

**A Web-Based Social Media Textual Posts Credibility Verification System using Natural
Language Processing**

**Derrick Nyaga Duncan
136788**

**An Information System Project 2 proposal submitted to the School of Computing and
Engineering Sciences in partial fulfilment of the requirements for the award of the
bachelor's degree in business information technology of
Strathmore University**

**School of Computing and Engineering Sciences
Strathmore University
Nairobi, Kenya.**

July 2023

Declaration

I declare that this work has not been previously submitted and approved for the award of a Bachelor's degree by this or any other University. To the best of my knowledge and belief, the proposal contains no material previously published or written by another person except where due reference is made in the proposal itself.

Student's signature:

..... [Signature]

..... [Date]

Approval

The Information System Project 2 proposal of **Derrick Nyaga Duncan** was reviewed and approved (*for examination*) by:

Supervisor's signature:

..... [Signature]

..... [Date]

Abstract

In recent years, social media has become the primary means by which people receive news and source for information, many relying on social media to get updates on current events and information inquiries. However, due to this dependence, people have become prone to misinformation and fake news, which has resulted in misleading beliefs and decisions, the spread of fear and panic, and polarisation. To address these negative effects, it's important to find ways to verify information quickly and accurately.

The study aims to verify the credibility of information shared on social media and classify them into different categories. The study will focus on Twitter, which has over 350 million active users worldwide. Although 80% of these users use Twitter as a source of information and news, it's often associated with the spread of misinformation. Therefore, the study will provide social media users with a safeguard against the negative impacts of fake news and misinformation by enabling fact-checking of information. To achieve this a transformer model will be fine-tuned and applied to the classification. The methodology used will be prototyping, starting with Requirement gathering, quick design, building a prototype, user evaluation, refining prototype and finally implementation and maintenance. Ultimately, this will raise awareness about the negative impacts of misinformation and fake news and promote accurate and trustworthy information sharing on social media platforms.

Keywords: *Social media, Misinformation, fake news, Transformer model*

Table of Contents

Declaration	ii
Abstract	iii
Table of Contents	iv
List of figures	vii
List of Tables	viii
List of Abbreviations	ix
Acknowledgment	x
Chapter 1: Introduction	1
1.1 Background	1
1.2 Problem Statement	2
1.3 Aim	2
1.4 Specific Objectives	2
1.5 Research Questions	2
1.5 Justification	3
1.6 Scope and Limitations	3
Chapter 2: Literature Review	4
2.1 Introduction	4
2.2 The State of Fake News and Misinformation on social media	4
2.2.1 Challenges Associated with Fake news and Misinformation on social media	5
2.3 Related Works	6
2.3.1 Fake news detection using a transformer based approach	6
2.3.2 Fake news Detection in social media with BERT-based deep learning approach	9
2.3.2 Fake news detection using a graph neural network based approach	11
2.4 Gaps in related works	13
2.5 Conceptual Framework	14
Chapter 3: Methodology	15

3.1 Introduction	15
3.2 Methodology	15
3.2.1 Requirements gathering and analysis.....	16
3.2.2 Quick Design	16
3.2.3 Build Prototype	16
3.2.4 User Evaluation.....	16
3.2.5 Refine Prototype.	16
3.2.6 Implementation and Maintain	17
3.3 System Analysis.....	17
3.3.1 Activity Diagram	17
3.3.2 Entity Relationship Diagrams	18
3.3.3 System Sequence Diagrams	18
3.3.4 Use Case Diagram.....	18
3.3.5 Class Diagram.....	19
3.4 Deliverables	19
3.4.1 System Proposal.....	19
3.4.2 System Design Diagrams	19
3.4.4 Pre-processed Data.....	19
3.4.5 Model	20
3.4.6 User Interface	20
3.5 Tools and Techniques	20
3.5.1 Scikit-Learn.....	20
3.5.2 Python	20
3.5.3 GitHub.....	21
3.5.4 JupyterLab.....	21
Chapter 4: System Analysis and Design	21
4.1 Introduction.....	21

4.2 System Requirements.....	21
4.2.1 Functional Requirements	21
4.2.2 Non-Functional Requirements	22
4.3 System Analysis Diagrams	23
4.3.1 Use Case Diagram.....	23
4.3.2 Sequence Diagram	25
4.3.3 Entity Relationship Diagram.....	26
4.3.4 Activity Diagram	27
4.4 System Design Diagrams	29
4.4.1 Class Diagram.....	29
4.3.4 Database Scheme	30
4.3.4 System Architecture.....	30
4.3.4 Wireframes.....	31
References.....	35
Appendices.....	37
Appendix 1:Timeline of Activities	37
Appendix 2: Marking Guide	38

List of figures

Figure 2.1 BART model architecture.....	7
Figure 2.2 Word clouds of news content	12
Figure 2.3 Conceptual Framework	14
Figure 4.1 Use Case Diagram	23
Figure 4.2 Sequence Diagram.....	26
Figure 4.3 Entity Relationship Diagram	27
Figure 4.4 Activity Diagram	28
Figure 4.5 Class Diagram	29
Figure 4.6 Database Scheme	30
Figure 4.7 System Architecture	31
Figure 4.8 Login Page.....	32
Figure 4.9 Registration Page.....	32
Figure 4.3 Make A Prediction Page	33

List of Tables

Table 2.1 Representation of BART confusion matrix	8
Table 2.2 Evaluation metrics table.....	8
Table 2.3 Evaluation of BART algorithm using a confusion matrix	8
Table 2.4 Attributes in the fake news dataset	10
Table 2.5 Evaluation for CNN with BERT confusion matrix	10
Table 2.6 Evaluation scores using graph neural networks.....	13

List of Abbreviations

ACC – Accuracy

AP – Average precision

AUC- Area Under a Curve

BART - Bidirectional and autoregressive transformer

BERT- Bidirectional encoder from transformers

CNN – Convolutional neural network

ERD – Entity Relationship Diagram

FN- False Negative

FP- False Positive

GNN- Graph Neural Networks

NLP- Natural language processing

OOAD- object-Oriented Analysis and Design

Prec – Precision

Rec- Recall

TN- True Negative

TP- True Positive

Acknowledgment

I'd like to show my gratitude and appreciation to everyone who made it possible for me to finish this project. First and foremost, I want to thank the Almighty God for providing me with the power, knowledge, and understanding to complete this undertaking. My sincere gratitude goes to my supervisor, Mr. Steve Kayugira, for his essential instruction, patience, and support from the beginning of this project. His helpful counsel, oversight, and critical participation when needed aided me immensely during this endeavour. I'd also like to express my gratitude to Strathmore University's School of Computing and Engineering Sciences for providing me with the opportunity to work on this project. It aided me in conducting extensive research, and I learned a great deal as a result. Finally, I'd like to express my gratitude to my family and friends for their unwavering support and assistance throughout this effort. Their insightful suggestions and encouragement have been extremely beneficial at various stages of this undertaking.

Chapter 1: Introduction

1.1 Background

Social media platforms in recent years have emerged as sources of information and news. These platforms enable users to consume and distribute information on a large scale as a result information spreads quickly and reaches a wide audience. However, the openness of social media results in the spread of misinformation which can cause significant socioeconomic and political repercussions (Allcot, 2017). Lewandowsky, (2020) states that deceptive or inaccurate information on social media can influence public opinion, election outcomes and as well as individual health choices.

False information and fake news on social media present a variety of challenges. The vast amount of information exchanged on these platforms makes it tedious and time-consuming for users to carefully check the validity of each post shared. The rate at which information circulates exacerbates this as false and misleading information can spread quickly without being sufficiently fact-checked (Guess, 2019).

The reliance on human resource to provide manual verification of posts on social media cannot keep up with amount and pace at which the information spreads as this becomes unfeasible, time consuming and expensive. Therefore, being able to automate systems to carry out the verification of textual content shared in an efficient and effective manner offers reliability (shu, 2017). Natural language processing (NLP) offers promising solutions that can aid in the combat against problems caused by false information present on social media, it includes numerous techniques and algorithms that provide computers with the ability to decipher, comprehend and generate human language (Jursafsky, 2020). By utilizing NLP techniques Research and developers can build automated systems that can analyze textual content, find trends, and extract valuable information (Potthast, 2018).

The issue of credibility verification has been investigated previously in NLP research using a variety of techniques. Through approaches such as sentimental analysis whereby emotional tone and sentiment represented in text are examined, plagiarism detection whereby similarities and potential instances of plagiarism or content manipulation, algorithms are used to compare textual content and known reliable sources of information (Horne, 2018). With the existence of fact checking sites such as PolitiFact that carry out credibility verification through a manual process, data gathered from it can' be used to deploy a model on the web that can analyze textual

content from Twitter posts therefore leveraging strengths of web scrapping and natural language processing techniques the verification process can be enhanced.

1.2 Problem Statement

With social media increasingly serving as people's main source of news and information, establishing the validity of the material posted on these platforms are a major concern. According to (Vosoughi, 2018) and (Pennycook, 2019), there is risk that people form false beliefs and make decisions based on inaccurate information due to absence of mechanisms for evaluating quickly and accurately textual content on social media.

The enormous amount of information published on social media platforms in real-time is too much for current fact checking techniques, including those used by organization such as PolitiFact (Guess, 2020) . This limitation makes it difficult to spot and verify fake news and misinformation in time thus quick dissemination, influencing the public's beliefs. To enable social media users, verify information quickly and accurately its crucial for an approach that is both effective and efficient at determining the reliability of information posted on Twitter.

1.3 Aim

To develop web-based social media textual post verification system using natural language processing.

1.4 Specific Objectives

- I. To study the state of fake news and misinformation on social media
- II. To identify challenges associated with fake news and misinformation on social media
- III. To investigate existing approaches in the verification of fake news and misinformation
- IV. To design and develop web-based social media textual post verification system using natural language processing
- V. To test the developed web-based social media textual post verification system

1.5 Research Questions

- I. What is the state of misinformation and fake news on social media?
- II. What are the challenges associated with fake news and misinformation on social media?
- III. How have the existing works approached verification of fake news and misinformation?
- IV. How will the system be used in verification of textual posts on social media?

V. How will the proposed system be tested?

1.5 Justification

The inundation of fake news and misinformation on social media platforms has raised urgent concerns throughout the globe. With consequences being severe on a variety of areas, including polarization, spread of fear and panic, poor health decisions and political outcomes. By mitigating this issue, textual post credibility verification system can foster a more educated society and rebuild credibility of information from social media.

While valuable and effective, the fact checking of information by organizations such as PolitiFact, the process becomes time consuming and resource intensive. By undertaking natural language processing techniques and enabling real-time analysis of textual posts on social media sites, the proposed system will streamline the credibility verification process and reduce time and effort required for verification.

