INFORMATICS INSTITUTE OF TECHNOLOGY

In Collaboration with

ROBERT GORDON UNIVERSITY ABERDEEN

Fake News Detection (COVID-19)

Twitter

A Final Thesis by

Student 1	Dilan Perera	2016274	20200124
Student 2	Chathuni Manage	2017915	20200997
Student 3	Chamathi Gunawardana	2017926	20201018
Student 4	Jenuk Giles	2017914	20200988

Supervised by

Mr. Nuwan Jayawardena

Submitted in partial fulfillment of the requirements for the BSc (Hons) in Artificial Intelligence And Data Science degree at the Robert Gordon University.

February 2022

Abstract

Fake news spreads rapidly and at a faster rate than actual news. This is most likely due to the fact that false news is frequently more unique or exciting than actual news and contains more superlatives. False tweets also had more rumor route propagation hops than true news, implying that fake news tweets are retweeted more frequently. Tweets about true news pieces, on the other hand, tend to flow slowly and steadily, reaching fewer people overall.
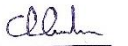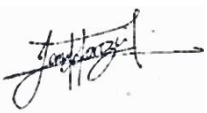
Social networking has become an inextricable element of modern life. Information is disseminated through one of several platforms that disseminate essential news at all hours of the day and night. This information might come from anywhere on the globe. Although such a linked display of communication has brought huge benefits, it may also have disadvantages. The propagation of disinformation on all social media platforms, particularly Twitter, is one such example. The effects of this disinformation range from innocuous to deadly, with global ramifications. Twitter, in particular, is fascinating to examine since it is utilized by so many people and can attract their attention with bite-sized, attention-grabbing messages. Though other social media sites may be able to look further into the disinformation, individuals who post misinformation do so for convenience rather than malicious purpose, making Twitter a better venue to investigate the topic. Limiting disinformation or recognizing and showing which tweets are false might help users learn the truth and prevent them from some of the potential dangers.

Key Words

Machine Learning, Python model, Twitter API

Declaration

I hereby certify that this project report and all the artifacts associated with it is my own work and it has not been submitted before nor is currently being submitted for any degree program.

Full Name of Student	RGU ID	IIT ID	Date	Signature
Kalthotage Dilan Gordon Perera	2016274	20200124	3th April,2022	
M.M.Chathuni Devindi Manage	2017915	20200997	3th April,2022	
Chamathi Lawanya Gunawardana	2017926	20201018	3th April,2022	
Jenuk Giles	2017914	20200988	3th April,2022	

Acknowledgement

We had to recruit the support and guidance of reputable persons who deserve my sincere appreciation in order to prepare for our assignment. We would like to express our appreciation to Mr. Nuwan Jayawardena (Supervisor) for providing us with a nice guidance for tasks throughout many consultations. We'd want to thank everyone who helped us write this assignment, both directly and indirectly. Several individuals, particularly our classmates have suggested important ideas, inspiring us to increase the quality of the work.

Table of Contents

CHAPTER 1 - INTRODUCTION

1.1. Chapter Overview

In this chapter we will be focusing on the main aspects of the problem we are trying to solve. It includes the motivation behind the selection of this domain and the work that already exists with regard to this project. This chapter also includes the entire scope of the project which includes the Software, Hardware and the Data that has been used for the project.

1.2. Problem Domain

Individual people' actions, guided by the quality of the information they have, are critical to the global response to this health crisis' success. The International Fact-Checking Network (IFCN), which brings together over 92 fact-checking organizations, had discovered over 7,623 unique fact-checked articles about the pandemic by July 18, 2020.

1.2.1. Significance of the Research

Fake news is not a recent concept. Before the rise of social media, news broadcasting was limited to radio, newspapers, and television, while journalists and other news producers were tasked with filtering out fake news. However, due to the fast growth of user-generated content, these typical filtering strategies are no longer effective. According to research, humans are not effective in identifying false fabrications in text-based content alone, thus there has been a push to automate news trustworthiness rating. If the project manages to be successful, it will expand people's critical evaluations of online media and spread awareness of its existence. The model will concentrate on limiting the spread of fake news and supporting users in judging the credibility of the content the general public comes across.

1.2.2. Magnitude of the Problem

Isolation is a form of torture that breaks even the strongest individuals. Due to this pandemic, people have been in isolation for over a year and a half, even more in some cases. If it weren't for social media and the ability to communicate with others

worldwide, most individuals would have developed various mental diseases and problems. The only way they keep in touch with reality and the outside world is through social media, and most of them believe what they see on social media to be true almost all the time. When society starts believing that all facts and news seen on social media is true it causes a great deal of untrustworthiness and credibility among the said media platforms. Which is why it is important that we focus on minimizing the coverage of fake news and increasing the spread of authentic news.

In 2020, Twitter grew dramatically, probably as a result of the COVID-19 pandemic. During the pandemic, Twitter observed an increase in the spread of misinformation about the disease. In March 2020, Twitter said that it will begin identifying tweets that may contain false or misleading information, and in some situations, it would provide links to fact-checking resources.

1.2.3. Project Issues and Difficulties

Obtaining a tagged dataset of tweets from Twitter is one of the project's primary challenges. More particular, because credibility analysis is subjective, it will be difficult to obtain labels that are accurate. Because the labels must be general enough to be used for various sorts of fake news, a wide range of tweets and users must be included in the crowdsourcing. Annotators who are unfamiliar with the content of the tweet may also submit incorrect labels. If we want to integrate the tweet's context and historical data in the assessment, it's challenging to provide credibility ratings in real time. This is due to the fact that there were additional tweets linked to the topic of the tweet.

Furthermore, the Twitter API greatly restricts the amount of study that can be done after the fact on how people react to and spread misinformation. The fact that the Twitter API does not allow for the retrieval of tweet replies older than seven days and limits the retrieval of retweets is one of the primary issues with gathering Twitter data.

1.3. Problem Definition

As a result, the goal of the project is to create a system that can evaluate the trustworthiness of tweets on Twitter using algorithms and display the results to the user. A classifier will be trained with annotated data from a crowdsourcing mobile app, and the results will be shown in the Twitter interface via a Web browser extension. The

focus of this research is to develop a more accurate method for detecting fake news about COVID 19 on Twitter.

The spread of fake news causes far more harm than good, always plunging the general public to panic and chaos. Although it is unclear what the objective of this propaganda is it will be difficult in finding solutions and working our way through the pandemic. In order to minimize the damage caused by COVID it must expunge all fake news and let the authentic news be widespread which would in turn help the general public.

1.3.1. Problem Statement

There is no accurate method to detect fake news about COVID 19 on Twitter.

1.4. Research Motivation

This research project's target is to create a highly accurate Fake News Detection System to detect fake news about COVID 19 on twitter. Due to the worsened mentality of the general public, it can be harder to detect fake news in their frame of mind. By identifying fake news and genuine news it will better help citizens make more suitable decisions. Moreover, if fake news happens to be widely disseminated, it will greatly influence the political balance of the given system and wreak chaos on the economy of the country, which will result in vast monetary loss.

1.5. Existing Work

Paper 1	Title	● Fake News Detection on social media using Geometric Deep Learning.
	Introduction	● This project mainly focuses on Twitter and uses fake news propagation patterns using geometric Deep Learning. Training has been done in a supervised manner.
	Features Extracted	● User profile , User activity. ● Network and spreading. ● Content
	Technology Used	● Geometric Deep Learning.

