BANGLA FAKE NEWS DETECTION USING MACHINE LEARNING ALGORITHMS

BY

DHIMAN SARKER
ID: 173-15-10396
AND

ZAHID HOSSAIN
ID: 173-15-10397

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

Supervised By

Narayan Ranjan Chakraborty
Assistant Professor
Department of CSE
Daffodil International University

Co-Supervised By

Raja Tariqul Hasan Tusher
Senior lecturer
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

DECEMBER 2021

APPROVAL

This Project titled “**Bangle Fake News Detection using machine learning algorithms**”, submitted by “**DHIMAN SARKER**” and “**Zahid Hossain**” 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 4th January 2022.

BOARD OF EXAMINERS

Chairman

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

Sadik

Internal Examiner

Md. Sadekur Rahman (SR)
Assistant Professor
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner

Afsara Tasneem Misha (ATM)
Lecturer
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Imran

External Examiner

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

DECLARATION

We hereby declare that, this project has been done by us under the supervision of **Narayan Ranjan Chakraborty, Assistant Professor, 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:

____

Narayan Ranjan Chakraborty
Assistant Professor
Department of CSE
Daffodil International University

Co-Supervised by:

____

Raja Tariqul Hasan Tusher
Senior lecturer
Department of CSE
Daffodil International University

Submitted by:

Dhiman Sarker
____

DHIMAN SARKER
ID: 173-15-10396
Department of CSE
Daffodil International University

Zahid
____

Zahid Hossain
ID: 173-15-10397
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

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

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

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

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

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

ABSTRACT

Advance modern technology and the growth of digitized content (using social media) manipulation users and turn to fake news, and its impact is deadly. Yellow news coverage is the cause of another big problem nowadays. Gain popularity and have profited through clickbait news publisher and social media-based news system circulated unauthorized news everywhere. The intention is to control religious, political, monetary, and other genuine things utilizing this simply get to strategy. The biggest concern is that it makes an annoyance and spreads savagery, even wage wars. Common people are not able to differentiate between fake and real news. The nature of fake news makes people suspects genuine news. Advance to use of NLP; it has ended up interesting to look for knowledge or designs within the era of fake news and thus find better prescient ways to discover fake news to categorize it from genuine news. In this paper, we propose an ML-based fake news detection strategy within the Bengali language. The proposed method uses a dataset on a LR, DT, RF, MNB, KNN, SVM algorithms. The calculation combination of TF-IDF-based content features (Unigram, Bigram, Trigram) and vectorizing to include selection. The accuracy of our proposed model is 92.13 on the Multinomial Naive Bayes algorithm, which is the highest accuracy from other algorithms. In expansion to this, we have performed a comprehensive analysis of different machine learning algorithms. At the same time, we have completed a comprehensive study where we have conducted a literature review with some questions related to fake news, which has helped us acquire the knowledge required for this research.

TABLE OF CONTENTS

CONTENTS **PAGE**

CHAPTER

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1
INTRODUCTION

1.1 Introduction

Now we are living in the time of modern information and technology. Everyone can be part of the online in a few seconds now. And it makes our daily life simple than before. For the most detail, people connected with the internet for the refreshment of the intellect as entertainment. It is also called the era of the internet-based smartphone since many people use it in their everyday lives. These days, it is very incomprehensible to think of a day without the internet and smartphone. In short, we can say that it gets to be an imperative portion of our life.

In this age of internet-based communication era, another popular term is Social Media (SM). People use this SM platform like Twitter, WeChat, WhatsApp, and Facebook to establish communication with their friends, check notifications, and upload (i.e., image, video, audio) content in their daily lives. These days, those platforms are also utilized for distributing news to their user. As the number of internet users is persistently expanding, the number of social media clients to boot amplified. As a result, the amount of information within the world became huge, called “Big Data” [1].

Now people love to get news from different SM platforms because of easy to get. For the vast number of users in the SM, the platform provided various types of information, but unfortunately, there is no process to check the validity of the data. So simple, when users get the information from SM platforms, they start to believe it, and the info makes them biased. On the other hand, authentic mass media as newspapers, magazines, journals, etc., in written form, and TV, Broadcast, Telecast, etc., become less powerful than SM platforms because of the fast and easy to access strategy. Particularly the young aged mostly depend on the Internet and Internet-based social media as the basic source for news since of their easy access, low cost, and 24/7 availability at the same time. The maturity level also very is low, so it is basic to inclination and makes them a portion of any fake effect.

According to analytics, the high spreading amount of social networking has become another concerning matter of fake news. Those social media become a section spreading fake news to a large amount of population. That fake news can diverse them in their own need. These sorts of news impact individuals within the most one-sided segments as open security, legislative issues, product survey, etc.[1]. Form 2016, the US election, the term 'Fake News' has become a common word worldwide. It has become the word of 2017 for using the term increasing rate by 365% from average time by Collin [62]. After that time, researchers, large companies, governments, and different organizations are more concerned about the word fake news. They start to combat this practice separately or collaborate, but their focus is to prevent it [2]. Another common example in our country 2012 Ramu incident where almost 26 thousand people attended to Buddhist Temple whose all are manipulated by some fake news. The real news is less quick at that point the fake news where comparing truth rarely extending among thousands of people, but fake news expanding between 1000-100,000 people groups within a particular time zone [3].

Different social network platforms have started to use a flag system to alarm users about any post. Fact-checking is a process of checking news is fake or real. Different large companies create these types of fact-checking websites to prevent this problem. There several people analysis the data survey them. After that, they record them. But there has an issue with quality assurance. Since humans do the survey, if anybody is biased, then the result also be one-sided. Solving this problem of biased intention Artificial Intelligence (A.I) has been a play an essential role against this Fake news. Intelligent systems work with automatic processes, so the rate of biased results is quite impossible—Generally, the Machine learning algorithm works with datasets. They analyze the dataset, discover different patterns from the datasets, correlate them, and predict them. Using those pattern machines can predict future data and make vital decisions depending on related data. However, the device can’t understand the human understand languages, so they must convert the data and make them numerical form using relevant extraction processes.

This extraordinary means of communication, Fake News, is easily spread worldwide, which quickly affects the users living in the society by mixing some wrong information with the correct information worldwide. News can usually be either true or false, but when in a news story, both truth and falsehood are revealed. Then it becomes difficult for the general newsreader or expert to find out the accuracy of that News properly. And to solve this problem, there is a need to create some tools that will help remove this fake information from the News and confirm the authenticity using knowledge-based prediction. In this context, review presents a Systematic Literature Review (S.L.R) whose objective was to know about Fake News, using different paper or articles from repositories.

After gathering knowledge from the RQ, we will learn about the changing of fake news and the creative perspective of contaminated news. Machine learning is necessary for false news detection and contains the personal benefits of spreading fake news in media. From the review, we get the highest used algorithms for detecting is Long Short-Term Memory (14.51%), Proposed framework (11.29%), Conventional Neural Network (9.67%) with most used feature extraction as 20.96% TF-IDF. 17.74% Word2Vector in different types of datasets between selected 62 articles from different repositories.

Different types of methods used fake news detection in social media and other sources. But there is a tremendous lack of advance within the Bengali language community and a gigantic gap since the syntactic structure. We try to find some processes and analysis Bangla text and detecting fake one. We work with datasets and then apply a natural language processing toolkit to develop a method for fake news detection. Our proposed strategy uses TF-IDF-based content features also utilizes vectorizing our content for way better exactness. Compare six machine learning classifiers. Among the different algorithms tested on the dataset, the Multinomial naive Bayes algorithm classifier outperforms the algorithm. And the accuracy of the model is 92.13% which is the best fit from other algorithms.

1.2 Motivation

Today we use online services almost directly or indirectly in almost every moment of our daily lives. As a result, we are now increasingly dependent on online. However, we rarely verify that the information online provides us is true or false. Various surveys prove that fake news spreads six times faster than true news. This problem has recently manifested in our country through various incidents. (A woman killed by the public as a child kidnapped or Padma bridge needs more heads). But fake news is no longer just a problem of any country, but it has become a global problem.

1.3 Rationale of the Study

Preventing this problem different social network platforms have started to use a flag system to alarm users about any post. Fact-checking is a process of checking news is fake or real. Different large companies create these types of fact-checking websites to prevent this problem. There several people analysis the data survey them. After that, they record them. But there has an issue with quality assurance. Since humans do the survey, if anybody is biased, then the result also be one-sided. Solving this problem of biased intention Artificial Intelligence (A.I) has been a play an essential role against this Fake news. Intelligent systems work with automatic processes, so the rate of biased results is quite impossible—Generally, the Machine learning algorithm works with datasets. They analyze the dataset, discover different patterns from the datasets, correlate them, and predict them. Using those pattern machines can predict future data and make vital decisions depending on related data. However, the device can’t understand the human understand languages, so they must convert the data and make them numerical form using relevant extraction processes. So, at the end of this research study, we can take a full idea about the impact, spread, and prevention of fake news.

1.4 Research Questions

A research question is embedded to give structure with the overview of the presenting area, set main aspects, and studies of a specific topic. In this study, some questions are set up which helps us know deeply about fake news, learn how news becomes false news, try to get an idea about the role of mass media, and the importance of detecting fake news. Together we will find out what kind of algorithm has been used for Fake News detection before to now, try to give a good idea of what algorithms and methods are being used the most with the help of accurate information and evidence. Different research methods can use the researcher, such as case studies, concepts, controlled experiments, and practical applications.

1.4.1 Question Standard

Case studies may inquire about a methodology that includes a strategy that envelops everything approaches information collection and investigation. This method is generally used in large and complex research where no work is done beyond the original data. Instead, the process starts from the root event, and various facts and evidence are gathered depending on the raw event using multiple sources [61].

Control Experiment is a type of experimental study where researchers control their central perspective. At the same time, they can work with independent variables if they need, and their main motive is gathering knowledge, ensuring main logic, and finally evaluating the predicted models. It presents a solution that reconciles with conjecture and reality [59].

Proof of Concept works with a practical model that can show a proven view of the theoretical knowledge established by the researchers or article. Therefore, it can be considered a summarized or not completed method or proposed idea, which is created to confirm that theory or hypothesis is an acceptable thesis in address is vulnerable to being abused helpfully. This method usually works with some approaches that have been proposed, but there are doubts about its integrity, and the primary goal of this method is to remove these doubts and make them usable. And the final method research is practical application where execution of any idea without any kind of scientific or empirical approval. In short, we can say that the proof concept, with zero percent of evaluation part.

1.4.2 PICO Model

The PICO (Population, Intervention, Control, and Outcomes) model was used by James, to guarantee research question quality. The main objective of the model is to demonstrate the capacity to characterization and classification the research question quality; at the same time, it's possible to evaluate the effects between a given population and intervention. A perfect control environment with pre-mapped articles can make finding a good outcome of any research question and searching terms easy [4]. The process is shown in Table-1.1, where intervention can help get a proper output from the population using control methods.

TABLE 1.1: PICO Model for Research Question Quality

Class	Description
Population	Fake news analysis using publications by the developers and analysts
Intervention	Automatic Fake News detecting in context using intelligent techniques, algorithms, framework and approaches.
Control	Automatic Fake News detecting system using intelligence. Search String base articles. Papers from the application that admit the intervention: * TI-CNN Convolutional neural networks for fake news detection [5] * MYTHYA Fake News Detector, Real Time News Extractor and Classifier[6]. * Detecting Fake News using Machine Learning and Deep Learning Algorithms [7] Papers from the application that not match with the intervention: * Fake News in Digital Media [8] * Fake news and online disinformation: a perspective of Thai government officials. [9] * History of Fake News [10]
Result	Prediction and detection Fake News in automatic process.

1.4.3 Selected Research Question

TABLE 1.2: Research Question (RQ)

	Description
RQ1	How the evolution of fake news came out?
RQ2	Why automated Fake News Detection is needed?
RQ3	What is the correlation between mass media and fake news?
RQ4	Which type of Fake News Detector was used in past?
RQ5	What is the current situation to identify fake news?

1.5 Expected Outcome

In this work, we will try to find the answer to those questions and discuss them with some evidence, example, pros, cons, and so on. In Table-1.2, research questions are shown. Question 4 and 5 characterize the algorithms, process, framework, and so on. Question 4 gives a brief about what kind of work has been done on this topic and what type of algorithm is used in the present time discussed in Question 5.

Moreover, various issues, including data sets, feature extraction, will be discussed briefly in this portion. Since it is incrementally, it will be possible to use the information in Question 5, permit surveying the development of the current research position, appearing on the off chance that there's the requirement to increase the utilization of the logical strategy. In this range, with replications of ponders that will permit to assessing on the off chance that other analysts freely will come up with the same comes about. Question 2 and 3 revile how big a problem fake news has become and how important it is to prevent it. At the same time, social media or mass media's important role in spreading Fake News. Question 1 will help the researcher know the changing process of news to false news, how easily the news becomes fake news and makes humans fool and biased as the false news creator needs. This type of question needs a perfect analysis of exterior insured of same kinds of works and a combination of different kinds of empirical proof. However, we try to present the best results in the current situation.

1.6 Report Layout

The rest of the review is organized as follows, discussing the foundation of this investigation and the related work of recognizing fake news in chapter 2. Chapter 3 describes the research methodology. We discuss dataset collection, cleaning, processing, and a model to detect fake news using a machine learning algorithm Discussion evaluate and the performance of the proposed model in chapter 4. Selected topics impact society and advise some processes to prevent the problem in chapter 5 and finally, Summarize the overview learned and future advice in chapter 6.

CHAPTER 2
BACKGROUND

2.1 Method

Systematic Literature Review (S.L.R) is a process whose primary purpose is to find out all the information about a particular research question similar to other available research papers. It reveals a well-organized infrastructure where research questions are explicitly reviewed [11], [12]. Where step by step, research questions are discussed, and results are revealed by making a specific decision based on the information and data obtained.