The proposed research fits with the expanding need for novel remedies to counter false information. For the challenges brought about by fake news and misinformation, there is a dire need for creative solutions that make use of natural language processing and machine learning capabilities, as social media is still evolving and playing a major role in information transmission.

1.6 Scope and Limitations

The goal of the study is to use natural language processing to create a web-based system for textual post credibility verification for social media. The system will mainly examine text that has been posted on social media sites mainly Twitter and offer credibility verification, other media such as photographs and videos will not be covered by the proposed research topic. The study has been limited to textual posts due lack of a dataset that encompasses other forms of media such as photographs and videos.

Chapter 2: Literature Review

2.1 Introduction

The proliferation of fake news and misinformation has become a pressing concern, particularly in social media. Extensive research has been conducted to tackle this problem especially in classification and detection. Techniques such as machine learning , natural language processing and deep learning have emerged as valuable tools when it comes to detection and classification. This chapter aims to look at the state of misinformation and fake news currently and various approaches that have taken place to try and combat the problem.

2.2 The State of Fake News and Misinformation on social media

Fake news and misinformation have become prevalent issues in the era of social media , posing significant challenges to the credibility and reliability of information shared on platforms like Twitter . With recent technological advancements, more individuals are spending time online and on social media. About 58.4% of the world's population use social media with a daily average time spent being 2 hours and 27 minutes (Chaffey, D., 2022) hence social media users high time spent will result to them highly encountering and being exposed to misinformation and fake news on regular basis. Studies

The scope and magnitude of fake news and misinformation on social media platforms is mind blowing. According to (Allcot, 2017) misleading news items travel rapidly and reach a larger audience than factual news stories. Another study carried out by (Guess, 2019) discovered that one in every four Americans visited websites that produced fake news during the United States 2016 election. Furthermore, (Vosoughi, 2018) discovered that bogus news reports were 70% more likely to be retweeted as compared to accurate stories, while Menczer (2020) highlights that false information on Twitter spreads six times faster than genuine information. Research conducted by shao, (2018) demonstrates that false information when it comes to dissemination of health and medical treatments results to negative impacts on people's well- being and decision-making processes.

The rapid spread of fake news and misinformation presents issues for both fact-checking organization and users. The quantity of disinformation appearing on social media platforms has risen drastically in recent years, according to the prominent fact checking organization PolitiFact (Guess, 2020). Moreover, recommendation systems and algorithms implemented on social

media platforms play a significant role in amplifying fake news and misinformation. These algorithms filter posts based on users' choices and behavior which results to creation of filter bubbles and echo chambers as a result more users are likely to be exposed to post and content that promotes their existing opinions, spreading disinformation and generating a polarized information environment (Guess, 2019).

According to Cook (2017) fake news and disinformation have implications that go beyond the transmission of misleading information as they result to potentially huge societal and political consequences. Misleading information can sway public opinion, influence election, and erode the trust institutions (Lewandowsky, 2020). Misinformation has been proven in studies to cause changes in behavior such as vaccine hesitation and adoption of conspiracy theories (Pennycook, 2019). This emphasizes the importance of mitigating the issue of fake news and misinformation as the implications can be far-reaching for people and the society as whole.

2.2.1 Challenges Associated with Fake news and Misinformation on social media

Despite efforts to solve dissemination of fake news and misinformation, various challenges remain. One of the difficulties is that flagging procedures are ineffective. Although social media sites have flagging techniques in place to allow users to report fake news and misinformation, these procedures frequently fall short of properly identifying and preventing propagation as a result false information easily evade detection and end up spreading to the masses.

Another challenge arises in manual fact-checking of content on social media platforms. Organizations such as PolitiFact have shown to be effective in refuting erroneous information however the process is time-consuming and a resource intensive (Guess, 2020). The huge volume of information published on social media daily makes it difficult for fact-checkers to keep up with the rapid spread of false information, hindering their ability to effectively identify and debunk misleading content. Furthermore, the spread of fake news and misinformation sometimes spreads at a fast rate outdoing the fact-checking process, resulting in widespread dissemination of inaccurate information. This leads to misinformation reaching large audiences before it can be fully disapproved.

Moreover, the ever-changing nature of fake news and misinformation presents a hurdle. Fact-checking procedures must constantly adapt to keep up with new strategies adopted by purveyors

of fake news as mechanisms used to propagate misinformation evolve. The manipulation of textual posts on social media, such as the use of deceptive language, inaccurate headlines, and forged quotes, complicates the fact-checking process and distinguishing between authentic and fake information more difficult (Lazer, 2018).

Additionally, social media platforms' recommendation systems and algorithms pose an obstacle in countering fake news and misinformation. These algorithms are intended for prioritizing social media posts based on users' habits and behavior, which frequently results in amplification of contentious content, including incorrect information (Guess, 2020)

This algorithmic bias can result in the spreading of misinformation and false news by exposing people to content that confirms their previous opinions, increasing confirmation biases and creating echo chambers.

2.3 Related Works

Some of the related works in fake news and misinformation classification include:

2.3.1 Fake news detection using a transformer based approach

This study by Shaina Raza and Cheng Ding (2022) proposed a transformer-based approach to classify news and social contexts on news into fake or real news. It implemented a modified bidirectional and autoregressive transformer (BART) architecture. The BART model ability to integrate text generation and comprehension in both bidirectional and autoregressive was used as bidirectional encoder from transformers (BERT) model excelled in discriminative tasks but is limited to generative tasks while GPT-2 is suited for generative tasks but falls behind when it comes to discriminative.

The modification of BART architecture was part of the methodology that was used, a comprehensive variety of elements from news articles and social contexts were added into the encoder part, employing a multi-head attention to balance the relevance of various of information. Second, forecasting of the next token by masking some of the input tokens and calculating the item that follows based on temporal position using the decoder to generate predictions based on prior text sequences and user actions. Finally, a linear transformation and SoftMax layer were used to generate the output i.e., the target label.

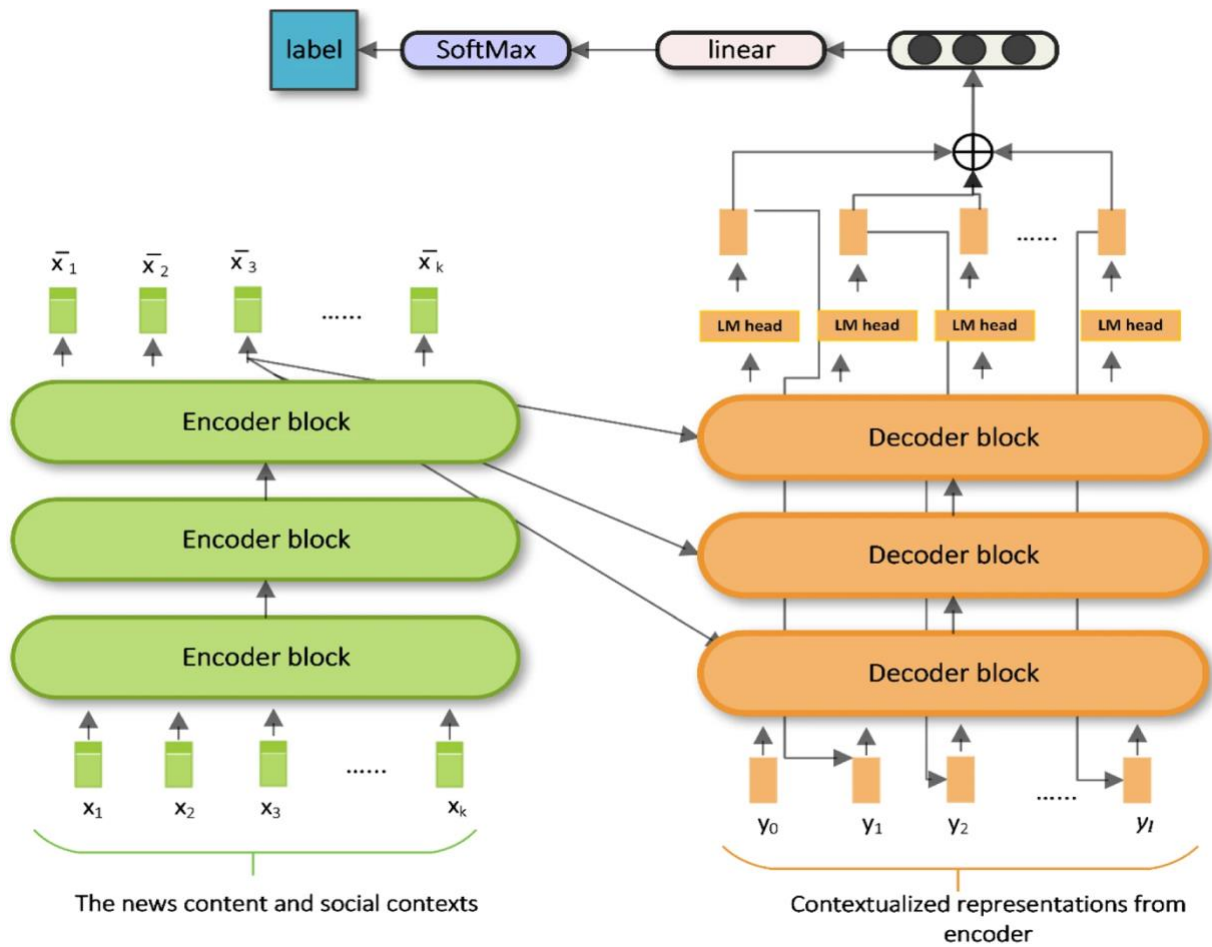


Figure 2.1 BART model architecture

The datasets used included target labels such as unreliable, mixed, reliable, true, satire, misleading content, imposter content, false and manipulated content. The first task carried out was to reduce the labels to real and fake to allow focus on binary classification. The datasets were then cleaned and pre-processed.

The model was evaluated using several key metrics. A confusion matrix was used to analyze the actual and expected classifications. True Positive (TP), True Negative (TN), False Positive (FP), False Negative (FN) values were represented in the matrix, offering insight into the model's performance.

Table 2.1 Representation of BART confusion matrix

	Actual fake	Actual real
Predicted fake	TP	FP
Predicted real	FN	TN

The True Positive (TP) represented number of false news stories detected as such. False Positive (FP) suggested the misclassification of authentic news stories as fake. True Negative (TN) indicated the correct identification of real news articles, whereas False Negative (FN) represented the misclassification of false news stories as real. Through the confusion matrix various evaluation metrics were calculated, including accuracy (ACC), precision (Prec), recall (Rec), F1-score (F1), area under the curve (AUC) and average precision (AP). These measures gave a thorough assessment of the model's performance, considering proper classifications, trade-offs between true positives and false positives and precision and recall relationships. With the evaluation framework a comprehensive understanding of the model's effectiveness in distinguishing between fake and real news was obtained.

