# Shahjalal University of Science and Technology
## Department of Computer Science and Engineering


# Unraveling the Web of Information: A Comprehensive Study on Bangla Misinformation, Disinformation, Satire and Fake News


HRITHIK MAJUMDAR SHIBU
Reg. No.: 2018331052
$4^{th}$ year, $1^{st}$ Semester

NASRULLAH SAMI
Reg. No.: 2018331036
$4^{th}$ year, $1^{st}$ Semester

Department of Computer Science and Engineering

**Supervisor**
MAHRUBA SHARMIN CHOWDHURY
Assistant Professor
Department of Computer Science and Engineering


21st October, 2023

# Unraveling the Web of Information: A Comprehensive Study on Bangla Misinformation, Disinformation, Satire and Fake News

A Thesis submitted to the Department of Computer Science and Engineering, Shahjalal University of Science and Technology, in partial fulfillment of the requirements for the degree of Bachelor of Science in Computer Science and Engineering.

By

Hrithik Majumdar Shibu
Reg. No.: 2018331052
$4^{th}$ year, $1^{st}$ Semester

Nasrullah Sami
Reg. No.: 2018331036
$4^{th}$ year, $1^{st}$ Semester

Department of Computer Science and Engineering

**Supervisor**

Mahruba Sharmin Chowdhury
Assistant Professor
Department of Computer Science and Engineering

21st October, 2023

# Recommendation Letter from Thesis/Project Supervisor

The thesis/project entitled *"Unravelling the web of information: A comprehensive study on Bangla Misinformation, Disinformation, Satire and Fake news Classification"* submitted by the students

1. Hrithik Majumdar Shibu
2. Nasrullah Sami

is under my supervision. I, hereby, agree that the thesis/project can be submitted for examination.

Signature of the Supervisor:

Name of the Supervisor: Mahruba Sharmin Chowdhury

Date: 21st October, 2023

# Certificate of Acceptance of the Thesis/Project

The thesis/project entitled *Unravelling the web of information: A comprehensive study on misinformation, disinformation, satire and fake news classification* submitted by the students

1. Hrithik Majumdar Shibu
2. Nasrullah Sami

on 21st October, 2023, hereby, accepted as the partial fulfillment of the requirements for the award of their Bachelor Degrees.

| | | |
|---|---|---|
| Head of the Dept. | Chairman, Exam. Committee | Supervisor |
| Md Masum | Dr.Sadia Sultana | Mahruba Sharmin Chowdhury |
| Professor | Associate Professor | Assistant Professor |
| Department of Computer Science and Engineering | Department of Computer Science and Engineering | Department of Computer Science and Engineering |

# Abstract

The modern information landscape is rife with misinformation, disinformation, satire and fake news, posing formidable challenges to information consumers and social harmony. This pervasive spread within the digital realm has ignited a demand for robust classification strategies to combat their deleterious effects on public discourse and decision-making. Moreover in the past works, researchers have classified if a Bengali news is fake or not, satire or real which concluded to binary classification process. In this work, We are presenting a Bangla dataset that includes multiple class and classifies information into categories such as misinformation, disinformation, satire, and real news. In addition to that, we are also proposing a multi-class classification model architecture named "Multi-Class Siamese Neural Network" for Bangla language which will classify if a news is misinformation, disinformation or satire which are all different categories of fake news and also it is classify real news. It enhances classification models and offers insights into the nature of deception in online information, which are crucial for media literacy efforts, content moderation strategies, and developing tools for critical evaluation. The findings of this research can empower individuals to navigate the digital world with greater discernment and contribute to a more informed and resilient information ecosystem. This research has also some advancements in the realm of Bangla fake news detection and classification process beyond the old traditional architectures and approaches.

**Keywords:** Misinformation, Disinformation, Satire, Multi-class, Siamese Neural Network

# Acknowledgements

We would like to thank the Department of Computer Science and Engineering, Shahjalal University of Science and Technology, Sylhet 3114, Bangladesh, for supporting this research. Our gratitude to our esteemed supervisor Mahruba Sharmin Chowdhury for her excellent guidance and assistance with our work is immeasurable. We are also indebted to our external advisor, Md Saiful Islam, for his unwavering support and guidance throughout the completion of this research.

# Contents

# List of Tables

# List of Figures

# Chapter 1

# Introduction

*"People like to say that the conflict is between good and evil. The real conflict is between truth and lies."*

The alarming surge in fake news propagation has become a pressing concern in today's digital age. With the rapid dissemination of information through various online platforms, the prevalence of satire, misinformation and disinformation has reached "unprecedented" levels. Detecting fake news has become a crucial endeavor, necessitating the development of advanced technological tools and critical media literacy education to empower individuals in distinguishing between accurate information and deceptive narratives. As society grapples with this escalating challenge, collaborative efforts are essential to safeguard the integrity of information and maintain a well informed society.

## 1.1 Getting Started

The internet has significantly changed the way people receive news, shifting from traditional media sources like radio and television to online sources like social networking sites. This has led to concerns about the spread of misleading material and the potential for profit for internet publishers. In today's world , information disorder has been an important issue and attracts increasing attention in recent years. The openness and anonymity of social media makes it convenient for users to share and exchange information, but also makes it vulnerable to nefarious activities. The term "fake news" describes news items or articles that are intentionally created false or produced and circulated through a variety of media outlets, including social media, news websites and even conventional printed newspapers.

The growth of digital platforms and social media has created an unprecedented problem for classifying Misinformation, Disinformation, Satire, and Fake News in today's quickly changing information environment. The border between truthful reporting and misleading content has blurred as society relies more and more on internet sources for news and information, potentially exacerbating the negative effects of these false narratives. This calls for a thorough comprehension of the subtle variations between these terminologies and their ramifications. In this discourse, we explore the complex classification of misinformation, disinformation, satire, and fake news, shining light on their distinctive traits, the inspirations behind their production, and the crucial function they serve in influencing discourse and public perception. We want to give people the tools they need to successfully navigate this challenging environment by exposing the complicated network of lies that permeates our information ecology.

## 1.2 Knowing The Difference

In an age defined by the rapid dissemination of information, the boundaries between fact and fiction have become increasingly blurred. The digital revolution has ushered in an era where news, opinions, and narratives traverse the globe in mere seconds, offering unprecedented access to knowledge, but also giving rise to a formidable challenge: the proliferation of Misinformation, Disinformation, Satire and Fake News.

Misinformation, Disinformation, Satire, and Fake News represent distinct yet interwoven threads in the intricate of information dissemination. *Misinformation* encompasses inaccuracies and false beliefs spread inadvertently, often due to incomplete understanding or genuine errors. On the other hand, *Disinformation* involves the deliberate spread of false or misleading information, often driven by ulterior motives such as political agendas, commercial gains, or ideological manipulation. While, *Satire* is intended to provoke thought and critique through humor, occasionally blurs the line between jest and reality, contributing to the challenge of discerning genuine information from satire. *Fake news* is a term that has gained considerable prominence, encapsulates a broad range of fabricated stories presented as genuine news, capitalizing on the trust invested in reputable sources. To stand against such phenomenon, a comprehensive understanding of the dynamics at play within this landscape is crucial to devising effective strategies for mitigating their adverse effects.

## 1.3 Problem Statement

This research paper embarks on a multidimensional exploration of misinformation, disinformation, satire and fake news, delving into their origins, mechanisms, consequences, and potential nuances. By unraveling the complex web woven by these phenomena, we aim to shed light on the psychological and technological underpinnings that drive their propagation. The previous studies were all based on specifically fake news or satire. This brings a shortcoming in the realm of information sharing which is that a fake news can be misinformation or disinformation or satire.

The difficulty of categorizing and differentiating between misinformation, disinformation, satire, and fake news has emerged as a crucial social concern in an era marked by the rapid distribution of information through digital platforms and social media. The widespread dissemination of false news in these categories has the potential to erode faith in reliable sources, decrease the ability to make informed decisions, and create social polarization. Our ability to handle the varied nature of the problem is hampered by the fact that current classification approaches frequently fail to capture the nuanced nuances and developing tactics used to distribute false narratives. Therefore, it is imperative to provide a thorough and complex framework for correctly classifying and differentiating between these many types of erroneous information. In order to better comprehend misinformation, disinformation, satire, and false news and to develop measures to lessen their negative effects on society, this study intends to investigate and establish a reliable classification approach.

## 1.4 Our Approach to the Solution

First of all, there is no existing work regarding the multi-dimensional and multi-class classification system for Bengali News. The novelty of our task lies in the making of a multidimensional dataset containing misinformation, disinformation and satire data in forms of fake news and real news for Bengali language.

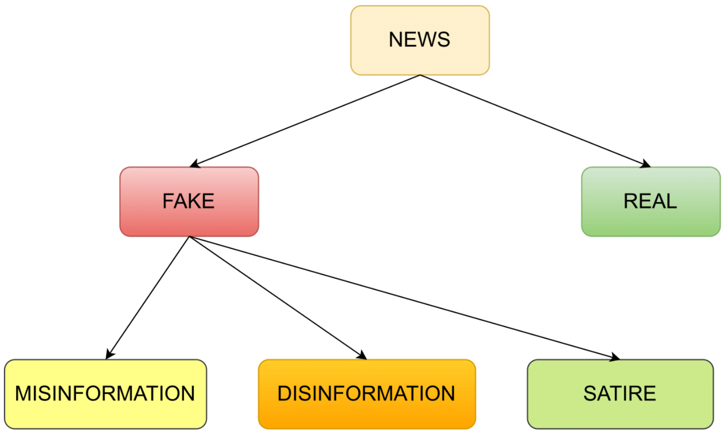


Figure 1.1: Fake News Classification

Next, based on the collected dataset, we are proposing a new multi-class classification architecture named as Siamese Neural Network [1] which is going to classify all fake news as misinformation, disinformation, satire and real news etc. The headline and the body text of a fake news will be considered as important features for the task. In our system, we are providing three kinds of features into the system architecture which includes news headlines, news bodies and headline-body relationships to justify the news as misinformation, disinformation and satire in forms of fake news.

As a whole, in this research, we are introducing a multidimensional dataset which is based on a proposed multi-class classification model. In a word, there is no such dataset or model which can classify if a fake news is misinformation or disinformation or satire. Siamese Networks are efficient for this kind of tasks due to their capacity to learn meaningful embeddings for text pairs, which is required for tasks involving similarity measurement, verification, and ranking. The Multi-Class Siamese Neural Network, a tested framework for determining similarity works to its full potential in a multi-dimensional setting. This design naturally grasps the deep links that distinguish texts across many categories by managing shared and task-specific routes. As we embark on this journey of classification, our aim is to illuminate the contours of an ever-evolving ecosystem of information, fostering a more informed and critically engaged society.

# Chapter 2

# Background Study

## 2.1 Definition

Misinformation, disinformation and satire are different kinds of fake news that everyone should be careful of. Regardless of where we find our news, there is a big possibility that we will be faced with unethical journalism, fake news and lies that pose as facts. That's why, it has become inevitable to know what satirical news, misinformation and disinformation are, what influence it has on the reader and how to spot it. This is because what we read in the media has a huge effect on how we look at the world around us. So, let's take a deeper look into what we are going to handle and how it will be done.

### 2.1.1 Misinformation

Misinformation refers to the content that is taken out of context and presented as facts without realizing it's factual impacts on society, a problem prevalent in the media industry, such as the pandemic [2]. It is easily shared, even if readers are not entirely certain, and contains false information without the intention to cause harm. For instance, a blogger may have read an article's headline and wrote an opinion piece without incorporating facts and research. There are different kinds of misinformation on social media as **Click-baits, Misleading Titles, Propaganda etc**. Click-baits are a kind of sensationalized material that are meant to focus readers' interest by appealing to their emotions or curiosity. Misleading titles refer to a well-written and factually accurate news, but that might convey the wrong impression if the headline is deceptive. Propaganda is biased and potentially misleading information that is disseminated through the media with an intention of persuading its target audience to adopt certain viewpoints.