We performed a systematic review of our study with some steps proposed. Main focus is to conduct scientific works using a smart automatic intelligent detection algorithm for Fake News, approaches, and methods. The rest of the following section defines research questions, scope, strategy, and selection standards. Figure-2.1 shows the steps we followed in the S.L.R process.

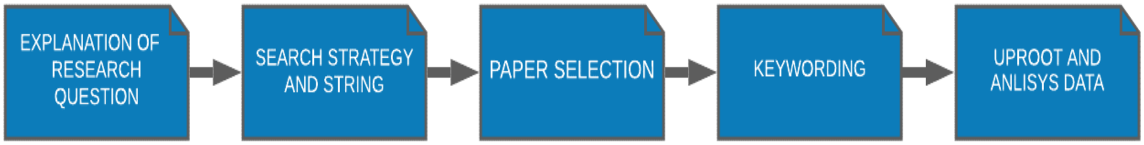


Figure 2.1: Systematic Literature review mapping

2.1.2 Searching

There are five databases have been used to perform the Systematic Literature Review: Google Scholler, IEEE, ResearchGate, Science Direct, and Microsoft Academic. Which are usually free accessible repositories, and researchers can collect research papers of their choice. Vast numbers of research papers are present there with categories and types. Using different searching methods, we search our needed publications article from those repositories. We will discuss our search method rest part of the review paper.

2.1.3 Method of searching publication

An accurate search in the digital repository needs an appropriate a search string, which can be in English or different terms. This search string will be done to connect other methods of fake news detection with our guesses, with the help of which we can get the required publications in separate repositories. PICO model control articles help identify these terms showed section 1.4.2, which later helps to create refined strings. Here we show before refining values in Table-2.1.

TABLE 2.1: Terms with categories with PICO model

Class	Description
Population	Fake News
Intervention	Intelligent computing, Text data mining, DM, Text mining, Automatic computing, Smart system, Natural Language processing, NLP, ML, Impact, Identification, Defeat, Modern Life, Social Media, Mass media.
Control	Analysis journal on Fake News
Results	Obtainment, Process, Techniques, Pre-processing, Algorithm, Detection, Arrangement, Prophecy, Outcome, Application, System, Tool, Framework.

After refining, the terms are much better organized which we have showed in Table-2.2 and Table-2.3, which is more constructed search strings.

The following is the highlight of the term obtained by the search string:

> *(Obtainment* OR Approach* OR Process* OR Techniques* Pre-processing* OR Algorithm* OR Detection* OR Arrangement* OR Prophecy* OR Outcome* OR Application* OR System* OR Tool* OR Framework) AND (“Data Mining” OR “Intelligent computing” OR “Text data mining” OR “Text mining” OR “Automatic computing” OR “Smart computing” OR “Natural Language processing” OR “NLP” OR “Machine Learning”) AND (“Fake News”)*

> *(“Impact” OR “Identification” OR “Defeat” OR “Modern Life” OR “Social Media” OR “Mass media”) AND (“Fake News”)*

TABLE 2.2: Searching strings with some terms

Strings terms		
Approach*		
Pre-Processing*	Data Mining	
Process*	Intelligent computing,	
Technique*	Text data mining	
Algorithm*	Text mining	
Detection	Automatic computing	Fake News
Arrangement*	Smart computing	
Prophecy*	Natural Language processing	
Tool*	NLP	
System*	Machine Learning	
Application*		
Outcome*		
Framework		

TABLE 2.3: Searching strings with some terms

Strings terms	
Impact	
Identification	
Defeat	Fake News
Modern Life	
Social Media	
Mass media	

2.1.4 Selection Criteria

In this Systematic Literature Review, we must filter our episodic document for the best outcome. That’s why we used elimination and inclusion criteria. After completing the searching, the process described in section 2.1.3, collected results were selected as regarded studies document which will evacuate step by step.

The selection process covered the abstract or summary, and analyzed the introduction of each papers. After filtering the summary, selected papers were read, analyzed, and transmitted in the extraction stage.

Inclusion process:

- The title, abstract, or keywords contains the subject of the articles;
- Fake news has been highlighted, how fake news is created and disseminated;
- Define some automatic process to detecting Fake News;
- Represent pre-processing, available data, effectuation fake news detection algorithm or process;
- Available online questions.

Elimination process used were:

- Publication that are not in English;
- Don’t fulfill inclusion demand-based publication;
- The distribution that doesn’t concern almost inquire about the field of computer science.
- Preliminary publication

This section will construct the review process from the search keyword from different bases, select works for analysis, and remove unnecessary articles from our primary studies until we reach our result comparing the research question we already defined. In Figure-2.2, we try to visualize our searching and selecting process of articles from the repository and step by step filtering them.

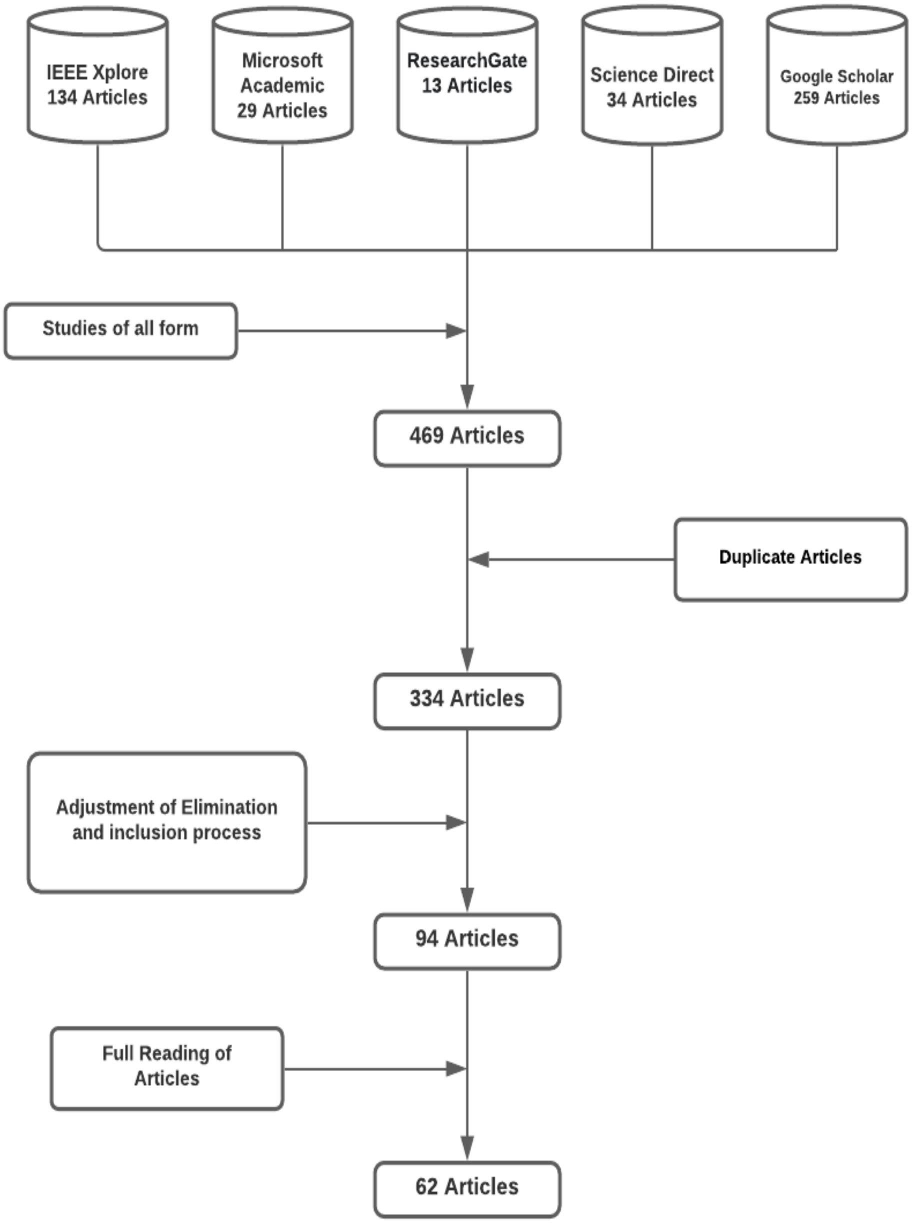


Figure 2.2: Selection process and searching process.

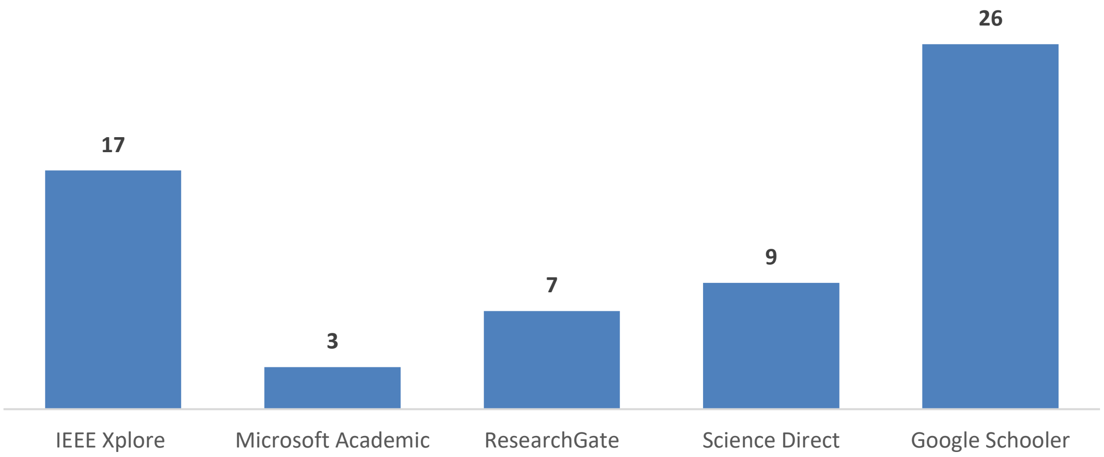


Figure 2.3: Formatting articles primary study in baseline

We collect a total of 469 articles from various databases. One hundred thirty-five (135) papers are omitted, which are mainly similar from different bases. After that, we used the elimination and inclusion process, and we removed 240 articles. The rest of the total 94 papers were thoroughly read and analyzed. After reading the entire article, we understand that 32 papers are unsuitable for our review work, which did not fulfill our required needs. Therefore, we rejected those papers according to the elimination process. Figure 2.3 shows the numbers of the article based on the repository after the selection process.

2.2 Question analysis

2.2.1 Evolution of Fake News

Fake news is intentionally false or deceiving or fabricated data presented as news that has no basis. It aims to harm the notoriety of a person or existence, making a profit through publicizing revenue. Mass request and published beneath the pretense of having an authentic see and feel of fake news [8]. It is a kind of yellow journalism where they use news for their own need and mislead the traditional printed news media or the new tech-based online social media readers/users. Not to say that fake news is any current object, it used for a long time as we saw in the "Great Moon Hoax" of 1835, where authors write some hoax news about the discovery of life and civilization in the Moon. But nothing happened then, but that article was able to convince the general reader that something like that had happened on the Moon [8], [13].

Presently a war between nations can be won or lost utilizing the complete quality data provided by the over-specified media. Media play a vital role in transmitting quality, comprehensive information to all. Last half-decade, the media's ratio of transmitting down quality information is increased. That's why the last decade is called the decade of fake news [8]. The rates of disclosure of low-quality information tends to increase many times compared to other times.

According’s to Narwal, Claire was the first research director who categorized news and drafted seven categories based on pattern, behavior or model, ordering, and impact, presented in Figure 2.4 with their flow and short characteristic.

Misleading news: Misinformation is untrue rumors, talk, and deceiving use of truths. Disinformation could be a subset of misinformation that's intentionally misleading. Communication can be distorted to create appealing features and little data. Almost all the information appears to readers within the primary news feeds. The most pattern of this news is to bias the readers. This sort of news makes a political impact, propaganda, or partisanship between a nation. This type of news impacts public opinions. Ex. Ops! covid-19 vaccine is full of saline water.

Manipulated news: A story where genuine vents are approximately uncovered, but this news is controlled through totally different information, news, or gossip. This type of news is used for its benefit. Ex. They are using different kinds of powerful manipulation tools of advanced technologies to open real digital photos or videos and create various fake news and changing minor elements, increasing color variation, removing the background, and adding new features into the images, making the content more eye-catching to audiences.

Parody or Satire: The primary purpose of this type of fake news is to parody or fun. There is no intention of harming anyone using this type of news. Different types of social media account or websites create commentary content for criticizing society wrongdoing, celebrities, and politicians for abusing the audience, but basically, those stories make them fools. Ex. A child that has been born with golden teeth from birth.

Twister content: False sources are narrated to authentic sources, whereas false news sources are made up quite like real news sites. It's hard to find out on a first look. This type of fake news is harmful to the market, industry, general users, and all of them because they think that getting info from a natural source but then founded cheated from a bit of mistake of them. Ex. bbc.com as bbc.com.lo. Maximum people focus on the title and headline of the website, but very few people look at the domain name of any website.

False Connection: This type of news content is different from the main headline. Here headlines, images, context display a story, but the main content body doesn't match that. Online networking is the best example of this type of news, and here low-quality journalists connect two different stories without any deep investigation.

Own Created news: A news depends on a piece of false or wrong information to manipulate the audience for a specific group of people, organization, political person success their agenda. Fabricated content generates profits, likes, comments and shares. Bots used widely circulated that to people. Gathering a vast number of audiences using this fabricated news creator makes a profit from other ads and, in real times, the traffic.