	Results	• This project has used URL and cascades. In the first setting they used the posts URL and the cascades that were linked to it and flagged the URL as true or false. In the second setting they focused on only one cascade and tried to flag it as true or false.
	Conclusion	• Naturally helps integrating heterogeneous data linked to the user profile and activity, social network structure, news spreading patterns and content. Can automatically learn task-specific features from the data.
Paper 2	Title	• Automated Fake News Detection Using Linguistic Analysis and Machine Learning
	Authors	• (Choudhary & Arora, 2021)
	Introduction	• In this project they created a Dataset with 345 valid news articles and used Linguistic Analysis and Word Count packages to obtain linguistic features of the articles. These features were normalized using Z-score normalization and they used an 80-20 approach when training the model
	Technology Used	• Text processing :- Based Machine Learning for automatic identification of Fake news. • Algorithms used :- Logistic Regression Support Vector Machine Random Forest Decision Tree K-Neighbors Classifier
	Limitations	• This model does not consider social media accounts or specific personal posts to compare. It focuses mainly on a broader picture when considering whether the article is real news or fake news.
	Conclusion	• Out of the algorithms used , the Support Vector Machine method gave the best prediction results.
Paper 3	Title	• Detection of Fake News on COVID-19 on Web Search Engines
	Authors	• Mazzeo, Rapisarda, & Giuffrida, 2021

	Introduction	●	When dealing with class imbalance through real-world data, they have applied re-sampling techniques such as over-sampling of fake news and under sampling of real news. To get satisfactory results the over-sampling technique was allowed.
	Technology Used	●	URL pattern analysis via phishing detection techniques can help ML algorithms' ability to detect the spread of fake news beyond the World Wide Web.
	Algorithms used	●	Machine learning workflow for training algorithms and classification of results from web search engines
	Conclusion	●	News has some specific textual properties that can be used to classify whether it is fake or real. When looking at search results such as title, snippets and links test features are related to news content, URL features are based on the source websites returned as search results on web search engines
Paper 4	Title	●	Fake News Detection Using Machine Learning Ensemble Methods
	Authors	●	Ahmad, Yousaf, Yousaf, & Ahmad, 2020
	Introduction	●	Social media platforms in their current state are extremely powerful and useful for their ability to allow users to discuss and share ideas. Some abuse this power using this for monetary gains, and other cases for creating biased opinions, manipulating mindsets, and spreading satire or absurdity. ● The rise of fake news has led to many problems in a wide range of domains. One such area affected by fake news is the financial markets. A rumor can have disastrous consequences and may bring the market to halt or even collapse, depending on who shares it.
	Algorithms	●	Linear SVM (LSVM). ● Convolutional Neural Network (CNN). ● Bidirectional Long Short-Term Memory Networks (Bi-LTSM).

	Datasets	● ISOT Fake News Dataset ● Kaggle (Total of 20,386 articles used for training and 5,126 articles used for testing) ● Kaggle (Total of 3,352 articles, both fake and true)
Paper 5	Title	● Fake news detection in social media.
	Author	● Kelly Stahl ● B.S. Candidate, Department of Mathematics and Department of Computer Sciences, California State University Stanislaus, 1 University Circle, Turlock.
	Introduction	● Fake news is the purposeful spread of false information through traditional news sources or social media. False information spreads quite quickly. People can download articles from websites and share the information, and re-share it again and again until it has spread so far from its source that it is indistinguishable from legitimate news.
	Technology Used	● Analytical Techniques ● Naive Bayes Classifier
	Conclusion	● Researchers are working to develop more reliable methods for detecting misleading information in social media. This study may be useful in assisting other researchers in determining which mix of methodologies should be utilized to accurately detect bogus news on Twitter, Facebook and other platforms. The proposed technique is a proposal for a more accurate fake news detection algorithm.

Table 1 : Existing Work

1.6. Research Gap

The COVID 19 pandemic was a colossal cork in the everyday world. Even though it was initially wreaking havoc in Wuhan, China after its worldwide spread it became extremely dangerous and numerous lives were lost to the pandemic. The virus, although a single strain, managed to evolve and mature eventually resulting in various strains of the virus. The only mode of communication among people is through various modes of social media. Technology services and various other

industries managed to create multiple platforms that would be dedicated to sharing information regarding the virus. The massive increase in working virtually and depending on the internet connectivity has laid bare the fragility of cyber communications systems and the chaos it entails. Apps may breed false confidence in technical fixes at the expense of more common sensical measures while leading us the society further down the slippery slope of a surveillance society.

The project will use Twitter as the base to focus on clarifying the deviations of fake news and authentic news. Currently Twitter differentiates between fake and factual news through a manual process of review and consultation of social media content. Conferring with various organizations and officials around the globe, Twitter actively mitigates the effects of fake news surfing the platform. This project will introducing machine learning to filter out the data and web crawl across multiple credited sources to efficiently and accurately distinguish between fake news and authentic news. The project will be focusing on the different types of COVID strands all the while detecting and dynamically altering fake news and real news.

This project will be presented in a web app to showcase real news as well as creating a web extension to gather data that will be used in cross referencing between the fake news and genuine news. This would prove to be extremely beneficial in effectively communicating COVID alerts among the citizens while clearing out the disruptive reports that would otherwise cause worldwide chaos.

1.7. Contribution to the Body of Knowledge

1. Technological contribution

    ● We will be introducing machine learning into the system to help with the detection of fake news.
    ● We will be developing a web extension to help cross reference with reputed sources.

2. Domain contribution

- We will be creating a web app to show the real and fake news.
- This app will categorize the news according to the strands of COVID that are there.

1.8. Research Challenges

Challenge	Reason	Solution
Insufficient time	It is difficult to manage time with another lecture and also difficult to find free time for all team members.	Fix a specific time for each week.
Lack of knowledge to understand some technical things	Because we are still learning, we have less experience	We have to self-study using external sources (You tube, LinkedIn)

Table 2 : Research Challenges

1.9. Research Questions

RQ1 : How to use AI and Machine learning techniques in fake news detection?

RQ2 : Who is most affected by the spread of misinformation (Age group, Country , Social status)?

RQ3 : What are the methods used to differentiate fake news and real news?

RQ4 : What is the impact of fake news on the different strands of the virus?

1.10. Research Aim

This proposed Fake News Detection System will identify fake news about COVID-19 among twitter and help the users to recognize whether the news is fake or real.

1.11. Research Objectives

Research Objectives	Explanation	Learning Outcome
Problem Identification	Fake news detection	LO1
Literature Review	RO1= To identify the most efficient method to identify fake news. RO2=To identify what subset of news the project should focus on. RO3=To identify what social media outlet to focus on.	LO1
Data Gathering and Analysis	Kaggle and other online sources	LO2, LO3

Table 3 : Research Objectives

1.12. Project Scope

1.12.1. In-Scope

No.	Description
1	To recognize false news on Covid - 19
2	Tag false medical news authors
3	Fact checking of the Covid – 19 virus.
4	Immunize users with true news of Covid - 19

Table 4 : In - scope

1.12.2. Out-Scope

No.	Description
1	Intention detection
2	Potential fake medical news topics detection
3	Remove malicious accounts
4	Suspicious Covid – 19 data source analysis

Table 5 : Out - scope

1.12.3. Prototype Diagram
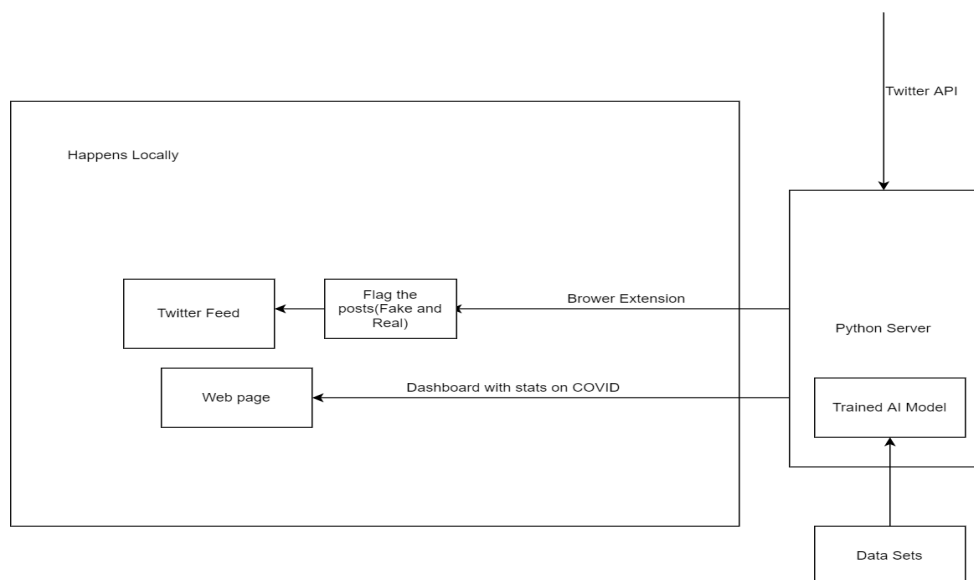
1.13. Resource Requirement

1.13.1. Hardware Requirements

- Intel Core i7 9th gen processor or higher. To carry out the basic functions required for the project.
- 16gb of ram- To process the data.

