content creators who are eager to spread false information renders such attempts useless. In light of this, businesses are spending more money to support the creation and development of better alternatives that can be applied to combat negative consequences of utilizing sensationalism to entice consumers to divulge their confidential info. However, despite being told it against, workers frequently go to these websites, putting respective companies on danger for internet [17]. To distinguish between false and true news, Horne and Adal [18] retrieve stylometry and psychological factors first from news titles. Bidirectional LSTM is used by Przybyla et al. [19] to construct a style-based text classifier that extracts style-based attributes from news stories. A neural network model is developed by Zellers et al. [20] to assess the accuracy of media from the news text. To extract the content-based characteristics from news writings, additional publications [21, 22] take into account lexicons, bag-of-words, TFIDF, and latent themes. An LSTM network-based false news detection method called DEFEND is proposed by Cui et al. [23]. The DEFEND takes into account comment threads to determine whether certain news is true or not. In order to learn the models of social situations, Nguyen et al [24] proposed false news detection technique FANG makes use of the graph learning architecture. The techniques covered above are recognized as gold standards in the study of false news. Instead of early false information identification, the country predominantly focuses on based on deep learning approaches. A few publications [25, 26] suggest early false news identification. Rubin et al. [27] was suggested to identify British news assessment. 360 satirical news articles from primarily 4 categories were examined and evaluated: civics, science, business, and delicate news. On the basis of their parody news research, they created an SVM classification model. The 2016 US surveys are the most noteworthy illustration of how fake news has spread swiftly during the previous 10 years [28]. The increasing spreading of false information online has

caused many problems, not just in politics but also in a wide range of other industries such sports, medicine, and academics [29]. In order to extract themes from the a media platforms corpus in 2016, a useful theme controller is designed on LDA was created [30]. A collection of 90,527 textual documents with an airline and airport management focus was employed in this investigation. Online platforms have surpassed print media as a major contributor of current affairs for the general public [31, 32]. Through these venues, people may now openly share their thoughts and leave comments. Just Facebook is the source of news for almost 35% of Americans [33]. Numerous individuals, groups, and sites on Facebook offer their opinions about current global events. The risk of creating fake news exists because of this freedom of speech. People have been exposed to false information since there is no verification of the veracity of news stories posted on Facebook, that has an effect on society [34]. The Twitter and Facebook also take the correct actions to combat false news in addition to machine learning techniques. Twitter and Facebook regularly delete users, webpages, and organizations that spread inaccurate info [35]. People can indeed be useful in identifying any types of misleading info they come across on social networking sites. Every posting, tweet, or remark which might seem deceptive to them can be reported. Tools and extensions have also been created to recognize and gather bogus news [36, 37]. Vosoughi et al. [38] take a novel technique to investigate the characteristics of news diffusion on social media; specifically, they explain how news (rumors) travel on Twitter and examine how false news varies from credible news in term of its Twitter dissemination. The paper explores the spread of disinformation online using a variety of analysis techniques, including depth, size, highest broadness, systemic popularity, actually imply broadening of correct and incorrect rumor spirals at different depths, quantity Twitter accounts attained at any deep, and time in mins for correct and incorrect rumor feedback loops to reach detail and amount of Twitter users. To conduct experiments, Iskandar et al. [39] gathered information from Fb, Instagram, and several blogging websites. They demonstrated how Nave Bayes is the ideal algorithm for their task since it deals with probability by critically analyzing several methods [40]. According to Johnston et al. [41], a machine - learning algorithm that can categorize Sunni-related misinformation. To identify bogus news in Bangla, Hossain et al. [42] create a dataset of 50 k occurrences. They have done a thorough analysis of both language and computer having to learn aspects. Utilizing SVM with such a linear kernel, a method proves how to find threatening and offensive Bengali terms in social media [43]. The model ran experiments on 5644 txt files, and it was able to predict words with a high precision

of 78%. To use a classification and boolean classification, Dinakar et al. [44] created a library of Comment sections for the purpose of identifying textual cyberbullying. Utilizing SVM, lexical, word and tf-idf characteristics, a new method for identifying hateful speech in Indonesian has been described [45]. a technique to identify offensive material and cyber abuse on Chinese social media. Their model, which combined LSTM with user-specific behavioral and character variables, had a 95 percent accuracy rate. Hammer [46] proposed a technique for identifying threats and acts of violence directed against minority groups in internet debates. This study took into account manually labeled sentences that included bigram properties of important words.

2.3 Research Summary

The study's information was acquired at randomly from multiple websites and platforms on the online. Extraneous data, numeric values, and special characters were removed from the dataset before to detection to produce a full and accurate detection result. The dataset included multiple repeat occurrences of various numbers and phrases, which were thoroughly evaluated and deleted for performance reasons. And we were hopeful to achieve the expected variable and outcome, which was very much likely to be helpful because of our research study.

2.4 Scope of the problem

Information, stories, or forgeries that are intended to mislead viewers in order to profit from their agreement or viewpoint on economic, social, religious, or other matters are referred to as fake news. These articles are frequently written to change people's opinions, forward a political goal, or place people in perilous circumstances, and they may bring in money for internet news outlets. On occasion, websites that have the same names and web links as reputable news sources may spread false information. In order to mislead readers and propagate false information through social media and the internet, deepfakes and channels employ their fake news material.

2.5 Challenges

We examined the slant assessment relying on voyager input in regards to carrier organizations in this investigation. According to our proposed method, both element determination and over-inspection procedures are equally relevant in terms of increasing our outcomes. Using highlight

selection methods, we were able to recover the best subset of highlights while reducing the number of calculations necessary to generate our classifiers. It has, however, decreased the lopsided appropriation of classes found in a large portion of our smaller datasets without causing overfitting. Our findings show that the proposed model has great grouping precision in anticipating events structure the two groups positive, negative data. Organizing Bengali text and processing it for model training was also a significant problem.

CHAPTER 3

Research Methodology

3.1 Introduction

We will go through the actions we followed to accomplish our research assignment in this chapter. We disguised many Python programming languages in order to conceal the complex machine learning methods. Because the dataset is the most important aspect of the Text Processing approach, the techniques were commonly chosen and placed on it. The fake news dataset is divided into two files, one containing actual news and the other bogus propaganda with comparable characteristics. Figure 1 depicts a flowchart method of viewing the procedure.