False Context: Contents without fact-checking are called false context. This type of news creates to break news or exclusive news to make TRP of viewing ratio of any content. Poor journalism, investigation gap is the main reason arise this type of problem. Without fact-checking, publishing any news is not a good practice of News providers in any news platform.

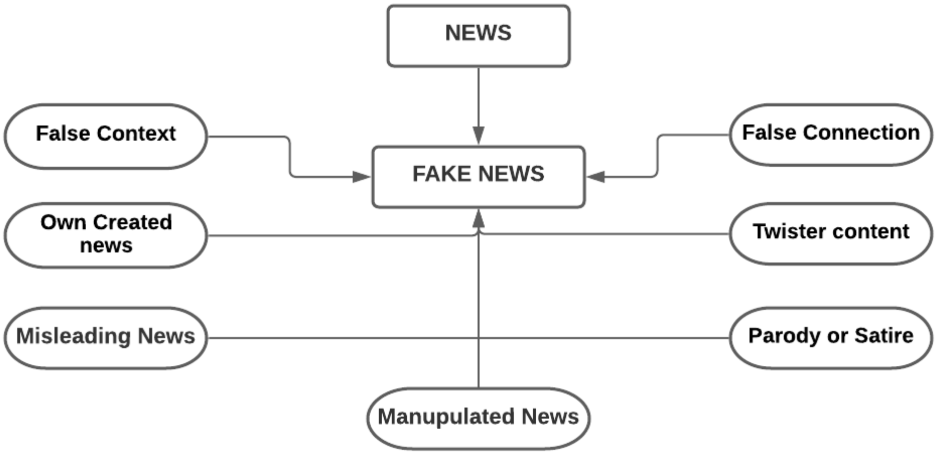


Figure 2.4: News to Fake news with its types

The intention of the fake news misleads reputed agencies, persons, financial gain, or political purposes with three vital parameters manipulation, mistrust, and misinformation with no valid sources, quotes. Two distinct motivations by Narwal behind writing fake news:

- Ideological: Create false news for favoring a specific person or to promote their intentional ideas.
- Financial: Clickbait revenue is a beautiful way to earn false news from any viral content.

With this motivation, creators also have some objectives to writing fake news to seek people's attention on social media platforms, entertainment, distract attention from current significant topic to non-sense matter, increase political influence, own influence think.

Before the Online networks came, everyone depended on printed media, radio, or television. Limitation and restriction of the audience, cost, and ease of access are relatively high, so the spreading rate is not alarming or substantial. After coming to the internet in our daily life or advancement in technology, ease access of internet and social media: content creation, curation, audience catching distribution, and consumption had made it cheaper and easier to spread fake news [14] . The maximum buying system is converted to an online base with a traditional shop system. It is also called the E-commerce era. In this system, people choose their product online, check the rating, overview, opinion of the product, and then purchase the items. As more and more people start spending time online than ever before, various fake newsmakers are taking advantage of that opportunity. Earlier, they used traditional methods, but now they have chosen online media to spread fake news.

2.2.2 Fake news as a double-edged sword

Nowadays, we are connected online in every way, and we meet a large part of the entertainment needs of daily life online. People spend a tremendous amount of time on social sites, and individuals are profoundly uncovered to a few false news. But now, this online is also vital in receiving the news, since the info from this medium is not always accurate, the readers deceive in various ways. For example, information weaponized to fulfill a person or group's motives and biased users/readers. Zhang et al., fake news has a significant variation with traditional ambiguous information like spam in various aspects:

Trouble in Identifications: Comparisons from unlimited standard messages in emails or survey websites, spams are usually easier to discover. But recognizing fake news from incorrect data is incredibly challenging. They required both evidence-collecting and careful fact-checking due to the need for other comparative news articles accessible.

Society impact: Spams exist in personal emails, websites and have a local impact on a small number of audiences. In contrast, social networks impact fake news tremendously due to globally massive users, then boosted by the wide information sharing and propagation among these users.

Initiative audience: Here users receive spam emails, and they start to share the news without any sense about its correctness.

Online has become the most pleasing way of spreading fake news because different types of fakes generator tools are available online where creators must upload images, content texts, headlines, and author's information. It’s also had opportunities to change the place information, multiple articles, date, and different information. Some popular web tools are “ClassTools Breaking News Generator”, “Break your news”, “World Gray News”, “Fodey” and so on [8]. Different SM platforms like Facebook, Twitter, Instagram, Reddit, Tumbler, etc., some posts become viral whose are created using these tools.

This fake news has started to use as the head of political interests. The most used example is the US presidential election results of 2016 with the election-related fake news in media. Social bots use to spearing this type of news [8], [13]. Generally, fake news is covered with different fields and languages to distort the truth. The false statement is hard to find based on their content. Even though the information is about facts but injected with wrong messages, a report shows that the fake news rate is only 54% without any guides [15], [16].

We use one or another media to get news in our daily life as websites, print media, and social media, in short, mass media. While there are some things we should be aware of, we do not notice that they are constantly cheating us:

- We never try to find out much about the author of an article. We don't even know if he wrote such a thing. If we have a platform with a name attached to a report, believe it and start sharing.

- Checking the duration of any articles or news released is important. Because various information is published in the media to mislead the current trading topic, we rarely check whether it is really from that time.

- Eye-catching headlines or supporting links misguided us in recent times mostly. Different headlines can attract our attention, and we also make other decisions on that headline without reading the body of the content, which is fooling the user psychologically. In the same way, various articles add many links as references to their information, but no relevant information is found when that link is inserted.

- Gain knowledge or information from reputed sites or truthful sources. Check articles misspelling, use of Capital letters, dramatic punctuation, low-quality grammatical usage. Quality sources always focus on their content and try to provide lack free articles.

- We never cross-checked or fact-checked any viral topic. But it is the best manually checking process to find any false or fake news. Everyone sharing any news in When everyone must share information, very few people come to verify the authenticity of that information, which easily deceives the user.

We must do all these methods manually, so never question the legitimacy of this information due to time constraints or the influence of social media or influencers. As a result, we unknowingly become the bearers of fake news in one way or another.

Unregulated sharing of information is driven by the marvel of fake news, which gotten to be a massive challenge in the online world. The rise of fake news has made our online lives much more insecure. A recent MIT study found that false information is six times faster than truth and gathers more audience than true stories [17]. Although it is possible to solve this problem through various methods, the question of its legitimacy remains. Various articles and blogs are composed to stop fake news from being exact news. But all are manually, so there is a very high probability of labeling any information biased. Because they can use their thought to provide those results or maybe be make them corrupt to a biased result. At the same time, it is pretty impossible to label data as true or false when the amount is enormous. At the same time, avoiding information contamination, opportune discovery, and control of unnatural substances are profoundly required. [18], [19]. The growth rate of data increases day by day, the need for automated systems through machines has become essential for fake news detection.

2.2.3 Mass media and Fake New

Establishing communication with a large audience via different types of media technologies is called Mass media. Like broadcasting media, they transmit information electronically to the audience via films, radio, recorded form or television, getting news from true sources, journalists, and media who take after the particular code of practice collecting unique data. On other hand, the advance of internet low editorial standards with little regulation like comments, sharing own fashion journalism is the foundation of Social Media But now, social media has somehow been able to emerge as a combination of all these things. And it has been made possible through online communication, through which people can know all the world's information from the comfort of their home with accurate time information. SM networks enable to replace traditional text-only news with image and video-based data, which provide a better experience and success to make many audiences. And it becomes the public stage of discussion, knowledge dissemination, emotions, talk about own ideology, sentiment sharing, business, and new experience gathering place.

But unfortunately, this online social network has become the primary means of spreading fake news, developing and biasing people with deceptive words. Facebook, Google, Twitter, unverified online new portal is the most common resources distribution of online fake news. SM is a low-quality news distributer source where false information, once posted, then spreads like floods caused by cyclones through likes, comments, sharing, and retweeting [8]. Various types of false news are disseminated through multiple online social media to influence the outcome of the 2016 US election, which plays a vital role in the vision of the subsequent election results. According to a post-election report, social media account for more than 41.8% of the fake news data traffic in elections, much greater than the data traffic shared by both traditional printed news/broadcast or social engines. Some popular phony news in that time is “Pizza gate Conspiracy”, “Endorses Donald Trump for president”, “Pope Francis Shocks World”, “Wikileak confirms Clinton sold the weapon to ISIS”. “Pizzagate” affected severely to the Democrat's images, social media flowed that Democrats run a child trafficking ring in Washington. This news drew the anger of thousands of people. The flow of the fake news in Reddit led to an actual shooting in the election time [8], [13]. The top twenty phony news stories were more ironic than the top twenty most-discussed true stories in the election time [20]. Not only the US election a report from freedom house over 65 countries 18 countries election online manipulated news, or misleading information played a significant role [17] . Online generated fake tools are another concerning matter for online networking because using those tools, the creator can make different categories fake news, publish and spread them online. Some of the posts became viral also means the audience takes that news seriously, so it is alarming for the users who take information online [8].

Social media is a communication platform that connects users using different electronic devices like smartphones, computers, laptops, etc. With the internet connection by SM tools like blogging, image, video, audio, messages, poll, voice message, stickers, gif to sharing personal thoughts, opinions, ideas on a topic. For any personal information consume social media is a perfect way. Everyone shared their daily personal data on the social platform without any second thought. Users can easily find everything they need on a single platform for the great combination of social media. According to the Pew, Research center [63] shows that *71% of Americans get at least some of their news input from social media platforms, and the service responded by more than 9,200 Americans.

Another online survey in nearly 1500 people from different ages of engineering and science background to understand the sources from daily news updates. Around 80% of people are found active on the Internet or SM and keep them up to date. A few of them collect news from traditional ways as newspapers or television or more reliable sources. But around 85% of people follow newspapers or television trustfulness of news content. At the age 15-35 years most active in SM and don't follow newspaper regularly. They trust the news from SM or the Internet without valid authentication [60]. Every like, comment, share, and reaction there can create the effect of emotions. Adolescents tend to use/trust social media more than other adults. Almost 65% of the US adult population is dependent on SM for their daily news [18]. According to Wiki [64], total population of world is *79 million, there are 46 million internet [65] user and 45 million social media user in 2021, comparing with the report 2019 Global digital changing rate of population and internet users is not so much but the social media users is almost 11 million increased which is huge [18].

Table 2.4: From the report of Statista (2021) top five social media list based on the user.

Name	Year	No. of User/months (Billion)	Features
Facebook	2004	2.8	Text, Image, video, live, stories, room required a valid email or phone to register or login—multiple communication systems by Meta.
YouTube	2005	2.2	Video sharing platform by Google. Community, Kids, Movies, Music, Shorts, Stories, TV primary services.
WhatsApp	2009	2.0	End-to-end encryption freeware, cross-platform centralized messaging, and voice-over-IP by Meta. Users can also use text, voice communication, audio/video call, Image,

			document, location, and other contents.
Instagram	2010	1.39	Photo and video sharing social network service owned by Meta.
Facebook Messenger	2011	1.3	Messaging app platform by Meta. Originally the only chatting form of Facebook. Text, video, document, emoji, Gif, Video/audio calling, etc., are the leading service to users.

2.2.4 Prevent fake news

Detecting fake news articles by identifying the writer, subjects is more important. Because it helps to altogether terminate fake news from the origins in the online network, using text or visual content or modeling user engagements to find fake news [21]. We can see them from different government databases using their profile information from social media platforms. News subjects or title is another critical factor to check the credibility of any articles compared/cross-validation to other authentic news sources. Articles body must check authentication and the title of the pieces [13]. Detect fake news using knowledge base methods; two approaches are generally used, i) Fact-checking and ii) Natural Language Processing (NLP) [22].

Fact-checking is a method to check the validation of any articles from online networks. Different companies launched the fact-checking process where checkers check reports manually with help from other databases that store similar information. Various valid sites, portals, and organization data are also used for the cross-validation of any content. Some popular fact-checking websites are FactCheck, PolitiFact, Snopes, TwitterTails, TweetCred, Hoaxy Emergent CreFinder, RumerLens, COMPA, InVID, ClaimBuster, TruthOrFiction, Full Fact, and Fake News Tracker [18].

According to Silva, the first social network that starts work against fake news uses an extra flag system where users can use it and detect any information as artificial in their news feed. If any news gets enough numbers of the fake flag, then another user can get a notification about the news and a warning notification that information contains unauthentic and misinformation [1]. Twitter is another well-known social network platform stricter its rules to decrease the spreading of fake news. Generally, it is a microblogging new content circulating strategic system. However, phony journalism is increasingly gaining using its fast and interactive real-time content spreading shortly.

Automated multimodal detection is an advanced process to find fake news. Here additional sub-task plays an essential role in finding relationships over modalities. In this process, outcomes heavily depend on it, which also revealed that lacking sub-task detection performance reduced by an average of 10% [21]. Fight against fake news Google launched Perspective API for finding toxic comments and troll-fighting using ML algorithm. Wikipedia located unlikely sources and made them notice to change their content if they have not changed it then ban their articles [23].

Machine learning algorithms are commonly used for the detection of fake news. In that process, the different algorithms first train up with some other datasets where many labeled data are stored already. After completing training, the datasets, test them with test datasets. If the algorithm gives good accuracy, then the algorithm defines as an excellent fake news detector. But in that process, data preprocessing plays a crucial role here. Because when the data are collected, it gathered some noise, which can decrease any algorithm's performance. But after cleaning the data in preprocessing, noise can be removed from the datasets. Now here we discuss some machine learning algorithms:

Support Vector Machine (SVM): Supervised learning demonstrates maybe a learning calculation that analyzes information within the classification and relapse preparation. A choice shape is used between fake and real news. When we need to check any information, we compare it with the choice shape already trained up. If the information is close to the actual word, it is defined as Real news; otherwise, Fake news.