### 2.1.2 Disinformation

Disinformation is a kind of piece of content that intentionally created to deceive and manipulate the public for an entity's agenda [2]. It can be found in blogs run by political parties without public awareness. People believe what they want to believe, leading to widespread disinformation surrounding politics, pandemics and racial issues. Readers should be cautious of this type of news, as it can cause significant damage. For example, "In 1946, **The R.J. Reynolds Tobacco Company** began making a bold claim in its advertisements, saying,'More doctors smoke Camels than any other cigarette!'" The main purpose of disinformation is to deceive people in general. There are different kinds of disinformation around us as **Hoaxes, falsified stories, click-bait** etc. Disinformation is harmful on social media because, the vast volume of information available and the short attention spans of readers can allow it to spread unchecked.

### 2.1.3 The Supply Chain

According to our conceptualization, the production and/or exchange of false information is the supply chain of misinformation and disinformation [3]. Both methods unintentionally spread false information. In other words, people who are involved in the production and/or transmission of false information are unaware that it is false. On the other hand, those who are involved in the production and/or spread of disinformation are aware that it is false and intend to mislead the audience. However, actors may disseminate purposely false information (i.e., misinformation) even if they are not aware that it is false. Therefore, having the knowledge that information is false at the time of creation or dissemination constitutes having the intention to spread false information. This suggests that the source of creation or dissemination affects intention. In other words, misinformation can become disinformation and vice versa because the intentions of the actors involved in its creation and dissemination can alter. The supply chain is represented below as the process of fabricating and disseminating lies, demonstrating that supply is always a multi-step process.

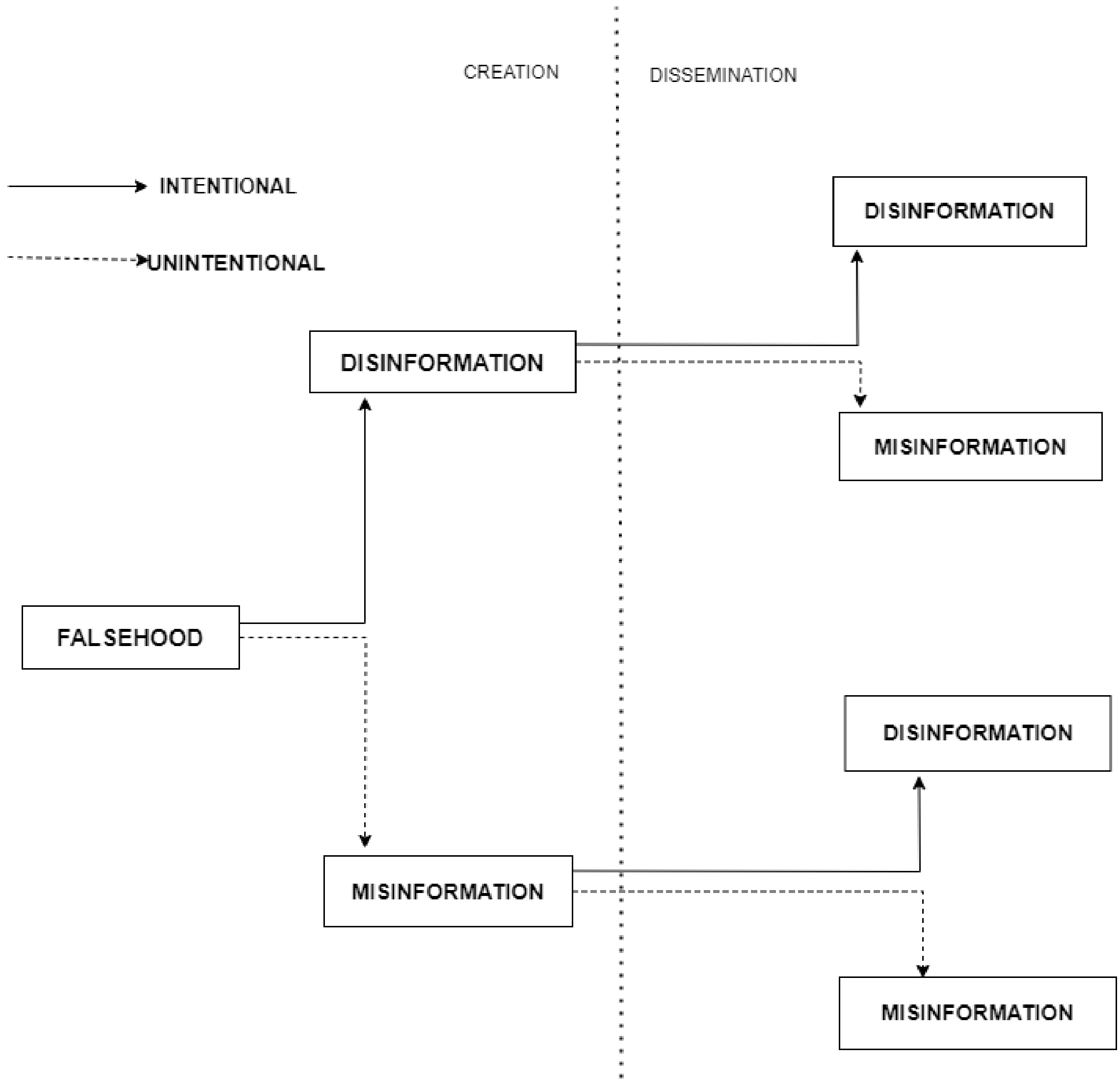


Figure 2.1: The Supply Chain
[3]

### 2.1.4 Satire

Satire is the humor of news which comes from its deadpan, sardonic tone and imitation of real news sources. Satire like this runs the risk of misinterpreted as reality because not every reader will get the irony. Despite not being intended to be manipulative, satire could, if interpreted incorrectly, have the same effect as false news. Satire is a form of literary, artistic, or rhetorical expression that uses humor, irony, exaggeration, or ridicule to criticize and comment on various aspects of society, politics, culture, or human behavior. Satire can be used as a creative tool to convey a message or critique a certain topic by mocking and frequently exaggerating its defects, contradictions or absurdities.

### 2.1.5 Fake News

Since misinformation, disinformation and satire are essentially false facts, one might wonder if it isn't just fake news as well. But, the actual situation is opposite. When it comes to fake news, the intention is all about pushing an agenda and **purposefully crafted, sensational, emotionally charged, misleading or totally fabricated information** that mimics the form of mainstream news. Fake news is also referred to as false information buy industry professionals and its main focus is to damage a person's reputation or evoke a specific emotion within the reader. Fake news often contains multi-modal information, making it crucial to improve detection performance. Current work focuses on extracting linguistic features like lexical and visual cues, learning neural language features using neural networks like CNNs, and extracting visual scene graphs to discover common sense knowledge. Advances also aims to extract visual scene graphs from images to enhance structured scene graphs.

## 2.2 Historical Context

The historical context demonstrate that misinformation, disinformation, satire and fake news are not new phenomena. Rather, they have existed and evolved throughout history. Studying their historical roots can provide valuable insights into their impact on society, information dissemination, and the role of media.

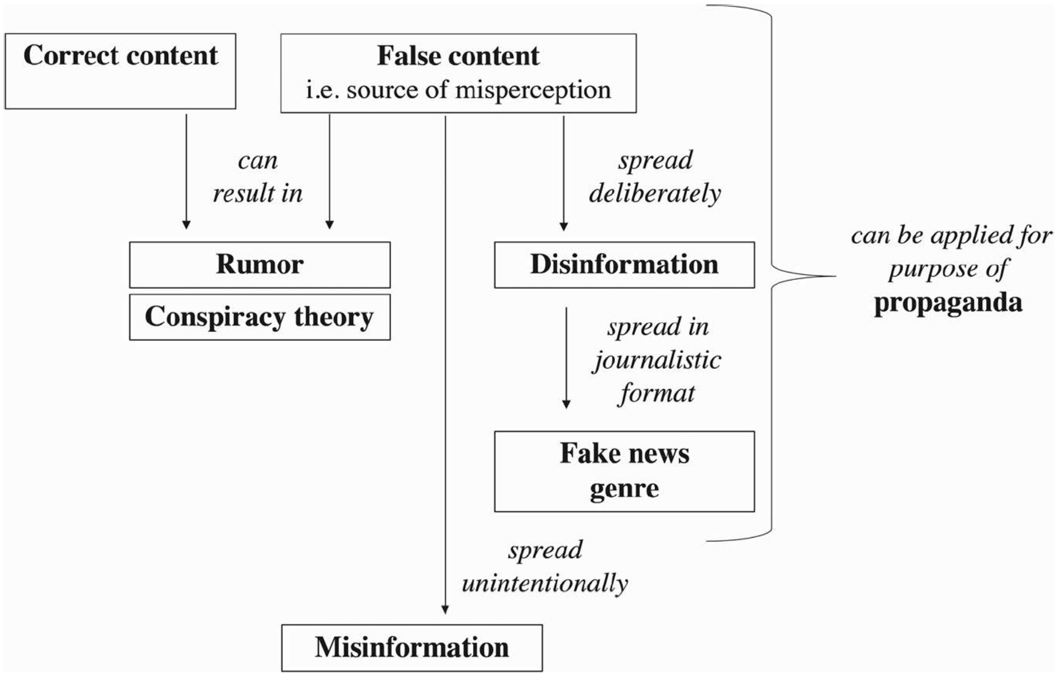


Figure 2.2: Relevant Concepts of Falsehood

### 2.2.1 Misinformation and Disinformation

Misinformation and disinformation have existed throughout the history for decades. Propaganda has been used to disseminate false information during times of war and conflict in order to sway to public opinion about the adversary and raise national spirit.

### 2.2.2 Satire

As a kind of of humor and social-satire , satire has been a well-established genre since past times to ridicule political figures and social conventions. Satire has frequently been as a communication style that is typically associated with aggression, judgement, mockery, play, laughter, and references to societal norms. [4]

### 2.2.3 Fake News

Although the phrase 'Fake News' has become more popular recently, the phenomenon itself has a long history. In the past, falsified information has been spread via a variety of methods, frequently for political or social reasons. In order to draw readers, so called "penny-press" publications in the US during the 19th century, sometimes ran sensational and exaggerated articles. Throughout numerous historical eras, the dissemination of falsified information has been a repeating topic.

## 2.3 Traditional Methods

It is a difficult undertaking to categorize misinformation, disinformation, false news, and satire, and doing so needs careful analysis of the textual and contextual characteristics that set these categories apart. To solve this problem, many machine learning techniques can be used. I'll list some of the methods and techniques that are frequently employed for text classification in this situation below:

### 2.3.1 Algorithms for Traditional Machine Learning

#### 2.3.1.1 Naive Bayes Classifier

Misinformation, disinformation, satire, and false news can all be accurately categorized using the probabilistic classifier known as the Naive Bayes method. Naive Bayes uses the independence premise to categorize text samples by modeling the conditional likelihood of a given text belonging to a particular category based on the appearance of its terms. This algorithm can examine the frequency of particular terms or phrases that are frequently indicative of erroneous information in the context of categorizing content linked to disinformation. Naive Bayes, for instance, may recognize linguistic patterns frequently connected to satire or hilarious purpose, assisting in the distinction between satirical content and really deceptive storylines. Naive Bayes may, however, have trouble capturing complicated word associations and sophisticated contextual information. Despite this drawback, Naive Bayes offers a simple and computationally effective method for the preliminary categorization of misinformation, disinformation, satire, and false news when combined with the right pre-processing methods and feature selection. They can be applied to binary or multi-class classification and they make the assumption that characteristics are independent. The basic structure and general equation of Naive Bayes classifier is given below:

$$P(C_k|x_1, x_2, ..., x_n) = \frac{P(C_k) \cdot P(x_1|C_k) \cdot P(x_2|C_k) \cdot ... \cdot P(x_n|C_k)}{P(x_1) \cdot P(x_2) \cdot ... \cdot P(x_n)}$$

The mathematical intuition of NB Classifier depends on the following terms:

1. The posterior probability of an event given another event

2. The likelihood of an event given another event

3. The prior probability of an event

4. The probability of another event

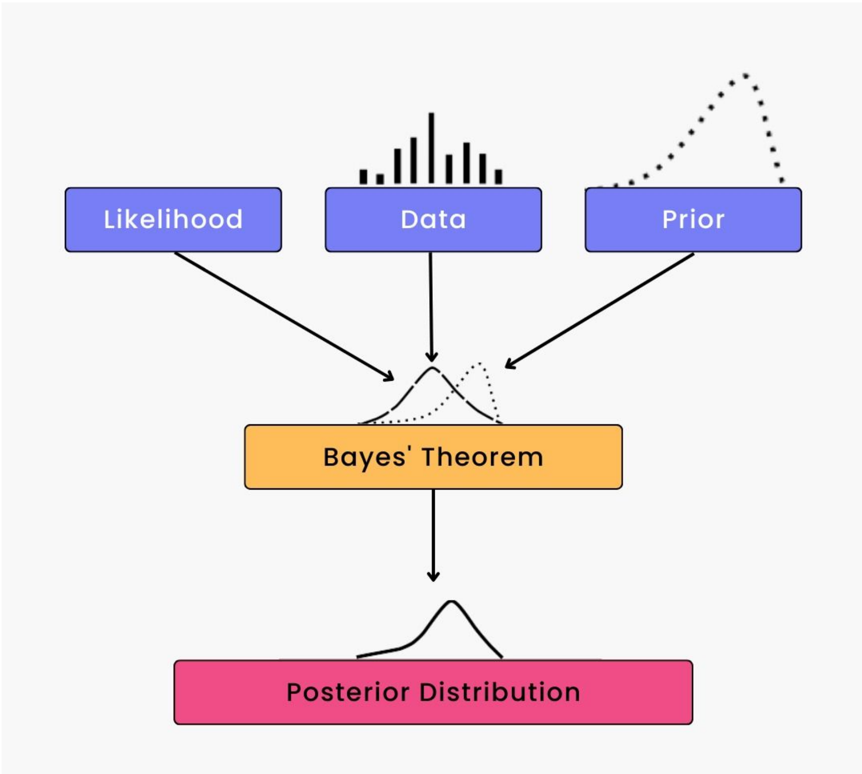


Figure 2.3: Naive Bayes Classifier

#### 2.3.1.2 Support Vector Machines

By successfully defining complex decision boundaries inside high-dimensional feature spaces, Support Vector Machines(SVMs) provide a potent method for categorizing misinformation, disinformation, satire, and false news. In the context of text classification, SVMs seek to identify a hyperplane that maximizes the margin between instances of various categories while optimally separating them. Using methods like TF-IDF or word embeddings, textual data can be converted into numerical feature vectors that SVMs can use to capture complex language patterns that distinguish between true news and false information. SVMs may generalize effectively to new, untested data and are particularly good at handling binary and multi-class classification jobs. SVMs are a useful tool for distinguishing between different types of information based on the inherent structures and linguistic characteristics present in the text due to their capacity for handling high-dimensional data and flexibility in using different kernel functions, such as linear, polynomial, or Radial Basis Function(RBF).

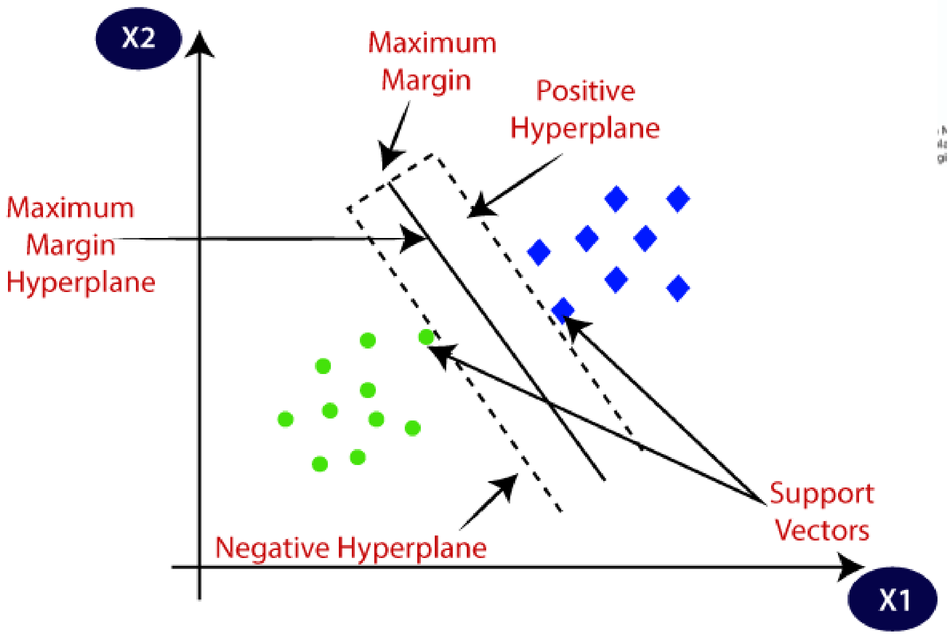


Figure 2.4: Support Vector Machine

#### 2.3.1.3 Decision Trees and Random Forest

Misinformation, disinformation, satire, and fake news may all be efficiently categorized using textual features using decision tree and random forest algorithms. Each node in a decision tree represents a feature, and each branch corresponds to one of its potential values. This results in a hierarchical structure of decisions. They can record linguistic clues like keywords and grammatical patterns to distinguish between various content categories. On the other hand, Random Forests are ensemble algorithms that combine different Decision Trees, lowering over-fitting and boosting generalization. Random Forests can collectively capture the different nuances and contexts that distinguish between misinformation, disinformation, satire, and false news by training on a variety of subsets of the data. These algorithms automatically handle feature selection, and their openness makes it possible to comprehend the reasoning behind choices, assisting in the discovery of textual cues particular to each category.

### 2.3.2 Feature Extraction and Representation

#### 2.3.2.1 Term Frequency-Inverse Document Frequency

The word relevance within a document in relation to its frequency across the entire corpus is captured by the feature extraction and representation method called TF-IDF (word Frequency-Inverse Document Frequency). It is a typical method of text representation. We can measure the importance of words within documents by using methods like TF-IDF(Term Frequency-Inverse Document Frequency), which helps us uncover key terms that could distinguish between content that is false and that is true. A typical diagram of TF-IDF is shown below:

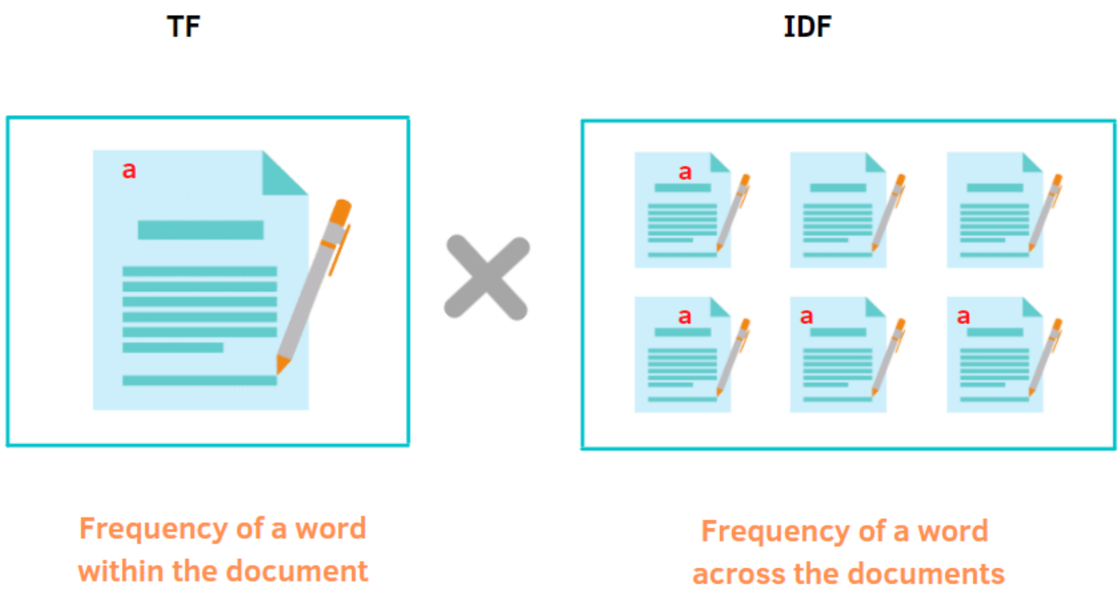


Figure 2.5: Term Frequency-Inverse Document Frequency

#### 2.3.2.2 Word Embeddings

Embeddings in Word Word2Vec, GloVe, or FastText are examples of pre-trained word embeddings that can give dense vector representations of words and capture semantic links between words. They provide dense vector representations of words that capture their semantic links, allowing the model to pick up on finer contextual details. Contextual embeddings like BERT(Bidirectional Encoder Representations from Transformers), which contain bidirectional information, can be used to better understand complex connections inside texts. These methods enable the model to classify material more precisely and subtlety by enabling it to comprehend not just the language subtleties but also the various tones, contexts, and linguistic structures that underlay misinformation, disinformation, satire, and fake news.

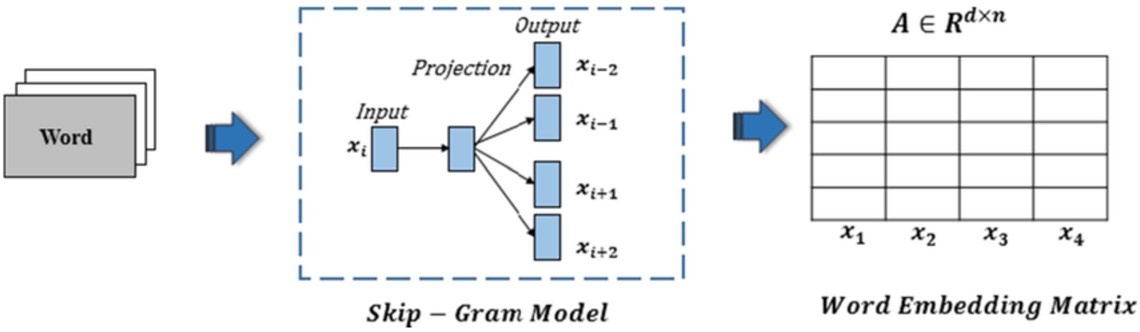


Figure 2.6: Word Embeddings Basic Diagram

1. **Continuous Bag Of Words(CBOW):** The CBOW method's objective is to anticipate a target word based on the context words around it within a predetermined window size. It works by instructing a neural network to reduce the distance between anticipated and actual phrases, so teaching the network to correlate words that frequently occur in similar circumstances. As a result, words with similar meanings or usage tend to be closer together in the vector space, creating embedding vectors. For applications like sentiment analysis, machine translation, and information retrieval, CBOW is particularly good at capturing the syntactic and semantic links between words.