Table 2.2 Evaluation metrics table

Metrics	BART
False Positive Rate	18
F1	16
Recall	15
Precision	14
Accuracy	17

Table 2.3 Evaluation of BART algorithm using a confusion matrix

	Actual fake	Actual real
Predicted fake	35,104	12,997
Predicted real	9793	32,893

Based on Table 2.1, Table 2.2 and Table 2.3 , the accuracy of the model was determined to be 74.89% suggesting that more than 74% of the outcomes were right. The precision obtained was 72.40% showing the occurrence of few false positives in which actual news predicated be false. The recall value obtained was 77.68% indicating that there more true positives than false negatives. With that it was discovered that false negatives, in which false news were predicted to be true, were more destructive fake news detection. The obtained F1-score was 74.95% suggesting a good level of performance.

2.3.2 Fake news Detection in social media with BERT-based deep learning approach

This study by Rohit Kumar, Anurag Goswami, Pratik Narang (2021) proposed a BERT based deep learning approach to detect fake news in social media. Bidirectional encoder from transformers (BERT) was used as it was considered to offer a bidirectional training approach as well as offer improvements in classification performance with ability to capture semantic and long-distance dependencies in sentences. The bidirectional encoder from transformers (BERT), with deep learning method based parallel block structure of the convolutional neural network (CNN) formed what was called FakeBERT which its aim was to be enhance the performance capabilities of the BERT model.

The dataset was collected during the 2016 US General election. It was made up of two files “train.csv ‘ and “test.csv”. The “train.csv” file contains labelled examples of fake and real news, but the “test.csv” was a testing dataset that contains that dint contain labels. While the source wasn’t cited, its highlighted that the dataset captures the real-world aspect. **Error! Reference source not found.** shows the independent variables and dependant variable contained on the fake news dataset.

Table 2.4 Attributes in the fake news dataset

Attribute	Number of Instances
ID (unique value to the news article)	20800
title (main heading related to particular news)	20242
author (name of the creator of that news)	18843
text (complete news article)	20761
label (information about that the article as fake or real)	20800

For data pre-processing data was truncated and segmented to fit within the length constraint.

Table 2.5 Evaluation for CNN with BERT confusion matrix

	Predicted negative	Predicted positive
Actual negative	1004 (TN)	63 (FP)
Actual positive	90 (FN)	942 (TP)

A confusion matrix was used to examine the CNN with BERT model evaluation results. True Negative (TN) had a value of 1004 in the confusion matrix, showing correct prediction of negative cases. The number of False Negative (FN) instances was 90, representing instances that were truly positive but were incorrectly labelled as negative. False Positive (FP) accounts for 63 cases when the outcome was negative but was wrongly projected as positive. True Positive (TP) has a rating of 942, indicating that positive instances were correctly predicted. The performance of the CNN with BERT model was evaluated using the confusion matrix in terms of its ability to reliably distinguish positive and negative instances.

The CNN with BERT model performed well in categorizing positive cases, according to the assessment criteria produced from the specified confusion matrix. With a recall value of 0.912, the model effectively detected a high proportion of genuine positive events in the previous evaluation. The precision score of 0.937 indicated that the model had a low rate of false positives, proving its ability to forecast positive events reliably. The high accuracy (ACC) of 0.927 indicated the model's general correctness across both positive and negative examples in the previous examination. Finally, the F1-score of 0.924 demonstrated a balanced performance

in correctly detecting positive cases while avoiding false positives and false negatives. These evaluation metrics revealed, the effectiveness of CNN with BERT in accurately classifying instances and showed its reliability in text classification.

2.3.2 Fake news detection using a graph neural network based approach

The study by Pallabi Saitaki, Kshitij Gundale, Ankit Jain, Dev Jadeja, Harvi Patel and Mohendra Roy (2022) proposed a hybrid approach for fake news detection. Their hybrid approach intended to improve the accuracy to improve the accuracy of fake news identification by incorporating a graph neural network for news propagation analysis with bi-directional encoder representations from the transformer model for text characteristics.

The dataset was collected from popular data repository FakeNewsNet particularly from PolitiFact and GossipCop. PolitiFact concentrated on American politics and validated the truth of remarks made by various political figures, whereas GossipCop fact-checked celebrity reportage. Articles having ground truth labels assigned by independent journalists were included in the datasets. The data gathering method outlined in a prior work was used, which entailed extracting relevant information such as news content, tweets, retweets, and user profiles from tweet IDs connected with each news story. News stories that lacked text were eliminated. Several pre-processing stages were carried out on the obtained datasets. Figure 2.2 shows representations of both types of news that were provided.

Table 2.6 Evaluation scores using graph neural networks

Convolutional Layer	Politifact				GossipCop			
	F1	Precision	Recall	Accuracy	F1	Precision	Recall	Accuracy
GraphConv	0.78861	0.755794	0.79047	0.788201	0.90702	0.96215	0.94935	0.93252
GATConv	0.76672	0.72966	0.79047	0.766926	0.90483	0.95616	0.95493	0.93186
GCNConv	0.794199	0.73147	0.866664	0.794632	0.9022	0.95372	0.95536	0.93021

The combination of GNN-based features and text yielded superior results, showcasing the importance of using a multimodal approach for accurate detection of fake news.

2.4 Gaps in related works

The existing works have extensively talked about the creation of models and algorithms for false news identification, but one key gap remains and that's development of a friendly user-interface to access the veracity of information. While the existing research studies have made significant advances in identifying and countering misinformation, there is still a need to transform this knowledge into practical tools that people can easily access and use.

Additionally, the existing research works its important to note that some of the studies have not comprehensively conducted comparative analysis. While these research works have contributed to the advancement of the field, the lack of having a robust comparative analysis limits the effectiveness as progress on how the research is progressing becomes unclear without benchmarks and comparison.

Moreover, the existing works focus on a binary classification as to whether the news are fake or real. While its useful in classification of information, it generalizes the complexity of information. Many social media textual posts contain nuanced or varied content that is difficult to simply divide into only two categories. It would be more advantageous if to have more categories that's able to capture a wide range of information types such mostly true, half-true, mostly false.

Given the rising importance of social media as a key source of news and information, its critical to fill these gaps to demonstrate the legitimacy of content provided on these platforms. There is a risk of people establishing incorrect beliefs and basing their decisions on inaccurate information if methods for quickly and accurately analyzing textual content are not in place.

Creating an effective and efficient method for determining the authenticity for platforms such as Twitter it's critical for enabling social media users to check information in timely and accurate manner.

2.5 Conceptual Framework

Data will be web scrapped from PolitiFact as it will be the primary source of data. The application will then fetch the data and store it in a database. The data will undergo pre-processing, training and fine-tuning then it will be sent to the model whereby from it classification will be done as per the users' inputs and the classification algorithm will classify the input. The category of the input given will be sent back to the application to be displayed on the user interface.

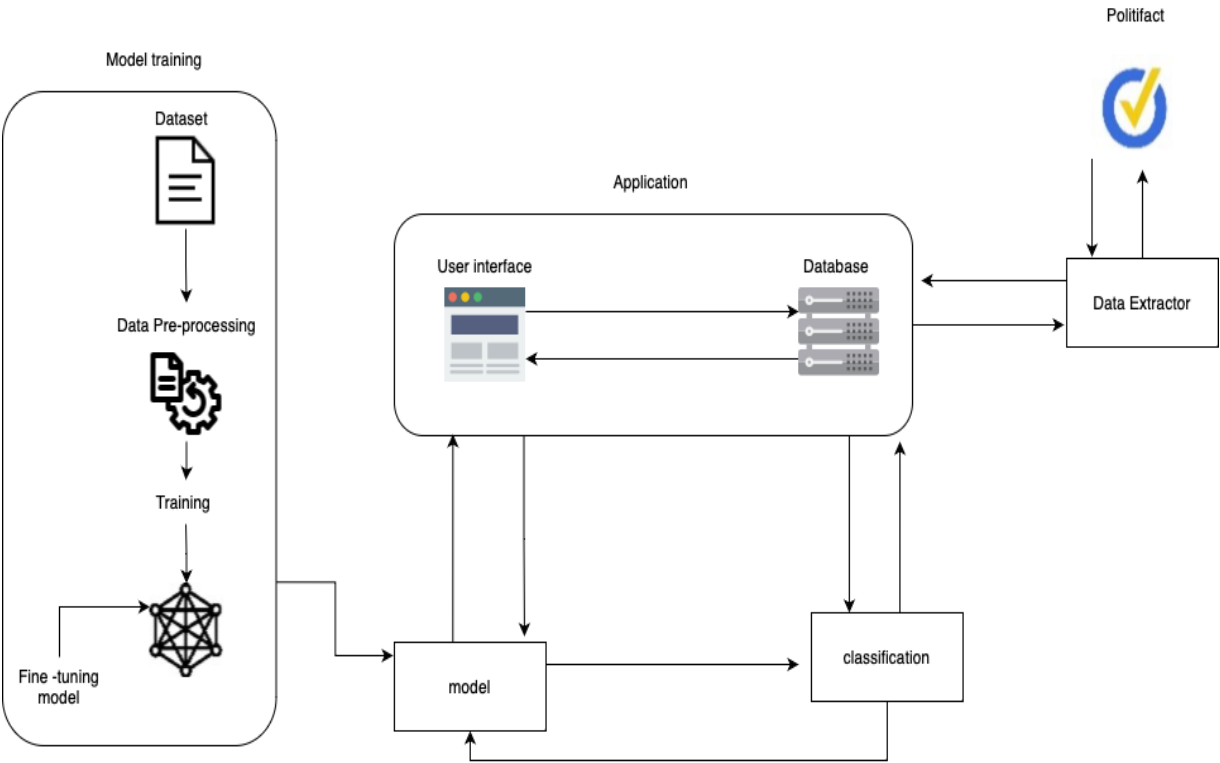


Figure 2.3 Conceptual Framework

Chapter 3: Methodology

3.1 Introduction

This chapter outlines the method used to generate the proposed solution. It emphasizes the system development methodology, system analysis and design approaches and system development tools and procedures. It also addresses the proposed solution deliverables. For development, the system will employ an object-oriented analysis and design methodology that is more data focused. OOAD is recommended because it allows a problem to be divided into smaller sections that may be modified separately without affecting the surrounding components.