K-nearest neighbor (KNN): Classification is the main procedure of this algorithm. A preparing set contains a case of news that makes a difference to discover real or fake news. Value of K will help to find K-closest news.

Decision Tree: Showing proactive approach where machine learning, data mining, and bits of knowledge are used. First, it makes an exhibit based on an information component that predicts the estimation of an objective variable. This broadly utilized calculation satchel the insatiable methodology at each portion and effectively makes a tree. Next, data are divided into several smaller subsets. Finally, the outcome is combined with the decision tree and node.

Random Forest Tree classifier: It, in addition, a tree-based classification performs by making a diverse Decision Tree but with an elective component structure. Finally, it combined all the results from different trees and produced a result.

Naïve Bayes Classifier: One of the primary characterization processes is to detect fake news. It follows the Bayes Theorem to check to approached news is real or false.

Ada Boosting Classifier: A meta indicator that begins with fitting a classifier and after that incorporate additional copies of the classifier. It is used to reduce bias variances.

Gradient Boosting classifier: It is a machine learning method used in classification and regression tasks. It is a forecast shown within the shape of an outfit of week predictions like decision trees.

Logistic Regression classifier: Assessing the parameters of the Logistic model with binomial backslide. Two possible qualities like '0' and '1' are represented. ex: good/bad, worst/best or short/tall.

Cultural algorithm (CA): A conventional genetic algorithm branch of evolutionary computation where knowledge component or belief space and population component are represented.

K (K Star):* Focus of the algorithm is to find the shortest path for K. It is a heuristic search algorithm.

Deep learning methods get tremendous success in pattern recognition. It works excellent in classification-based work like text, image—this process has been extensively used in NLP (Natural Language Processing) in recent years. Because of the enormous data, the machine learning process is hard to maintain, so researchers are starting to use different types of deep learning algorithms. Here we try to discuss some of them.

Artificial Neural Network (ANN): This algorithm works like a human mind. Like humans learn from a different source and store the memory into the system when need they use the memory and decide. This algorithm also does the same work using learning information and comparing that with the news data, and making a new decision.

Recurrent Neural Network (RNN): Class of ANN working with the association between hubs from a coordinated chart along with a temporal arrangement. It can utilize its inside memory to handle the variable-length sequence of inputs.

Long-short-term memory (LSTM): A feedforward neural network with artificial recurrent neural network design is utilized for deep learning, not as it formed a single information point as a picture. Still, arrangements of information like video discourse can be handled. It is also used in unsegmented handwriting, speech recognition. It solves the issue of long-range conditions and evacuates the slope issues.

Convolutional Neural Network (CNN): A deep neural network, primarily used in visual imagery.

Feature Extraction is an integral part of getting higher accuracy for different types of algorithms. For example, the algorithm finds fake or real news, but feature extraction increases the accuracy rate to solve inner processes. Now we discuss some features which are used for detecting fake news.

BERT: Bidirectional Encoder Representation from Transformers, a transformer-based machine learning strategy for Natural Dialect Processing pre-trained created by Google.

TF-IDF: "Term Frequency-Inverse Document Frequency" is a numeric statistic reflecting how important the word is to a document. This process is mainly used in text mining and information retrieval.

BoW (Bag of Word)**:** Its disregarded grammar and even word order but kept multiplicity. A simplifying representation NLP, it also used for computer vision.

Word2Vec: Method of natural language processing distributed 2013 where the machine can learn word affiliations from a broad collection of text. After preparing the model, can distinguish synonymous words or propose other words for partial sentences.

Various algorithms and feature extraction methods are used for fake news detection through machine learning and subsequent deep learning. From which initially good accuracy was not obtained, but gradually becoming more effective. This is because the medium of various feature extraction, algorithms, and pre-processing changes with time, and the accuracy rate is also increasing. Figure 2.5 shows a graphical representation of a much-used algorithm in a specific time zone by Silva. From the expression, the most used algorithm is Long Short-Term Memory with 17.14%. Naïve-Bayes and Similarity algorithm (11.43%), SVM, RF and Harmonic Boolean Label Crowdsourcing (8.57%), Stochastic Gradient Descent, Multilayer Perceptron and Least Squares Temporal Differences (5.71%) and LM-BFGS, KNN, DT(J48), Binary LR, Barabasi-Albert and AdabostM1(2.86%) [1].

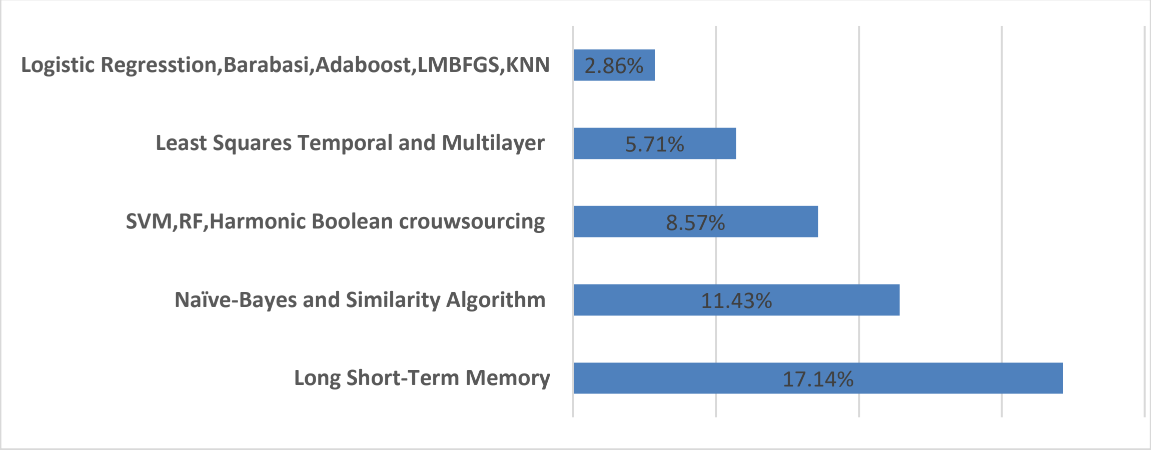


Figure 2.5: Primary study by algorithms of Silva

2.2.5 New Techniques

Figure 2.6. represents a graphical show of our current work algorithms in recent times. This mapping is made by each assessment from the articles and observing the most used algorithm. Long Short-Term Memory (14.51%), Proposed framework (11.29%), Conventional Neural Network (9.67%), Neural Network and Support Vector Machine (4.83%), Random Forest (3.22%), Decision Tree, Gray wolf, K-mean, Cultural, Harmonic BLC algorithm, K-NN, Support Vector Clustering (1.61%).

From 62 articles in our selected paper, we can see that 23.53% used lane algorithms for fake news detection, and 76.47% used machine learning algorithms. But this usage rate is much higher than using deep learning algorithms before.

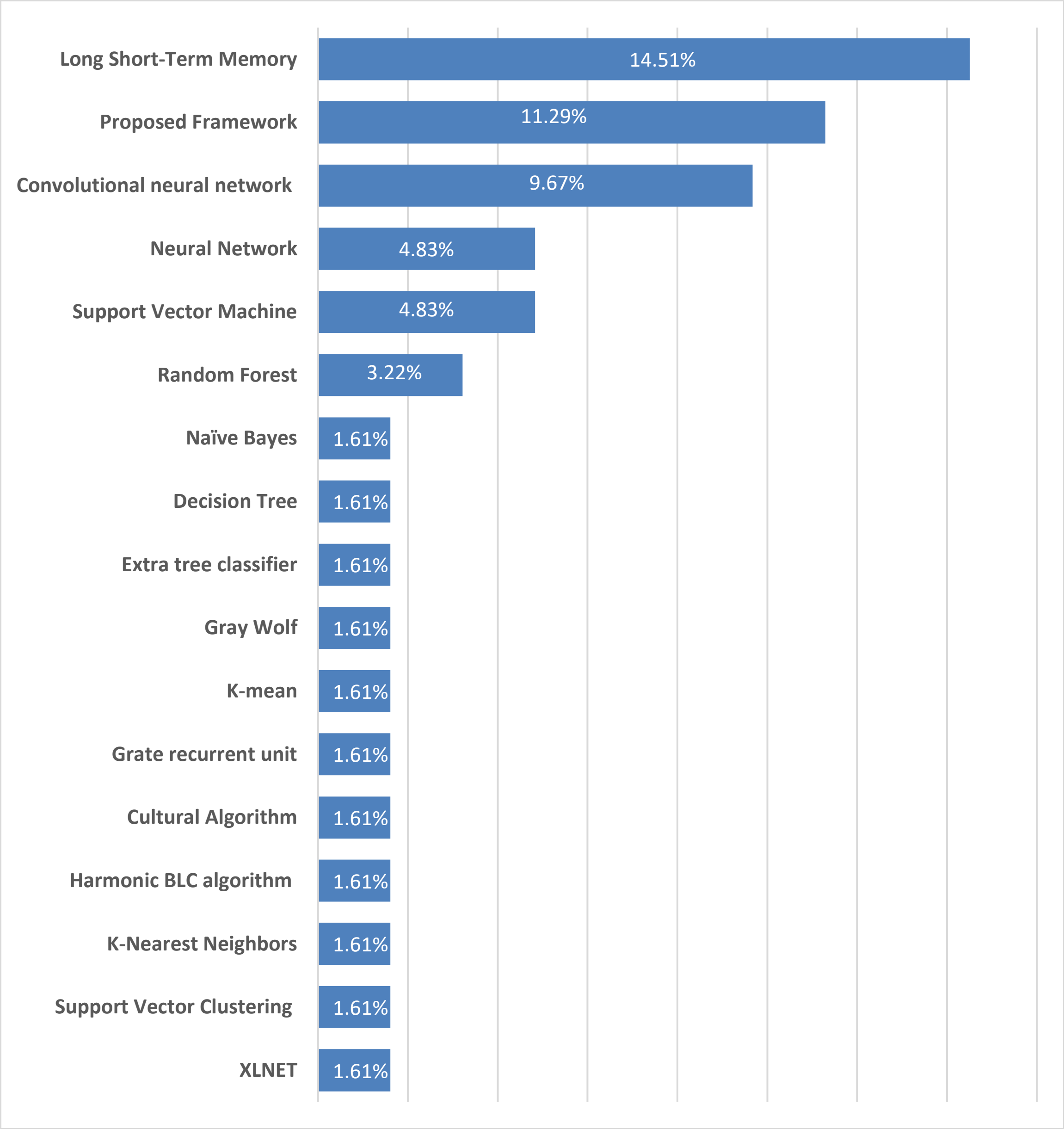


Figure 2.6: Algorithm's used in the selected articles (Base most used)

Table 2.5 a summarization of the work with proper references respectively where an article has revealed which algorithm has performed the best.

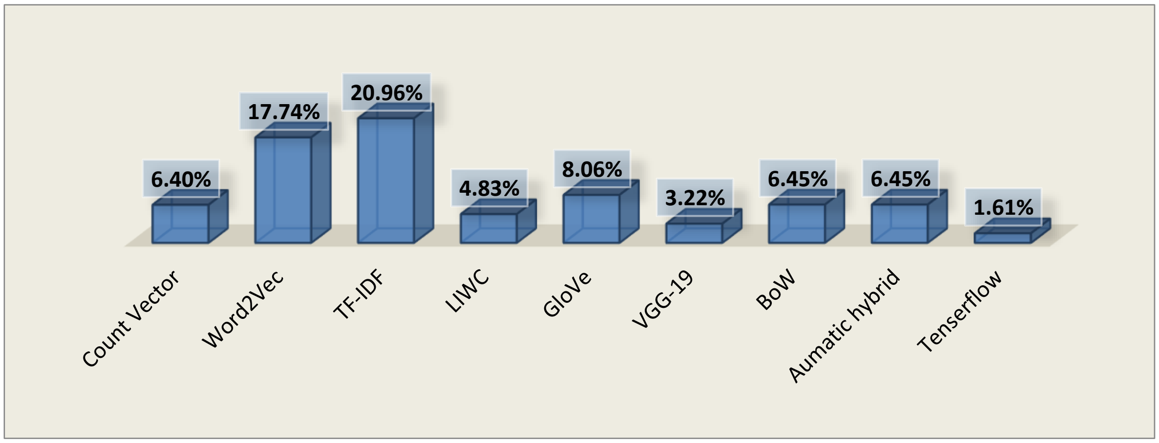


Figure 2.7: Bar graph of the Feature extraction from selected articles

Datasets and feature extraction are significant to make the algorithm work properly. If a good cleaning dataset is taught to machine algorithms, it is possible to get good accuracy results. Also, various feature extraction methods such as C2V, W2V, TF-ID, BoW, LIWC helps algorithms to bring more effective to work and get better accuracy than usual time. Figure 2.7 and Figure 2.8, based on our selected articles in Table 2.5, show the most-used datasets and feature extraction to find fake news with reasonable accuracy.