1.13.2. Software Requirement

● Python- For the programming aspect of the project
● Google Co-labs- To train the AI model.
● JavaScript-For the web application.

1.13.3. Data Requirement

● We will be gathering datasets from Kaggle and other online sources.

1.13.4. Skill Requirement

● Knowledge in Google Co-labs.
● Python Knowledge
● Web Development Knowledge.

1.14. Chapter Summary

This document has given a clear overview of the project idea of the Fake News Detection System. Very first part describes this chapter. After that, the problem domain and problem definition are clearly defined. Furthermore Research Motivation, Existing work and Research Gap are available. After that technological and domain contributions were defined. Then includes Research Challenges, Research Questions, Research Aim and Research Objectives. Finally, describe Project Scope and resource requirements of the Fake News Detection System under the Introduction Chapter.

CHAPTER 2 - LITERATURE REVIEW

2.1. Chapter Overview

This chapter covers the Domain of the problem this project solves. And developers do the research about existing works related to the project. Then this chapter clearly defines technologies, algorithms, tools and techniques used by previous developers for existing projects.

2.2 Concept Map
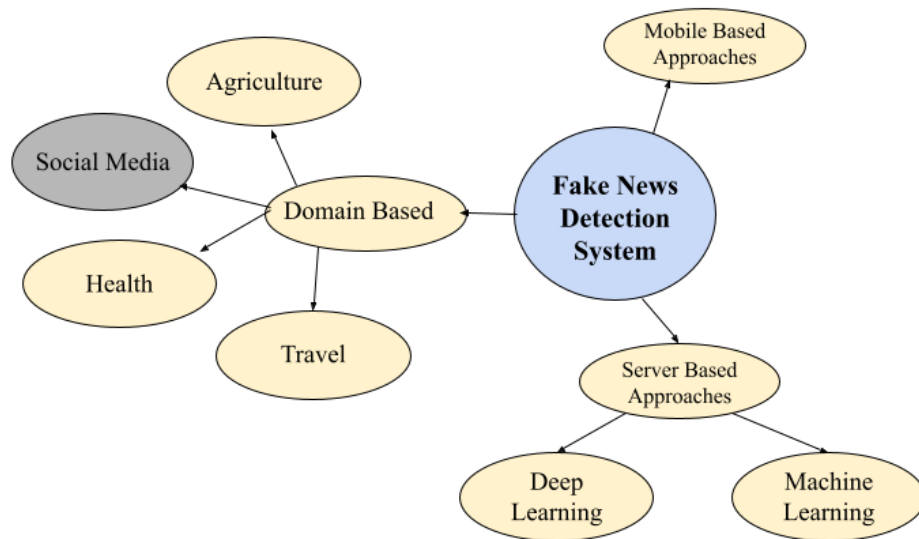
2.3 Existing Work

2.3.1. Fake News Detection on social media using Geometric Deep Learning

This project mainly focuses on Twitter and uses fake news propagation patterns using geometric Deep Learning. Training has been done in a supervised manner.

Methods and Materials Used

Features extracted:-

- User profile (glocalization and profile settings, language, word embedding of user profile self-description, date of account creation, and whether it has been verified), User activity (number of favorites, lists, and statuses).
- Network and spreading (social connections between the users, number of followers and friends, cascade spreading tree, retweet timestamps and source device, number of replies, quotes, favorites and retweets for the source tweet).
- Content (word embedding of the tweet textual content and included hashtags).

Technology Used:-

- Geometric Deep Learning

Results

This project has used URL and cascades. In the first setting they used the posts URL and the cascades that were linked to it and flagged the URL as true or false. In the second setting they focused on only one cascade and try and flag it as true or false.

Conclusion

Naturally helps integrating heterogeneous data linked to the user profile and activity, social network structure, news spreading patterns and content. Can automatically learn task-specific features from the data.

2.3.2. Automated Fake News Detection Using Linguistic Analysis and Machine Learning

In this project they created a Dataset with 345 valid news articles and used Linguistic Analysis and Word Count packages to obtain linguistic features of the articles. These features were normalized using Z-score normalization and they used an 80-20 approach when training the model.

Technology Used

Text processing-based Machine Learning for automatic identification of Fake news.

Algorithms used: -
- Logistic Regression
- Support Vector Machine
- Random Forest
- Decision Tree
- K-Neighbor Classifier.

Limitation

This model does not consider social media account or specific personal posts to compare. It focuses mainly on a broader picture when considering whether the article is real news or fake news.

Conclusion

Out of the algorithms used , the Support Vector Machine method gave the best prediction results. When considering the results of the project it was clear that fake news articles were shorter in length, conveyed less confidence, appeared in a more negative tone and has less analytical thinking but surprisingly the fake news articles seemed to be more authentic and displayed a more personal disclosing tenor.

2.3.3. Detection of Fake News on COVID-19 on Web Search Engines

When dealing with class imbalance through real-world data, they have applied re-sampling techniques such as over-sampling of fake news and under sampling of real news. To get the satisfactory results the over-sampling technique was allowed. The under-sampling method was not able to increase the model performance because it shows a very poor result due to the small sample size.

Technology Used

URL pattern analysis via phishing detection techniques can help ML algorithms' ability to detect the spread of fake news beyond the World Wide Web.

Algorithms used: -

Machine learning workflow for training algorithms and classification of results from web search engines.

Conclusion

News has some specific textual properties that can be used to classify weather it is fake or real. When looking at search results such as title, snippets and links test features are related to news content, URL features are based on the source websites returned as search results on web search engines.

2.4 Chapter Summary

This chapter focuses on finding the most appropriate algorithm/technique/approach to achieve critical research on the problem domain. First, it outlines a concept map that gives an overview of how data is collected and how the algorithms are selected. Then describes the algorithms, Techniques used in existing systems.

CHAPTER 3 - METHODOLOGIES

3.1. Chapter Overview

This chapter covers the methods used in the areas of research and project development. And describes the methodology of project management. Under the project management methodology presents project plan, deliverables, risk associated with project and mitigation plan.

3.2. Research Methodology

Research Methodology	Philosophy	Interpretivism will be used since our project depends on peoples posts.
	Approach	Deductive since there are past systems that have been deployed.
	Strategy	A quantitative survey was used in this project.
	Choice	Multi-Method had been used in this project.
	Time Horizon	The project will be using a longitudinal time horizon since it will be considering a global area in gathering data.

3.3. Development Methodology

Development methodology is an important aspect of any project. It helps organize the project and make the work flow more efficient. This project will be using an agile development methodology. Agile

methodology can help us minimize risks because we can go back and fix defects that we hadn't fixed. This will make the project more flexible and it would help us reduce the defects in the final product. Since this methodology does not have a certain fixed set of rules we can mold it into making it so that the group will be able to work efficiently.