2. **Skip-Gram Method** In problems involving natural language processing, the Skip-gram method is a well-liked approach for creating word embeddings, a representation of words in a continuous vector space. The Skip-gram technique, created as a component of the

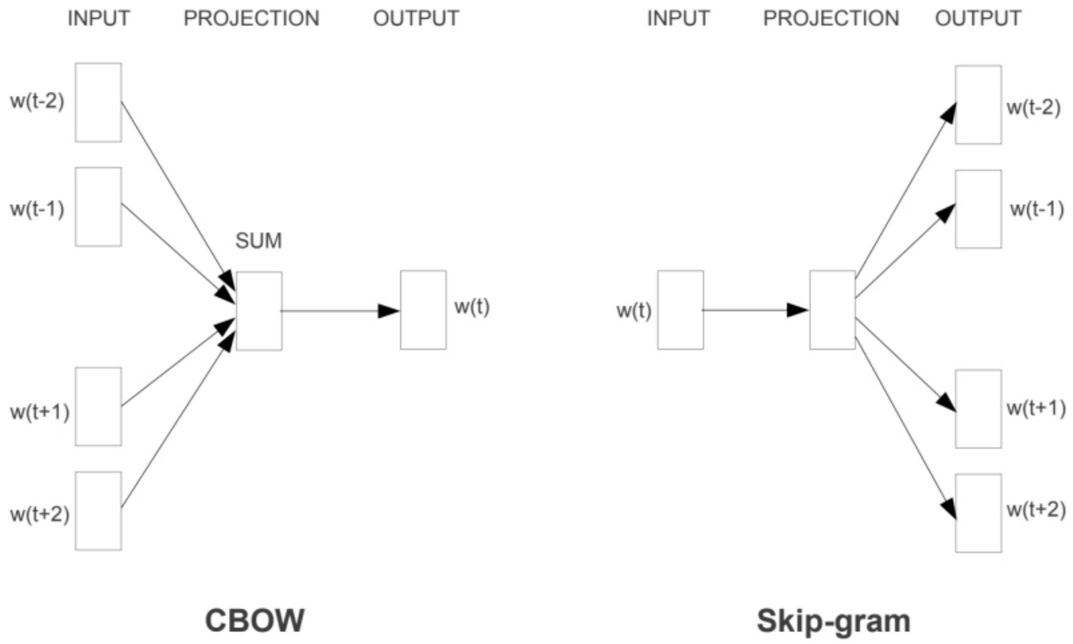


Figure 2.7: CBOW and Skip-Gram Method

Word2Vec paradigm, focuses on anticipating the context words given a target word. Skip-gram seeks to capture the contextual links between words by learning to anticipate the surrounding words based on a center word, in contrast to the Continuous Bag of Words (CBOW) technique, which predicts a target word from its context words. The model can learn representations that encode semantic links and similarity between words using this technique. The Skip-gram model captures the distributional semantics of words by training on huge text corpora, placing related words close to one another in the vector space. Due to its capacity to capture subtle semantic correlations between words, Skip-Gram embeddings have demonstrated effectiveness in a variety of natural language processing applications, including machine translation, sentiment analysis, and named entity recognition.

## 2.4 Neural Networks

Due to the complexity involved, categorizing various sorts of information, such as fake news, misinformation, disinformation and satire, call for a sophisticated approach. For such tasks, neural networks are frequently used, and the following list of neural network types can be used:

### 2.4.1 Recurrent Neural Network(RNN

RNNs and their variations, such as Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU), operate well with sequential data, such as text. They are appropriate for language-related tasks because they can record contextual information and word associations. RNNs can identify various forms of content by taking into account the coherence and flow of the information, which may be essential in separating satire from misinformation, for example. RNNs, in contrast to conventional feed-forward neural networks, feature connections that loop around on themselves, enabling them to keep track of the context or memory of prior inputs. RNNs may detect temporal connections and patterns in the data because of this looping mechanism. A simple architecture of Recurrent Neural Network is given below:

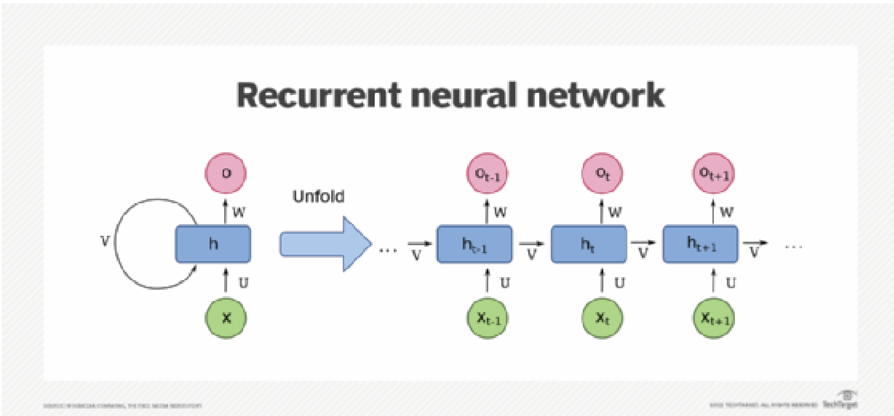


Figure 2.8: Recurrent Neural Network

### 2.4.2 Transformers

The field of natural language processing has been transformed by the use of Transformers, particularly models like BERT(Bidirectional Encoder Representations from Transformers) and its variations. These models can recognize complex links between words and phrases since they have already been pre-trained on vast volumes of text data. A BERT model can be accurately classified by understanding the intricacies of various forms of information by fine-tuning on a particular classification assignment. Transformers use attention mechanisms to allow the model to focus on specific parts of the input sequence. This can be beneficial for tasks where different parts of the text hold varying degrees of importance in determining the content type.

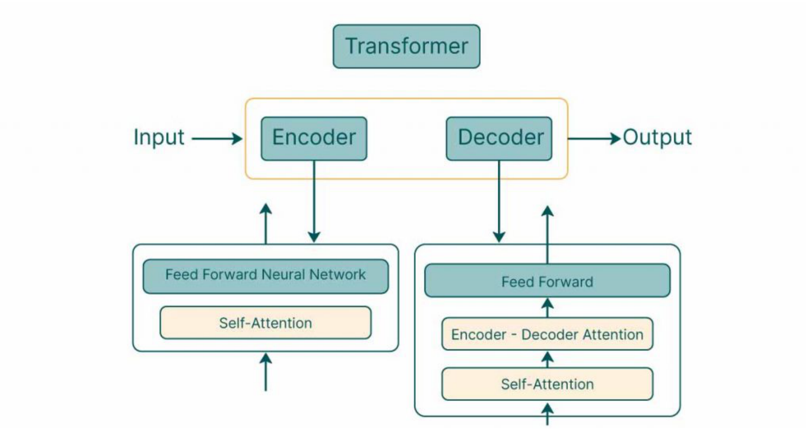


Figure 2.9: Transformers Architecture

### 2.4.3 Hybrid Models

Combining various neural network types can produce results that are more reliable and precise. For instance, a hybrid model might combine an LSTM to collect contextual information with a CNN to capture local features in text. These models can benefit from the advantages of various architectures to execute categorization tasks more effectively. Moreover, to increase classification accuracy, ensemble methods combine many neural networks (or other models). Each network may specialize in identifying particular traits of various content kinds, enabling a deeper comprehension of the input data. A perfect hybrid model architecture for text classification has been done by Priyanka et al. [5]

### 2.4.4 Convolutional Neural Networks

Convolutional Neural Networks (CNNs) have attracted a lot of attention in recent years due to their effectiveness in a variety of natural language processing applications. By considering words or letters as "filters" and using convolution operations to detect regional patterns and relationships, CNNs, which were initially developed for image analysis, have been repurposed to handle sequential

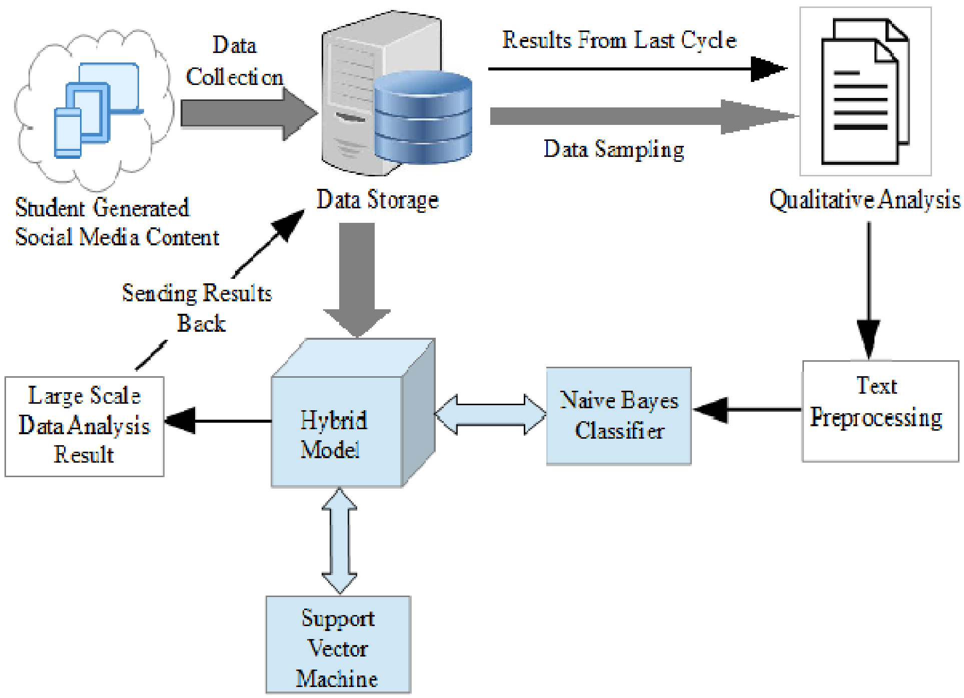


Figure 2.10: Hybrid Models

data, such as text. Although their main application is image recognition, CNNs can also be used to classify text. By considering text as an image and using each word or character as a pixel, CNNs can be used, for example. This method can discover regional patterns in the text, which may help in identifying the distinctive traits of satire, fake news, or other sorts of content. Kawsari et al. [6] has done a tremendous work in text classification using CNN.

## 2.5 Rule-Based Methods

Misinformation, disinformation, satire and fake news are classified using rule-based systems by Yulkani et al. [7] that use predetermined sets of rules, patterns, heuristics or keywords to distinguish between these categories. Even though rule-based techniques might not be as adaptable as machine learning models, they can be good at identifying particular linguistic or stylistic traits unique to each form of content.

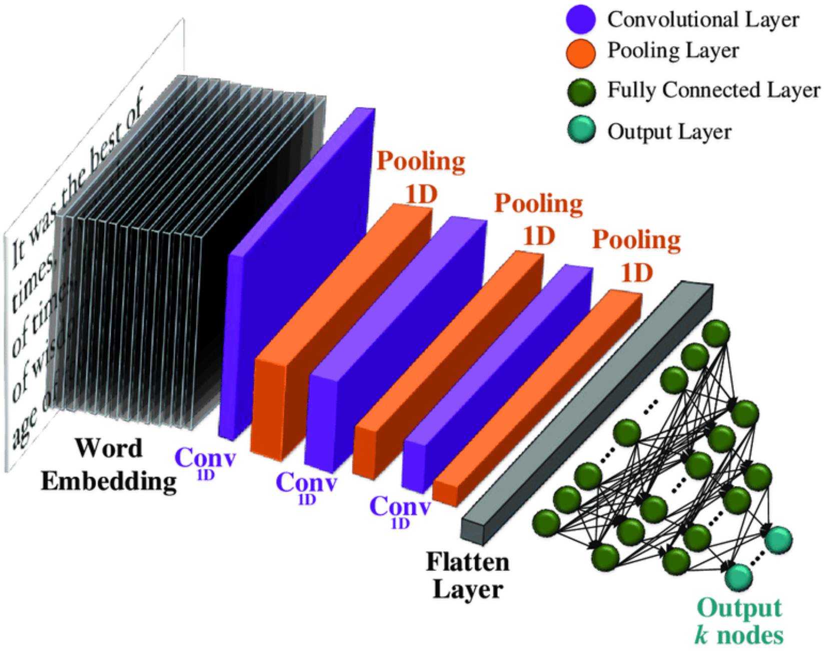


Figure 2.11: Convolutional Neural Network (CNN)

### 2.5.1 Source Credibility Rules

Specifically in the context of media, news, and content analysis, source credibility norms are a collection of guidelines and criteria used to evaluate the dependability and trustworthiness of information sources. These guidelines assist in determining whether the information is accurate, unbiased, and reliable by assisting in the evaluation of the trustworthiness of the sources from which the information originates. To distinguish between trustworthy sources and those that might be linked to false information, deception, or untrustworthy content, source credibility standards are frequently used.

### 2.5.2 Contextual Analysis

In order to understand information's meaning, relevance, and ramifications on a deeper level, contextual analysis involves looking at and analyzing it in the perspective of a larger context. It entails taking into account the external situations, circumstances, and variables that affect how the information is interpreted. In many fields, including as linguistics, literature, the social sciences, and data analysis, contextual analysis is used to unearth subtleties and insights that might not be immediately obvious.

### 2.5.3 Fact Checking Patterns

Take into account the patterns fact-checkers use to spot erroneous or misleading information. Classifying and recognizing misinformation, disinformation, satire, and fake news depend heavily on fact-checking patterns. Verifying the veracity of claims, assertions, or information offered in various kinds of material is known as fact-checking. We may create a framework to evaluate the dependability and credibility of information sources and content categories by looking at fact-checking patterns.