3.2 Methodology

Prototyping methodology will be used in development of the system. This methodology concurrently executes the analysis, design, and implementation phases to swiftly construct a simplified version of the proposed system and get reviews and feedback from users. The prototype of the system is a rapid version of the system that offers limited features. As shown in Figure 3. (Martin, 2020), based on feedback received the developers subsequently engage in reanalysis, redesign, and reimplement a second prototype, rectifying any flaws and incorporating additional features. The iterative cycle continues until the users reach a consensus that the prototype delivers sufficient functionality to be implemented (Pressman, 2020). The methodology will be utilized due to its inherent flexibility in design, enabling easy modification and adjustments according to the users or developers' preferences and needs. It also allows for the convenient identification of missing functionalities and early detection of errors, resulting in significant savings in effort and cost.

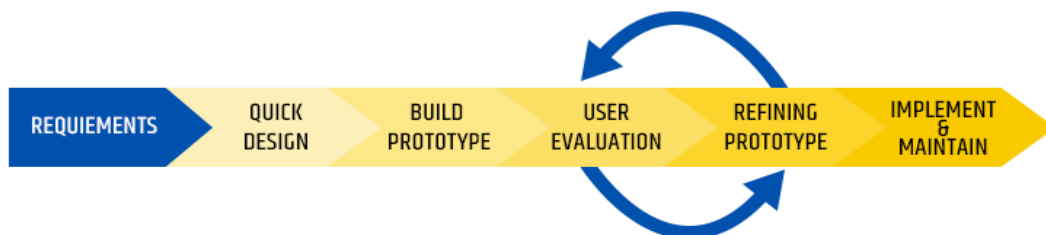


Figure 3.1 Prototyping methodology

3.2.1 Requirements gathering and analysis.

The prototyping model commences with requirements analysis, where software requirements are defined as descriptions of the desired features and functionalities of the target system. These requirements specifically identify the functionalities that the system needs to possess to meet user's satisfaction. (Bahill, 2019).

The requirements for the proposed system will be gathered from Secondary data web scrapped from PolitiFact and hosted on Kaggle. The dataset will serve as the training data for the model utilized in the proposed system.

3.2.2 Quick Design

During the second stage, a preliminary or rapid system design is developed once requirements are established. The design is not comprehensive but offers a concise representation of the system, focusing on the essential aspects that are visible to the user. The purpose of this stage is to provide the user with a high-level understanding of the system. Creating a quick design in the development of the prototype that the user can test and give feedback.

3.2.3 Build Prototype

During this stage, a preliminary prototype is designed using the information gathered from quick design. This prototype typically serves as a simplified version of the system, providing an approximation of key features and functionalities of the final product. It serves as a functional model of the desired system.

3.2.4 User Evaluation

The users conduct a comprehensive evaluation of the first prototype, carefully assessing its strengths and weaknesses. They identify elements that need to be added or removed based on their feedback. The developer collects and analyses this feedback from users to gain valuable insights.

3.2.5 Refine Prototype.

Following the user evaluation, if the user expressed dissatisfaction, the prototype will be refined based on the requirements, user feedback, and suggestions. The refined prototype is then evaluated by the user, like the previous iteration. This iterative process continuous as the

prototype is adjusted to meet the user's needs, allowing the developer to gain deeper understanding of the necessary changes. Once the user is satisfied the developed prototype, a final system is constructed based on the final prototype, a final system is constructed based on the final prototype.

3.2.6 Implementation and Maintain

Once all the reequipments are fully met and the user is entirely satisfied the final product is designed and developed, considering the final prototype. The system then undergoes rigorous testing, if deemed ready, it is deployed into production. After implementation, the system goes thorough routine maintenance to minimize downtime and prevent significant failures on a large scale.

3.3 System Analysis

System analysis involves the decomposition of a system into constituent parts, aiming to comprehend the function, nature, and interdependencies of those parts (Jeffrey Hoffer, 2021). Its purpose is to examine a system or its components to identify objectives. In the case of the proposed system, object-Oriented Analysis and Design (OOAD) will be utilized, which employs several tools and system analysis. These tools include activity diagrams, entity relationship diagrams, system sequence diagrams, use case diagrams and class diagram.

3.3.1 Activity Diagram

An activity diagram provides a graphical representation of the flow or sequence of actions are performed to complete an activity or process (Brahma Dathan, 2016). This diagram showcases the workflow of a system, starting form a designated start point and leading to finish point, while also outlining the various decision paths that they may be taken as events occurs.

In the context of the proposed system, activity diagram will be utilized to describe the sequential flow of activities in a more detailed manner, providing visualization of use cases at a granular level. Additionally, it will be used to visualize the sequence of activities and depict workflows within the system.

3.3.2 Entity Relationship Diagrams

An entity relationship diagram (ERD) is diagrammatic representation that showcases the relationship between entities within a system. Entities can encompass people, objects, concepts, or events for which data can be stored (John Satzinger, 2015).

The ERD will play a crucial role in identifying entities that will be present in the proposed system and how they relate to the objects and functions within the system, as well as to each other. It aids in the design and construction of system's database since ERDs can be easily translated into relational tables. This diagram will facilitate the understanding of data organization and the establishment of relationships between entities, contributing to the effective implementation of the system's databases.

3.3.3 System Sequence Diagrams

A system sequence diagram (SSD) is a specific type of sequence diagram that focuses on illustrating input and output events. It depicts the sequence of use cases and defines their order, along with events taking place inside the system. The SSD tracks how function and use cases are performed within a system.

In the context of proposed solution, a system sequence diagram will be employed to model the interaction of the system with events. It will illustrate the events generated by actors, their proper flow, and the potential connections and relationships between events. The diagram help visualize and understand the sequence of events within the system and ensure that they are correctly executed and interconnected.

3.3.4 Use Case Diagram

A use case diagram provides a high-level overview of the interactions between users who are the actors and the system itself. It shows how different actors (users or external systems) interact with the system through various use cases. The diagram helps in understanding the roles of different users and how they engage with the system to achieve their goals or perform specific actions. It is an essential tool for visualizing and communicating the overall structure and functionality of a system in a clear and straightforward manner (Dinesh Batra, 2019).

3.3.5 Class Diagram

A class diagram provides a comprehensive depiction of the object types present in a system and the static relationship that exists among them. It showcases the methods, attributes, relationships, and operations of objects as well as any constraints imposed on the system (Pressman, 2020).

In the proposed system, a class diagram will be utilized to visually represent the various types of objects in the system and illustrate the relationship that exist between them. Furthermore, the class diagram will aid in the construction of executable and accurate code for the model application. It serves as a valuable tool for designing and implementing the system by providing a clear representation of the object structure and their interactions.

3.4 Deliverables

System deliverables are the services or outputs that are expected to be supplied at various phases of system development as well as at the end of a project. These deliverables are critical in keeping projects on schedule and ensuring efficient time and resource allocation.

3.4.1 System Proposal

A system proposal is a document that explains the goals and needs of a project and seeks approval prior to its formal start. It provides critical information regarding the value of the system, associated risks, validation requirements, and implementation scope.

3.4.2 System Design Diagrams

System design diagrams are visual representations of a system's structure and components. They will offer a high-level overview of the system's architecture, enabling an understanding of how the various components interact and work together to achieve the set systems objective.

3.4.4 Pre-processed Data

Pre-processed data is data that undergoes a series of transformation to make it suitable for analysis or training of a model. This provides high quality, usable and effective data that will be used for training the model.

3.4.5 Model

A model refers to a mathematical or computational depiction of a real-world system or problem. It will be developed using an algorithm that captures the patterns and correlations of the pre-processed data, thereby undergoing training.

3.4.6 User Interface

A user interface refers to a graphical interface that through it users interact with a software application, website or any digital system. The model will be integrated with a user interface enabling users to interact with it.

3.5 Tools and Techniques

The tools and techniques that will be used in developing the proposed system will include:

3.5.1 Scikit-Learn

Scikit-learn, often known as Sklearn, is an open-source machine learning framework that provides developers with a complete collection of tools and functions. It provides a user-friendly and legible syntax, making it easier to properly employ programming resources. (Géron, 2019) Sklearn will be used in the suggested solution because of its wide collection of pre-built functions and advanced procedures that simplify the process of generating multiple models. Sklearn is well-known for its versatility and extensive range of machine learning algorithms, which enable quick model creation, evaluation, and fine tuning for the specified categorization task.

3.5.2 Python

Python is an extremely powerful programming language that's known to be high-level, interactive, and object-oriented. It is distinguished from other languages by its readability and coherence (Géron, 2019). Python has a large ecosystem of tools and frameworks that help shorten the development time for machine learning projects. These resources contain ready-to-use functions and technologies that help developers working on machine learning tasks be more efficient and productive.

3.5.3 GitHub

GitHub is a web-based platform and cloud service that assist developers in storing, monitoring, and tracking changes to their codebase. (Mudholkar, 2017). The proposed solution code will be pushed to GitHub to ensure effective version control. Using GitHub UI, developers can easily track changes and manage file revisions, making it a vital throughout the development process of the proposed solution.

3.5.4 JupyterLab

JupyterLab is a popular open-source web based integrated development environment (IDE) for data analysis, machine learning, and scientific research. It supports a variety of programming languages and has an easy-to-use interface for creating and sharing interactive notes. JupyterLab wide ecosystem and plugins enable users to personalize and enhance their working environment to meet their individual requirements. (Jupyter Team, 2021).

Chapter 4: System Analysis and Design

4.1 Introduction

This chapter outlines how the system functional and non-functional requirements of the system and the different diagrams showing how the system will be designed and analyzed.

4.2 System Requirements

The functionalities that a system needs to meet user needs are identified by system requirements. They lower implementation costs and make it possible to suit user needs very effectively. The following are a few of the system requirements reviewed for the project:

4.2.1 Functional Requirements

Functional requirements explain the behaviour of the system under specified conditions and are the features that must be implemented for a user to be able to complete their tasks. They include:

- i. User interface module - This module serves as the interface through which users engage with the system. Its purpose is to streamline user-system interactions, ensuring a user-friendly experience that demands minimal effort to achieve desired outcomes.
- ii. Predict feature- This feature enables users to complete a form, which is subsequently utilized by the classifier to generate predictions.

- iii. Database – The user details are stored here together with the predictions made for statements.
- iv. Classifier module- This module is fed with the statements that have passed by user to be verified as true, mostly true, half true, barely true, false, pants on fire

4.2.2 Non-Functional Requirements

Non-functional requirements are a set of specifications that specify the system's operation capabilities and limits, with the goal of improving its functionality. They define a system's quality qualities and describe how it functions. Although they are not required, they are necessary for validating a system's performance.