3.4. Project Management Methodology

3.5.1. Project Plan
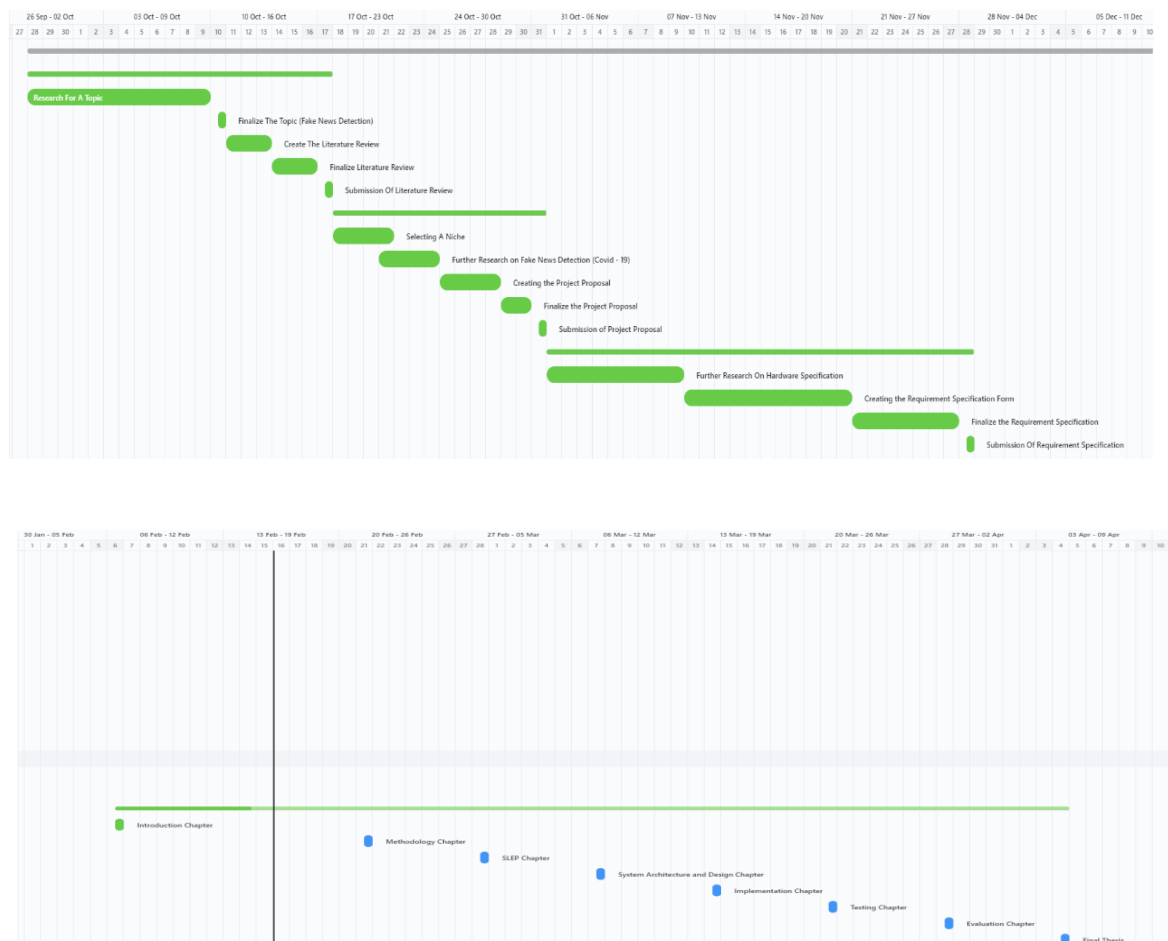
3.5.2. Deliverables

Deliverable	Date
Literature Review	17th October 2021
Project Proposal	31st October 2021

Software Requirement Specification	28th November 2021
Introduction Chapter	6th February 2022
Methodology Chapter	21st February 2022
SLEP Chapter	28th February 2022
System Architecture and Design Chapter	7th March 2022
Implementation Chapter	14th March 2022
Testing Chapter	21st March 2022
Evaluation Chapter	28th March 2022
Final Thesis	4th April 2022

3.5.3. Risk associated with project and mitigation plan

Obtaining a tagged dataset of tweets from Twitter is one of the project's primary challenges. More particular, because credibility analysis is subjective, it will be difficult to obtain labels that are accurate. Because the labels must be general enough to be used for various sorts of fake news, a wide range of tweets and users must be included in the crowdsourcing. Annotators who are unfamiliar with the content of the tweet may also submit incorrect labels. If we want to integrate the tweet's context and historical data in the assessment, it's challenging to provide credibility ratings in real time. This is due to the fact that there were additional tweets linked to the topic of the tweet.

Furthermore, the Twitter API greatly restricts the amount of study that can be done after the fact on how people react to and spread misinformation. The fact that the Twitter API does not allow for the retrieval of tweet replies older than seven days and limits the retrieval of retweets is one of the primary issues with gathering Twitter data.

3.5. Chapter Summary

This document has given a clear overview of the Methodology Chapter. This chapter describes the Research Methodologies, Development Methodologies and Project Management Methodologies.

CHAPTER 4 - SOFTWARE REQUIREMENT SPECIFICATION

1.1 Chapter Overview

This section centers around gathering framework prerequisites and examining the data gathered. In the first place, the stakeholders are recorded, and their obligations are characterized. The document includes a description of the techniques and methods that have to be used for this project. There is a brief summary of similar previous projects that is presented in tabular format. There is a context diagram and a use case diagram to describe the functionality of the projects. The document includes a list of functional requirements and non-functional requirements.

1.2 Rich Picture
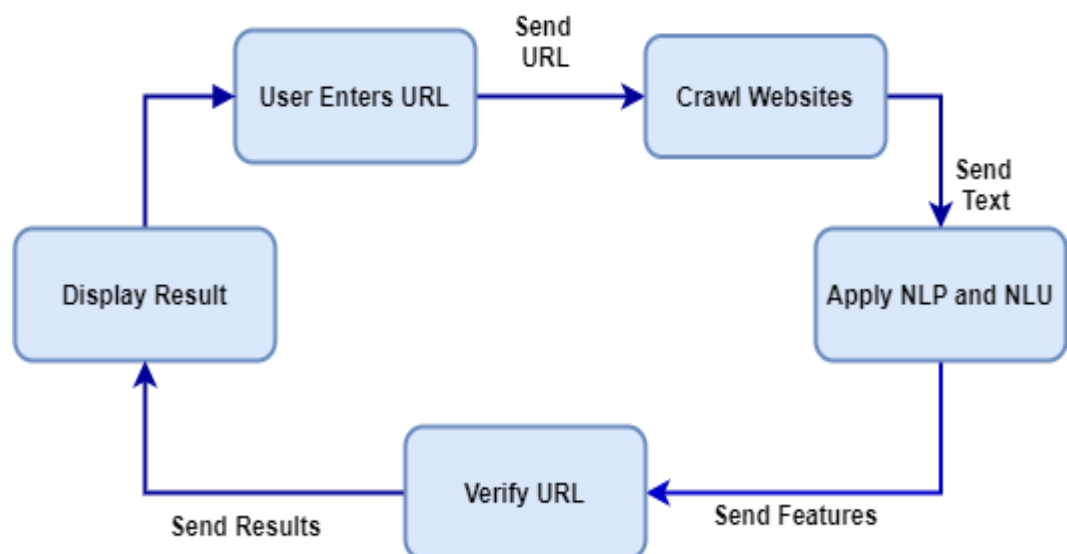
Figure 1- Rich Picture (Self Composed)

1.3 Stakeholder Analysis

When it comes to any organizational project, all of the internal people and teams who the project will involve or affect are called its stakeholders. A stakeholder analysis is a process of identifying these people before the project begins; grouping them according to their levels of participation, interest, and influence in the project; and determining how best to involve and communicate each of these stakeholder groups throughout.