# Chapter 3

# Related Works

The proliferation of misinformation, disinformation, satire and fake news on digital platforms has become a pressing concern within the realm of Bengali language. The issue of accuracy and authenticity of the content and information shared holds particular significance in linguistically diverse regions, where nuances of language, culture and context can impact the spread and perception of false information. In the context of Bengali language, which boasts a rich linguistic heritage and a vast online presence, the challenges associated with detecting and categorizing false information into misinformation, disinformation, satire and fake news are both unique and multifaceted. But in recent times, many researchers have been trying to prevent Bengali Fake News in some smart ways using machine learning and neural network based architectures. There are a handful number of works based on Bengali fake news detection which gives output as fake or non fake. But, there are no such works in our researched literature that includes the multi-class classification of fake news into multiple categorizes as misinformation, disinformation and satire.

## 3.1 Conventional Approaches

In recent years, many machine learning and deep learning approaches have been used to detect whether a news is fake or not. This approaches typically involves feature engineering and the use of various classifiers like Support Vector Machines(SVM), Logistic Regression(LR), Naive Bayes(NB), Random Forests(RF), K-nearest Neighbors(KNN), Decision Trees(DT) etc.

Hossain et al. [8] created *BanFakeNews* which is an annotated collection of over 48 thousand read and roughly 1300 false news stories in Bengali. This dataset on Bengali Fake News is the only one that is publicly accessible that we are aware of, and it has paved the way for research on Bengali Fake News Detection. They have gathered reliable news stories from 22 reputable sources and three different types of fake news- fake, click-bait and satire-from some well known websites. They tested classic machine learning models including SVM, LR and RF, and empirically discover that SVM that incorporates all language variables outperforms the other two models, reaching 0.91 F1-score on a false class. Additionally, they have noted that lexical features outperform other language features. Furthermore, they contend that more usage of punctuation makes a news fake frequently.

A benchmark research on fake news was done by Khan et al. [9] using three different English datasets, the largest and most varied of which was their own creation, to access the efficacy of several practical approaches. Both conventional machine leaning-based and neural network-based techniques are tested in their research. They utilized SVM, LR, DT, MNB, and KNN as standard models, nevertheless, Multi-nominal Naive Bayes using N-gram features outperformed all of the other models.

Multinomial Naive Bayes(MNB) and Support Vector Machines(SVM) are utilized to develop a Bengali Fake News detection model in the study conducted by Hussain et al. [10] employing count vectorizer and TF-IDF(Term frequency- Inverse Document Frequency) as feature extraction techniques. They developed their own collection of roughly 2500 new articles, of which 993 were fake and 1548 were legitimate. SVM outperformed MNB in their testing, with ans accuracy of 0.9664 compared to 0.9332 for MNB.

The BanfakeNews Dataset [8] and another dataset of 2500 news articles [10] were used in the research by Sraboni et al. [11] in order to train many well-known machine learning algorithms, including RF, PAC, MNB, SVM, LR, DT using TF-IDF as a feature extraction process. To reduce bias, they used 3500 real data and 2300 made-up data instead of all nearly 52000 available data. Additionally, they tested several train-test-split ratio including 50-50, 60-40, 80-20,and 70-30, and empirically discovered that the latter split ratio produces the greatest results. PAC and SVM models outperformed the other models they trained in terms of accuracy, coming in at 0.938 and 0.935 respectively.

## 3.2 Neural Network Based Approaches

A neural network is a form of computing model inspired by the structure and operation of the human brain. It is a layer-based hierarchy neurons or interconnected nodes. A deep neural network is formed when a neural network has multiple layers. Deep learning is the use of neural networks with multiple hidden layers to learn hierarchical data representations. In contrast to typical machine learning, deep learning algorithms can distinguish relevant features and comprehend the semantic context of textual input. The following section discusses the relevant works based on neural network models.

Along with more conventional techniques, Khan et al. [9] experimented with deep learning models such CNN, LSTM, Bi-LSTM, C-LSTM, HAN, Conv-HAN, and char level C-LSTM. Since no model consistently outperforms every other model on all three of their datasets, they interpreted the results as meaning that no deep learning model is inherently better than others. Additionally, they point out that while neural network based models show acceptable accuracy and F1-Score on a sizable dataset, they show overfitting on a smaller dataset(LIAR).

Mridul et al. [12] implemented a research on Bangla Satirical News detection process using traditional CNN(Convolutional Neural Networks) architectures. The goal of their study was to identify Bangla Satire news propagated through internet news portals and social media. They presented a hybrid technique for extracting features from texts that combines Word2Vec and TF-IDF methodologies. With conventional CNN architecture and their proposed feature extraction techniques, they successfully detected whether a Bangla text piece is satire or not with an accuracy of more than 0.96.

In order to categorize and identify misinformation on a dataset, Aditya et al. [13] combined the neural network-based models and Natural Language Processing(NLP). They have combined LSTM(Long Short Term Memory) and BERT(Bidirectional Encoder Representations using Transformers) models to classify misinformation on a dataset. They have used tokenization, lemmatization and stop words removal techniques as data preprocessing steps. After splitting the dataset in a ratio of 80:20, they have processed the dataset into the LSTM and BERT model respectively. They showed that, BERT can capture the context, and semantics of text data, making it particularly effective for tasks that require a deep understanding whereas, LSTM can learn long term dependencies in the data and can be trained with a relatively small amount of data. However, in their study, BERT model outperformed the LSTM model in classifying misinformation with an accuracy of 0.6488 whereas, for LSTM, it is 0.6059.

To improve the fake news identification process, Zahin et al.[14] advised using a multichannel strategy, which suggests using a variety of data sources or representations. In this instance, it is probably used to efficiently capture both spatial information (using CNN) and sequential dependencies (using LSTM) within the Bangla text data to combine CNN and LSTM structures. While LSTMs are built to handle sequential data, making them suited for jobs like natural language processing, CNNs are well recognized for their ability to identify spatial patterns within data. This capacity is frequently applied to applications like image categorization. This study combines these two architectures and makes use of their complementing advantages to increase the precision of fake news identification in the context of Bangla language with an accuracy of 75.05 percentage.

## 3.3 Ensemble and Transformer Based Approaches

Some research tried to merge multiple methods together to achieve better results named as Ensembling. Pre-trained transformer models , such as BERT , have revolutionized the field of natural language processing by demonstrating astonishing performance on a number of NLP tasks. These models capture intricate relationships and meanings as they learn contextualized word representation from massive datasets. Researchers have effectively fine-tuned pre-trained transformer models for the tast of detecting fake news, satire or misinformation etc.

In the field of pre-trained transformers, Aditya et al. [13] proposed a BERT model which is based on transformer based model. They introduced a number of feature extraction techniques as the data pre-processing methods. After that, they fed their collected data to a transformer based BERT model to classify the misinformation. In their research, they got the best performance of 0.6488 in terms of accuracy compared to the other model LSTM(Long Short Term Memory) which is a natural language processing model.

Banfakenews [8] is now still the only sufficiently large publicly available Bengali News Dataset, but it is very much imbalanced and weighted towards real news. Keya et al. [15] proposed to utilize text augmentation techniques to mitigate the imbalance. Their research investigates classical, deep learning and transformer methodologies, and they offer a pre-trained transformer-based solution.

They optimized the BERT base uncased architecture with a balanced dataset, and their construction model - AugFakeBERT outperforms existing techniques with an accuracy score of 0.9245.

## 3.4 Siamese Neural Networks

The ability of Siamese networks to learn meaningful embeddings for text pairs, which is necessary for tasks involving similarity measurement, verification, and ranking, makes them effective for text classification tasks[1]. It is highly suited for text classification since the Siamese architecture encourages the network to capture the underlying semantic links between texts. Because they can learn to map texts into a common space where like texts are closer together and dissimilar texts are farther apart, Siamese networks excel in the setting of text classification.

Jonas et al.[16] in their research have suggested a Siamese Recurrent Network (SiamRNN) for calculating text similarity. Recurrent neural networks (RNNs) are used by the network to encode each pair of sentences that it receives as input. On the basis of the encoded representations, the Siamese architecture then learns a similarity score. They have shown an excellent illustration of the successful application of Siamese architectures to text classification tasks, demonstrating their capacity to learn useful text embeddings for similarity assessment.

# Chapter 4

# Data Collection and Preprocessing

Researchers from a variety of fields, including computer science, social science, and journalism, have been working on developing techniques and technology to recognize and battle fake news as the spread of misinformation and fake news has become a global concern. We collected existing datasets on fake news, real news, and satirical news and merged them together to create a new benchmark dataset in order to aid in study and analysis in this crucial subject.

This extensive dataset will be a useful tool that incorporates a variety of news articles from across the credibility spectrum, from true, factual reporting to misinformation, disinformation and satirical reporting. It has been painstakingly chosen to help scholars, analysts, and data scientists explore the diverse terrain of internet news information.

## 4.1 Data Collection

### 4.1.1 Real News

We collected real news from the following datasets:

1. Hossain et al. [8] created *BanFakeNews* dataset that contents 48678 authentic news and 7201 labeled authentic news.

2. Sadik Al Jarif posted a dataset on kaggle on bangla fake news detection that consists of 10000 real news

3. Hussain et al. [10] provided a dataset with 1548 real news

### 4.1.2 Fake News

We collected fake news from the following datasets:

1. Sadik Al Jarif posted a data set on kaggle consisting of 4537 fake news.
2. Hussain et al. [10] provided a dataset with 922 fake news
3. Hossain et al. [8] created *BanFakeNews* dataset that contents 1217 fake news.
4. We crawled 500 data from jachai.org and bdfactcheck.com and found fake news articles.

### 4.1.3 satirical News

We collected satirical news from the following datasets:

1. Sharma et al. [12] provided a dataset consisting of 1480 satire news.
2. Hossain et al. [8] created *BanFakeNews* dataset that consists of 1217 satire news.
3. We crawled 500 data from jachai.org and factcheck.org and found satirical news articles.

## 4.2 Data representation

### 4.2.1 Total data by *Class*

| Class | Number of data |
|---|---|
| Satire | 2773 |
| Fake | 5540 |
| Real | 67428 |
| Total | 75989 |

Table 4.1: Number of data in each class

### 4.2.2 Image representation

#### 4.2.2.1 Newly gathered data

| | Title | Content | Class |
|---|---|---|---|
| 0 | টিটুর পাছায় ফোঁড়া হয়েছে ! সবার দোয়াপ্রার্থী | ল্টানো তানপুরার মত একটি অঙ্গ। অত্যন্ত তুচ্ছতাচ... | 0 |
| 1 | পাঁচ মাস বগলের লোম না কাটায় বউ চলে গেলো যশোরে... | নেক আয়োজন করে বিয়ে করে বাসর ঘরে আনন্দ উল্লাস... | 0 |
| 2 | সবার আগে চাঁদ দেখতে পেয়েছে যশোরের আরাফাত | ন্ধ্যার পর চাঁদ দেখা কমিটি থেকে জানানো হয়েছিলো... | 2 |
| 3 | নোয়াখালীতে চাঁদ দেখা গেছে মোট তিনটা ! | রাদেশে কোথাও চাঁদ দেখা না গেলেও নোয়াখালীর আবু ... | 0 |
| 4 | বৃষ্টিতে সিগারেট ভিজে যাওয়ায় অঝোরে কাঁদছে পথ... | রের আলো না ফুটতেই অগ্নিগর্ভ রাজধানীতে নেমে এসে... | 0 |
| ... | ... | ... | ... |
| 495 | ইসলামী বেংক ছাড়া আমাদের চলেই না: মালাই লামা | বাজেটের পুর্বে বাজেট নিয়ে আলচনা সভার আয়জন করেছ... | 2 |
| 496 | ওপারে তুমি রাধে এপারে আমি, মাঝে মদী বহে রে: মু... | ভারতের লোকসভা নির্বাচনে বৃহত্তর বিজেপির খানকির... | 2 |
| 497 | গৌতম বুদ্ধের বানী, জানি মাগার নাহি মানি: ফখা | বৌদ্ধ ধর্মাবলম্বীদের পবিত্র দিবস বুদ্ধ পুর্নিম... | 2 |
| 498 | বংগবন্ধু দুস্টু একটা: খালেদা | সোমবার মধ্য রাত্রে গুলশানে নিজ কার্যালয়ে আয়জিত... | 1 |
| 499 | সময় হইলেই ধরব: কামাল | নারায়নগঞ্জের চাঞ্চল্যকর সাত খুন মামলায় অভিযুক্... | 2 |

500 rows × 3 columns

Figure 4.1: Newly annotated dataset

We are continuously crawling more and more data from different online sources of fake news. We hope to amass good amount of data that can help us and future researchers to embark in the field of multi-class labelling of fake news.