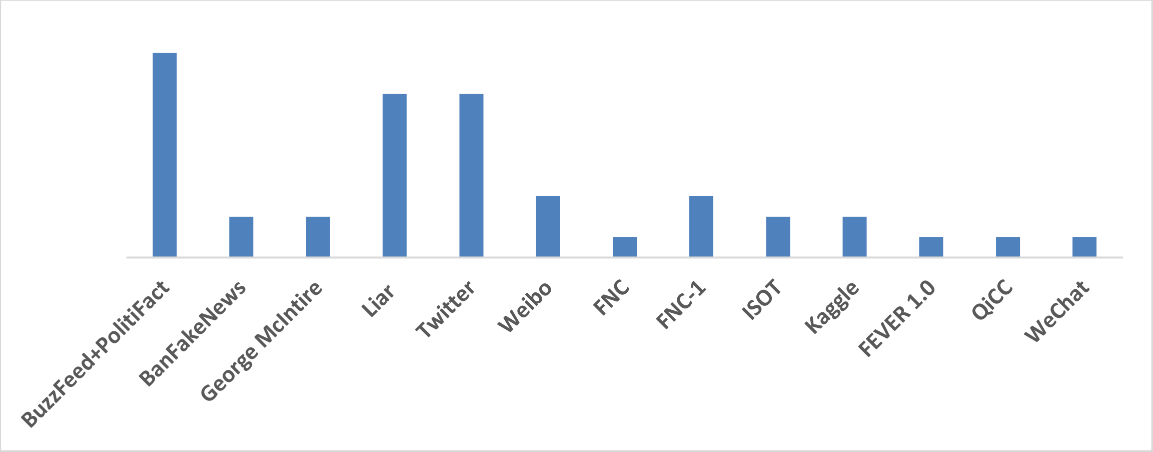


Figure 2.8: Used Datasets from selected articles in graphical representation.

2.3 Summary of RQ

Despite the fact that the few replications of the considers, anticipating more profound evaluations of predisposition and outside legitimacy, as well as the combination of test evidence, the regard prerequisite guided the examination and starting presentation of the calculations that gotten the foremost refined accuracy execution, since the Colossal Data setting requires arrangements that consider the tradeoff ampleness and adequacy. Underneath, we list the three best comes about and the calculation that accomplished the most noticeably awful execution.

The evaluation shows that the highest accuracy achieved by Decision Tree, Random Forest and Extra tree in publicly available dataset ISOT contains 23.481 fake news articles and 21,417 true news articles where true reports taken from a reliable source as Reuters and fake news from Wikipedia and PolitiFact flagged articles [24]. In the nonstatic process, a convolutional neural network with a margin loss system to detect fake news achieved 99.9% accuracy with Word to Vector (W2V) [25]. Then Long Short-Term memory by Sahoo contracted with baseline dataset from raw data using some filtering process. The data is labeled as fake or real. The dataset was created by Facebook spam content like images, links, hashtags, URLs, etc., and the model achieved 99.4% accuracy [26]. A combined proposed method with Gradient Boosting Decision Tree (GBDT) and CNN in Kaggle fake news dataset contains 4009 entries with URL, headline, body, and label where label 1 denote as real and 0 as fake—achieved accuracy 97.59%. Then Support Vector clustering and machine achieved 91% accuracy in final datasets from PolitiFact data with 4022 users' profiles on Twitter. The total data is 62,367, where fake is 34,429 and real 29,938 [27]. Ajao work with five rumor stories with 5,800 tweets with PHEME datasets can achieve 82.29% accuracy using the LSTM algorithm [28]. Finally, the worst accuracy is 41% using LSTM; also, the LIAR dataset contains 12,836 short statements from 3,341 speakers, which covered 141 topics of Politifcat.com [29]

2.4 Related work

Various researchers have been conducting their research for a long time to solve the problem of fake news. They have used different machine learning methods to solve this problem. Some of them have been able to detect fake news much better and some less. But to solve this problem, they have carried out various studies at different times.

Mugdha, Shafayat Bin Shabbir, et al., working with multiple algorithms to their proposed datasets. And the datasets are a combination of news from different websites as

BDFactchek. Mainly they collect the title, body, date, URL, and label for the datasets. Define the date into two labels Fake and Real. For getting good accuracy, they tokenized their data and stemmed it. Extracting features used TF-IDF and different tree classifiers. Between the other algorithms, Gaussian Naïve Bayes Classifier (GNB) gets a good accuracy of 85.52%, and the F1-Score is 82.1% [22].

Ensemble voting classifiers used multiple different types of machine learning algorithms and selected the top three highest accuracy algorithms to ensemble voting classifier. Used a datasets combination of 6500 data where 3252 data were used as Fake data, and 3259 data were used as actual data. From the most top highest algorithm (MLP, LR, XGB), getting accuracy 94.47 for soft voting and 93.99 for hard voting [15].

Arnab and Maruf et at., collected their data from popular newspaper online websites. They balanced their data in a simple format in label satire or not satire news. They were vectorizing their data for similar work accessing, which are used to train their data. Preprocessing data ignoring stop words and punctuation, stemming data. Mainly focused on the CNN architecture and the accuracy is 96.4% from the dataset [30].

Zobaer, Saiful et at., collect real news from Bangladesh's 22 most widespread and trusted portals. The main types are Misleading/False Context, Clickbait, and Satire/Parody. Data are categorized in some form. They used their datasets in popular machine learning algorithms like SVM, RF, and LR, and they got 53%, 46%, 53% F1-Score, respectively, and BERT, a NN model, gets 68% in the deep learning process [31] .

For collection fake Bangla news, they mainly focused on Facebook, covering the most popular social media sites in BD, which work for political parties, companies, and regular people broadly communicating and broadcasting news. They collected 726 news articles, gave them the tag Fake or real, and used text mining to extract the datasets. Document Term Matrix (DTM) is used for text mining, and TF-IDF vectorizing removes the word commonly used but has no strong meanings. The proposed method for detecting fake news is a web interface based with a random forest classifier, and the accuracy for the title is 83.5%, body 82.9%, and combination of title and body is 85% [3]

DETECTING FAKE NEWS USING DEEP LEARNING APPROACHES, M. S. Hasan, R. Alam, and M. A. Adnan. They proposed a multi-model Ensemble Neural Network for solving this problem. Three different types of NN-connected models. They trained five different types of datasets at a ratio of 80:20. TF-IDF features extraction method to extract features. And the proposed model get accuracy for different datasets 94.0%, 61.0%, 99.0%, 94.13% respectively [32].

Hussain, Md Gulzar, et al. used a machine learning algorithm for detecting fake news. They collect data from various newspapers but have not used any previous datasets. Collect around 2500 articles, and all are public datasets. All data are divided into two parts real news and fake news, for preprocessing removing unnecessary special characters, numerical numbers, punctuation marks, special symbols. Extracting features, they used TF-IDF vectorizing and finally split the dataset into two portions training and testing in 70% and 30%, respectively. They got a good accuracy of 96.4% for the SVM algorithm [33]

Soma and Sanjay et al. proposed a hybrid model and trained the model with the Kaggle Bengali news and Online Bengali news datasets. Total 25k news covering with seven different domains represent here. They preprocess their data in three main concepts: Cleaning raw data, Feature extraction, and Synthetic new classification. They remove phrases, English sentences, Stop words, Word Stemming, emotion symbols, Symbols, pictographs. The proposed model gain 86% accuracy from the combined datasets [34].

2.5 Scope of the Problem

Looking closer, various algorithms have been used for fake news detection, but mainly on English words. Different algorithms have been used, from machine learning to deep learning. But the amount of fake news detection work on Bengali characters is much less than that. It is no longer possible to say that fake news is just a foreign problem because it has also started to affect our country.

2.6 Challenges

It is impossible to do any work without problems, and it may lack adequate explanations. It may be a lack of research. The work must face various issues which need to be solved to move the work forward. We also had some problems, which are given below:

2.6.1 Construction Validity

Detecting fake news is not easy to expand string, solve this problem, we used synonyms or related terms. PICO model helped to identify and refine the articles that were related to the search. With the PICO model, we placed the word, and a search string was used to find the papers, and articles not related to our screening were excluded.

- The researchers extracted and classified the algorithms, causing bias and data extraction problems that made characterizing valid. Numerous articles have been inaccurately enlisted concurring to the researchers' judgment. Selection and extraction methods have been used to solve this problem by the researchers involved in this research. Some of the approved articles can’t fulfill the transparent methodology as research, evaluation, or validation. Solving this problem involves the researcher reading articles fully to find their needed characteristics.
- Even though the inquiry was completed on an essential computerized premise, it is inconceivable that the results of this systematic literature review secured all the works on the subject. In any case, this considered displayed prove of the advancement, methods used, and gaps to be investigated, serving as a direction for future work in this work field.

CHAPTER 3
RESEARCH METHODOLOGY

3.1 Dataset Utilized

According to the author ' Ethnologue ', Bangla is the sixth most spoken language [67]. Because of the structural complexity, Bengali languages are hard to find. So, for collecting data, we take help from a famous dataset repository, Kaggle [66]. This dataset consists of the same types of news which is represented in next:

*Parody/Satire***:** Here, news represented comedic entertainment.

Misleading context: News with untrusted information represented there. This type of news can be misleading the audience.

Clickbait: This type of news uses sensitive headlines to attract audiences, but when clicked, body news is different from the headline news.

They collected data from different types of websites and then removed duplicate news from the datasets. Total 49k data contain where 48k used as real data and 1k data used as fake data.

3.2 Statistical Analysis

3.2.1 Creation of new Datasets

In this research, we used labeled datasets from the primary 48k datasets, consisting of 8k data. Labeled data was divided into two portions: real 7k data and fake 1k data. There are eleven (11) categories of news represented in the fake labeled datasets, and the categories are Miscellaneous, Entertainment, Lifestyle, National, International, Politics, Sports, Crime, Education, Technology, and Finance, respectively. And twelve types of categories of news are represented in the real labeled datasets: National, Sports, International, Politics, Editorial, Entertainment, Miscellaneous, Crime, Finance, Education, Technology and Lifestyle, respectively. Table-3.1 represented the number of categories for the labeled fake or real news.

Table 3.1: Category ways Real and Fake Data

Category	Real	Fake
Miscellaneous	290	654
Entertainment	302	106
Lifestyle	90	102
National	3501	99
International	844	91
Politics	385	90
Sports	889	54
Crime	224	42
Education	109	30
Technology	96	29
Finance	136	2
Editorial	336	-

3.2.2 Label

Due to the combination of the received label dataset, we have prepared the dataset to our advantage, where we have labeled the information contained in the dataset's content. The labels through which the machine can learn what is right and what is wrong. And we express this label with two numbers, '0' and '1'. The label by '0' is indicated fake, '1' is real. The Table 3.2 and Figure 3.1 are based on dataset data:

Table 3.2: Overview of Datasets

Total Reviews	8497
Real News	7200
Fake News	1297

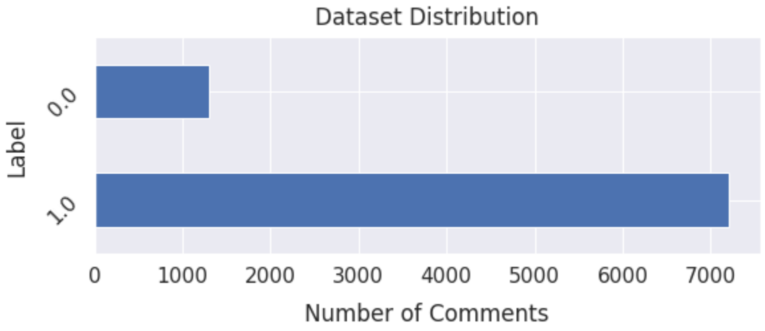


Figure 3.1: Overview of Datasets

3.3 Proposed system

Our proposed methodology is described step by step now; first, we must collect novel datasets that can be used to detect fake and real news. As we know that available raw data is not easy to use first, we must process the data we collected. We collect our data in two different datasets individually, so after cleaning and preprocessing, we must combine the data and focus on the features extracting our data. After that, we put the data into classifiers for our accuracy. Here we used different classifiers for detecting fake ones and comparing the classifiers, and finding out the best one—figure 3.2 visualizing our proposed methodology step by step.

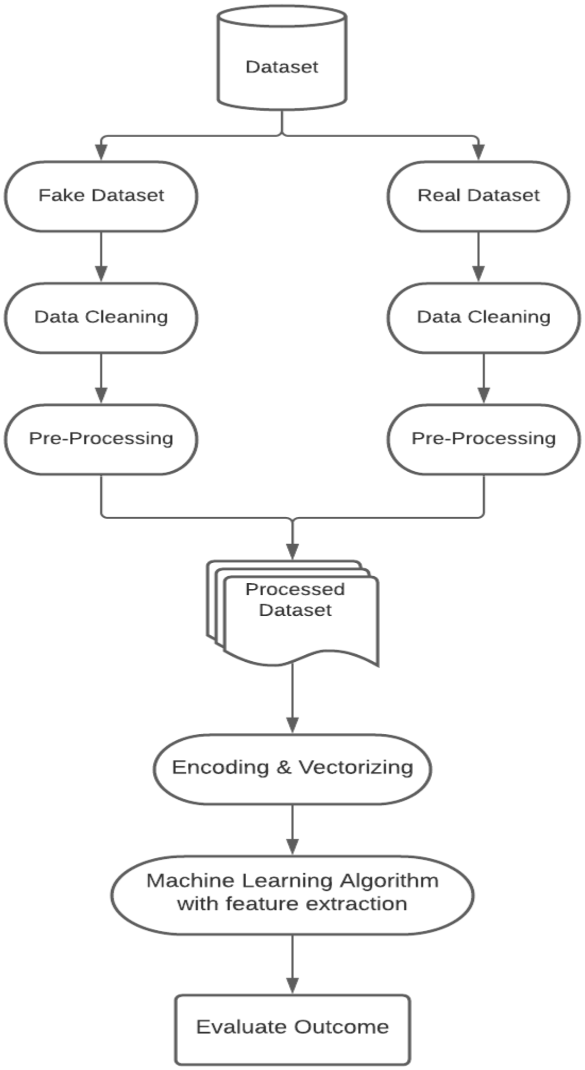


Figure 3.2: Process architecture using machine learning

3.4 Implementation

3.4.1 Cleaning Data

Use of spacing, ordinary signs (called accentuation marks), and specific typographical devices helps to the understanding and adjust perusing of composed content, whether perused noiselessly or audition called punctuation. Different types of punctuation are present in datasets. So, we remove unnecessary punctuation from bangle text using the range is '[^\u0980-\u09FF]'.