1.3.1 Onion Model
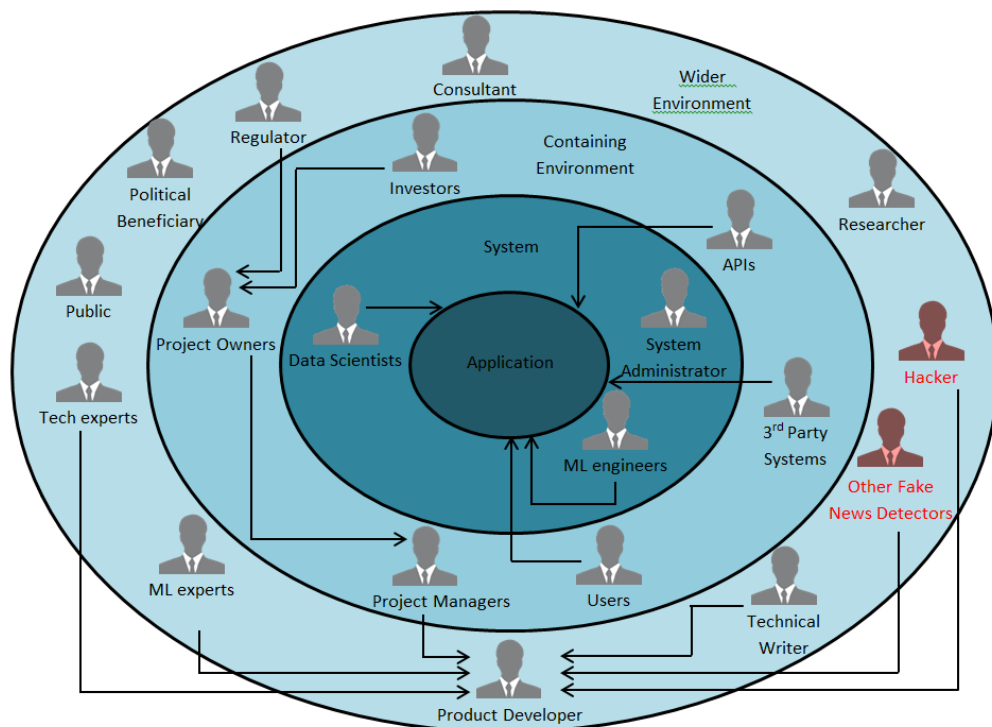
Figure 2- Onion Model (Self Composed)

1.3.2 Stakeholder Viewpoints

STAKEHOLDER	ROLE	BENEFITS

Data scientists, ML Engineers	Operational Maintenance	Design and develop models
System Admin	Operational Administration	Implement the application
Sponsor	Financial Beneficiary	Financial support
Product Owner	Functional Beneficiary	Owner of the Fake News Detection system
Product Manager	Managerial Support	Manage the process of application to ensure smooth flow
Users, 3rd Part Systems, APIs	Functional Beneficiary	Using the Fake news Detector
Regulator	Quality Regulator	Make sure the system works to the needed standards
Tech Experts	Expert	Check if platforms supports the system
ML Experts	Expert	For making a more advanced detector
Public	Functional Beneficiary	Using the Fake news Detector

Product Developer	Developer, Operational Maintenance	Creates and maintains the system
Technical Writer	Operational Support	Needed for document creation of the system
Consultant	Educational Beneficiary	Gets reviews on the current system
Political Beneficiary	Functional Beneficiary	For political campaigns
Researcher	Educational Beneficiary	To research on other systems and topics related to fake news detection
Other Fake News Detectors	Negative Stakeholder	Creates a competition for the current system
Hacker	Negative Stakeholder	May corrupt the system

1.4 Selection of Requirement Elicitation Techniques/Methods

An online survey was carried out to gather information on how effective the project would be in the real-world day to day. We weigh and discuss our methodologies, examining various options.

1.4.1. Surveys & Questionnaires

We discuss utilizing online questionnaires to gather data in order to detect fake news on Twitter concerning Covid – 19 proved fruitful as there were a variety of rewards as well as certain hindrances in making sure the data was precise and credible.

Advantages	Disadvantages
Easily accessible to anyone through the network Clearly identifiable with time stamped records of data Precise results with accurate statistical significance	Unanswered or misinterpreted questions May not feel comfortable providing answers that present themselves in an unfavorable manner Inability to further clarify various concerns and difficulties

1.4.2. Observation of existing Systems and Literature Review

Through the observation and analysis of previous existing systems it is possible to clearly outline the aspects that must be changed and made more efficient.

Advantages	Disadvantages
Helps find the most efficient way of carrying out the project. Can find variation to car you the project. Have a certain level of guidance on what aspect to look out for.	Takes time to understand and find relevant articles related to the project. May narrow the approaches if individuals focus on only a few articles.

1.4.3. Interviews

Conducting interviews are extremely time consuming and need to managed and looked after till the very end. Interviewing individuals would have not been the most efficient way to collect the necessary data for our initiative, which is why we decided to.
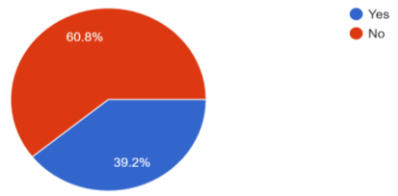
Advantages	Disadvantages
Open-ended questions so that in-depth information will be collected

Allows for natural and honest opinions | Time consuming.

Can only reach few people. |

1.4.4. Follow Requirement Gathering Method

When considering interviews and Survey Questionnaires this project will greatly benefit from survey and questionnaire approach towards gathering data on how impactful the proposed project would be in the real world. It is possible to get a broader scope of individuals with a wide variety of ideas. This method is also less time consuming and has a better community outreach.

1.5 Discussion of Results

1.5.1. Questionnaire Findings

Question - Do you use Twitter?
Aim o the Question - To identity amount of the people who use Twitter
Observation -
According to this observation 39.2%
people are using Twitter. 60.8% are do

not use Twitter.

Conclusion - Although the percentage of Twitter users is less, the number of users is significant.

Question - Do you get informed about COVID through social media?

Aim of the Question - To identify the percentage of people who gather information about COVID through social media.

Observation -

According to this observation 86.3% people get informed about COVID through social media. But 13.7% people not use social media to get informed about COVID.
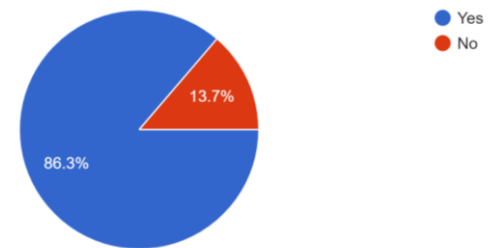
Conclusion - Many people nowadays use social media to stay up to date with the latest topics. As a result, its reliability is critical.

Question - Have you been a victim of fake news?

Aim of the Question - To identify the percentage of people who have been victims of fake news

Observation -

According to this observation 66.7% people had been a victim of fake news and 33.3% people were not affected by fake news.
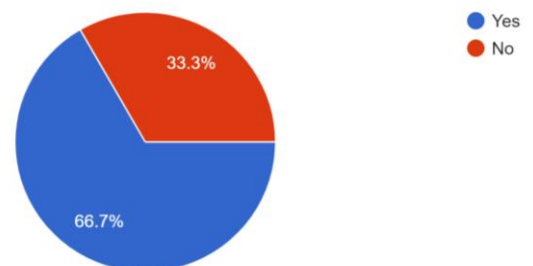
Conclusion - Fake news affects a large number of individuals, and it should be controlled to a

minimum.
Question - Do you know of any methods that fake news is being detected these days?
Aim of the Question - To identify whether the user already has a familiar system or not.
Observation - 88.2% of the participants have responded that they have not come up with an application which identifies fake news while 11.8% have said that they have come up with applications. Yes ● No ● 88.2% 11.8%
Conclusion - According to these records, it is worthwhile to create a system to detect fake news.
Question - Will a platform that collects real news be helpful to you?
Aim of the Question - To identify in advance how important it is to create such a system.
Observation - 96.1% percentage of people wants to try fake news detection system. A very less percentage of people such a system is not required. Yes ● No ● 96.1%
Conclusion - A large number of people need a high precision system to detect fake news.
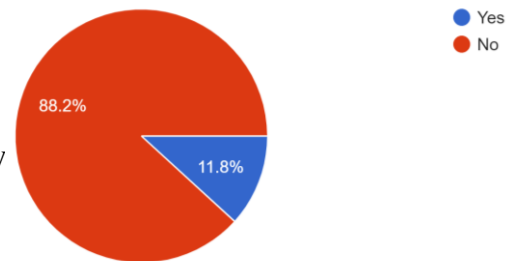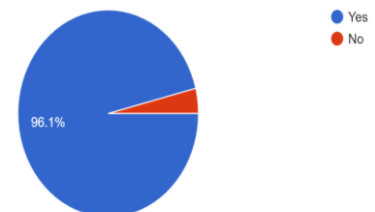
1.6 Summary of Findings

