

---

# Fake News Detection Classification Using Machine Learning

---

By

Abidur Rahman (011 171 275)  
Ranabir Devnath (011 171 288)  
Shubhradev Chakrabarty (011 171 301)  
Asif Ahammed (011 171 183)  
Susmita Debnath (011 171 105)

March 24, 2021



DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING  
UNITED INTERNATIONAL UNIVERSITY

---

# Abstract

---

The abstract should contain a summary of the work presented in this report in a single paragraph. The paragraph should provide a general background of the problem, methodology and results.

---

# Acknowledgements

---

This work would have not been possible without the input and support of many people over the last two trimesters. We would like to express my gratitude to everyone who contributed to it in some way or other.

First, we would like to thank my academic advisors, ...

Our sincere gratitude goes to ...

We are also thankful to ...

Last but not the least, We owe to our family including our parents for their unconditional love and immense emotional support.

---

# Publication List

---

[Optional] The main contributions of this research are either published or accepted or in preparation in journals and conferences as mentioned in the following list:

## Journal Articles

- 1.

## Conference Papers

- 1.

## Additional Publications

- 1.

# Table of Contents

<b>Table of Contents</b>	<b>v</b>
<b>List of Figures</b>	<b>vi</b>
<b>List of Tables</b>	<b>vii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Project Overview . . . . .	1
1.2 Motivation . . . . .	1
1.3 Objectives . . . . .	2
1.4 Methodology . . . . .	2
1.5 Project Outcome . . . . .	3
<b>2 Background</b>	<b>4</b>
2.1 Preliminaries . . . . .	4
2.2 Literature Review . . . . .	4
2.2.1 Related Research . . . . .	7
2.3 Gap Analysis . . . . .	7
2.4 Summary . . . . .	7
<b>3 Project Design</b>	<b>8</b>
3.1 Requirement Analysis . . . . .	8
3.1.1 Functional and Nonfunctional Requirements . . . . .	8
3.1.2 Context Diagram . . . . .	8
3.1.3 Data Flow Diagram Level 1 . . . . .	8
3.1.4 UI Design . . . . .	8
3.2 Detailed Methodology and Design . . . . .	8
3.3 Project Plan . . . . .	8
3.4 Task Allocation . . . . .	8
3.5 Summary . . . . .	8
<b>4 Implementation and Results</b>	<b>9</b>
4.1 Environment Setup . . . . .	9

---

4.2	Testing and Evaluation . . . . .	9
4.3	Results and Discussion . . . . .	9
4.4	Summary . . . . .	9
<b>5</b>	<b>Standards and Design Constraints</b>	<b>10</b>
5.1	Compliance with the Standards . . . . .	10
5.1.1	Software Standards . . . . .	10
5.1.2	Hardware Standards . . . . .	10
5.1.3	Communication Standards . . . . .	10
5.2	Design Constraints . . . . .	10
5.2.1	Economic Constraint . . . . .	11
5.2.2	Environmental Constraint . . . . .	11
5.2.3	Ethical Constraint . . . . .	11
5.2.4	Health and Safety Constraint . . . . .	11
5.2.5	Social Constraint . . . . .	11
5.2.6	Political Constraint . . . . .	11
5.2.7	Sustainability . . . . .	11
5.3	Cost Analysis . . . . .	11
5.4	Complex Engineering Problem . . . . .	11
5.4.1	Complex Problem Solving . . . . .	11
5.4.2	Engineering Activities . . . . .	12
<b>6</b>	<b>Conclusion</b>	<b>13</b>
6.1	Summary . . . . .	13
6.2	Limitation . . . . .	13
6.3	Future Work . . . . .	13
	<b>References</b>	<b>14</b>

# List of Figures

1.1	Simple Machine Learning model . . . . .	2
-----	---	---

# List of Tables

5.1	Mapping with complex problem solving. . . . .	11
5.2	Mapping with complex engineering activities. . . . .	12



# Chapter 1

## Introduction

Social media for news consumption can be compared with a double-edged sword. It is comparatively low cost, easy to use and access, and rapid dissemination of information leads people to consume news from social media. But this however comes at the cost of questionable trustworthiness and significant risk of exposure to ‘fake news’, i.e., low quality news with intentionally false information.