3.4.1 Pre-Processing

In the raw datasets, we have various noisy data, symbols, characters, and stop words, which can decrease our accuracy. So we have to get rid of our datasets in the preprocessing steps.

- Feature selection and classification: In the datasets, we have different types of attribute for real news: articleID', 'domain', 'date', 'category', 'source', 'relation', 'headline', 'content', 'label', 'publisher'. And attributes for fake news: 'articleID', 'domain', 'date', 'category', 'source', 'relation’, ‘headline', 'content', 'label'.

- In both labeled fake and real datasets, we work with the attribute’s ‘label’ and ‘content’ attributes. We combine our labeled data in single datasets and total data represented in Table 6.

3.4.2 Unique Data

In the vast dataset, we must calculate unique words for good accuracy. From datasets, we find out the total word and unique word. If we categories the date, then we get 79k word for the actual data and 41k word for the fake data. Table 3.3 showed the documentation.

	Real	Fake
Number of Documents	7202	1299
Number of words	1833006	366826
Number of Unique Words	77674	41008

Table 3.3: Documentation of Datasets

3.4.3 Label Encoding

In this case, we converted our labels into numeric form as we know that the machine can’t understand our language. Using this label encoding makes or datasets machine-readable form. So, when we use a machine-learning algorithm to train our data, it can easily machine understand and predict accurately.

3.4.4 TF-IDF

Lastly, we used Term Frequency–Inverse Document Frequency or TF-IDF, a popular feature for numeric representation of transformed text data. It can categorize the importance of a particular word in the document. It is a very widely used feature for NLP. In this case, we use the n-gram TF-IDF feature extraction process where the value of n is three and presents Unigram, Bigram, Trigram, respectively.

CHAPTER 4

EXPERIMENTAL RESULTS AND DISCUSSION

4.1 Performance Evolution

In this work, we train our datasets in multiple classification algorithms: Logistic Regression (LR), Decision Tree (DT), Random Forest (RF), Multinomial Naive Bayes (MNB), K-Nearest Neighbors (KNN), Linear (SVM-Support Vector), Machine and Radial Basis Function Kernel (SVM).

4.1.1 Unigram features

A one-word sequence is called unigram. The unigram features one word vectorizing each time. For example, the Given word is “আমার সোনার বাংলা” here unigram feature extraction will be “আমার”, “সোনার”, “বাংলা”. So, using this unigram feature extraction, our performance table shown in the Table-4.1.

Table 4.1: Performance table for Unigram Feature

Model Name	Accuracy	Precision	Recall	F1-Score
LR	87.18	86.77	100.00	92.92
DT	90.35	94.27	94.27	94.27
RF	90.47	89.82	100.00	94.64
MNB	**92.13**	92.01	99.86	95.44
KNN	84.59	84.52	100.00	91.55
Linear SVM	84.59	84.52	100.00	91.61
RBF SVM	87.88	87.41	100.00	93.28

4.1.2 Bigram features

A two-word sequence is called Bigram. The Bigram features two-word vectorizing each time. As example: Given word is “আমার সোনার বাংলা আমি তোমায় ভালোবাসি” here unigram feature extraction will be “আমার সোনার”, “বাংলা আমি”. So, using this Biigram feature extraction, our performance table shown in the Table-4.2

Table-4.2: Performance table for Bigram Feature

Model Name	Accuracy	Precision	Recall	F1-Score
LR	87.18	86.77	100.00	92.92
DT	90.35	94.27	94.27	94.27
RF	90.47	89.82	100.00	94.64
MNB	**92.13**	92.01	99.86	95.44
KNN	84.59	84.52	100.00	91.55
Linear SVM	84.59	84.52	100.00	91.61
RBF SVM	87.88	87.41	100.00	93.28

4.1.3 Trigram features

A three-word sequence is called Trigram. The Trigram features tree word vectorizing each time. As example: Given word is “আমার সোনার বাংলা আমি তোমায় ভালোবাসি” here unigram feature extraction will be “আমার সোনার বাংলা”, “আমি তোমায়ত ভালোবাসি”. So, using this unigram feature extraction, our performance table shown in the Table-4.3

Table-4.3: Performance table for Trigram Feature

Model Name	Accuracy	Precision	Recall	F1-Score
LR	87.18	86.77	100.00	92.92
DT	90.35	94.27	94.27	94.27
RF	90.47	89.82	100.00	94.64
MNB	**92.13**	92.01	99.86	**95.77**
KNN	84.59	84.52	100.00	91.55
Linear SVM	84.59	84.52	100.00	91.61
RBF SVM	87.88	87.41	100.00	93.28

4.2 Analysis and Discussion

According's to the performance tables; we extract our datasets three times as unigram, bigram, and trigram. And the datasets are trained by multiple algorithms LR, DT, RF, MNB, KNN, Linear SVM, and RBF SVM. Now we visualize our accuracy and F1-score for all the algorithms for three different features, shown in Figure-4.1, Figure-4.2, and Figure-4.3, respectively.

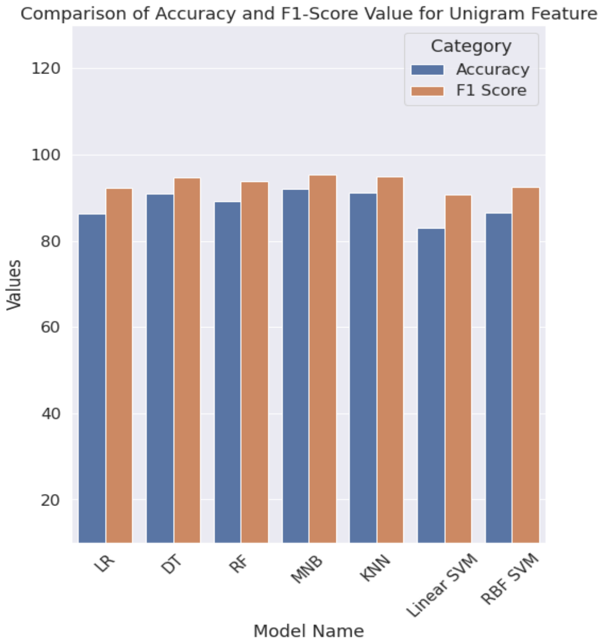


Figure 4.1 Unigram Performance

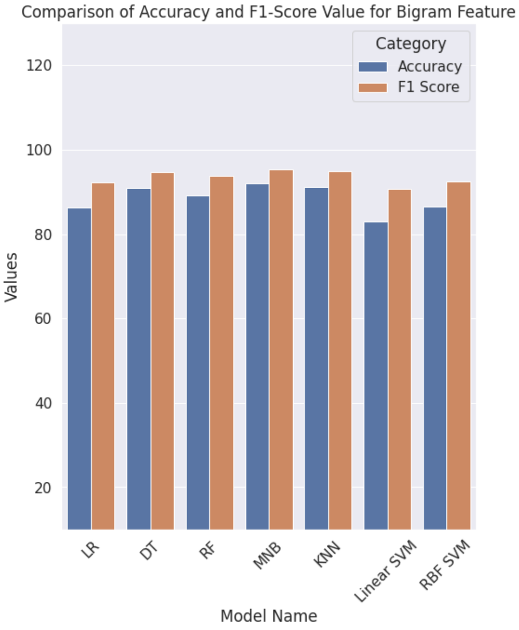


Figure 4.2 Bigram Performance

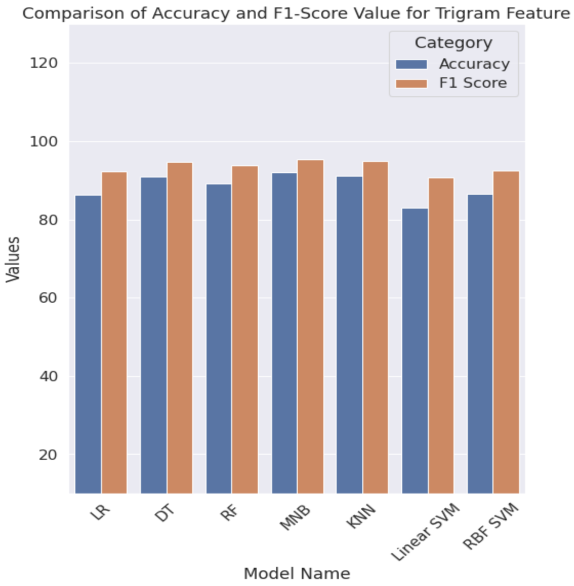


Figure 4.3 Trigram Performance

From the figure we see that accuracy and F1-Score for LR 86.25% and 92.33%, DT is 90.95% and 94.60%, RF is 89.19% and 93.87%, MNB is 92.13% and 95.44%, KNN is 81.07% and 94.82%, SVM (linear) is 82.96% and 90.66% and SVM (RBF) is 86.60% and 92.51%. From the values, we can see that Multinomial Naive Bayes (MNB) is better than all other algorithms.

CHAPTER 5
IMPACT ON SOCIETY, ENVIRONMENT AND SUSTAINABLILITY

5.1 Impact of Society

There are very few critical events in our life since the online internet because this online has made the medium of communication. Because of online communication and business, news and other fields have got a new look. Because of online, everyone can meet their needs at home, so evil people have made this online the central place of fulfilling their interests. They are influencing, persuading, and harming people everywhere in online news, business, communication. Although various online fact-checking or manual methods have been tried to solve this problem, given the increasing number of online usages, it would be almost impossible to do so manually.

5.2 Sustainability plan

Over time the method of exchanging such quality information has changed. In the past, a few tons of equipment were needed to store a little information. However, there are currently trillions upon trillions of data being generated in just a few seconds. And with this ever-increasing information development comes a growing collection of fake information that's difficult and time-consuming to partition manually. Machine learning plays an essential part in solving this issue, and its significance is getting to be gigantic within the future. Because a machine can't find the exact fake information manually, the most important thing is to teach the machine properly so it can find the fake news exactly.

CHAPTER 6

SUMMARY, CONCLUSION, RECOMMENDATION AND IMPLICATION FORM FUTURE RESEARCH

6.1 Conclusion

In the literature review, a step-by-step study was appeared to distinguish and analyze fake news, how it was evaluated, the necessity of automated detection, relation with media, intelligent computing techniques used now or before. The survey mapping was conducted utilizing the subject look and determination strategy used. Information was extricated and analyzed from 62 articles to get our research goal.

In the past, human communication was minimal. So, people used to receive news through very few mediums. And those who used to publish news also used to publish that information only after verifying enough info. The wrong information was published in the past, but the amount was meager unintentional mistakes. But gradually, that unintentional mistake began to take the form of intentional forgetfulness. Hoax, Satire, Clickbait, etc., are an example. Mass media, of course, is one of the means of receiving news, which used to include TV, radio. Still, with the development of information technology, it has now come into the hands of people with the help of smartphones or computer laptops. Since we get the information very quickly, it goes without saying that we do not tend to sort out that information, which makes us follow some personal ideology. We have discussed various algorithms used to detect this fake news; different manual methods were used earlier. Later researchers focused on machine learning because a person can be biased because of his own opinion in verifying any information as true or false. Therefore, the possibility of giving biased results through the machine is reduced. As a result, this machine learning method for fake news detection has also significantly increased. As a result, most used algorithms for detecting fake news Long Short-Term Memory (14.51%), Proposed framework (11.29%), Conventional Neural Network (9.67%), Neural Network and Support Vector Machine (4.83%), Random Forest (3.22%), Decision Tree, Gray wolf, K-mean, Cultural, Harmonic BLC algorithm, K-NN, Support Vector Clustering (1.61%).

Based on our study, the primary focus is to detect fake news from the selected datasets BanFake. That’s why we collect our data and combine the fake and real data for training our ML algorithm. After collecting data, we clean our data, preprocess them and make them useable for our work. Encoding, vectorizing data and training the data in multiple machine learning algorithms. After properly learning machine will predict and show the accuracy depending on the datasets. Using different algorithm types s, we get MNB ML algorithm gain the highest accuracy 92.13%, and F1-Score 95.44%, respectively.

6.2 Future work

Working with text data is critical. Because vast amounts of data are created every day, we need to train our model for more data to increase accuracy. But machine learning algorithm-based models can't work with much data, so we must use the deep learning model for more accuracy and work a considerable number of structured or unstructured data. We will work on more data and use a deep learning model to handle the vast data in the future. Deep learning can play a leading role in solving this problem, but we need to work harder. Because fake news is no longer just a problem of one nation or country, it has become a global problem.

REFERENCE

[1] C. V. M. Silva, R. Silva Fontes, and M. Colaço Júnior, "Intelligent Fake News Detection: A Systematic Mapping," *J. Appl. Secur. Res.*, vol. 16, no. 2, pp. 168–189, Apr. 2021, doi: 10.1080/19361610.2020.1761224.

[2] E. C. Tandoc, Z. W. Lim, and R. Ling, "Defining 'Fake News': A typology of scholarly definitions," *Digit. Journal.*, vol. 6, no. 2, pp. 137–153, Feb. 2018, doi: 10.1080/21670811.2017.1360143.

[3] F. Islam *et al.*, "Bengali Fake News Detection," in *2020 IEEE 10th International Conference on Intelligent Systems (IS)*, Varna, Bulgaria, Aug. 2020, pp. 281–287. doi: 10.1109/IS48319.2020.9199931.

[4] K. L. James, N. P. Randall, and N. R. Haddaway, "A methodology for systematic mapping in environmental sciences," *Environ. Evid.*, vol. 5, no. 1, p. 7, Dec. 2016, doi: 10.1186/s13750-016-0059-6.