Findings	Literature Review	Questionnaire	Existing System

1.	The model achieves very high accuracy and robust behavior in several challenging settings involving large-scale real data, pointing to the great potential of geometric deep learning methods for fake news detection.	✓	✓	
2.	The system extracted different textual features from the articles using an LIWC tool and used the feature set as an input to the models.	✓		
3.	Graph theory and machine learning techniques can be employed to identify the key sources involved in the spread of fake news.	✓		
4.	Fake news detection deep learning results are presented with hyper parameter tuning to achieve improved performance.	✓	✓	
5.	Linguistic Features Based Fake News Detection and Classification approach is proposed	✓		✓
6.	SVM and Naive Bayes classifiers tend to rival each other due to the fact they are both supervised learning algorithms that are efficient at classifying data.	✓		
7.	Trained different machine learning algorithms using both textual and URL properties of data with pre-processing methods. (bag-of-words , TF-IDF)	✓		
8.	Malicious URL detection technique also considered for fake news detection.	✓		
9.	URL pattern analysis via phishing detection techniques can enhance ML algorithms' ability to	✓		

detect and mitigate the spread of fake news across the World Wide Web.			
10. The system considered two different settings of fake news detection (URL-wise and cascade-wise) Using the same architecture for both settings.	✓	✓	

1.7 Context Diagram
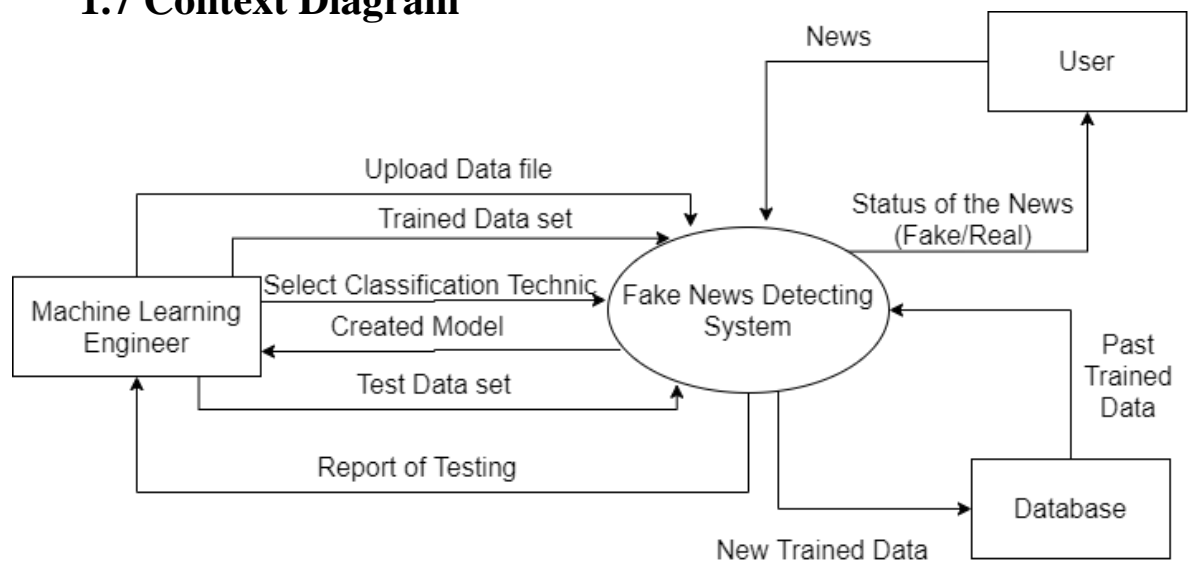
Figure 3- Context Diagram (Self Composed)

1.8 Use Case Diagram
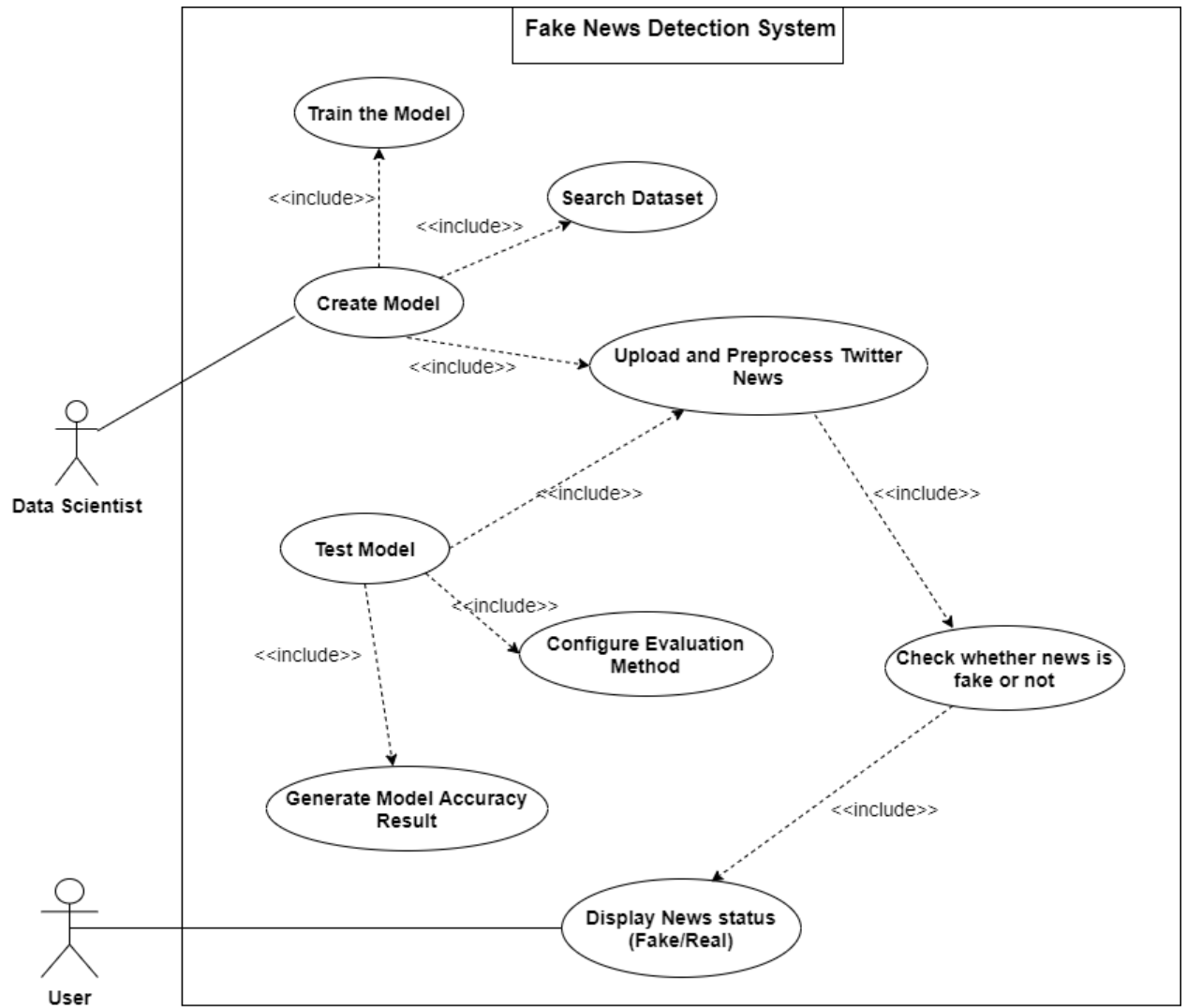
Figure 4- Use Case Diagram (Self Composed)

Requirement Specification

In this section, the system needs are established and prioritized at various levels based on the Priority.

Priority Level	Description
Critical	The most important features and functions of the system
Important	It's not required, but it's regarded to be useful
Non-important	Requirements that are outside the scope of the project

1.9 Functional Requirements

The table below lists the system's functional requirements as well as their priority level.