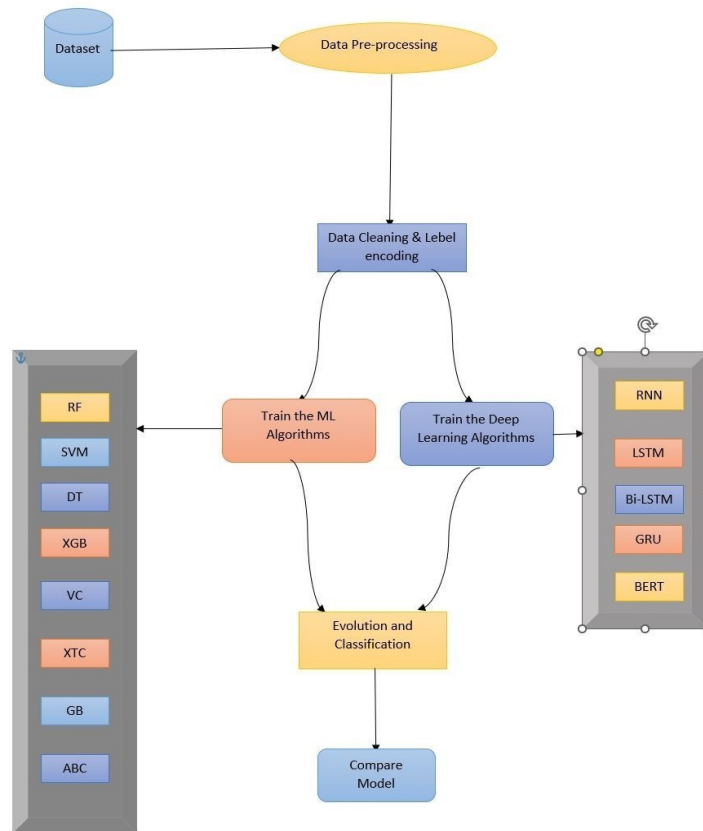


Figure 1. Method ML Flow Chart

And if we look at the deep learning approach, the identification of fake news is a binary classifier issue. By analyzing the bias that is inherent in authored news stories and examining the relationship

between both the title with body of the item, the suggested approach for spotting fake news evaluates the authenticity of the findings in the piece.

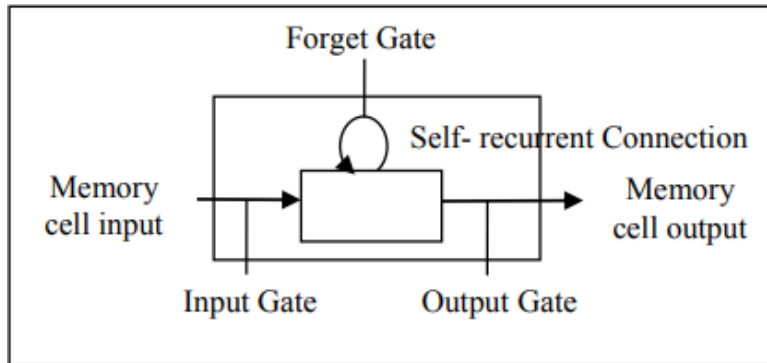


Fig 2. LSTM cell structure

The model is given the word embedding output. The ml model used in this application is a combination of iterative and incremental with a first layer that embeds values for vocabulary size, number of features, and sentence length. The next step is LSTM, which has 100 neurons per layer, followed by a dense layer with sigmoid function as we only need one output in the end. In order to prevent overfitting, we have employed binary cross variance to calculate loss, Adam algorithm for adaptive estimation, and ultimately adding a fall out layer in between. The model was then trained and tested. For both of the pre-assembled testing data sets, the result is predicted. If the predicted value is larger than 0.5 and less than 0.5, it is classified as 1 and as less as 0. $(TP+TN)/Total$ is a measure of accuracy. The following terminology was employed: True Negative (TN), which signifies that the test cases and prediction were in fact unfavorable; True Positive (TP), which denotes that the forecast and test instances were both accurate; False Negative (FN), which happens when a prediction is made but the results of the testing process are positive; When a prognosis is produced but the testing process turn out be negative, it is known as a false positive (FP). The architecture employed in LSTM is as follows: input consists of input shape, embedding vector features, and vocabulary size. Dropout value 0.2 to prevent overfitting, SpatialDropout1D layers using parameter 0.4, 256 LSTM units, a thick layer containing two neurons with soft max activation. And in case of Bi-LSTM the Dropout value was same, Bi directional LSTM units were 64. The RNN was comprehended on the similar consequences and was tuned according to the datasets preference.

3.1 The Classifier Algorithms

A preceding classification technique for literature and data gathering is called the Decision Tree Classifier (DTC). In several disciplines, DTC is a powerful tool for categorization. Making a tree which supported a attribute for classified data points is the main idea. The major problem with a DTC is that some features or attributes may be more appropriate for children than for parents. To remedy this drawback, a quantitative modeling was employed to feature selection inside the tree. For a training data set that consists of n negative and p positive values.

$$H\left(\frac{p}{n+p}, \frac{n}{n+p}\right) = -\frac{p}{n+p} \log_2 \frac{p}{n+p} - \frac{n}{n+p} \log_2 \frac{n}{n+p}$$

The training set E splits into the prefixes of "E1," "E2," "...," and "Ek," and you select K in the characteristic with the unique value. The anticipated entropy (EH) will continue to stay after the effort inside the attribute (branches $I = 1, 2, \dots, k$), including

$$EH(A) = \sum_{i=1}^K \frac{p_i + n_i}{p + n} H\left(\frac{p_i}{n_i + p_i}, \frac{n_i}{n_i + p_i}\right)$$

MNB is a variant of NB classifiers, which are refers generally built on the Bayes theorem. It takes into account how many times a team has practiced using the training dataset. For text categorization, MNB outperforms Bernoulli Naive Bayes (BNB) [49].

In the period of instruction, RF generates several random decision trees. It gathers the data out of each tree throughout projection and then delivers the outcome that most of the trees have correctly anticipated. The RF classifiers utilized in this writer's studies have a forest of 10 trees.

Famous machine learning (ML) algorithm Support Vector Machine produces an ideal hyperplane to classify fresh examples from testing phase. Various kernels are used by SVM classifiers to create the best hyperplane. Kernels are processes that look for patterns. The studies in this paper employ one of the SVM kernels.

There are two distinct node kinds in the decision tree: interior and exterior. While internal nodes hold the characteristics necessary for classification, outer nodes indicate the decision class. The clustering algorithm was assessed using a top-down method that divided homogenous data into subgroups. The uniformity of samples is defined by its entropy, which is computed using the equation.

$$E(S) = \sum_{i=1}^n p_i \log_2 p_i$$

Here, p_i is the likelihood that a sample will belong to the training course, while $E(S)$ stands for the graph's entropy. Entropy was utilized to assess the split's quality. Every characteristic was taken into account while deciding on the appropriate split for each node. Possible combination of the characteristics is controlled by randomized state 0.

The decision trees that make up the Random Forest (RF) each function independently. The most likely conclusion line is identified using the "Gini index" from each branch. This index was determined using equation.

$$Gini = 1 - \sum_{i=1}^c (P_i)^2$$

Here, c stands for the overall number of classes, while p_i denotes the likelihood of each class. Inside the forest, where the 'Gini' scale is used to assess the level of split, we utilized 100 trees. If there are at least two inner nodes as well as all 's extremely taken into account within every component, the network nodes are separated.