[5] Y. Yang, L. Zheng, J. Zhang, Q. Cui, Z. Li, and P. S. Yu, "TI-CNN: Convolutional Neural Networks for Fake News Detection," *ArXiv180600749 Cs*, Jun. 2018, Accessed: Sep. 18, 2021. [Online]. Available: http://arxiv.org/abs/1806.00749

[6] A. Thakur, S. Shinde, T. Patil, B. Gaud, and V. Babanne, "MYTHYA: Fake News Detector, Real Time News Extractor and Classifier," in *2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184)*, Tirunelveli, India, Jun. 2020, pp. 982–987. doi: 10.1109/ICOEI48184.2020.9142971.

[7] Abdullah-All-Tanvir, E. M. Mahir, S. Akhter, and M. R. Huq, "Detecting Fake News using Machine Learning and Deep Learning Algorithms," in *2019 7th International Conference on Smart Computing & Communications (ICSCC)*, Sarawak, Malaysia, Malaysia, Jun. 2019, pp. 1–5. doi: 10.1109/ICSCC.2019.8843612.

[8] B. Narwal, "Fake News in Digital Media," in *2018 International Conference on Advances in Computing, Communication Control and Networking (ICACCCN)*, Greater Noida (UP), India, Oct. 2018, pp. 977–981. doi: 10.1109/ICACCCN.2018.8748586.

[9] P. Sa-nga-ngam, T. Mayakul, W. Srisawat, and S. Kiattisin, "Fake news and online disinformation. a perspectives of Thai government officials," in *2019 4th Technology Innovation Management and Engineering Science International Conference (TIMES-iCON)*, Bangkok, Thailand, Dec. 2019, pp. 1–4. doi: 10.1109/TIMES-iCON47539.2019.9024427.

[10] S. Burshtein, "The True Story on Fake News," p. 1.

[11] B. Kitchenham *et al.*, "Robust Statistical Methods for Empirical Software Engineering," *Empir. Softw. Eng.*, vol. 22, no. 2, pp. 579–630, Apr. 2017, doi: 10.1007/s10664-016-9437-5.

[12] K. Petersen, R. Feldt, S. Mujtaba, and M. Mattsson, "Systematic Mapping Studies in Software Engineering," presented at the 12th International Conference on Evaluation and Assessment in Software Engineering (EASE), Jun. 2008. doi: 10.14236/ewic/EASE2008.8.

[13] J. Zhang, B. Dong, and P. S. Yu, "FAKEDETECTOR: Effective Fake News Detection with Deep Diffusive Neural Network," *ArXiv180508751 Cs Stat*, Aug. 2019, Accessed: Sep. 18, 2021. [Online]. Available: http://arxiv.org/abs/1805.08751

[14] S. A. Khan, M. H. Alkawaz, and H. M. Zangana, "The Use and Abuse of Social Media for Spreading Fake News," in *2019 IEEE International Conference on Automatic Control and Intelligent Systems (I2CACIS)*, Selangor, Malaysia, Jun. 2019, pp. 145–148. doi: 10.1109/I2CACIS.2019.8825029.

[15] A. Mahabub, "A robust technique of fake news detection using Ensemble Voting Classifier and comparison with other classifiers," *SN Appl. Sci.*, vol. 2, no. 4, p. 525, Apr. 2020, doi: 10.1007/s42452-020-2326-y.

[16] S. Shabani and M. Sokhn, "Hybrid Machine-Crowd Approach for Fake News Detection," in *2018 IEEE 4th International Conference on Collaboration and Internet Computing (CIC)*, Philadelphia, PA, Oct. 2018, pp. 299–306. doi: 10.1109/CIC.2018.00048.

[17] M. Choraś, M. Pawlicki, R. Kozik, K. Demestichas, P. Kosmides, and M. Gupta, "SocialTruth Project Approach to Online Disinformation (Fake News) Detection and Mitigation," in *Proceedings of the 14th International Conference on Availability, Reliability and Security*, Canterbury CA United Kingdom, Aug. 2019, pp. 1–10. doi: 10.1145/3339252.3341497.

[18] P. Meel and D. K. Vishwakarma, “Fake news, rumor, information pollution in social media and web: A contemporary survey of state-of-the-arts, challenges and opportunities,” *Expert Syst. Appl.*, vol. 153, p. 112986, Sep. 2020, doi: 10.1016/j.eswa.2019.112986.

[19] K. Shu, D. Mahudeswaran, and H. Liu, “FakeNewsTracker: a tool for fake news collection, detection, and visualization,” *Comput. Math. Organ. Theory*, vol. 25, no. 1, pp. 60–71, Mar. 2019, doi: 10.1007/s10588-018-09280-3.

[20] K. Shu, D. Mahudeswaran, S. Wang, D. Lee, and H. Liu, “FakeNewsNet: A Data Repository with News Content, Social Context, and Spatiotemporal Information for Studying Fake News on Social Media,” *Big Data*, vol. 8, no. 3, pp. 171–188, Jun. 2020, doi: 10.1089/big.2020.0062.

[21] P. Shah and Z. Kobti, “Multimodal fake news detection using a Cultural Algorithm with situational and normative knowledge,” in *2020 IEEE Congress on Evolutionary Computation (CEC)*, Glasgow, United Kingdom, Jul. 2020, pp. 1–7. doi: 10.1109/CEC48606.2020.9185643.

[22] S. B. S. Mugdha *et al.*, “A Gaussian Naive Bayesian Classifier for Fake News Detection in Bengali,” in *Emerging Technologies in Data Mining and Information Security*, vol. 1300, A. E. Hassanien, S. Bhattacharyya, S. Chakrabati, A. Bhattacharya, and S. Dutta, Eds. Singapore: Springer Singapore, 2021, pp. 283–291. doi: 10.1007/978-981-33-4367-2_28.

[23] M. Choraś, A. Giełczyk, K. Demestichas, D. Puchalski, and R. Kozik, “Pattern Recognition Solutions for Fake News Detection,” in *Computer Information Systems and Industrial Management*, vol. 11127, K. Saeed and W. Homenda, Eds. Cham: Springer International Publishing, 2018, pp. 130–139. doi: 10.1007/978-3-319-99954-8_12.

[24] S. Hakak, M. Alazab, S. Khan, T. R. Gadekallu, P. K. R. Maddikunta, and W. Z. Khan, “An ensemble machine learning approach through effective feature extraction to classify fake news,” *Future Gener. Comput. Syst.*, vol. 117, pp. 47–58, Apr. 2021, doi: 10.1016/j.future.2020.11.022.

[25] M. H. Goldani, R. Safabakhsh, and S. Momtazi, “Convolutional neural network with margin loss for fake news detection,” *Inf. Process. Manag.*, vol. 58, no. 1, p. 102418, Jan. 2021, doi: 10.1016/j.ipm.2020.102418.

[26] S. R. Sahoo and B. B. Gupta, “Multiple features based approach for automatic fake news detection on social networks using deep learning,” *Appl. Soft Comput.*, vol. 100, p. 106983, Mar. 2021, doi: 10.1016/j.asoc.2020.106983.

[27] G. Sansonetti, F. Gasparetti, G. D’aniello, and A. Micarelli, “Unreliable Users Detection in Social Media: Deep Learning Techniques for Automatic Detection,” *IEEE Access*, vol. 8, pp. 213154–213167, 2020, doi: 10.1109/ACCESS.2020.3040604.

[28] O. Ajao, D. Bhowmik, and S. Zargari, “Fake News Identification on Twitter with Hybrid CNN and RNN Models,” in *Proceedings of the 9th International Conference on Social Media and Society*, Copenhagen Denmark, Jul. 2018, pp. 226–230. doi: 10.1145/3217804.3217917.

[29] Y. Long, Q. Lu, R. Xiang, M. Li, and C.-R. Huang, “Fake News Detection Through Multi-Perspective Speaker Profiles,” p. 5.

[30] A. S. Sharma, M. A. Mridul, and M. S. Islam, “Automatic Detection of Satire in Bangla Documents: A CNN Approach Based on Hybrid Feature Extraction Model,” in *2019 International Conference on Bangla Speech and Language Processing (ICBSLP)*, Sylhet, Bangladesh, Sep. 2019, pp. 1–5. doi: 10.1109/ICBSLP47725.2019.201517.

[31] M. Z. Hossain, M. A. Rahman, M. S. Islam, and S. Kar, “BanFakeNews: A Dataset for Detecting Fake News in Bangla,” *ArXiv200408789 Cs*, Apr. 2020, Accessed: Dec. 03, 2021. [Online]. Available: http://arxiv.org/abs/2004.08789

[32] Md. S. Hasan, R. Alam, and M. A. Adnan, “Truth or Lie: Pre-emptive Detection of Fake News in Different Languages Through Entropy-based Active Learning and Multi-model Neural Ensemble,” in *2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, The Hague, Netherlands, Dec. 2020, pp. 55–59. doi: 10.1109/ASONAM49781.2020.9381422.

[33] M. G. Hussain, M. Rashidul Hasan, M. Rahman, J. Protim, and S. Al Hasan, “Detection of Bangla Fake News using MNB and SVM Classifier,” in *2020 International Conference on Computing, Electronics & Communications Engineering (iCCECE)*, Southend, United Kingdom, Aug. 2020, pp. 81–85. doi: 10.1109/iCCECE49321.2020.9231167.

[34] S. Das and S. Chatterji, “Identification of Synthetic Sentence in Bengali News using Hybrid Approach,” p. 8.

[35] N. R. Bhowmik, M. Arifuzzaman, M. R. H. Mondal, and M. S. Islam, "Bangla Text Sentiment Analysis Using Supervised Machine Learning with Extended Lexicon Dictionary:," *Nat. Lang. Process. Res.*, vol. 1, no. 3–4, p. 34, 2021, doi: 10.2991/nlpr.d.210316.001.

[36] E. Qawasmeh, M. Tawalbeh, and M. Abdullah, "Automatic Identification of Fake News Using Deep Learning," in *2019 Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS)*, Granada, Spain, Oct. 2019, pp. 383–388. doi: 10.1109/SNAMS.2019.8931873.

[37] R. Oshikawa, J. Qian, and W. Y. Wang, "A Survey on Natural Language Processing for Fake News Detection," *ArXiv181100770 Cs*, Mar. 2020, Accessed: Sep. 18, 2021. [Online]. Available: http://arxiv.org/abs/1811.00770

[38] E. A. Hassan and F. Meziane, "A Survey on Automatic Fake News Identification Techniques for Online and Socially Produced Data," in *2019 International Conference on Computer, Control, Electrical, and Electronics Engineering (ICCCEEE)*, Khartoum, Sudan, Sep. 2019, pp. 1–6. doi: 10.1109/ICCCEEE46830.2019.9070857.

[39] S. Vinit Bhoir, "An Efficient FAKE NEWS DETECTOR," in *2020 International Conference on Computer Communication and Informatics (ICCCI)*, Coimbatore, India, Jan. 2020, pp. 1–9. doi: 10.1109/ICCCI48352.2020.9104177.

[40] T. Saikh, B. Haripriya, A. Ekbal, and P. Bhattacharyya, "A Deep Transfer Learning Approach for Fake News Detection," in *2020 International Joint Conference on Neural Networks (IJCNN)*, Glasgow, United Kingdom, Jul. 2020, pp. 1–8. doi: 10.1109/IJCNN48605.2020.9207477.

[41] T. Zhang *et al.*, "BDANN: BERT-Based Domain Adaptation Neural Network for Multi-Modal Fake News Detection," in *2020 International Joint Conference on Neural Networks (IJCNN)*, Glasgow, United Kingdom, Jul. 2020, pp. 1–8. doi: 10.1109/IJCNN48605.2020.9206973.

[42] N. X. Nyow and H. N. Chua, "Detecting Fake News with Tweets' Properties," in *2019 IEEE Conference on Application, Information and Network Security (AINS)*, Pulau Pinang, Malaysia, Nov. 2019, pp. 24–29. doi: 10.1109/AINS47559.2019.8968706.

[43] S. Girgis, E. Amer, and M. Gadallah, "Deep Learning Algorithms for Detecting Fake News in Online Text," p. 6.

[44] D. S and B. Chitturi, "Deep neural approach to Fake-News identification," *Procedia Comput. Sci.*, vol. 167, pp. 2236–2243, 2020, doi: 10.1016/j.procs.2020.03.276.

[45] R. K. Kaliyar, P. Kumar, M. Kumar, M. Narkhede, S. Namboodiri, and S. Mishra, "DeepNet: An Efficient Neural Network for Fake News Detection using News-User Engagements," in *2020 5th International Conference on Computing, Communication and Security (ICCCS)*, Patna, India, Oct. 2020, pp. 1–6. doi: 10.1109/ICCCS49678.2020.9277353.

[46] K. Shu, L. Cui, S. Wang, D. Lee, and H. Liu, "dEFEND: Explainable Fake News Detection," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, Anchorage AK USA, Jul. 2019, pp. 395–405. doi: 10.1145/3292500.3330935.

[47] C. Zhang, A. Gupta, C. Kauten, A. V. Deokar, and X. Qin, "Detecting fake news for reducing misinformation risks using analytics approaches," *Eur. J. Oper. Res.*, vol. 279, no. 3, pp. 1036–1052, Dec. 2019, doi: 10.1016/j.ejor.2019.06.022.

[48] H. Liu, L. Wang, X. Han, W. Zhang, and X. He, "Detecting Fake News on Social Media: A Multi-Source Scoring Framework," in *2020 IEEE 5th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA)*, Chengdu, China, Apr. 2020, pp. 524–531. doi: 10.1109/ICCCBDA49378.2020.9095586.