The system non-functional requirements of the system include:

- i. System reliability- Denotes the likelihood that the programme will operate without interruption for a particular amount of time. The system's reliability is assured through preventative maintenance, in which the hardware and software are subjected to regular and routine maintenance to keep them functioning and avoid unscheduled downtime due to unexpected failure.
- ii. System security - Ensures that the software and its data are safe from unauthorised access. Any user information in the system is encrypted, and the database is constantly backed up.
- iii. System usability - Describes how difficult it will be for a user to learn and utilise the system. The system's user interface is user-friendly, allowing the user to readily comprehend and navigate the system. Because the interface is simple, the user can do most tasks without assistance.
- iv. System performance – This is a quality attribute that describes the system's responsiveness to various user interactions with it. Poor performance results in a negative user experience. For optimal performance, the web pages in the system should load in less than five seconds and have an efficient throughput and reaction time.
- v. System availability – This is determined by how long the system's features and services are available for use across all operations. The system's availability is ensured through shortening recovery times in the event of a failure. This includes debugging issues, replacing components as needed, and restarting the database service. The system is also fault-tolerant, which means that it can continue to operate without

interruption even if one or more of its components fail. The information database is constantly backed up so that if the primary database fails, operations can be diverted to the backup database.

4.3 System Analysis Diagrams

The diagrams used for system analysis include:

4.3.1 Use Case Diagram

A use case diagram shows how various actors interact with the system.

Figure 4.1 shows the social media user, the administrator, and the tasks that each of the actors can perform. The user first registers and then logs in before making predictions by filling out a form that captures statements and their properties. Following that, a verdict for each statement is displayed. Users can also generate reports that display the verdict for their submitted statements. The administrator, on the other hand, gets access to a dashboard from which they manage users and view various statements submitted by users through reports generation.

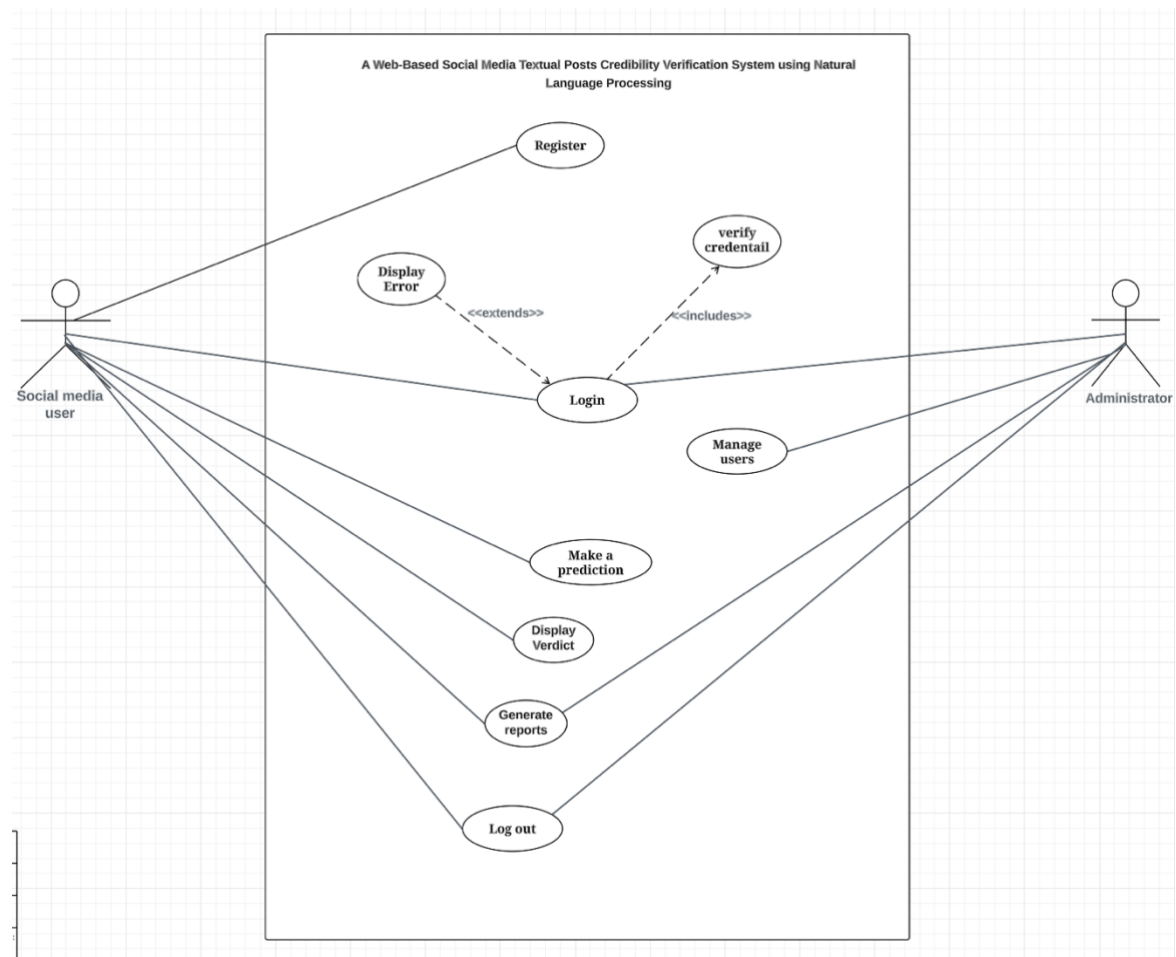


Figure 4.1 Use Case Diagram

Use Case Scenario one: Initiate Credibility Verification

Use case	Initiate Credibility Verification
Actor	user
Pre-condition:	Be logged into the system All the data fields must have input
Post-condition:	The statements are categorized into one of 6 categories: true, mostly true, half true, mostly false, false, and pants on fire
Alternate condition/action:	All fields must have data otherwise the credibility verification will not be initiated when the button is pressed pressed.

Use Case Scenario Two: Input Values

Use case	Input Values
Actor	user
Pre-condition	-Be logged into the system
Post-Condition	All the data fields should have a relevant value input before moving onto initiation of credibilty Verification

Alternate condition/action	All fields must have data otherwise credibilty Verfication will not be initiated when the button is pressed
----------------------------	---

Use Case Scenario Three: View verdict

Use case	View verdict
Actor	user
Pre-condition	Be logged into the system A credibilty verdict must be present
Post-Condition	The result shoud be displayed
Alternate condition/action	The system should prompt the user to input values for the credibilty verdict to be given

4.3.2 Sequence Diagram

A sequence diagram shows the interaction between objects in a sequential order.

They describe how and in what order the objects in a system work together to achieve the goal of the system. The social media user first fills a from that initiates a classification, the form's details are subsequently saved in the database and submitted to the classifier for classification. Following classification, a verdict is generated and displayed to the user, as well as being saved in the database.

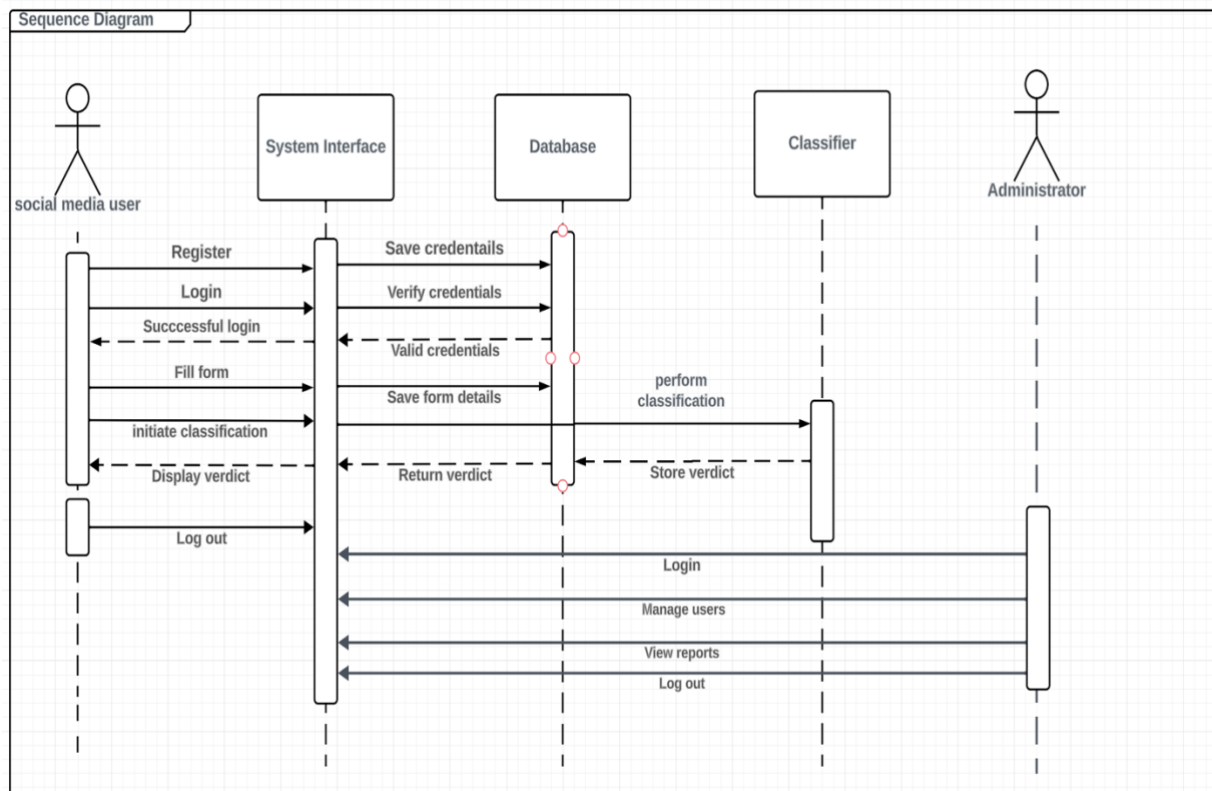


Figure 5.2 Sequence Diagram

4.3.3 Entity Relationship Diagram

An Entity Relationship Diagram (ERD) is a graphical representation of the many entities in a system and the relationships that exist between them. Figure 4.4 shows the entities in the system how they relate to each other and their interaction. A user fills one or many forms that have one statement that are the classified by a model, each statement receives a one verdict.