	Requirement and Description		Priority Level
FR01	**Detect the posts that are related to covid**		Critical
	This is the input for the model to clarify whether the news is fake or true.		
FR02	**Filter out the unnecessary parts of the post**		Critical
	This is done to ensure that only Covid -19 related text within a post will be used in the model.		
FR03	**Detect the intent of the post**		Critical
	This is done to check whether the posted details are a question or whether an individual is making a statement regarding Covid.		
FR04	**Separate depending on the variant of Covid**		Not important
	The system will detect which variant of Covid the news belongs to.		

FR05	**Use trusted sources to compare**		Important
	The system will use trusted sources like the WHO web page to cross reference the facts on COVID.		
FR06	**Present the news**		Not important
	A web interface will be used to display the fake news and the real news that was detected for consumer reference.		

1.10 Non-Functional Requirements

Performance

How quickly does the system react to user actions, or how long does a user have to wait for a given activity to take place.

Accuracy

The system's accuracy is important, and it must correctly identify between fake and real news. More incorrect information in the detection of fake news shows that the system or model is unusable.

Usability

Usability can be defined as how easy it is for a user/customer to use the system.

	Requirement and Description	Specification	Priority Level

NFR01	System should be able to handle multiple simultaneous requests.	Performance	Important
NFR02	The system should respond to a user query and return a result in less than 10 seconds.	Performance	Not important
NFR03	Time taken by Machine Learning algorithms should be in a few seconds.	Performance	Important
NFR04	The system should be user friendly and easy to use.	Usability	Not important
NFR05	The user should be able to learn the system in less than 5 minutes.	Usability	Not important
NFR06	System's accuracy level should be greater than 95%.	Accuracy	Impotant

1.11 Chapter Summary

This document has given a clear overview of the Software Requirement Specification. Very first part describes the overview of this chapter. After that Rich Picture Diagram and Stakeholder Analysis are clearly defined. Furthermore discussion of results and summary of findings are available. Then displayed the Context Diagram and Use Case Diagram of Fake News Detection System. Finally, describe Functional and Non- Functional Requirements of the Fake News Detection System under the requirement specification.

CHAPTER 5 - SOCIAL,LEGAL, ETHICAL AND PROFESSIONAL ISSUES

5.1. Chapter Overview

This chapter mentioned social, legal, ethical and professional issues related to the system of Fake News Detection on Twitter.

5.2. SLEP Issues and Mitigation

Legal	As this project is related to fake news detection on Twitter this system needs access to Twitter API.
Ethical	All users are notified and given instructions on how to protect the system's data. They are also restricted from disclosing any information.
Professional	By using a coding standard and analyzing the code the quality of the code can be ensured. Commenting, documentation and code grouping can be done to maintain a good code practice.
Social	The dataset of this system is in the csv format. The system quality is maintained by using a good coding standard. All users are informed regarding the Information Rights and Obligations.

5.3. Chapter Summary

The chapter displayed clear SLEP analysis and its mitigation form related to the project.

CHAPTER 6 - SYSTEM ARCHITECTURE & DESIGN

6.1. Chapter Overview

This chapter will be focusing on the Architecture and Design of the Fake News Detection System on Twitter. First part describes the goals of designing this system and architecture of the system. Then the system design part describes Choice of design paradigm, Component diagram, Class diagram, Sequence diagram, UI Design, User experience and Process Flow Chart.

6.2. Design Goals

Design Goals come up with the Non Functional requirements of the system.

Design Goal	Description
Performance	Performance means how quickly does the system react to user actions, or

	how long does a user have to wait for a given activity to take place. If a user is going to check the status of news, the Fake News Detection system should show the results in a few seconds. That is the goal of the system.
Accuracy	The system's accuracy is important, and it must correctly identify between fake and real news. More incorrect information in the detection of fake news shows that the system or model is unusable. In the Fake News Detection system accuracy level should be more than 95% percent.
Usability	Usability can be defined as how easy it is for a user to use the system.

6.3. System Architecture Design
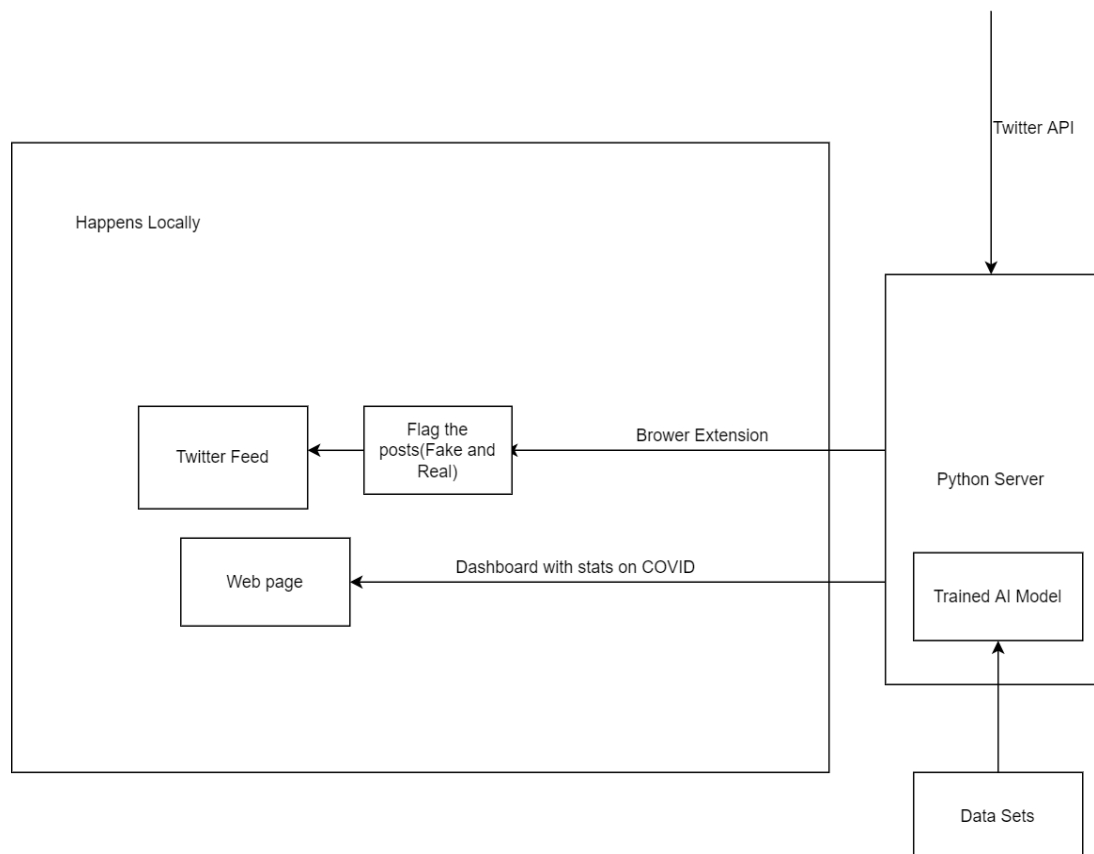
6.4. System Design

6.4.1. Choice of Design Paradigm

Oriented Analysis and Design (OOAD) was chosen to design the fake news detection system in Twitter on Covid-19, taking into consideration both OOAD and SSADM features. Class diagrams, activity diagrams, and sequence diagrams all benefit from it. System architecture design is also less complex than Structured Systems Analysis and Design (SSAD).

Development methodology is an important aspect of any project. It helps organize the project and make the work flow more efficient. This project will be using an agile development methodology. Agile methodology can help us minimize risks because we can go back and fix defects that we hadn't fixed. This will make the project more flexible and it would help us reduce the defects in the final product. Since this methodology does not have a certain fixed set of rules we can mold it into making it so that the group will be able to work efficiently.