#### 4.2.2.2 Merged dataset

| | Title | Content | Class |
|---|---|---|---|
| 0 | ১০ লাখ টাকা ক্ষতিপূরণ পেল রাজীব-মীমের পরিবার | রাজধানীর কুর্মিটোলায় বাসচাপায় নিহত শিক্ষার্থী ... | 1 |
| 1 | আসিফের পুট্টু মেরে কুয়াকাটা বানাব: মুহিত | বিশেষ মতিনিধিঅর্থনীতি ও রাজনৈতিক পরিস্থিতি নিয... | 0 |
| 2 | খালেদা জিয়ার রাষ্ট্রদ্রোহসহ ১১ মামলার শুনানি ২... | বিএনপি চেয়ারপারসন খালেদা জিয়ার বিরুদ্ধে দায়ের ... | 1 |
| 3 | বাংলাদেশ-আফগানিস্তান ‘গুরুত্বহীন’ ম্যাচ আজ | নিছক মজার ছলেই মাশরাফি বিন মুর্তজা সাংবাদিকদের... | 1 |
| 4 | রাত ১১টার পর ফেসবুক বন্ধের দাবি রওশন এরশাদের | বিরোধীদলীয় নেতা বেগম রওশন এরশাদ যুবসমাজকে রার ... | 1 |
| ... | ... | ... | ... |
| 75984 | উ. কোরিয়ার একমাত্র পথ কূটনীতি, নিরস্ত্রীকরণ: প... | এ পথ থেকে সরে গেলে তারা আরো বেশি একঘরে হবে এবং... | 1 |
| 75985 | ‘বিতর্কিত সিদ্ধান্তে আমরা ক্রিকেটে হেরে গেলাম’ | নিজস্ব প্রতিবেদক : ‘এশিয়া কাপের ফাইনালে বাংলাদ... | 1 |
| 75986 | দিনে এশা আর রাতে জোহর নামাজ যেসব দেশে! | ঢাকা: বিশ্বের কয়েক দেশে দিনে এশা আর রাতের বেলা... | 1 |
| 75987 | প্রাণ ঝরবেই? | বেপরোয়া চালকের দৌরাত্ম্য থামছে না। ফলে সড়কে ... | 1 |
| 75988 | ইন্দোনেশিয়ায় ভূমিকম্পের পর সুনামিতে নিহত প্রায়... | ইন্দোনেশিয়ায় গতকাল শুক্রবার আঘাত হানা ভূমিকম্প... | 1 |

Figure 4.2: Part of our full dataset

## 4.3 Data Preprocesssing

### 4.3.1 Planning

Our new merged dataset consists of 3 attributes :

1. Title
2. Content
3. Class

Our *Class* attribute can have 4 different values ranging from 0 to 3. Class value and associated meaning is shown below:

| Value | Classification |
|---|---|
| 0 | Satire |
| 1 | Disinformation |
| 2 | Misinformation |
| 3 | Real |

Table 4.2: Classification of news in our dataset

### 4.3.2 Procedure

We preprocessed our data using the following procedure:

1. Remove all the unwanted attributes.
2. Rename the corresponding attributes as *Title*, *Content*, and *Class*.
3. Remove all the unwanted characters/words/parts of text from *Title* and *Content*.
4. Replace the Classes according to values given in table 4.2.
5. Repeat steps 1 to 4 on each individual dataset.

6. Merge all the datasets into one large dataset.

7. Drop any duplicate rows that may occur during the merging process.

### 4.3.3 Data Cleaning

We encountered unnecessary empty spaces, tabs and newlines ('\n', '\t', ' ') which had to be removed from the dataset. Some of the data consisted of unwanted suffix in their *Content* attribute, consisiting of english characters and words. Those were removed as well. As the datasets were previously processed by their respective owners, there was not much cleaning work to be done.

# Chapter 5

# Methodology

Modern deep learning models like BERT, GANs(Generative Adversarial Networks) and U-Nets are able to execute tasks like image recognition, picture segmentation, and language modeling at the cutting edge. There is hardly a day that passes without a fresh development in machine learning. Tech behemoths like Google, Microsoft, and Amazon are developing intricate deep learning architectures that perform in as with humans. However, these models have the drawback of requiring a large amount of labelled data. or a particular task there may not always be a lot of data available. The deep learning model will perform badly if there is insufficient data to adequately model distinct classes. **Siamese Neural Networks** can help to a great extent in this situation. Even when there are fewer samples per class and an imbalanced class distribution, it helps to develop models with acceptable accuracy.

## 5.1 Siamese Neural Networks

A group of neural network structures called Siamese Neural Networks(SNN) have two or more identical sub-networks [1]. The term "identical" in this context refers to their shared configuration, including their parameters and weights. The two sub-networks mirror parameter updating. These networks are employed in many applications because they can compare feature vectors to determine how similar two inputs are. In a Siamese network, two sub-networks are often utilized to process two inputs, and a separate module is used to integrate the output of the sub-networks to produce the final result. The two neural sub-networks, which are both feed-forward perceptrons, work concurrently and compare their results at the end, typically using a cosine distance. They both use error back-propagation during training. A Siamese Neural Network's output can be thought of as the degree of semantic similarity between the projected representations of the two input vectors.

Because of the following factors, Siamese network structures excel at certain tasks:

- Exchanging weights amongst sub-networks reduces the number of parameters that need to be trained and the amount of data needed. Additionally, Over-fitting can be less likely to occur;
- The sub-network effectively serves as a representation of the input. Since similar input types, such as sentences or signatures, are processed using similar models, it makes sense in each of the aforementioned scenarios.

The basic architecture of Siamese Neural Network is given below:

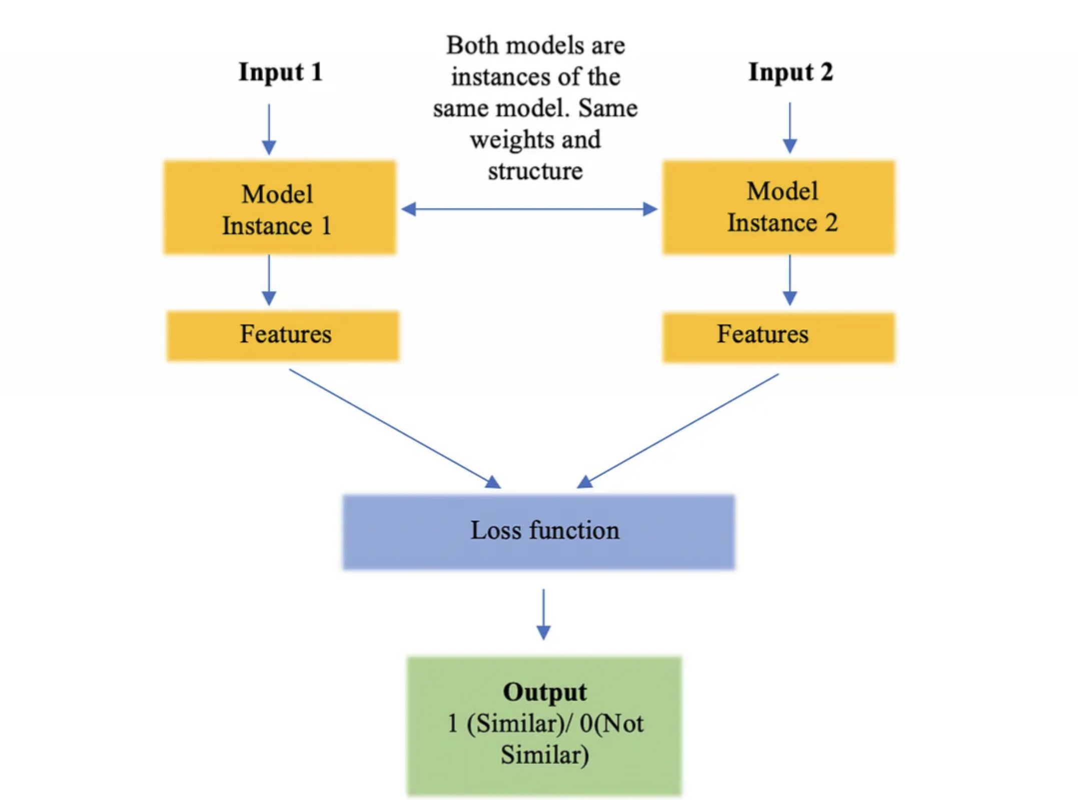


Figure 5.1: Siamese Neural Network

## 5.2 Model Selection

Due to the distinctive architecture and advantages, Siamese Neural Networks are increasingly used for applications including text classification, similarity testing, and recommendation systems. For text classification, there are a number of reasons to pick a Siamese Neural Network over other machine learning, deep learning, and neural network methods. Let's contrast Siamese Neural Networks with a few alternative strategies and discuss the factors that led to their selection:

### 5.2.1 Traditional Machine Learning Methods

Traditional machine learning methods have limitations when it comes to capturing complicated correlations in text data, such as logistic regression, decision trees, and random forests. They frequently want hand-crafted features and have trouble comprehending context and semantic significance.

### 5.2.2 Single Neural Networks

Text categorization can be done using single neural network models, such as feed-forward or recurrent neural networks (RNNs). To be effective, these models need a lot of labeled data, though. When tackling similarity-based tasks, they may not adequately capture complex relationships between text samples.

### 5.2.3 Deep Learning Networks

Recurrent neural networks (RNNs) are excellent at handling sequential data, whereas convolutional neural networks (CNNs) are adept at collecting local patterns in data. However, RNNs can be computationally expensive and have vanishing gradient issues, whereas CNNs may overlook global dependencies when used for text similarity tasks.

### 5.2.4 Siamese neural networks

Siamese Networks are made to learn the similarities or differences between pairs of inputs using contrastive learning. The network is encouraged to reduce the distance between similar pairs and maximize the distance between dissimilar pairs in the embedding space during training via contrastive loss.

### 5.2.5 Benefits and Justifications for Siamese Networks

1. Siamese Networks do not share weights between the sub-networks, in contrast to other topologies. As a result, they can extract separate features from both inputs and recognize subtle distinctions.

2. Siamese networks easily manage unbalanced data because they place more emphasis on relative comparisons than absolute forecasts.

3. Siamese networks are good at capturing semantic similarity because they can be trained to map inputs into a common space where related items are placed closer together and dissimilar items are placed farther apart.

4. Effective Training: Siamese Networks can reach convergence more quickly than standard networks since they simply need to learn relative distances.

In conclusion, Siamese Neural Networks are particularly effective at tasks like text categorization that require measuring the similarity or dissimilarity between pairs of inputs. They excel at identifying semantic links and handling skewed data, outperforming conventional machine learning models and single neural network models. Siamese Networks stand out for their design, contrastive learning, and capacity to generalize from little amounts of input, despite the fact that CNNs and RNNs each have their own advantages.

## 5.3 How It Works

A particular sort of neural network architecture called a Siamese network is made for jobs requiring the computation of similarity or distance between pairs of inputs by Chicco et al [17]. Applications like face recognition, text similarity, signature verification, and others frequently use it. Sharing weights across two or more identical sub-networks enables a Siamese network to process various inputs while developing a useful representation of similarity.

A fundamental Siamese network operates as follows:

### 5.3.1 System Architecture

A Siamese network is made up of two or more identical sub-networks that are connected by a similarity metric layer and are also known as arms or towers. An input, like as an image or a passage of text, is processed independently by each sub-network to produce a feature representation. Each sub-network is composed of layers, such as embedding layers, convolutional layers, and recurrent layers, to extract textual features. The outputs of the two sub-networks are merged and transmitted to a final classification layer following processing. The system architecture of a Siamese Neural Network Model is given below.