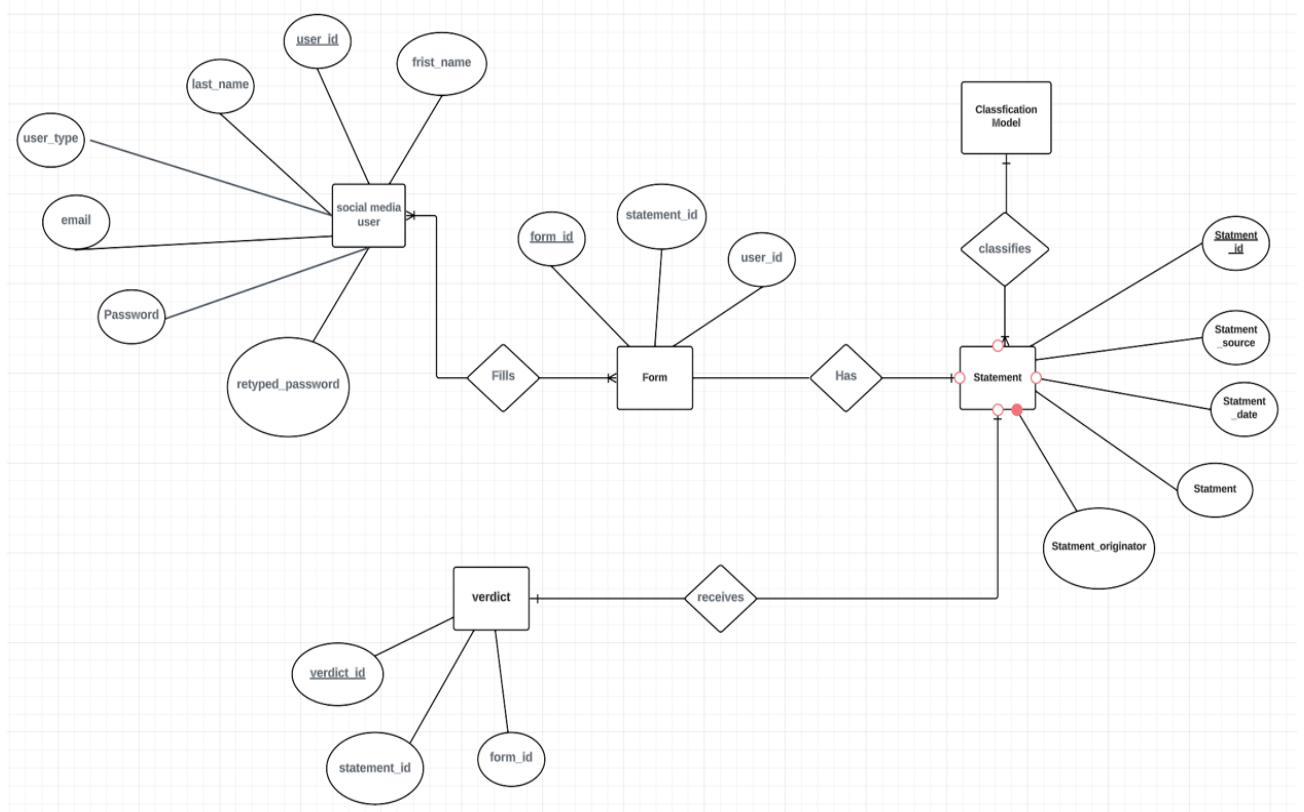


Figure 6.3 Entity Relationship Diagram

4.3.4 Activity Diagram

An activity diagram depicts the flow of control in a system and refers to the steps involved in executing a use case. They are used to represent the sequential and concurrent actions that occur in the system. They aid in the visual representation of system workflows. In figure 4.3 we see the flow of activities in the system. A user first as to login and then fills a form that is used to make a prediction the form details are stored in the database then they are processed by a classifying model, the model then gives a verdict that concurrently its stored in a database.

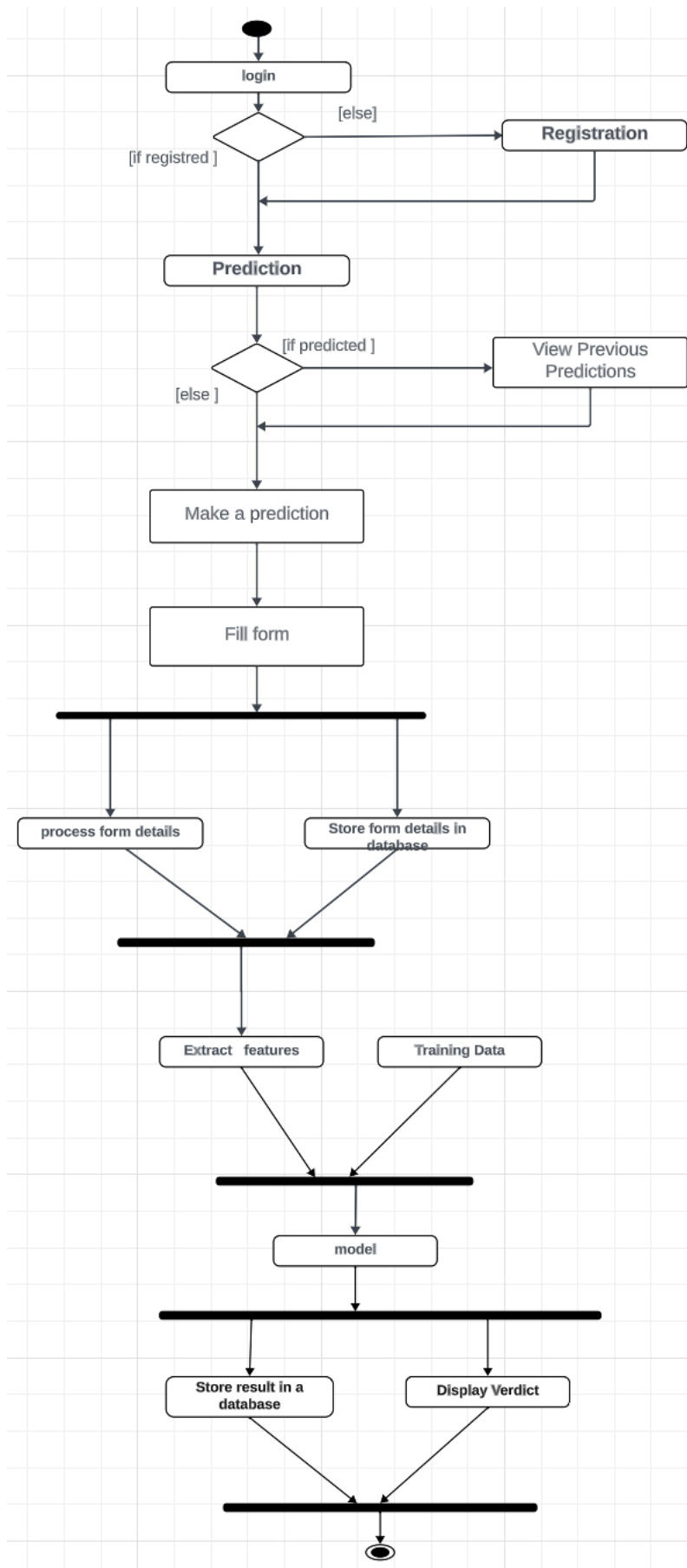


Figure 7.4 Activity Diagram

4.4 System Design Diagrams

The diagrams used for system design include:

4.4.1 Class Diagram

A class diagram is a sort of static diagram that depicts the structure of a system by displaying the system's classes, characteristics, processes, and object relationships. Figure 4.5 shows the classes in the system which are the social media user, the form, the statement, the classifier and verdict. The characteristics and processes of each class are shown together with the relationships between classes.

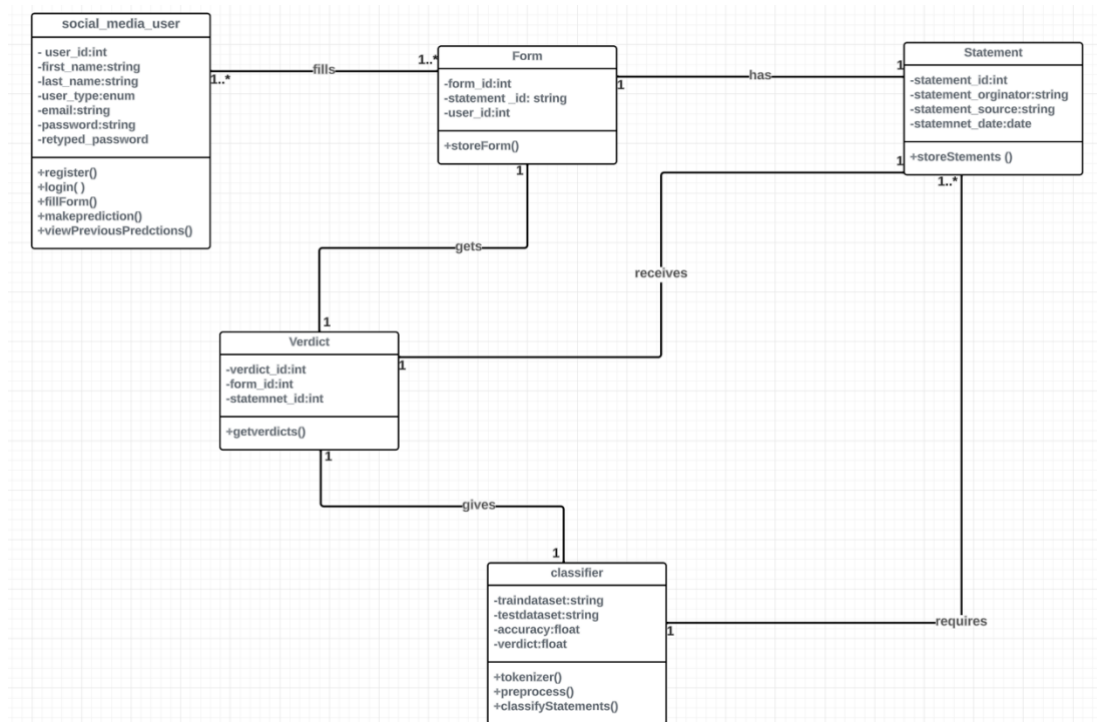


Figure 8.5 Class Diagram

4.3.4 Database Scheme

A database schema is the skeleton structure that represents the database's logical perspective. It outlines how the data is organised and how it may be related to other tables or data models.

As shown in figure 4.6 when the user registers the details are stored on the social_media_users table which are used for logging in. The form table is used to take account of the forms being filled and acts as a link for all the other tables as it contains the primary key for the other tables as foreign keys. Once a form is filled the details of the form are then stored in the statement tables and once a verdict is carried out for a statement the verdict table takes account of this.