3.2 Research Subject and Instrumentation

The title I've picked is "Fake News Detection Using ML Approach." This is a critical topic of research in Natural Language Processing. So far, I have investigated the approach for conducting estimation research in Bangla using a specific and theoretical strategy. A superb learning model requires a powerful computer and a plethora of equipment. An example of a concept analyzer is shown below the main instrument for this model.

Hardware and Software:

- 2GB RAM and Intel core i7 7th generation 2.4ghz.
- 500 GB Hard Disk.

Tools:

- Windows 11
- Python 3.10
- Jupyter Notebook
- NLTK
- Pandas
- Numpy

3.3 Preprocessing

We wanted to refresh the entire dataset by deleting the trash character in order to prepare the data for the algorithms' train instance. What more characters are there except special characters ("!", "@", "#", "\$", "%", "&", "'", "*"), number characters (1, 2, 3, 4, 5, 6, 7, 8, 9, 0), white space, and duplicate characters? Because the character a particular was reconciled numerous times when adopting the data collection, the duplicated character may be kept. In order for a computer to discern between classes, the data must be raw.

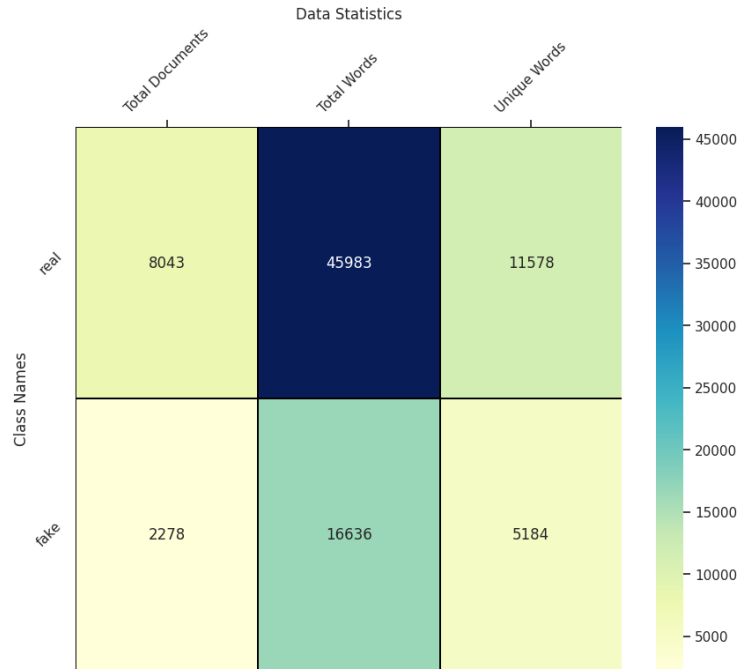


Fig 3. Data Statistics

As shown in the fig the dataset statistics was determined for processing the models capacitation.

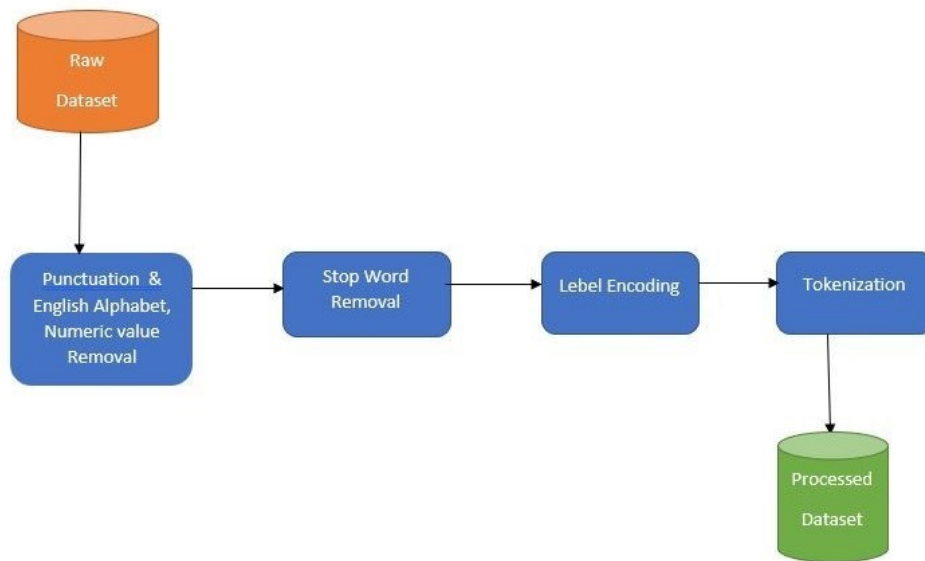


Fig 4. Data Preprocessing Method

3.4 Stop word remove

A stop word is a widely used phrase, such as ".", ",", "'", "|", and so on, that a web index has already been designed to disregard while sorting portions for viewing and retrieving those as the result of a pursuit query. I don't need these keywords to eat up valuable database space or time to process. I may efficiently evacuate individuals for this purpose by keeping a list of keywords that we consider are stop words. Python's NLTK (Nltk Toolkit) has a collection of stop words from 16 distinct dialects. They may be found in the nltk information search.

3.5 Tokenization

Tokenization is the process of isolating the arguments from the phrase and referring to these single arguments as tokens. In such approaches, tokenization is necessary for learning the algorithm's input. The data was divided into two groups for labeling: positive and negative.

3.6 Implementation Requirements

We preferred Python as a programming language to use the machine learning technique. The Panda library is used to load the data, while a NLTK package is used for preprocessing. The complete implementation is written in Python in a jupyter environment.

CHAPTER 4

Experimental Results and Discussion

4.1 Introduction

In the instance of our identification of negative and positive outcomes, the algorithms were able to achieve a very volatile and appreciating level of result. The methods were chosen and placed on our dataset frequency because, in Natural Language Processing, the dataset is the most important aspect of the entire operation. The algorithms we used were Linear Regression, Decision Tree, Random Forest, Multinomial Naive Bayes, KNN, Linear SVM, and RBF SVM. These are all advanced machine learning approaches that produced the desired outcome. And then for comparison purpose we have experimented with the deep learning approaches which were LSTM, Bi-LSTM, RNN, GRU, BERT methods. Then we have demonstrated a comparison perspective.