### 1.1 Project Overview

We spent a significant amount of time in our lives interacting with people through social media platforms. We tend to seek out and consume news from social media rather than traditional news organizations. The reason behind this change in consumption behaviors can be derived as:

- (i) it is less expensive to consume news on social media compared with traditional news media, such as newspapers or television.
- (ii) it is easier to further share, comment on, and discuss the news with friends or other readers on social media.

Despite having these advantages, the authenticity of news on social media is significantly lower than traditional news organizations. Because of its easiness and cheapness, large volumes of fake news, i.e., those news articles with intentionally false information, are produced online for a variety of purposes, such as financial and political gain. One individual or society can be badly impacted by this immense spread of fake news. Nowadays, it is very difficult to identify a news whether it is fake or real. So, we will build a machine learning model that will automatically detect if a news is true or not.

### 1.2 Motivation

Fake news can be extremely dangerous and misleading in a society. According to stats in the last ten years, there are several brutal incidents reported due to the spread of falsified

news in social network platforms which are considered as the easiest media for spreading rumors. Mostly clumsy and innocent people are used to lead this violence because they are prone to believe these rumors. To help mitigate the negative effects caused by fake news-both to benefit the public and the news ecosystem-It's critical to develop methods to automatically detect fake news on social media

### 1.3 Objectives

The objective of this project is to discover the viability and impediments of language-based systems for detecting any type of fake news which is detected using machine learning algorithms, AI calculations.

- The aim of this project is to provide a systematic evaluation of the machine learning algorithm for fake news prediction
- This can save time for people. Online news readers would be more benefited by using this project.

### 1.4 Methodology

- **Manually curate corpus:** Preprocessed dataset on this field especially in Bangla is very rare. So, for training this system, curating data from different social media is needed.
- **Manually preprocess data:** The dataset, which has been curated manually, is real-life data. First of all, they have to be labeled as “Real” or “Fake”, which is very difficult to insure. Also, the data-set is incomplete, unstructured, and noisy. So, the dataset has to be manually preprocessed.

After that a machine learning model can learn without being explicitly programmed. it can improve its performance by gaining more data. Then we can use our trained machine learning model to classify fake or true news.

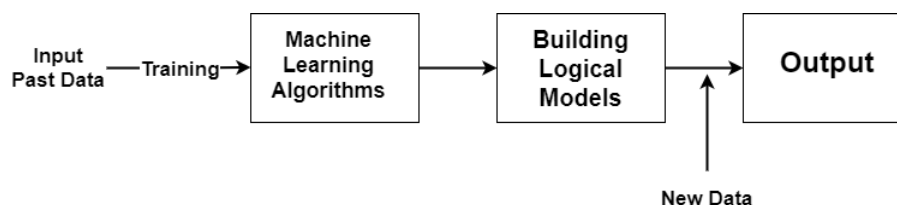


Figure 1.1: Simple Machine Learning model

## 1.5 Project Outcome

Our goal is to identify fake news from social media. For reaching our goal, we are expecting to make a web service with the properties of identifying fake news. Which will work as an extension or plugin with social media websites. The service will take keywords from a specific post or news from a particular media and run different algorithms over it. Then it will finally predict whether the **social media post** or news is possibly fake or not.

## Chapter 2

# Background

### 2.1 Preliminaries

We have reviewed several research paper based on our project topic and gathered a lot of knowledge about the topic.