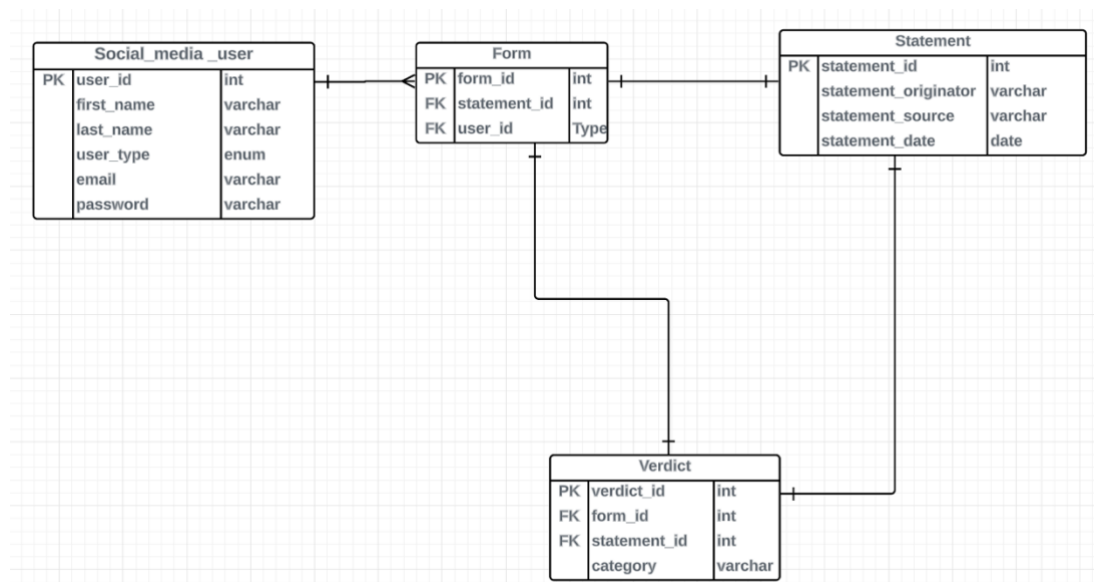


Figure 9.6 Database Scheme

4.3.4 System Architecture

A system architecture diagram is an abstract representation of the component architecture of a system. It provides a quick overview of the system's component architecture to aid in component-component interactions and system functionality. It depicts the system's key operations as well as the relationships between the various system components. The figure 4.7 shows the user can make a request to access service from the system by send a request from the

web browser after being connected to the internet. The web browser then requests the web server for the request via the internet from which the request is taken to the database server that responds accordingly to the request made.

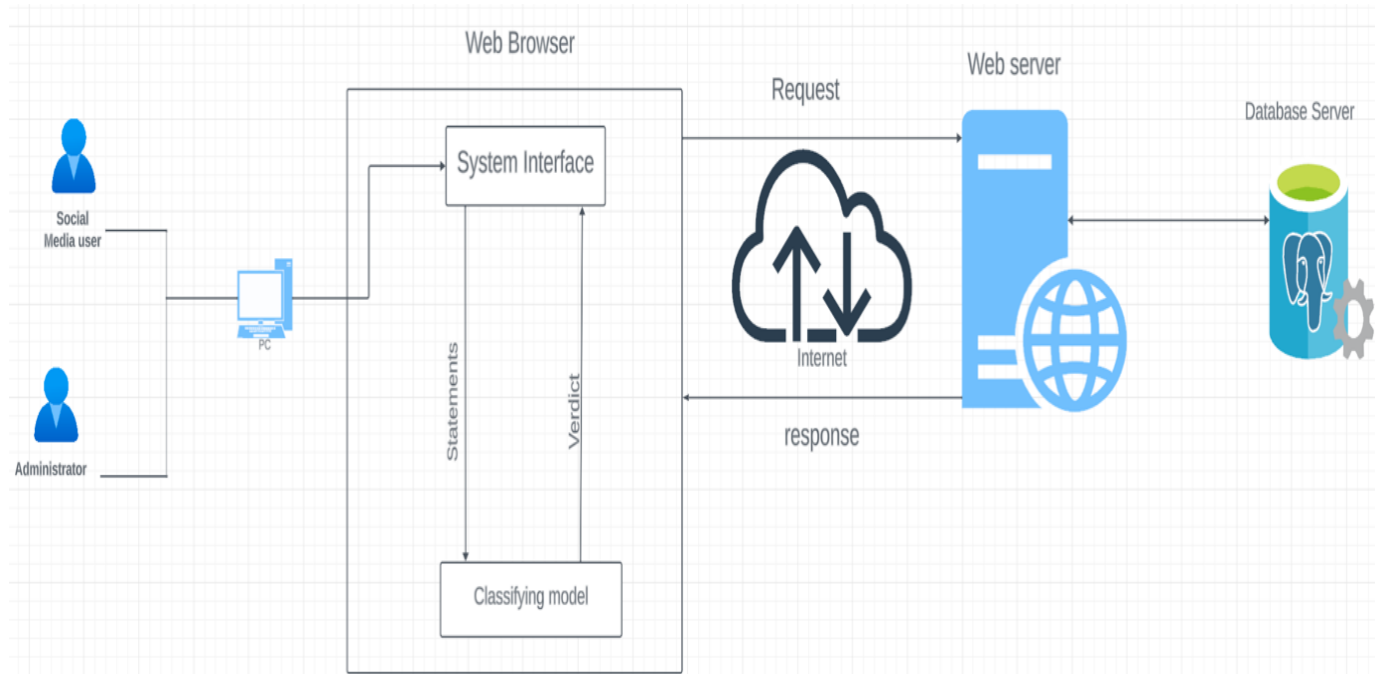
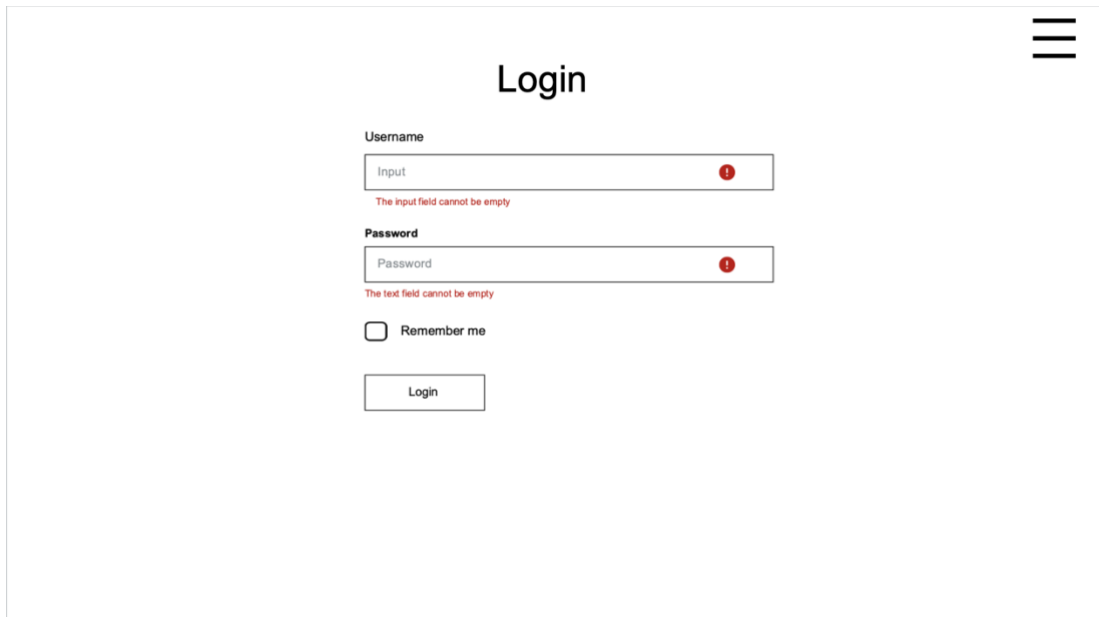


Figure 10.7 System Architecture

4.3.4 Wireframes

Wireframes give a visual guide of how the web application will look like. Figure 4.8 shows the login page that a user inputs their username and password.



≡

Login

Username

Input !

The input field cannot be empty

Password

Password !

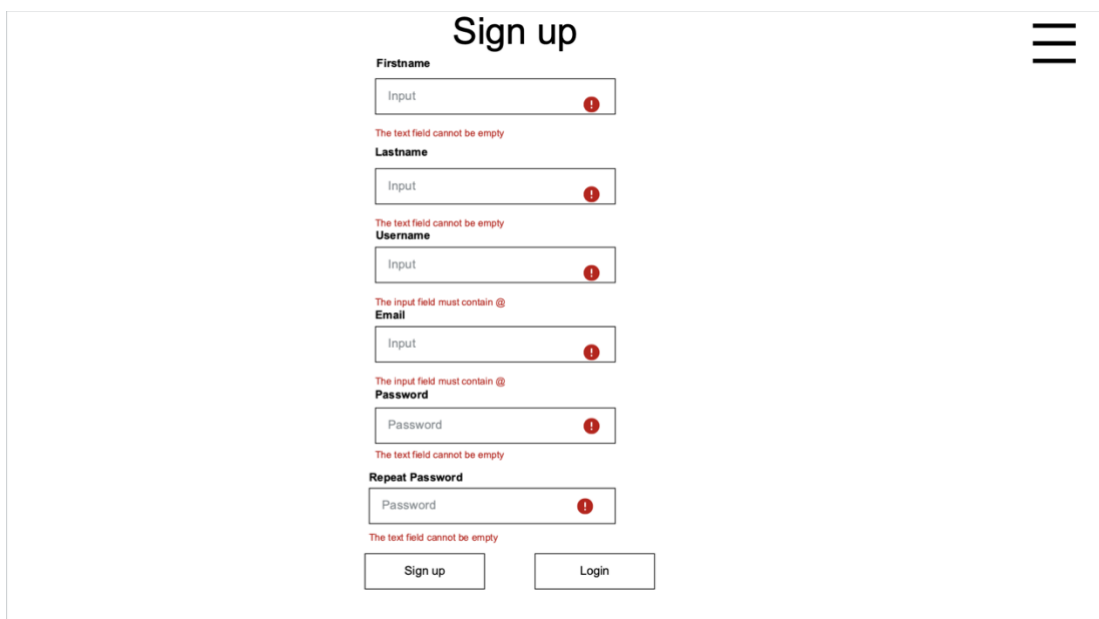
The text field cannot be empty

☐ Remember me

Login

Figure 11.8 Login Page

Figure 4.9 shows a registration form that a user fills to be able to login into the system to make a prediction.



≡

Sign up

Firstname

Input !

The text field cannot be empty

Lastname

Input !

The text field cannot be empty

Username

Input !

The input field must contain @

Email

Input !

The input field must contain @

Password

Password !

The text field cannot be empty

Repeat Password

Password !