4.2 Model Performance

In the instance of our identification of negative and positive outcomes, the algorithms were able to achieve a very volatile and appreciating level of result. The methods were chosen and placed on our dataset frequency because, in Natural Language Processing, the dataset is the most important aspect of the entire operation. The algorithms we used were Linear Regression, Decision Tree, Random Forest, Multinomial Naive Bayes, KNN, Linear SVM, and RBF SVM. These are all advanced machine learning approaches that produced the desired outcome. And then for comparison purposes we have experimented with the deep learning approach. Here, we have demonstrated our result diagrams, in which we implemented our dataset in order to achieve the highest accuracy. The dataset was contemplated into training for 80/20 split. The test case contains in total 20% of the total dataset. With the dataset preference we figured some of the most re-categorized curved which would represent the performance evaluation.

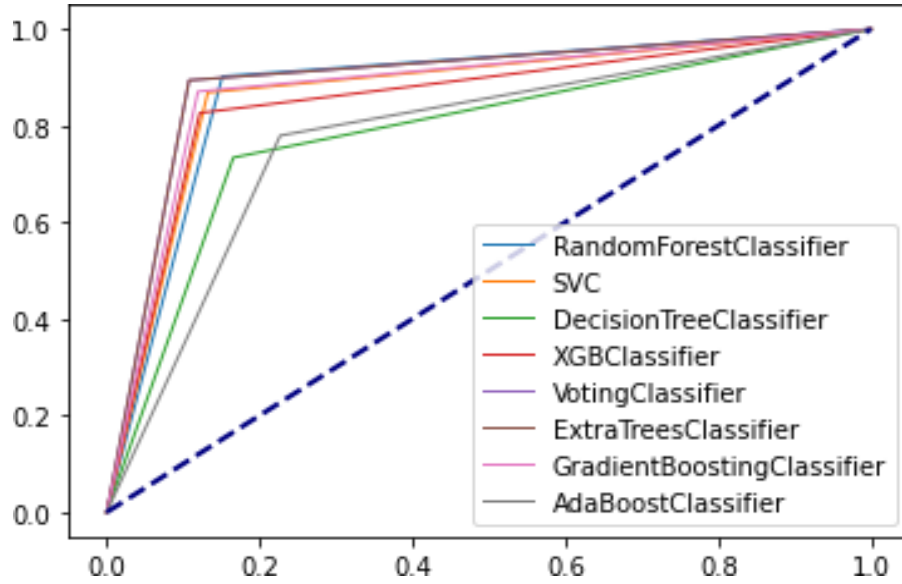


Fig 5. ROC Curve model performance

As different ML algorithms were used for the approach, we were able to acquire a quite fascinating result outcome. Each of the algorithms were working towards separating the real news from the fake news perspective.

Table 1.0 Highest accuracy evaluation

RF	SVM	DT	XGB	VC	XTC	GB	ABC
0.87	0.86	0.77	0.85	0.88	0.89	0.89	0.77

The accuracy evaluation is measured by evaluating a certain mathematical convolution measuring between the true negatives, true positive and the false negative, false positive. The term is given below,

$$Accuracy = \frac{TrueNegatives + TruePositive}{TruePositive + FalsePositive + TrueNegative + FalseNegative}$$

The model names were then divided while the precision, recall, F1-score and Accuracy wise, the divication was then recalled onto the table given below.

Table 1.1 The comparison between classifiers

Model Name	Precision	Recall	F1-score	Accuracy
RF	0.88	0.87	0.88	0.87
SVM	0.87	0.87	0.87	0.86
DT	0.78	0.78	0.77	0.77
XGB	0.88	0.82	0.85	0.85
VC	0.89	0.90	0.89	0.88
XTC	0.89	0.89	0.89	0.89
GB	0.88	0.88	0.88	0.89
ABC	0.78	0.78	0.78	0.77

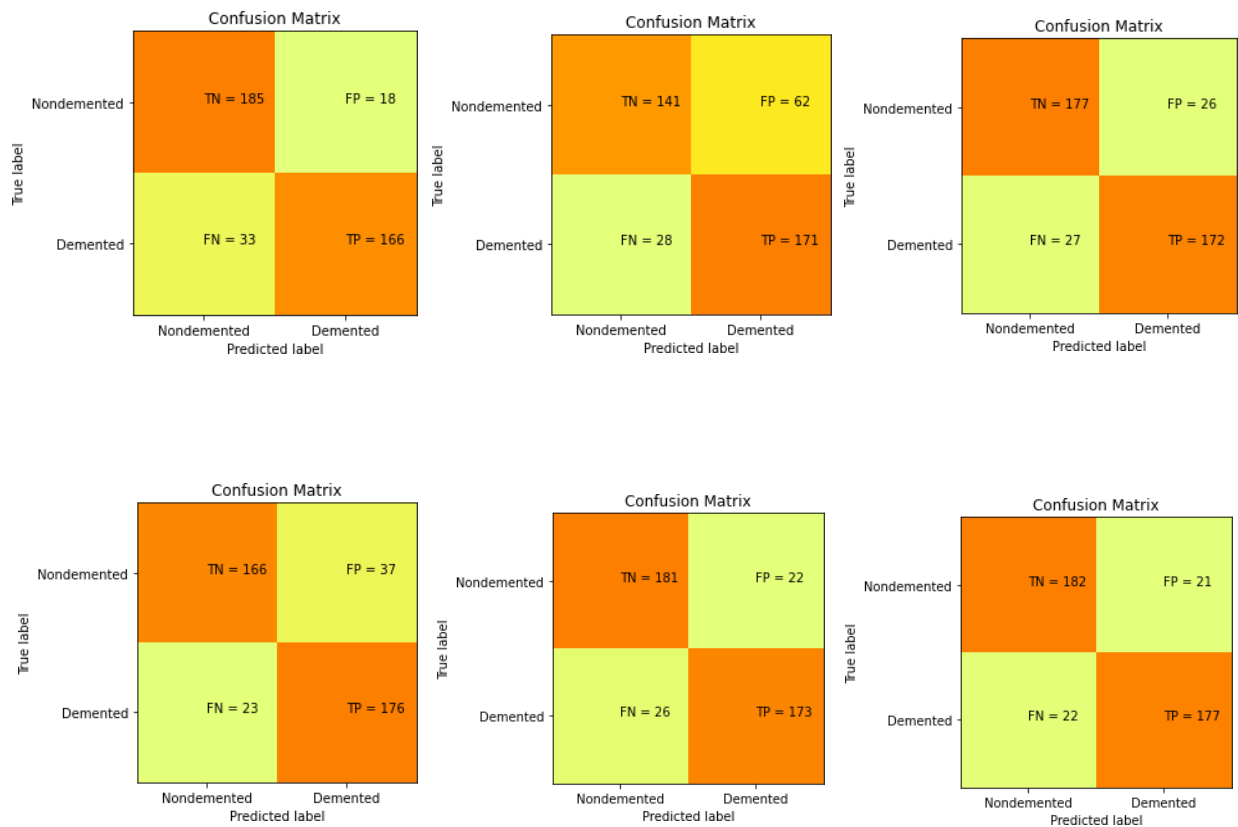
Table 1.2 The comparison between Deep learning model