[49] A. Uppal, V. Sachdeva, and S. Sharma, "Fake news detection using discourse segment structure analysis," in *2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, Noida, India, Jan. 2020, pp. 751–756. doi: 10.1109/Confluence47617.2020.9058106.

[50] R. K. Kaliyar, A. Goswami, P. Narang, and S. Sinha, "FNDNet – A deep convolutional neural network for fake news detection," *Cogn. Syst. Res.*, vol. 61, pp. 32–44, Jun. 2020, doi: 10.1016/j.cogsys.2019.12.005.

[51] S. Esmaeilzadeh, G. X. Peh, and A. Xu, "Neural Abstractive Text Summarization and Fake News Detection," *ArXiv190400788 Cs Stat*, Dec. 2019, Accessed: Sep. 18, 2021. [Online]. Available: http://arxiv.org/abs/1904.00788

[52] Y. Fang, J. Gao, C. Huang, H. Peng, and R. Wu, "Self Multi-Head Attention-based Convolutional Neural Networks for fake news detection," *PLOS ONE*, vol. 14, no. 9, p. e0222713, Sep. 2019, doi: 10.1371/journal.pone.0222713.

[53] E. Tacchini, G. Ballarin, M. L. Della Vedova, S. Moret, and L. de Alfaro, “Some Like it Hoax: Automated Fake News Detection in Social Networks,” *ArXiv170407506 Cs*, Apr. 2017, Accessed: Sep. 18, 2021. [Online]. Available: http://arxiv.org/abs/1704.07506

[54] S. Singhal, R. R. Shah, T. Chakraborty, P. Kumaraguru, and S. Satoh, “SpotFake: A Multi-modal Framework for Fake News Detection,” in *2019 IEEE Fifth International Conference on Multimedia Big Data (BigMM)*, Singapore, Singapore, Sep. 2019, pp. 39–47. doi: 10.1109/BigMM.2019.00-44.

[55] W. Antoun, F. Baly, R. Achour, A. Hussein, and H. Hajj, “State of the Art Models for Fake News Detection Tasks,” in *2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT)*, Doha, Qatar, Feb. 2020, pp. 519–524. doi: 10.1109/ICIoT48696.2020.9089487.

[56] Ghinadya and S. Suyanto, “Synonyms-Based Augmentation to Improve Fake News Detection using Bidirectional LSTM,” in *2020 8th International Conference on Information and Communication Technology (ICoICT)*, Yogyakarta, Indonesia, Jun. 2020, pp. 1–5. doi: 10.1109/ICoICT49345.2020.9166230.

[57] I. Kareem and S. M. Awan, “Pakistani Media Fake News Classification using Machine Learning Classifiers,” in *2019 International Conference on Innovative Computing (ICIC)*, Lahore, Pakistan, Nov. 2019, pp. 1–6. doi: 10.1109/ICIC48496.2019.8966734.

[58] T. Islam, 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)*, Dhaka, Bangladesh, May 2019, pp. 1–4. doi: 10.1109/ICASERT.2019.8934841.

[59] Wohlin, C., Runeson, P., Host, M., Ohlsson, M. C., Regnell, B., & Wessle’n, A. (2012). Experimentation in software engineering. Springer Science & Business Media.

[60] Barua, R., Maity, R., Minj, D., Barua, T., & Layek, A. K. (2019, July). F-NAD: an application for fake news article detection using machine learning techniques. *In 2019 IEEE Bombay Section Signature Conference (IBSSC)* (pp. 1-6). IEEE.

[61] Yin, R. K. (2001). Estudo de caso: planejamento e metodos. 2a edicao. SAGE

[62] Collins 2017 word of the year shortlist, available at << https://www.collinsdictionary.com/word-lovers-blog/new/collins-2017-word-of-the-year-shortlist,396, HCB.html >>, last accessed on 02-12-2021 at 11:44 AM.

[63] Pew Research Center, available at << https://www.pewresearch.org/journalism/2021/01/12/news-use across-social-media-platforms-in-2020/>>, last accessed on 02-12-2021 at 11:45 AM.

[64] World population- Wikipedia, available at << https://en.wikipedia.org/wiki/World_population >>, last accessed on 02-12-2021 at 11:55 AM.

[65] Internet users in the world 2021|Statista, available at <<https://www.statista.com/statistics/617136/digital-population-worldwide/>>, last accessed on 02-12-2021 at 11:57 AM.

[66] BanFakeNews A Dataset for Detecting Fake News in Bangla from Kaggle, available at << https://www.kaggle.com/cryptexcode/banfakenews >>, last accessed on 02-12-2021 at 12:57 AM.

[67] Ethnologue, “List of 200 most spoken languages,”, available at << https://www.ethnologue.com/guides/ethnologue200 >>, last accessed on 07-12-2021 at 11:57 AM.

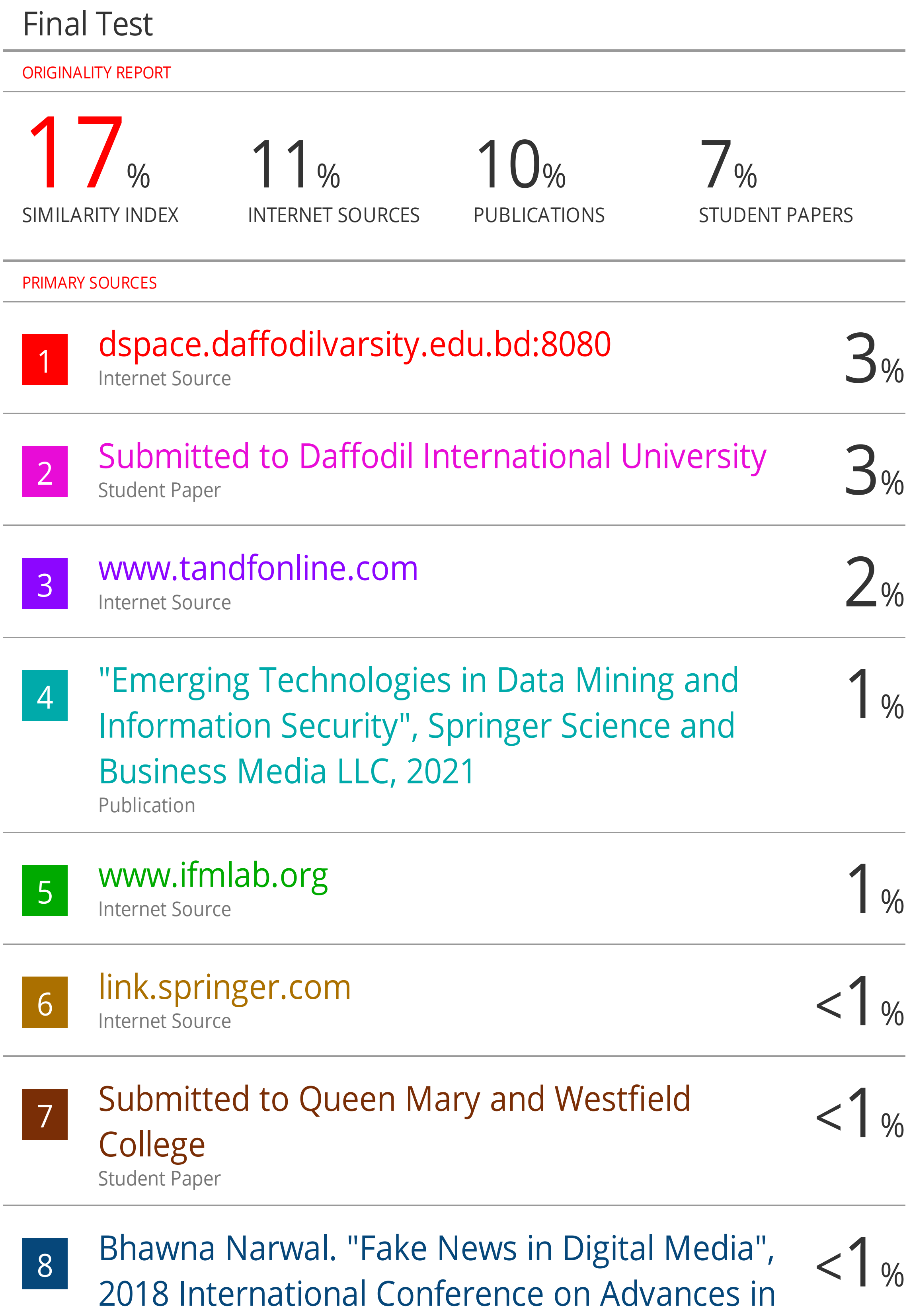
APPENDIX

Table 2.5. Articles, algorithm, references

Article	Algorithm	Reference
A Gaussian Naive Bayesian Classifier for Fake News Detection in Bengali	Gaussian Naive Bayes Classifier	[22]
A substantial procedure of fake news detection utilizing Ensemble Voting Classifier and comparison with other classifiers	Ensemble Voting Classifier [Top 3 ML with best accuracy MLP, LR and X-Gradient Boosting]	[15]
Programmed Detection of Parody in Bangla Archives: A CNN Approach Based on Crossbreed Include Extraction Show	Convolutional neural network	[30]
Bangla Content Assumption Investigation Utilizing Supervised Machine Learning	Proposed Method (BTSC algorithm (Bangla Text Sentiment Analysis)	[35]
Detection of Bangla Fake News using MNB and SVM Classifier	Support Vector Machine	[33]
Identification of Synthetic Sentence in Bengali News using Hybrid Approach	Proposed Hybrid method	[34]
Automatic Identification of Fake News Using Deep Learning	Long Short-Term Memory	[36]
A Study on Natural Language Processing for Fake News Discovery	Long Short-Term Memory	[37]
A gathering machine learning approach through effective feature extraction to classify fake news	Decision Tree Random Forest, Extra tree classifier	[24]
A Study on Programmed Fake News Recognizable proof Strategies for Online and Socially Delivered Information	Convolutional neural network	[38]
An Efficient FAKE NEWS DETECTOR	Grey Wolf Optimization Algorithm	[39]
A Deep Transfer Learning Approach for Fake News Detection	Long Short-Term Memory	[40]
Convolutional neural organize with edge misfortune for fake news discovery	Convolutional neural network	[25]
BDANN: BERT-Based Space Adjustment Neural Network for Multi-	Artificial Neural Network	[41]

Modal Fake News Detection		
Identifying Fake News utilizing Machine Learning and Deep Learning Algorithms	Support Vector Machine	[7]
Detecting Fake News with Tweets' Properties	Random Forest	[42]
Deep Learning Algorithms for Detecting Fake News in Online Text	Convolutional neural network	[43]
Deep neural approach to Fake-News identification	Long Short-Term Memory	[44]
DeepNet: An Proficient Neural Network for Fake News Location utilizing News-User Engagements	Artificial Neural Network	[45]
dEFEND: Explainable Fake News Detection	Proposed Method	[46]
Recognizing fake news for decreasing deception risks using analytics approaches	K-mean	[47]
Identifying Fake News on Social Media: A Multi-Source Scoring System	Proposed method	[48]
Fake News Discovery Through Multi-Perspective Speaker Profiles	Long Short-Term Memory	[29]
Fake news discovery utilizing talk section structure investigation	Gated recurrent unit	[49]
Fake News Recognizable proof on Twitter with Hybrid CNN and RNN Models	Long Short-Term Memory	[28]
FAKEDETECTOR: Viable Fake News Detection with Deep Diffusive Neural Network	Proposed method	[13]
FakeNewsTracker: an apparatus for fake news collection, detection, and visualization	Support Vector Machine	[19]
FNDNet – A deep convolutional neural organize for fake news location	Proposed method	[50]
Crossover Machine-Crowd Approach for Fake News Location	Neural Network	[16]
MYTHYA: Fake News Locator, Real-Time News Extractor, and Classifier	Proposed method combined (GBDT+CNN)	[6]
Neural Abstractive Content Summarization and Fake News	Long Short-Term Memory	[51]

Discovery		
Multimodal fake news location employing a Cultural Algorithm with situational and regulating information	Cultural Algorithm	[21]
Different features-based approaches for automatic fake news discovery on social systems utilizing deep learning	Long Short-Term Memory	[26]
Self Multi-Head Attention-based Convolutional Neural Systems for fake news location	Convolutional neural network	[52]
Some Like it Hoax: Automated Fake News Detection in Social Networks	Harmonic BLC algorithm	[53]
SpotFake: A Multi-modal System for Fake News Discovery	Proposed multimodal framework	[54]
State of the Craftsmanship Models for Fake News Discovery Errands	XLNET	[55]
Synonyms-Based Enlargement to Progress Fake News Discovery utilizing Bidirectional LSTM	LONG SHORT-TERM MEMORY	[56]
Pakistani Media Fake News Classification utilizing Machine Learning Classifiers	K-nearest Neighbor	[57]
Untrustworthy Clients Location in Social Media: Profound Learning Methods for Programmed Discovery	Support Vector Clustering Support Vector Machine	[27]
Utilizing Social Systems to Distinguish Pernicious Bangla Content Substance	Multinomial Naive Bayes	[58]
TI-CNN: Convolutional Neural Systems for Fake News Discovery	Convolutional neural network	[5]



Final Test

ORIGINALITY REPORT

17%	11%	10%	7%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	3%
2	Submitted to Daffodil International University Student Paper	3%
3	www.tandfonline.com Internet Source	2%
4	"Emerging Technologies in Data Mining and Information Security", Springer Science and Business Media LLC, 2021 Publication	1%
5	www.ifmlab.org Internet Source	1%
6	link.springer.com Internet Source	<1%
7	Submitted to Queen Mary and Westfield College Student Paper	<1%
8	Bhawna Narwal. "Fake News in Digital Media", 2018 International Conference on Advances in	<1%