The text field cannot be empty

Sign up Login

Figure 12.9 Registration Page

Figure 4.9 Shows make a prediction page where the user fills a form and then does a prediction and

Credibility Verification

Statement

The text field cannot be empty

Statement Originator

The text field cannot be empty

Statement Source

The text field cannot be empty

Statement Date

The text field cannot be empty

Predict

Figure 13.10 Make A Prediction Page

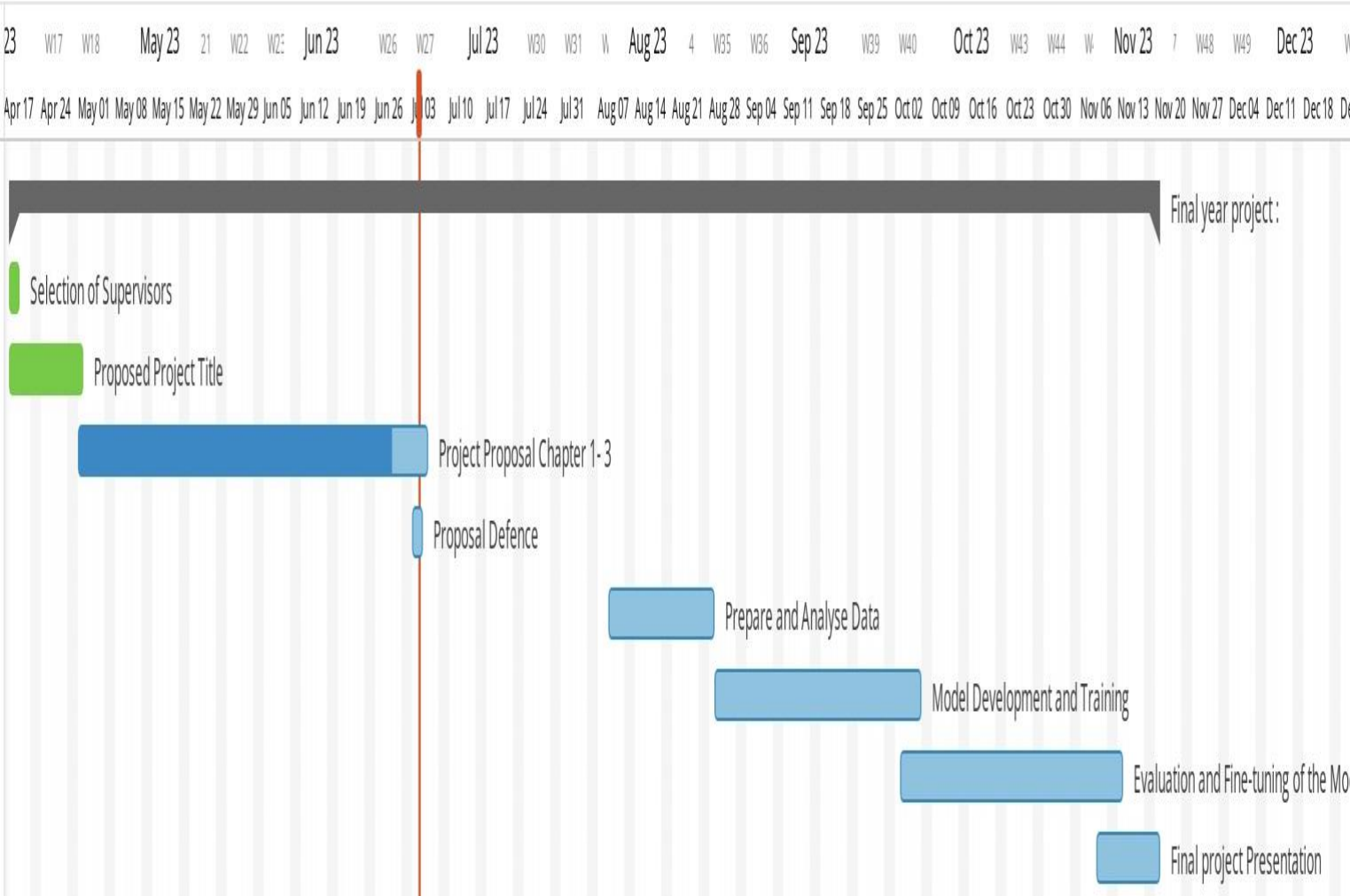
References

- Pennycook, G., & Rand, D. G. (2019). The Implied Truth Effect: Attaching Warnings to a Subset of Fake News Stories Increases Perceived Accuracy of Stories Without Warnings. *Managemnet Science* , 67(11), 4944-4957.
- Vosoughi, S., Roy, D., & Aral, S. (2018). The Spread of true and false online. *Science*, 359(6380), 1146-1151.
- Guess, A., Nagler, J., & Tucker, J. (2020). Less than you think: Prevalence and Predictors of fake news . *Science Advances* , eaay3539.
- Allcot, H., & Gentzkow, M. (2017). Social media and fake news in 2016 election . *Journal of Economic Perspectives*, 31(2), 211-236.
- shu, K. D. (2017). FakeNewsNet: A Data Repository with News Content, Social Context and Dynamic Information for Studying Fake News on Social Media.
- Guess, A. N. (2019). Exposure to untrustworthy websites in the 2016 US Election. *Nature Human Behaviour*, 3(4) , 308-313.
- Jursafsky, D. &. (2020). Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition (3rd ed.). Pearson.
- Potthast, M. H. (2018). A Plan for Actionable.
- Horne, B. D. (2018). This just in: Fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news. *Proceedings of the International AAAI Conference on Web and Social Media*, 12(1), 607-616.
- Wardle, C., & Derakhshan, H. ., (2019). Information disorder: Toward an interdisciplinary framework for research and policy making. *Council of Europe report*, 1(0), 1-63.
- Lazer, D. M. (2018). The science of fake news. *Science*. 359(6380), 1094-1096.
- Chaffey, D. (2022, June 1). *Global social media statistics research summary*. Retrieved from Smart Insights: <https://www.smartinsights.com/social-media-marketing/social-media-strategy/new-global-social-media-research/>
- Menczer. (2020). The science of fake news. *Science*. 359(6380), 1094-1096.
- Lewandowsky, S., Ecker, U. K., & Cook, J. . (2020). Beyond Misinformation: Understanding and Coping with the “Post-Truth” Era. *Applied Research in Memory and Cognition*, 9(4), 434-440.

- Shaina Raza and Cheng Ding. (2022). Fake news detection based on social contexts: a tranformer based approach . *International Journal of Data Science and Analytics* .
- Rohit Kumar Kaliyar, A. G. (2021). FakeBERT: Fake news detection in social media with a BERT-based deep learning approach. *Multimedia Tools and Applications*, 80:11765–11788.
- Pallabi Saitaki, K. G. (2022). Modelling Social Context for Fake News Detection: A Graph Neural Network Based Approach. *Pandit Deendayal Energy University*.
- Shao, C., Ciampaglia, G. L., Varol, O., Yang, K. C., Flammini, A., & Menczer, F. (2018). The spread of fake news by social bots.
- Martin, M. (2020). *Prrototyping in Software Engineering: Methodolgogy Process*. Retrieved from Ace Infoway: <https://www.aceinfoway.com/blog/project-scope-and-prototype-model>
- Pressman, R. S. (2020). *Software Engineering: A Practitioner's Approach*. New York : McGraw-Hill Education.
- Bahill, A. T. (2019). Tradeoff decisions in system design. Springer. *Science+Business Media*.
- Jeffrey A. Hoffer, J. G. (2021). *Modern Systems Analysis and Design* . Pearson Education .
- Brahma Dathan, S. R. (2016). *Object-Oriented Analysis, Design and Implementation: An Integrated Approach*. Springer .
- John W. Satzinger, R. B. (2015). *Systems Analysis and Design in a Changing*. Cengage Learning.
- Dinesh Batra Joey F. George, J. S. (2019). *Object-Oriented System Analysis and Design*. Ingram.
- Géron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 2nd Edition*. O'Reilly Media, Inc.
- Mudholkar, M. &. (2017). *A Beginner's Guide to Git and GitHub*. Retrieved from ResearchGate: <https://doi.org/10.13140/RG.2.2.20126.89927>
- Jupyter Team. (2021). *JupyterLab: A web-based interactive development environment for Jupyter notebooks, JupyterLab 3.0*. Retrieved from <https://jupyterlab.readthedocs.io/en/stable/>

Appendices

Appendix 1:Timeline of Activities



Appendix 2: Marking Guide

Student Number(s)

Working Title:

Evaluation Points	Weight	Score	Notes
Title page: Informative, concise and appropriate	2 pts		
Abstract To have background, problem, solution, methodology (approach data and tools) outcomes and expectations <i>Check on Completeness and correctness</i>	3 pts		
Introduction Background (2) <i>A clear illustration of issue, context and audience</i> Problem Statement (2) <i>Pain points, audience, who is affected and how solution comes in to fix the pain. What facts support this</i> Objectives (S.M.A.R.T and Linked to Problem Statement) (3) Research questions (3) <i>Alignment of questions with objectives</i> Justification (2) <i>Should be research supported.</i> Scope of Project (2) <i>Specify boundaries of people process, HW/SW, data etc</i> Limitations (1.5) <i>Challenges Expected</i> Delimitation (1.5)	(17 pts)		

<p><i>To do to counter anticipated challenges</i></p> <p><i>Check for correctness, completeness and citation of work</i></p>			
<p>Literature Review/Related Work</p> <p>Literature objectives mapping as aligned with research questions (2)</p> <p>Critique of content adequacy of What it is, how it presents, its implications, Citations of content align with work (4)</p> <p>Review of at least 3 systems comprehensively the working behind it (2)</p> <p>Gaps identification, analysis relative to the proposed solution (3)</p> <p>Conceptual Framework clear to communicate how it works, data flows, processing, actors (2)</p> <p>Describe input process output storage boundaries Emerging technologies contextualization (5)</p>	(18 pts)		
<p>Intended Approach/ Methodology</p> <p>Research Design (2) <i>experimental, casual etc to determine type of data to be used, Variables etc</i></p> <p>Research Methodology (5) <i>Methodology (1), Correct process (1), Design and Development tools (1), Research Paradigm (2)</i></p> <p>Deliverables and milestones (2) <i>Examinable bits from ideation</i></p>	(9 pts)		

<i>Proposal, design, test cases documentation doc</i> <i>Proof of concept- modules</i>			
Proposal Presentation Table of Contents and List of Figures (2) Are relevant references provided and formatted correctly? (1) Is there a clear and proper use of language? (1) Effective report structure (chapters and sections) and layout (2)	(6 pts)		
Total Marks	55		

Verdict (Please tick)

☐

Accepted

☐

Reject

Comments (**Especially if verdict is reject**)