Model	Real				Fake			
	Acc	P	R	f1	Acc	P	R	f1
RNN	0.94	0.96	0.96	0.96	0.94	0.86	0.85	0.85
LSTM	0.93	0.96	0.94	0.95	0.93	0.80	0.87	0.84
Bi-LSTM	0.94	0.96	0.95	0.96	0.94	0.83	0.87	0.85
GRU	0.92	0.96	0.94	0.95	0.92	0.78	0.85	0.81
Bert	0.95	0.96	0.98	0.97	0.95	0.85	0.72	0.78

As from the table 1.1 we can count that the GB and XTC were able to acquire the highest of the accuracy of 89%. As for the F1 score the XTC altogether with VC with the 0.89. In case of recall the VC were capable enough to acquire the 90. In the concern of precision VC and XTC both scored 89%. In overall performance VC and XTC were able to perform best in all average. In case of acquiring such result the reason behind it was in the dataset customization as the VC Dimension is a characterization algorithm, which can easily distinguish between such data. The client-server paradigm is used by XTC. Server are either used as data collection interfaces for monitoring by routinely asking lists of signals or as process control, i.e., establishing specific process signals.

Any client connecting over TCP can access XTC daemons, and messages are sent in ASCII format. Therefore, XTC must be viewed as highly vulnerable from a security perspective. The client-server paradigm is used by XTC. Servers are used either as data collection interfaces for monitoring by routinely asking lists of signals or as process control, i.e., establishing specific process signals. Any client connecting over TCP can access XTC daemons, and messages are sent in ASCII format. Therefore, XTC must be viewed as highly vulnerable from a security perspective.

The models' different confusion matrices were then assembled for a better understanding of the algorithm's performances.



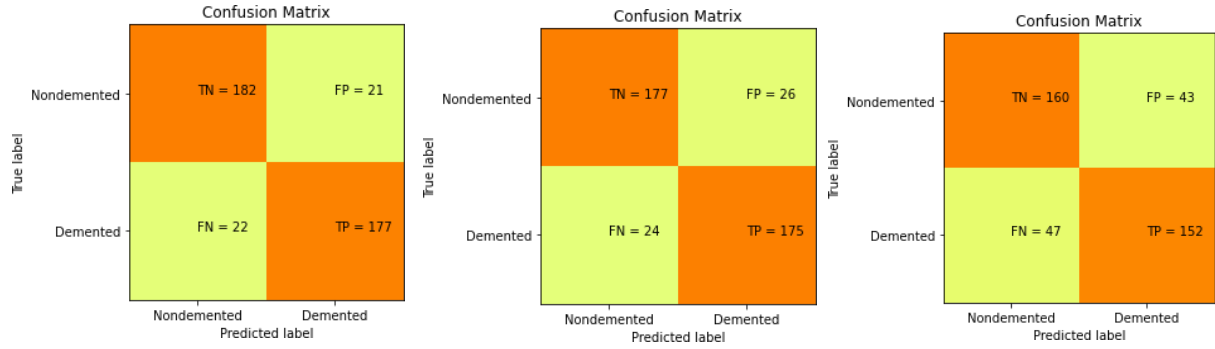


Fig 6. Confusion matrixes of different Classifiers.

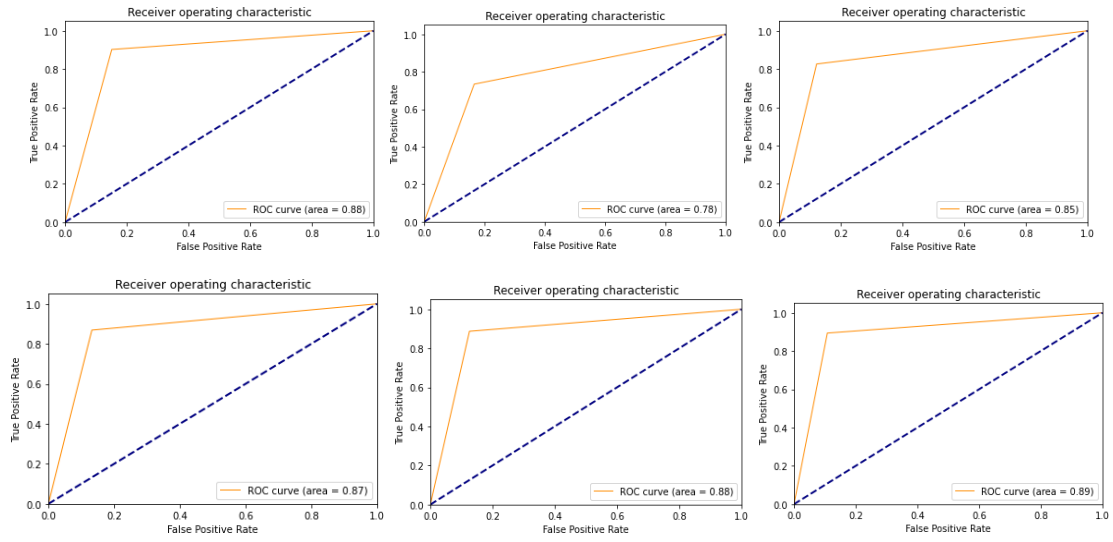


Fig 7. ROC curves of ML Classifiers

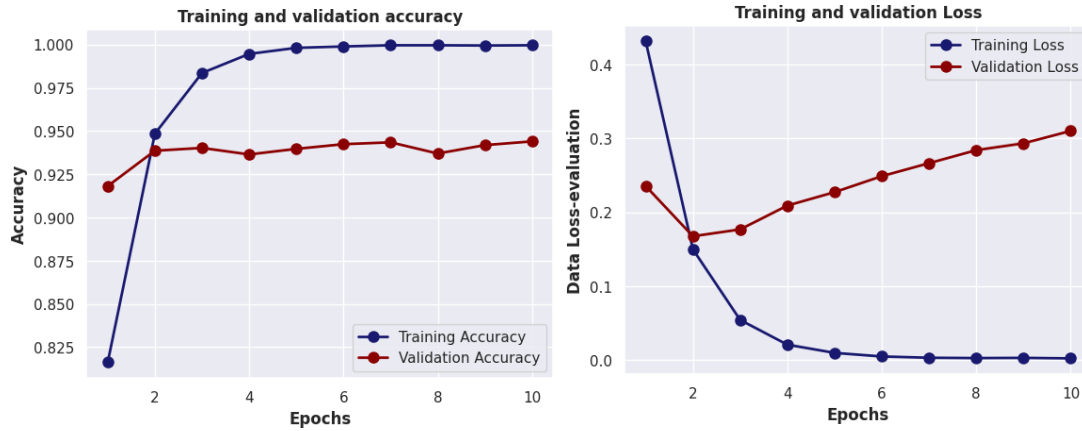


Fig 8. Training and Validation Accuracy & Loss Of RNN

In some cases, the validation accuracy drops below 85%, which is constant of the datasets multiclass maneuver. But as in our case we were working on binary classification. Which and why we got quite the acceptable training and validation accuracy,