### 2.2 Literature Review

#### **Event Adversarial Neural Networks for Multi-Modal Fake News Detection**

Published year: 2018

Author: Wang, Yaqing and Ma, Fenglong and Jin, Zhiwei and Yuan, Ye and Xun, Guangxu and Jha, Kishlay and Su, Lu and Gao, Jing

Publisher: Association for computing machinery(ACM)

Summary: Since news reading on social media becomes widely popular, fake news becomes a major issue concerning the public. It can take advantage of multimedia content to mislead readers which can cause a huge negative impact on one person or society. Most of the existing approaches do not handle the fake news on newly emerged events. The machine learning model tends to learn event-specific features that can not be transferred to unseen events. So in this paper, they proposed an end-to-end framework named Event Adversarial Neural Network (EANN), which can derive event-invariant features and thus benefit the detection of fake news on newly arrived events. Their framework consists of three components: the multi-modal feature extractor, the fake news detector, and the event discriminator. The multi-modal feature extractor is responsible for extracting the textual and visual features from posts. It cooperates with the fake news detector to learn the discriminable representation for the detection of fake news. The role of the event discriminator is to remove the event-specific features and keep shared features among events. They collected data from Weibo and Twitter. They claimed that their result outperformed

state-of-the-art-methods. [1]

### **Fake News Detection on Social Media A Data Mining Perspective**

Published year: 2017

Author: Shu, Kai and Sliva, Amy and Wang, Suhan and Tang, Jiliang and Liu, Huan

Publisher: Association for computing machinery(ACM)

Summary: Social media for news consumption can be compared with a double-edged sword. It is comparatively low cost, easy to use and access, and rapid dissemination of information leads people to consume news from social media. But however, this comes at the cost of questionable trustworthiness and a significant risk of exposure to ‘fake news’, i.e., low-quality news with intentionally false information. In this paper, the main focus was on the understanding of fake news. The characterization of fake news. In this survey, they presented a comprehensive review of detecting fake news on social media, including fake news characterizations on psychology and social theories, existing algorithms from a data mining perspective, evaluation metrics and representative datasets. They used some widely popular datasets such as “BuzzFeedNews”, “LIAR”, “BS Detector”, “CREDBANK”. They used Knowledge-based approach and Style based approach for solving this problem.[2]

### **Fake News Detection on Social Media using Geometric Deep Learning**

Published year: 2019

Author: Federico Monti, Fabrizio Frasca, Davide Eynard, Damon Mannion, Michael M. Bronstein

Publisher: Unknown

Summary: Context-based analysis to detect fake news poses some challenges. Like, often political, social commonsense, or ethical knowledge is required to interpret the news, which current natural language processing algorithms are still missing. From recent studies, it’s come to life that, fake news and real news spread differently on social media. Forming a propagation pattern, and that could be used for automatic fake news detection. In this paper, a novel automatic fake news detection model based on geometric deep learning has been introduced. The model was trained and tested on news stories, verified by professional fact-checking organizations, that were spread on Twitter. In this paper, a geometric deep learning approach has been presented for fake news detection on social network like Twitter or other. [3]

### **Defending against neural fake news**

Published year: 2019

Author: Zellers, Rowan and Holtzman, Ari and Rashkin, Hannah and Bisk, Yonatan and Farhadi, Ali and Roesner, Franziska and Choi, Yejin

Publisher: Unknown

Summary: as technology progress in recent years in machine learning, specifically in natural language processing, it has raised a dual-use concern. As the fake news classification is becoming more and more popular, the technology might also enable adversaries to generate neural fake news (targeted propaganda that closely mimics the style of real news). This paper investigates the threats posed by adversaries seeking to spread disinformation. They presented a model for controllable text generation called Grover. It can generate the rest of the article from a given article like “Link Found Between Vaccines and Autism”. Also, they found that the best current discriminators can classify neural fake news from real, human-written, news with 73 percent accuracy, assuming access to a moderate level of training data. They claimed that the best defense against Grover turns out to be Grover itself, with 92 percent accuracy. They are going to release the Grover publicly soon.[4]