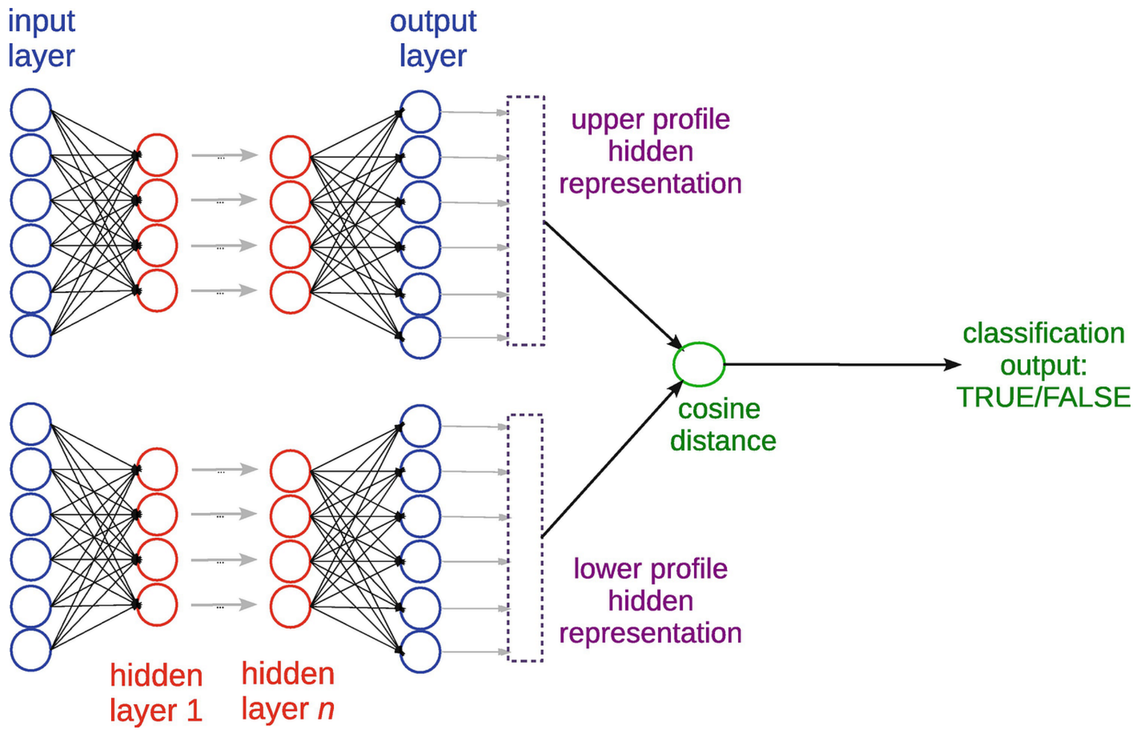


Figure 5.2: Structure of SNN
[17]

### 5.3.2 Combined Weights

The sub-networks' weights are coupled, which means they are the same and are changed collectively throughout training. By doing this, it is made sure that both networks pick out comparable features from their respective inputs. Weight sharing entails initializing the weights of one twin sub-network and then directly copying those weights to the other twin. Back-propagation computes gradients for both twins simultaneously as they each process their separate input data during training. The weight updates are applied to both networks after being averaged across the two twins.

### 5.3.3 Extracting Features

To extract meaningful feature representations, the sub-networks process the data coming from their separate inputs. The substance of the supplied data is captured by these features. Each twin sub-network receives input information (such as pictures) on its own. In order to identify patterns like edges, textures, and forms, convolutional layers evaluate the input. By performing element-wise actions on each neuron's output, activation functions introduce non-linearity. Pooling layers narrow the network's field of reception while reducing its geographic dimensions which in terms help extracting features.

### 5.3.4 Layer of Similarity Metrics

The feature vectors from both sub-networks are run through a similarity metric layer after feature extraction. The similarity or distance measure between the two feature vectors is calculated by this layer. In order to create a single vector that represents the pair of data points, the outputs from both twin sub-networks are concatenated. Euclidean Distance, Cosine similarity, and Contrastive loss[17] are some frequently employed similarity measures.

The following common operations are used to construct this composite representation:

1. Absolute Difference: The difference between the twin outputs' elements.

2. Squared Difference: The difference between the twin outputs' elements.

3. Concatenation: A new axis is created by concatenating the outputs.

4. Element-wise Multiplication: In terms of elements, multiply the twin outputs.

### 5.3.5 Loss Mechanism

The sub-networks in the network are encouraged to produce comparable feature representations for similar inputs and dissimilar representations for different inputs during training using a loss function. The task determines the precise loss function. In the feature space, the contrastive loss seeks to increase the distance between different pairings and reduce the distance between similar pairs. The squared difference between the computed distance and the expected distance (based on the label) for a given pair of data points is determined by the loss function to penalize the model for inaccurate predictions. The network learns embeddings that are closer for similar pairings and farther apart for dissimilar pairs as a result of the loss. The cross-entropy loss, which operates on a class prediction basis, cannot assist in obtaining the information about similarity or dissimilarity. Mean squared errors also don't provide us with the details we need to achieve our objective. A Contrastive loss function and a Triplet loss function are the two most often utilized loss functions. Let's take a closer look at each of them.

1. **Contrastive Loss Function:** The contrastive loss function is a distance-based loss function that minimizes the second term when two feature vectors are similar (y=0) and when they are dissimilar (y=1) in order to maximize the distance between them. By doing this, the vectors are guaranteed to be at least m units apart, saving time if they are already that far away. The equation for this kind of loss function is as follows:

   $$L(y, d) = y \cdot d^2 + (1 - y) \cdot \max(margin - d, 0)^2 \tag{5.1}$$

   Here, y is the label for the pair of inputs, where 0 means the inputs are similar and 1 means the inputs are dissimilar.
   d is the distance between the embeddings of the two inputs.
   margin is a hyper-parameter that controls the degree of separation between the similar and dissimilar inputs.

2. **Triplet Loss Function:** Using triplet loss and a comparison of an anchor image with both a positive and a negative image, the Siamese Network learns similarity ranking. To determine how similar and far one picture is to other classes, the model calculates the intra-distance for each pair. According to the approach of minimizing similar pairs closer and dissimilar pairs further away, the term with n is maximized and the term with p is minimized to minimize the loss function. The equation associated with it is given below:

$$L(a, p, n) = \max(d(a, p) - d(a, n) + margin, 0) \tag{5.2}$$

Here, a, p, and n are the variables in the equation.

d(a, p) and d(a, n) are the distances between the embeddings of the given input and the positive input(which is similar to given input), and the given input and the negative input(which is dissimilar to given input), respectively.

margin is the hyper-parameter that controls the degree of separation between the positive and negative inputs.

max is a function that returns the maximum value between two numbers.

0 is the number that is used to truncate the value of d(a, p) - d(a, n) + margin when it is less than 0

### 5.3.6 Training

Input pairs are fed into the Siamese network along with their accompanying labels (whether they are similar or dissimilar) during training. The network learns to create similar features for similar inputs and dissimilar features for various inputs by adjusting its weights through back-propagation to minimize the loss function. Gradient descent updates the shared weights while back-propagation calculates gradients in relation to the loss. Training procedure repeats until the model converges after several epochs.

### 5.3.7 Inference

Once trained, the Siamese network may be used for a variety of tasks, including determining text similarity, which involves feeding it passage of texts and using the trained similarity measure to determine how similar they are. In order to assess the degree of similarity between the texts, the output embeddings are then compared using a similarity metric. Based on the similarity score, a threshold score, a threshold can be established to determine if the texts are similar or dissimilar.

Siamese Neural Networks are highly suited for similarity-based tasks because they learn to build strong embeddings that capture the underlying structure and relationships between pairs of data points [18]. Moreover, Siamese networks are adaptable and can be used for a variety of tasks and inputs. They are helpful for situations where straightforward classification may be difficult, such as when working with little labelled data or intricate data distributions, because they learn to encode similarity information in the shared feature space.

## 5.4 Our Proposed Architecture

We are going to describe a ground-breaking strategy built around the Siamese Neural Network architecture in our effort to improve the precision and level of detail of multi-class text classification like misinformation, disinformation, satire in the form of fake news and the real ones. Our innovative framework uses the inherent strength of similarity measurement to tackle the challenges of classifying various textual inputs. This architecture, which is based on the idea of learning semantic links, has the power to fundamentally alter how humans understand and classify texts from various classes. This section describes the core of the Siamese Neural Network we propose, highlighting its unique layers, embeddings, and strategies for capturing nuanced inter-textual details. We set out to redefine the limits of multi-class text classification by taking advantage of the network's capacity to capture textual relationships within a continuous vector space. The implications go beyond accuracy to include interpretability and adaptability across a variety of domains and languages.

### 5.4.1 Modified Siamese Neural Network

Our suggested Siamese Neural Network architecture includes a dual-path framework that handles two input text sequences simultaneously. It uses advanced methods to create a multidimensional representation of the input texts, including embeddings, encoders, sequential analysis, distance measures and softmax classifiers. A holistic knowledge of text semantics is facilitated by the architecture's inherent capacity to capture both global and local information, opening up a viable avenue for precise text classification.

**Our proposed Multi-Task Siamese Network architecture is provided below:**

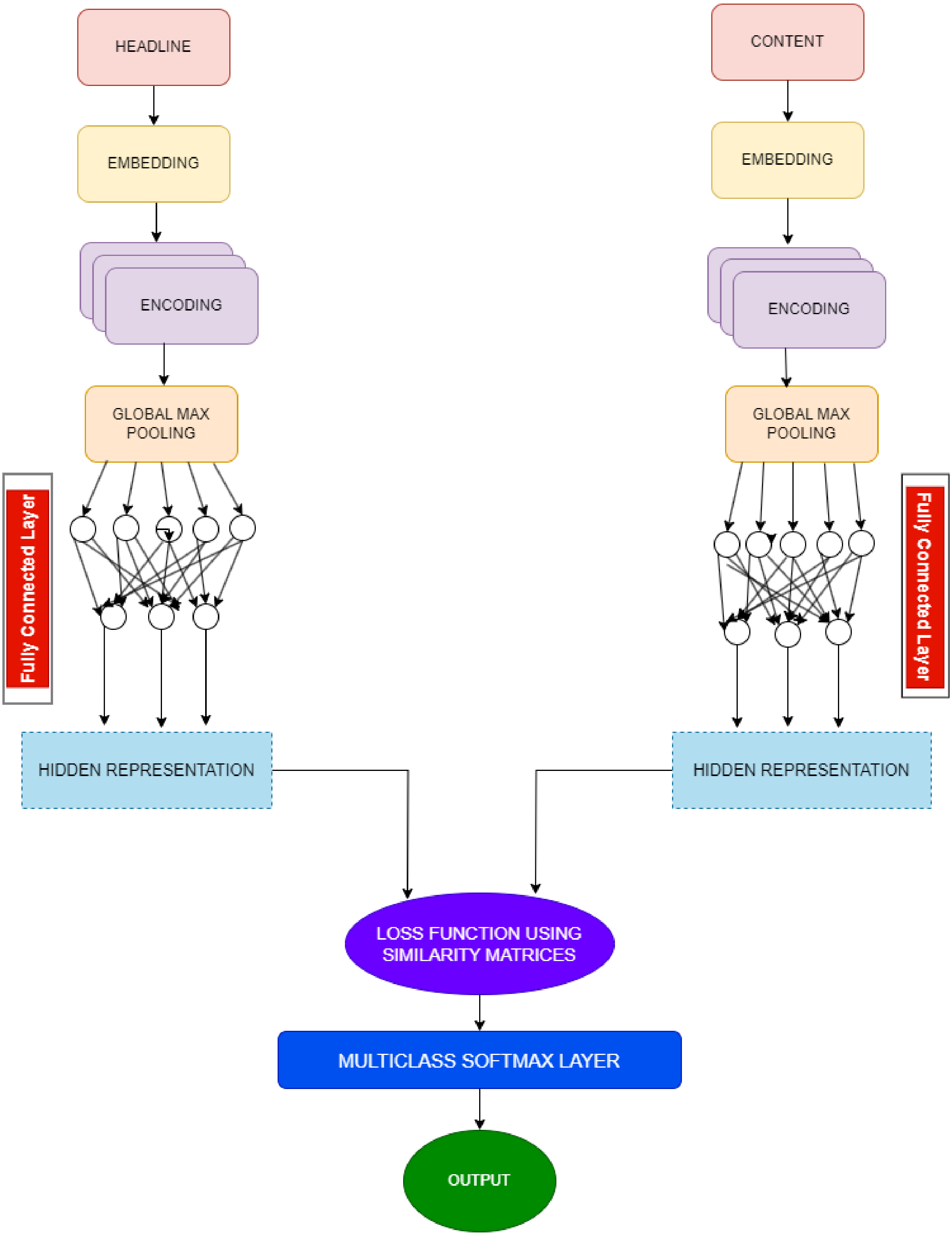


Figure 5.3: Architecture of our proposed Siamese Neural Network

### 5.4.2 Description of our model architecture

In the context of a multi-class Bengali text classification problem utilizing a Siamese Neural Network, let's now dive into more thorough explanations of each of the above steps :

1. **Embedding:** Words are represented as dense vectors in a continuous space when they are embedded. The network can comprehend the meaning and context of words thanks to these vectors, which capture semantic links. Pre-trained word embeddings (such as Word2Vec, GloVe) can be used for a multi-class text classification task, or embeddings can be taught as part of the model during training. This operation is improved using BERT model(Bidirectional Encoder Representations from transformers).