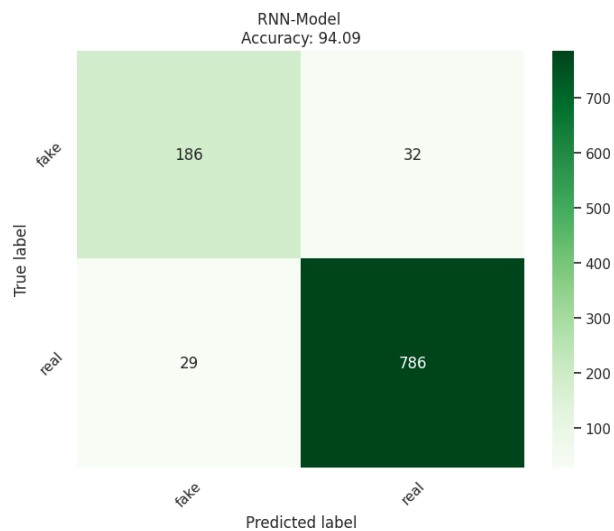


Fig 9. Predicted Label

Table 1.3 Comparative Analysis (CA)

Paper	Algorithm	Accuracy
[42] Hossain et al. worked a elaborate work on 50k dataset to detect fake news.	SVM	91%
[47] Tanvirul et al. worked to detect spam from Bengali text using.	MNB	82%
[48] In this research work Dense Neural Network was used to take out fake news.	DNN	85%
[50] In this paper they used a fifty thousand of data to analyze with a CNN-LSTM model.	CNN-LSTM	75%
[51] Shafayat et al. used a GNB method the evaluate the fake news detection.	GNB	87%
Our Proposed Model	RNN	94%

The RNN was able to get 94% of accuracy which is far more than the genuine machine learning algorithms. The main reason for such occurrences can be considered because of the dataset size. The neural network works as a net which filters the target class via layers. If the data size is greater then it can acquire more consistency. Which lacks more frequently the machine learning algorithms. These methods, however, have little chance of being widely used because they can't be adapted to other connections without a major reconfigure and a lot of labor. Instead, the primary driving force is to minimize such effort.

to employ learning-based techniques. a method that keeps track of protocol-specific sequences of thoroughly examined protocol communications employs separate Markov chains, much as we do,

to identify the same kinds of attacks which we are interested in. This finite-state machine-like model needs to be further tweaked using process- or p2p information, such the so-called "worth of relevance," to achieve an acceptable false-positive rate.

4.3 Summary

Many words were rendered incomprehensible once special characters were removed, as these were also responsible for assessment in the event of sentiment expression. In some circumstances, I needed to take a different approach by re-processing them into a more positive perspective. Such circumstances compels that the model we are proposing the RNN to be our proposed model because of its high quantified accuracy achievement. The analysis report shows that with 94% accuracy the RNN model is standing atop the other analysis methods.

CHAPTER 5

Impact on Society, Environment and Sustainability

5.1 Impact on Society

Each individual feeling today can be linked to the words we view on a daily basis on various digital platforms. In this instance, these platforms must have a framework in place to differentiate among real feelings and manufactured antagonism. As a result, we've chosen to focus our efforts on one of the most important genres of all time: false news. We intend to do this by bringing in a more distinct and diverse digital future. By which we can initiate the different concerns and rumors that impacts highly on negative basis on our contempered society.

5.2 Impact on Environment

The harm is caused by the massive distribution of bogus stories on social media. You should be more cautious if you are a person or company that delivers a lot of information, potentially with the help of life online management. It takes only seconds to increase your social media feeds. Take a look at the beginning of the story. And if you've never heard of it, search it up to see whether it's credible. If you don't have time, disregard it, especially if it appears to be satire, misleading information, or deliberate public relations. By not spreading false information, you may assist to minimize the spread of deception and fake news.

One of the key advantages of web-based living, as previously said, is the ability to offer information to a large number of people in a short period of time. Although this appears to be a great benefit in a crisis, perhaps it is a substantial burden because incorrect material may be disseminated in a split second. This can lead to major deception and terror. When news of the Queen's death spread, it was because queen having skipped a Yuletide banquet due to a normal sickness. This, along with other hoaxers spreading phony news, led many individuals to realize the Queen had dead.

5.3 Ethical Aspects

Web-based living is one of the best ways to meet and interact with new people who share your interests since it allows you to search for groups that are interested in your perks and pastimes. This is excellent for meeting new people, but it's also wonderful for love interests and web dating, which has grown more popular than traditional face-to-face encounters owing to online life and Tinder-like characteristics.

The internet is a fantastic tool for quickly sharing news throughout the world, with "breaking news" tweets getting thousands of retweets in minutes. This may be extremely useful for updating people about necessary news such as weather reports and missing peoples.

As previously mentioned, internet connection has had a good impact on society in a number of ways, all at little added charge because all critical web-based life phases are free. Evaluate another thing or institution that has ever revolutionized your life as a result of the internet, but then consider their cost.

5.4 Sustainability

- There are around 2.4 billion total dynamic web-based daily customers.
- 90% of major companies have at least dual web-based life cycles.
- In case they are unable to went into the online life catagories, 65 percent of individuals feel uneasy and uncomfortable.

CHAPTER 6

Conclusion and Future Work

6.1 Conclusion

We believe that this study will add to the production idea of BNLN's continuous development age in a huge world where emotional intelligence is becoming more associated with all of the turbulence between our everyday life structure and habits. In cases of recent decades of works where Identification of fake news has long been a source of controversy among experts in this field, and we hope that our study will encourage others. In the event of a disaster, we may imagine a future where science augments one's intelligence. In our research the deep learning approach was proven to acquire more reliable result and accuracy estimation than its predecessor ML models. We hope to go forward to achieve in more accurate and reliable result in further. In cases of overcoming limitations in cases natural language processing there are many points where researchers face the issue of dataset training, because of Bengali contexts are more difficult to train to machine, now in cases of such researchers are trying to overcome the possible problems to supplementary overcome this essences.