### **CSI: A Hybrid Deep Model for Fake News Detection**

Published year: 2017

Author: Ruchansky, Natali and Seo, Sungyong and Liu, Yan

Publisher: ACM

Summary: Fake news has drawn attention both from the public and the academic communities. It provides malicious parties to manipulate the outcomes of public events such as elections and others. Because such high stakes are at play, automatically detecting fake news is an important issue, yet challenging problem that is not yet well understood. In this paper, it is mentioned that there are three generally agreed-upon characteristics of fake news. the text of an article, the user response it receives, and the source users promoting it. In this project, they proposed a model that combines all three characteristics for a more accurate and automated prediction. Specifically, they incorporated the behavior of both parties, users and articles, and the group behavior of users who propagate fake news. Their model CSI stands for Capture, Score, and Integrate. The first module based on RNN to capture the temporal pattern of user activity on a given article. The second module learns the behavior of users and the third module classifies an article as fake or not. Experimental analysis on real-world data shows that CSI achieves higher accuracy than existing models, and extracts meaningful latent representations of both users and articles.[5]

### **Detection of Bangla Fake News using MNB and SVM Classifier**

Published year: 2020

Author: Hussain, Md Gulzar and Hasan, Md Rashidul and Rahman, Mahmuda and Protim, Joy and Al Hasan, Sakib

Publisher: IEEE

Summary: interest in research in this area(Fake News Classification) has risen because people can get easily contaminated by these fake pieces of news and their effects on the offline community. A significant amount of research has been conducted on the detection of fake news from English texts and other languages but a few in Bangla Language. So, In this paper, they offered a solution for fake news detection in the Bengali language. They used two supervised machine learning algorithms, Multinomial Naive Bayes (MNB) and Support Vector Machine (SVM) classifiers to detect Bangla fake news. Their approaches had good accuracy. SVM with linear kernel had 96.64 percent accuracy and MNB had 93.32 percent accuracy.[6]

Put a summary of similar web applications, mobile apps similar to your work.

### **2.2.1 Related Research**

Here report the summary of the investigation of the research literature.

## **2.3 Gap Analysis**

Here summarise the gap where you intend to work.

## **2.4 Summary**

## **Chapter 3**

# **Project Design**

[Must be present in FYDP-1 Report and also in Final Report]

Every chapter should start with 1-2 sentences on the outline of the chapter.

### **3.1 Requirement Analysis**

#### **3.1.1 Functional and Nonfunctional Requirements**

#### **3.1.2 Context Diagram**

#### **3.1.3 Data Flow Diagram Level 1**

#### **3.1.4 UI Design**

### **3.2 Detailed Methodology and Design**

You have to mention alternate solutions that you have considered. Why you have selected the specific solution, etc.

### **3.3 Project Plan**

### **3.4 Task Allocation**

### **3.5 Summary**



## Chapter 4

# Implementation and Results

[Must be present in Final Report. Incomplete version might be included in FYDP-1 Report, however it is optional.]

Every chapter should start with 1-2 sentences on the outline of the chapter.

### 4.1 Environment Setup

### 4.2 Testing and Evaluation

### 4.3 Results and Discussion

### 4.4 Summary

## **Chapter 5**

# **Standards and Design Constraints**

[Must be present in FYDP-1 Report and also in Final Report]

Every chapter should start with 1-2 sentences on the outline of the chapter.

### **5.1 Compliance with the Standards**

Only mention the standards that are related to your project. This list is not complete. For each of the standards discuss the alternates with pros and cons and rationale of selection.

#### **5.1.1 Software Standards**

#### **5.1.2 Hardware Standards**

#### **5.1.3 Communication Standards**

### **5.2 Design Constraints**

Only mention the constraints that are related to the design of your project. This list is not complete.

**5.2.1 Economic Constraint****5.2.2 Environmental Constraint****5.2.3 Ethical Constraint****5.2.4 Health and Safety Constraint****5.2.5 Social Constraint****5.2.6 Political Constraint****5.2.7 Sustainability****5.3 Cost Analysis**

Provide a cost analysis in terms of budget required and revenue model. In case of budget, you must show an alternate budget and rationales.