2. **Encoding:** Word embeddings are processed by encoding layers, usually LSTM or GRU, to identify textual context and sequential dependencies. Memory cells are incorporated into LSTM networks in order to address the vanishing gradient issue and capture longer-range relationships. GRU networks are less complex alternatives that are more efficient in terms of computation.

3. **Global Max Pooling:** Following encoding, the LSTM/GRU layers' outputs are subjected to global max pooling. Selected is the maximum value over all time steps for each feature map (dimension). By performing this technique, the sequence length is decreased while maintaining the key characteristics in a fixed-size representation. Important details about the text are encapsulated in the pooled representation.

4. **Fully Connected Layers:** After that, a number of fully connected layers receive the pooled representation. The extraction and manipulation of features are carried out by these levels. The network may learn complicated correlations and patterns because neurons in completely linked layers are connected to every neuron in the layer before them. The concealed representations gradually become a more ethereal feature space.

5. **Hidden Representation:** Hidden Representation: The input text is hiddenly represented by the output of the completely connected layers. The most important information taken from the original text and encoded in this representation. The contextual features required for classification are carried by the concealed representation.

6. **Loss Function using Similarity Matrices:** In multi-class text classification using a Siamese network, the objective is to capture semantic links between various classes in addition to learning the class labels. Cross-entropy loss, contrastive loss, or triplet loss are frequently employed as the loss function. Similar samples are located closer in the embedding space than dissimilar ones thanks to contrastive loss. The goal of triplet loss is to make sure that the distance between the anchor-positive pair and the anchor-negative pair is less.

7. **Multi-class Softmax Layer:** Class scores are generated by the last fully connected layer. Every neuron belongs to a certain class. The scores are subjected to the softmax activation function, which converts them into class probabilities. The final categorization is determined by the network and is based on the class with the highest probability. Because the final layer takes into account all classes at once, the architecture naturally allows multi-class classification.

This architecture gives the network the ability to effectively distinguish between multiple classes when performing a multi-class text classification task. The Siamese Neural Network can learn to accurately categorize texts into a variety of categories by combining sequential analysis, global max pooling, fully connected layers, and appropriate loss functions.

The Siamese Neural Network architecture incorporates embeddings, encoding, pooling, fully linked layers, and sophisticated loss functions in a comprehensive manner for multi-class text classification. This architecture transforms raw features into abstract representations while concentrating on the essential elements of each text and ignoring noise. The substance of the original text is captured in the hidden representation, whilst the separation and convergence of dissimilar cases are highlighted in the loss function. The multi-class softmax layer orchestrates various classes inside a single space and generates class probabilities that highlight the classification that is most appropriate. This architecture paves the path for useful information ecosystems in the digital age by providing accuracy and insight that transcend language and domain borders.

# Chapter 6

# Evaluation

Establishing a baseline for assessing the effectiveness of Bengali fake news classification is a crucial first step in our effort to battle the ubiquitous influence of fake news in today's digital ecosystem. This chapter provides the fundamental basis on which we evaluate the performance of our categorization strategies and any ensuing improvements.

The baseline assessment acts as a standard against which we will compare the improvements and novel approaches to the identification of fake news that we examine in the future works. By laying a strong foundation, we seek to evaluate the early efficacy of our approaches, highlight potential difficulties, and create a benchmark for fair comparisons as we delve further into the complexities of fake news identification. By following this methodical procedure, we hope to open the door for a thorough and informed study of our classification efforts, advancing the larger goal of reducing the spread of misinformation and promoting a more reliable information ecosystem.

## 6.1 Models

We experimented mostly on the basic models SVM(Support Vector Machine) and LR(Logistic Regression) given by hossain et al. [8] along with their *BanFakeNews* dataset. We replaced their dataset with our own dataset and experimented with the following models.

### 6.1.1 SVM - Unigram

| **Real** | | | **Fake** | | |
|---|---|---|---|---|---|
| Precision | Recall | f1-score | precision | recall | f1-score |
| 0.95 | 0.98 | 0.97 | 0.82 | 0.59 | 0.69 |

Table 6.1: Results of the model SVM - Unigram

### 6.1.2 SVM - Bigram

| **Real** | | | **Fake** | | |
|---|---|---|---|---|---|
| Precision | Recall | f1-score | precision | recall | f1-score |
| 0.95 | 0.98 | 0.96 | 0.78 | 0.55 | 0.64 |

Table 6.2: Results of the model SVM - Biigram

### 6.1.3 SVM - Trigram

| **Real** | | | **Fake** | | |
|---|---|---|---|---|---|
| Precision | Recall | f1-score | precision | recall | f1-score |
| 0.94 | 0.98 | 0.96 | 0.75 | 0.49 | 0.60 |

Table 6.3: Results of the model SVM - Trigram

### 6.1.4 SVM - C3-gram

| **Real** | | | **Fake** | | |
|---|---|---|---|---|---|
| Precision | Recall | f1-score | precision | recall | f1-score |
| 0.95 | 0.99 | 0.97 | 0.83 | 0.59 | 0.69 |

Table 6.4: Results of the model SVM - C3-gram

### 6.1.5 LR - Unigram

| **Real** | | | **Fake** | | |
|---|---|---|---|---|---|
| Precision | Recall | f1-score | precision | recall | f1-score |
| 0.94 | 1.00 | 0.97 | 1.00 | 0.48 | 0.65 |

Table 6.5: Results of the model LR - Unigram

## 6.2 Discussion

Hossain et al. [8] in their *BanFakeNews* dataset has 7 attributes whereas our dataset consists of 3 attributes. We have excluded *articleID*, *domain*, *date* and *category*. Due to that some functions had to be deleted from their source code of the model. This may have resulted negatively on the performance of the model when ran on our dataset. Also due to the inclusion of new data, the previous *stop-words* and *word embeddings* are not enough to process the data thoroughly. That might also have contributed to the poor results. Our future work includes finding suitable baseline models and improving the performance of our dataset to generate more pragmatic baseline results.

# Chapter 7

# Conclusion

In the end, using the Multi-Class Siamese Neural Network, we are trying to improve the multi-class categorization between Bengali fake news and real news , and we are hopeful of producing encouraging results that will highlight the potential of this novel methodology. By making a more robust and various class labelled dataset on Bengali Misinformation, Disinformation, Satire in form of fake news and real news. The Siamese Neural Network framework's use of embeddings, various encoders, and task-specific heads will not only demonstrate the flexibility of the design but also unlock fresh perspectives on Bengali text comprehension.

The future works of this dissertation includes the following key points:

1. We target to collect $10K+$ fake data and classify them based on table-4.2
2. We will preprocess our dataset using the following steps:
    - Remove all numbers
    - Remove all punctuations
    - Remove leading and ending whitespaces
    - Tokenize the *Title* and *Content*
    - Remove stop words from *Title* and *Content*
    - Stem and lemmatize *Title* and *Content*

3. Implement our model architecture to adapt the Bengali linguistic nuances and multi-class categorization requirements.
4. Compare the outcome of our proposed model architecture with the existing baseline models.

# References


[1] Z. Zhou, Y. Yang, and Z. Li, "Apsn: Adversarial pseudo-siamese network for fake news stance detection," *Electronics*, vol. 12, no. 4, 2023. [Online]. Available: https://www.mdpi.com/2079-9292/12/4/1043

[2] K. Shu, S. Wang, D. Lee, and H. Liu, *Disinformation, Misinformation, and Fake News in Social Media Emerging Research Challenges and Opportunities: Emerging Research Challenges and Opportunities*, 01 2020.

[3] S. Lecheler and J. L. Egelhofer, "Disinformation, misinformation, and fake news: understanding the supply side," *Knowledge resistance in high-choice information environments*, pp. 69–87, 2022.

[4] A. Day, *Satire and dissent: Interventions in contemporary political debate*. Indiana University Press, 2011.

[5] P. Ingole, S. V. Bhoir, and A. V. Vidhate, "Hybrid model for text classification," *2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, pp. 7–15, 2018. [Online]. Available: https://api.semanticscholar.org/CorpusID:52913101

[6] K. Kowsari, K. Jafari Meimandi, M. Heidarysafa, S. Mendu, L. Barnes, D. Brown, L. Id, and Barnes, "Text classification algorithms: A survey," *Information (Switzerland)*, vol. 10, 04 2019.

[7] S. Yuliani, M. Abdollah, B. Abdollah, S. Sahib, and Y. Wijaya, *A Framework for Hoax News Detection and Analyzer used Rule-based Methods*, 01 2019.

[8] M. Z. Hossain, M. A. Rahman, M. S. Islam, and S. Kar, "Banfakenews: A dataset for detecting fake news in bangla," *arXiv preprint arXiv:2004.08789*, 2020.

[9] J. Y. Khan, M. T. I. Khondaker, S. Afroz, G. Uddin, and A. Iqbal, "A benchmark study of machine learning models for online fake news detection," *Machine Learning with Applications*, vol. 4, p. 100032, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S266682702100013X

[10] 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)*, 2020, pp. 81–85.

[11] T. Sraboni, M. R. Uddin, F. Shahriar, R. A. Rizon, and S. I. S. Polock, "Fakedetect: Bangla fake news detection model based on different machine learning classifiers," Ph.D. dissertation, Brac University, 2021.

[12] 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)*, 2019, pp. 1–5.

[13] A. Harbola, M. Manchanda, and D. Negi, "Misinformation classification using lstm and bert model," in *2023 International Conference on Innovative Data Communication Technologies and Application (ICIDCA)*, 2023, pp. 1073–1077.

[14] M. Z. H. George, N. Hossain, M. R. Bhuiyan, A. K. M. Masum, and S. Abujar, "Bangla fake news detection based on multichannel combined cnn-lstm," in *2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, 2021, pp. 1–5.

[15] A. Keya, M. A. Wadud, M. Ph. D., M. Alatiyyah, and M. A. Hamid, "Augfake-bert: Handling imbalance through augmentation of fake news using bert to enhance the performance of fake news classification," *Applied Sciences*, vol. 12, p. 8398, 08 2022.

[16] P. Neculoiu, M. Versteegh, and M. Rotaru, "Learning text similarity with siamese recurrent networks," 01 2016.

[17] D. Chicco, *Siamese Neural Networks: An Overview*, 08 2020, vol. 2190, pp. 73–94.

[18] W. Yang, J. Li, F. Fukumoto, and Y. Ye, “HSCNN: A hybrid-Siamese convolutional neural network for extremely imbalanced multi-label text classification,” in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Online: Association for Computational Linguistics, Nov. 2020, pp. 6716–6722. [Online]. Available: https://aclanthology.org/2020.emnlp-main.545