6.2 Recommendations

Experimental analysis is the most latest phenomenon in understanding public needs; it's a cheaper and more smart tool for studying how people feel about a particular problem and the brand influence of smaller blogging. In this example, we examined people's attitudes toward the aviation industry, as well as Airways' continuous troubles and how the public perceived them. The research confirmed our assumptions about how effective a Twitter assumption research approach is. The computation's Logistic Regression and Random Forest classifiers, together with two coding for better results, convincingly establish the mass group premise and, as a result, The airliner may easily evaluate the data to profit from this by striving to improve characteristics that are uncomfortable or disliked by the lead audience. This assignment has several choices, including:

6.3 Implication for Further Study

Because of the sensitive growth to info on the internet and in internet-based living areas, firms may employ conclusion examination to obtain insight into consumers' perspectives about their products or administrations. In contemporary literature, fake news investigation is typically carried out using a few months' worth of user data. Unless social media content is regularly read, this coating protects factually important and serious repercussions from being recognized. In order to deliver a genuinely massive client evaluation, a detailed review of tweets must fulfill a few criteria.

REFERENCES

- [1] J. Carson, (2020). “Fake news: What exactly is it – and how can you spot it?” vczx. Retrieved 4 January 2020, from <https://www.telegraph.co.uk/technology/0/fake-news-exactly-has-reallyhad-influence/>.
- [2] Why is fake news invented? - 30sec to check it out. (2020). Retrieved 4 January 2020, from <https://30secondes.org/en/module/why-is-fakenews-invented/>.
- [3] Research Guides: Fake News: How to Spot Misinformation: Introduction. (2020). Retrieved 4 January 2020, from https://guides.library.queensu.ca/fake_news. [
- [4] The globalization and commercialization of fake news | Diplomatic Council. (2020). Retrieved 4 January 2020, from <https://www.diplomatic-council.org/node/377>.
- [5] The spread of true and false news online. (2020). Retrieved 4 January 2020, from <https://centerforvaccineethicsandpolicy.net/2018/03/11/thespread-of-true-and-false-news-online/>.
- [6] Better Language Models and Their Implications. (2020). Retrieved 4 January 2020, from <https://openai.com/blog/better-language-models/>.
- [7] Grover - A State-of-the-Art Defense against Neural Fake News. (2020). Retrieved 4 January 2020, from <https://grover.allenai.org/>.
- [8] A.P.S. Bali, M. Fernandes, S. Choubey, M. Goel, (2019). “Comparative Performance of Machine Learning Algorithms for Fake News Detection”. In: Singh M., Gupta P., Tyagi V., Flusser J., Ören T., Kashyap R. (eds) Advances in Computing and Data Sciences. ICACDS 2019. Communications in Computer and Information Science, vol 1046. Springer, Singapore.
- [9] H. Ahmed, I. Traore, & S. Saad, (2017). “Detecting opinion spams and fake news using text classification”. Security And Privacy, 1(1), e9. doi: 10.1002/spy2.9.
- [10] J. Norregaard, B. Horne, & S. Adali, (2020). NELA-GT-2018: “A Large Multi-Labelled News Dataset for The Study of Misinformation in News Articles”. Retrieved 4 January 2020, from <https://arxiv.org/abs/1904.01546>.
- [11] K. Shu, A. Sliva, S. Wang, J. Tang, & H. Liu, (2017). “Fake News Detection on Social Media”. ACM SIGKDD Explorations Newsletter, 19(1), 22-36. doi: 10.1145/3137597.3137600.
- [12] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, “Fake news detection on social media,” ACM SIGKDD Explorations Newsletter, vol. 19, no. 1, pp. 22–36, 2017.
- [13] S. Vosoughi, D. Roy, and S. Aral, “The spread of true and false news online,” Science, vol. 359, no. 6380, pp. 1146–1151, 2018.
- [14] H. Allcott and M. Gentzkow, “Social media and fake news in the 2016 election,” Journal of Economic Perspectives, vol. 31, no. 2, pp. 211–236, 2017.
- [15] H. Ahmed, I. Traore, and S. Saad, “Detection of online fake news using n-gram analysis and machine learning techniques,” in Proceedings of the International Conference on Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments, pp. 127–138, Springer, Vancouver, Canada, 2017.
- [17] Nah, F.F.H., 2015. Fake-website detection tools : Identifying elements that promote individuals use and enhance their performance 1 introduction.
- [18] Aldwairi, M., Hasan, M., Balbahaith, Z., 2017b. Detection of drive-by download attacks using machine learning approach. Int. J. Inf. Sec. Priv. 11, 16–28. URL: <https://doi.org/10.4018/IJISP.2017100102>, doi:10.4018/IJISP.2017100102.

- [19] Horne, B., Adali, S.: This just in: Fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news. In: Proceedings of the International AAAI Conference on Web and Social Media (2017)
- [20] Przybyla, P.: Capturing the style of fake news. Proc. AAAI Conf. Artif. Intell. 34, 490–497 (2020). <https://doi.org/10.1609/aaai.v34i01.5386>
- [21] Silva, R.M., Santos, R.L.S., Almeida, T.A., Pardo, T.A.S.: Towards automatically filtering fake news in Portuguese. Expert Syst. Appl. 146, 113199 (2020). <https://doi.org/10.1016/j.eswa.2020.113199>
- [22] Zhou, X., Wu, J., Zafarani, R.: SAFE: similarity-aware multi-modal fake news detection. Adv. Knowl. Discov. Data Min. 12085, 354 (2020)
- [23] Zellers, R., Holtzman, A., Rashkin, H., Bisk, Y., Farhadi, A., Roesner, F., Choi, Y.: Defending against neural fake news. Neurips (2020)
- [24] Nguyen, V.H., Sugiyama, K., Nakov, P., Kan, M.Y.: FANG: leveraging social context for fake news detection using graph representation. Int. Conf. Inf. Knowl. Manag. Proc. (2020). <https://doi.org/10.1145/3340531.3412046>
- [25] Shu, K., Wang, S., Liu, H.: Beyond news contents: The role of social context for fake news detection. In: WSDM 2019—Proceedings of 12th ACM International Conference on Web Search Data Mining, vol. 9, pp. 312–320 (2019). <https://doi.org/10.1145/3289600.3290994>
- [26] Liu, Y., Wu, Y.F.B.: FNED: a deep network for fake news early detection on social media. ACM Trans. Inf. Syst. (2020). <https://doi.org/10.1145/3386253>
- [27] Rubin, Victoria, L., et al.: Fake news or truth? Using satirical cues to detect potentially misleading news. In: Proceedings of NAACL-HLT (2016)
- [28] Ashraf Abdul, Jo Vermeulen, Danding Wang, Brian Y Lim, and Mohan Kankanhalli. 2018. Trends and trajectories for explainable, accountable and intelligible systems: An hci research agenda. In Proceedings of the 2018 CHI conference on human factors in computing systems. 1–18
- [29] [n.d.]. Coronavirus update: Health worker arrested for spreading fake news. ([n. d.]). <https://www.businesstoday.in/latest/trends/story/coronavirusupdate-health-worker-arrested-for-spreading-fake-news-252182-2020-03-16>
- [30] Vala Ali Rohani, Shahid Shayaa, and Ghazaleh Babanejaddehaki. 2016. Topic modeling for social media content: A practical approach. In 2016 3rd International Conference on Computer and Information Sciences (ICCOINS). IEEE, 397–402.
- [31] Alfred Hermida. 2016. Social media and the news. The SAGE handbook of digital journalism (2016), 81–94.
- [32] Rasmus Kleis Nielsen and Kim Christian Schrøder. 2014. The relative importance of social media for accessing, finding, and engaging with news: An eightcountry cross-media comparison. Digital journalism 2, 4 (2014), 472–489.
- [33] Elisa Shearer and Amy Mitchell. 2021. News use across social media platforms in 2020. (2021).
- [34] Margaret Van Heekeren. 2020. The curative effect of social media on fake news: A historical re-evaluation. Journalism Studies 21, 3 (2020), 306–318.
- [35] Yoel Roth and Del Harvey. 2018. How Twitter is fighting spam and malicious automation. Twitter [blog], June 26 (2018).
- [36] Agata Giełczyk, Rafał Wawrzyniak, and Michał Choraś. 2019. Evaluation of the existing tools for fake news detection. In IFIP International Conference on Computer Information Systems and Industrial Management. Springer, 144–151.