**5.4 Complex Engineering Problem****5.4.1 Complex Problem Solving**

Table 5.1: Mapping with complex problem solving.

P1 Dept of Knowl- edge	P2 Range of Con- flicting Require- ments	P3 Depth of Analysis	P4 Familiarity of Issues	P5 Extent of Applicable Codes	P6 Extent of Stake- holder Involve- ment	P7 Inter- dependence
✓	✓	✓	✓			✓

P1: This project requires the study of existing models with similar goals(K8), corpus collection from social media(K4), statistical knowledge of data analysis(K3), knowledge of designing of the machine learning based model(K3, K4), integration of different components(K5, K6).

P2: Conflicting technical requirements: Collecting data from mainstream social media like Facebook is very difficult. To achieve higher accuracy from our model, we need a huge amount of data. We have to collect data that will be very tough to handle because they will be unstructured, incomplete, and noisy. So balancing good accuracy within limited data as well is a tough job.

P3: Due to the quantity and the quality of data of social media, no obvious formula can be applied to pre-process or classify as a Machine learning problem. Depth of analysis is needed to find a way to pre-process and select a specific algorithm.

P4: Since there is no agreed definition of the term “Fake News”, it is very tough to characterize Fake news. It has a complex pattern. Many psychological and social theories are related to fake news. So, As a Computer Science and Engineering student, “Fake News Characterization” is very burdensome.

P7: The dependency of sub-systems in this project is common. Like data collection, labeling, and pre-processing, model training, building a user interface, etc.

### 5.4.2 Engineering Activities

Table 5.2: Mapping with complex engineering activities.

A1 Range of re- sources	A2 Level of Interac- tion	A3 Innovation	A4 Consequences for society and environment	A5 Familiarity
✓	✓		✓	✓

A1: In this project, our stakeholders are various kinds of social media users. We have to scrape data from mainstream social media like Facebook or Twitter. Our motive is to build an extension on social media.

A2: We have to interact with news experts and traditional newspapers to identify the authenticity of the news.

A4: The main consequence of classification social media news, is the false classification of news. As this is a machine, it’s possible to get the wrong output. And to mitigate that we think we have to learn our machine more and more data so that our machine accuracy may increase.

A5: This project has been done previously in English and another language. But in Bangla, very few projects have been done, and also which are not widely popular.

## **Chapter 6**

# **Conclusion**

[Must be present in FYDP-1 Report and also in Final Report. Might be incomplete in FYDP-1 Report.]

Every chapter should start with 1-2 sentences on the outline of the chapter.

### **6.1 Summary**

### **6.2 Limitation**

### **6.3 Future Work**

# References

- [1] Yaqing Wang, Fenglong Ma, Zhiwei Jin, Ye Yuan, Guangxu Xun, Kishlay Jha, Lu Su, and Jing Gao. Eann: Event adversarial neural networks for multi-modal fake news detection. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, KDD '18, page 849–857, New York, NY, USA, 2018. Association for Computing Machinery.
- [2] Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. Fake news detection on social media: A data mining perspective, 2017.
- [3] Federico Monti, Fabrizio Frasca, Davide Eynard, Damon Mannion, and Michael M. Bronstein. Fake news detection on social media using geometric deep learning, 2019.
- [4] Rowan Zellers, Ari Holtzman, Hannah Rashkin, Yonatan Bisk, Ali Farhadi, F. Roesner, and Yejin Choi. Defending against neural fake news. In *NeurIPS*, 2019.
- [5] Natali Ruchansky, Sungyong Seo, and Yan Liu. Csi. *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, Nov 2017.
- [6] Md Gulzar Hussain, Md Rashidul Hasan, Mahmuda Rahman, Joy Protim, and Sakib Al Hasan. Detection of bangla fake news using mnb and svm classifier, 2020.