- [37] Kai Shu, Deepak Mahudeswaran, and Huan Liu. 2019. FakeNewsTracker: a tool for fake news collection, detection, and visualization. *Computational and Mathematical Organization Theory* 25, 1 (2019), 60–71.
- [38] H. Ahmed, I. Traore, and S. Saad, “Detection of online fake news using n-gram analysis and machine learning techniques,” in *Proceedings of the International Conference on Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments*, pp. 127–138, Springer, Vancouver, Canada, 2017.
- [39] Iskandar, B. Terrorism detection based on sentiment analysis using machine learning. *J. Eng. Appl. Sci.* 2017, 12, 691–698
- [40] Sarker, I.H. A machine learning based robust prediction model for real-life mobile phone data. *Internet Things* 2019, 5, 180–193. [CrossRef]
- [41] Johnston, A.H.; Weiss, G.M. Identifying Sunni extremist propaganda with deep learning. In *Proceedings of the 2017 IEEE Symposium Series on Computational Intelligence (SSCI)*, Honolulu, HI, USA, 27 November–1 December 2017.
- [42] Hossain, M.Z.; Rahman, M.A.; Islam, M.S.; Kar, S. BanFakeNews: A Dataset for Detecting Fake News in Bangla. arXiv 2020, arXiv:2004.08789.
- [43] Chakraborty, P.; Seddiqui, M.H. Threat and Abusive Language Detection on Social Media in Bengali Language. In *Proceedings of the 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT)*, Dhaka, Bangladesh, 3–5 May 2019.
- [44] Dinakar, K.; Reichart, R.; Lieberman, H. Modeling the detection of textual cyberbullying. In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, Barcelona, Catalonia, Spain, 17–21 July 2011
- [45] Aulia, N.; Budi, I. Hate Speech Detection on Indonesian Long Text Documents Using Machine Learning Approach. In *Proceedings of the 2019 5th International Conference on Computing and Artificial Intelligence*, Bali, Indonesia, 19–22 April 2019.
- [46] Hammer, H.L. Detecting threats of violence in online discussions using bigrams of important words. In *Proceedings of the 2014 IEEE Joint Intelligence and Security Informatics Conference*, The Hague, The Netherlands, 24–26 September 2014.
- [47] Islam, T., S. Latif, and N. Ahmed. Using social networks to detect malicious bangla text content. in 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT). 2019. IEEE.
- [48] Hossain, M.R. and M.M. Hoque, Automatic bengali document categorization based on deep convolution nets, in *Emerging Research in Computing, Information, Communication and Applications*. 2019, Springer. p. 513-525.
- [49] A. McCallum and K. Nigam, —A Comparison of Event Models for Naive Bayes Text Classification,| *AAAI/ICML-98 Work. Learn. Text Categ.*, pp. 41–48, 1998.
- [50] George, M. Z. H., Hossain, N., Bhuiyan, M. R., Masum, A. K. M., & Abujar, S. (2021, July). Bangla Fake News Detection Based On Multichannel Combined CNN-LSTM. In *2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT)* (pp. 1-5). IEEE.
- [51] Mugdha, S. B. S., Ferdous, S. M., & Fahmin, A. (2020, December). Evaluating machine learning algorithms for bengali fake news detection. In *2020 23rd International Conference on Computer and Information Technology (ICCIT)* (pp. 1-6). IEEE.

FAKE NEWS DETECTION USING ML APPROACH

ORIGINALITY REPORT

14 %
SIMILARITY INDEX

12 %
INTERNET SOURCES

7 %
PUBLICATIONS

7 %
STUDENT PAPERS

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	4 %
2	Submitted to Daffodil International University Student Paper	4 %
3	Farzana Islam, Mohammad Minhazul Alam, S. M. Shahadat Hossain, Abdul Motaleb, Sabrina Yeasmin, Mehedi Hasan, Rashedur M. Rahman. "Bengali Fake News Detection", 2020 IEEE 10th International Conference on Intelligent Systems (IS), 2020 Publication	2 %
4	"Sentimental Analysis and Deep Learning", Springer Science and Business Media LLC, 2022 Publication	1 %
5	arxiv.org Internet Source	<1 %
6	aclanthology.org Internet Source	<1 %

7	Nourin Islam, Nasrin Akter, Abdus Sattar. "Sentiment Analysis on Food Review using Machine Learning Approach", 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS), 2021 Publication	<1%
8	ebin.pub Internet Source	<1%
9	www.mdpi.com Internet Source	<1%
10	Submitted to South Bank University Student Paper	<1%
11	Submitted to Midlands State University Student Paper	<1%
12	Submitted to Kingston University Student Paper	<1%
13	researchspace.csir.co.za Internet Source	<1%
14	tud.qucosa.de Internet Source	<1%
15	www.irjmets.com Internet Source	<1%
16	mohdsaifuzzaman.com Internet Source	<1%

17	Internet Source	link.springer.com	<1%
18	Internet Source	m.scirp.org	<1%
19	Publication	"Disinformation, Misinformation, and Fake News in Social Media", Springer Science and Business Media LLC, 2020	<1%
20	Internet Source	norma.ncirl.ie	<1%
21	Internet Source	oapub.org	<1%
22	Internet Source	staging-medinform.jmir.org	<1%
23	Internet Source	www.ijraset.com	<1%
24	Publication	"Computing Science, Communication and Security", Springer Science and Business Media LLC, 2020	<1%
25	Publication	M. F. Mridha, Ashfia Jannat Keya, Md. Abdul Hamid, Muhammad Mostafa Monowar, Md. Saifur Rahman. "A Comprehensive Review on Fake News Detection with Deep Learning", IEEE Access, 2021	<1%

Exclude quotes

Off Exclude bibliography

On