6.4.2. Component Diagram

6.4.3. Class Diagram

Dataset

+ attribute1:type = String
+ attribute2:type
- attribute3:type

+ preprocess()
+ trasform()
+ Model()
+ Validation()
+ update()

AI Tool

+ Algorithms: String
+ attribute2:type

+ InsertNews()
+ Preprocess()
+ predict
+ viewResults()
+ saveResults()

supply data ▶

Dashboard

+ input: String

+ getDetails(): object
+ display()

Supply
data

DataBase

+ dbConnection: String

+ getDetails(): object
+ saveDetails(): object
+ updateDetails(): object

Train ▶

use
▼

Model

+ Algorithms: String

+ Train()
+ Test()

supply data
▼

Supply
data

Twitter API

+ newsPost: String
+ DBConnection:String

+ getNews()
+ format()

Extensions

+ input: String

+ getDetails(): object
+ display()

6.4.4. Sequence Diagram

6.4.5. UI Design
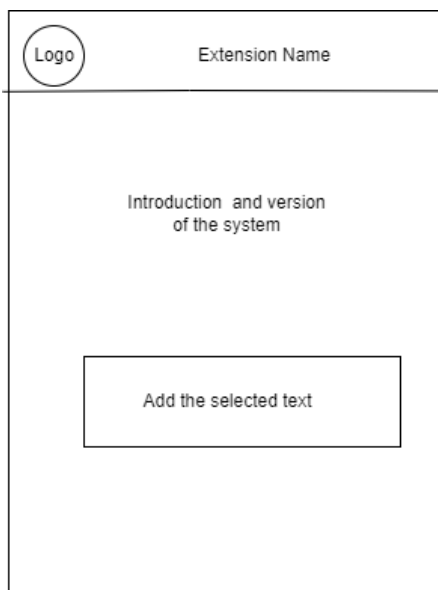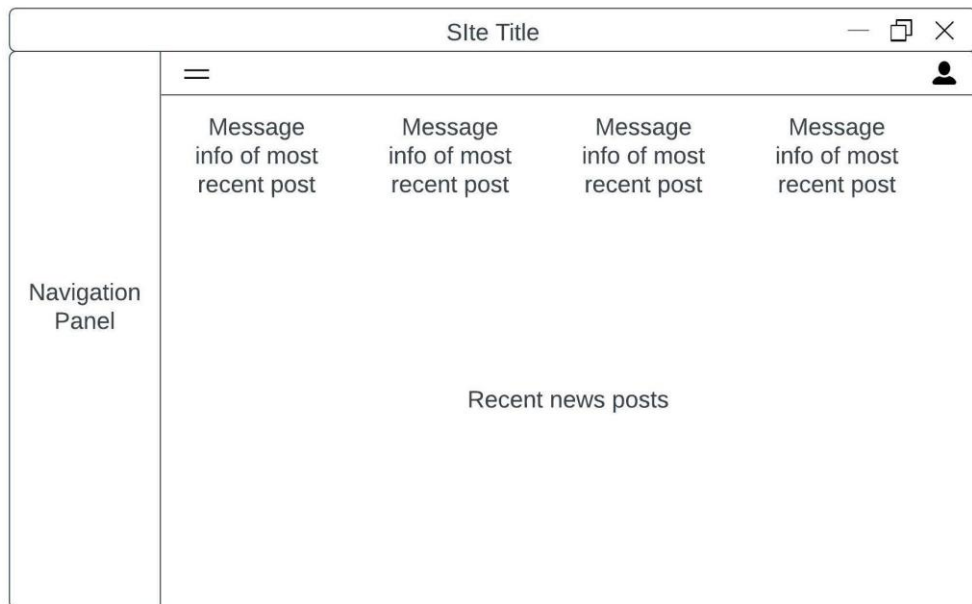
6.4.6. User Experience

The project will help twitter users to be able to know whether the tweet they are seeing about COVID is fake or true and this will help them avoid various issues that the news may cause them.It will use an extension to help communicate with the twitter feed.

6.4.7. Process Flow Chart

```
┌─────────────────────┐
│  Language used;     │
│  PYTHON             │
└─────────────────────┘
          │
          ▼
┌─────────────────────┐
│     OpenFile        │
└─────────────────────┘
          │
          ▼
┌─────────────────────┐
│  Using Trained      │
│  Model              │
└─────────────────────┘
          │
          ▼
┌─────────────────────┐
│  Method of          │
│  implementation     │
└─────────────────────┘
          │
          ▼
┌─────────────────────┐
│  File recieved      │
└─────────────────────┘
          │
          ▼
┌─────────────────────┐
│  File               │
│  preprocessing      │
└─────────────────────┘
          │
          ▼
┌─────────────────────┐
│  File analysis      │
└─────────────────────┘
          │
          ▼
┌─────────────────────┐
│  News Post          │
│  classification     │
└─────────────────────┘
          │
          ▼
┌─────────────────────┐
│  Status of news     │
│  post (Output)      │
└─────────────────────┘
```
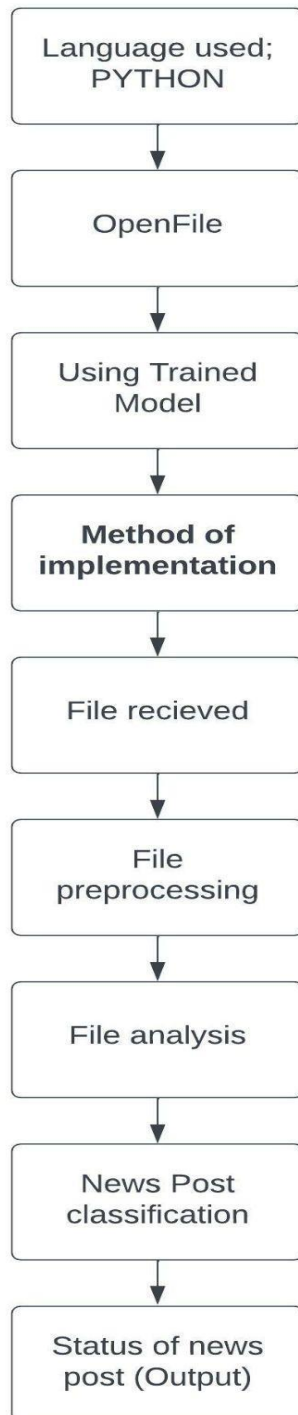
6.5. Chapter Summary

This chapter has given a clear overview of the Architecture and Design of the Fake News Detection System. It describes designs of the system's front end and back end using Class diagram, Sequence diagram, Ui Design and Process Flow diagram.

CHAPTER 7 - IMPLEMENTATION

7.1. Chapter Overview

This chapter covers the technologies used for developing the Fake News Detection system. Then describes programming Languages, libraries, IDEs and Tools used in this project. And also the used data set is mentioned in this chapter. Finally pseudo codes for core functionalities are mentioned clearly.

7.2. Technology Selection

7.2.1. Technology Stack

Below technology stack was used for the system.

Figure 7.1 : Technology Stack

7.2.2. Data Selection

When starting a data science project we need a high accuracy data set. We got Fake News Detection on Twitter data sets from Github. But some data sets are unable to be used, because this project mainly focuses on fake news about Covid-19. Finally the project continues with the below Data Set.

https://github.com/malomodaniels/COVID-19-Fake-Tweets-Classification

7.2.3. Selection of Development Framework

7.2.3.1. Backend

For fake news detection systems, Python is used as a web framework. Python libraries must be established so that developers may devote more time to the system's development. It will also be very simple to refactor framework code. Using the Python

language and Python files, the core section of the system may be easily executed in python. When designing apps, the Python web framework always follows the best coding principles.

7.2.3.2. Frontend

For frontend designing, HTML, CSS and Javascript are used.

7.2.4. Programming Language

Programming language that was used to implement this system is python. Python language is easy to use and maintain.

Python is an interpreted language. An interpreter executes the statements of code one-by-one and it helps to identify eros easily.

Python has so many libraries. Then developers don't need to create basic functions from scratch.

7.2.5. Libraries

This project used flask, Express, Node.Js, Nodemon, Mongoose and Ejsused for designing the web application. Ejs is a templating language used to generate html markup with plain javascript. Nodemon is more like a tool to further develop applications that use node.js.

7.2.6. IDE

The IDE used in this project is VS code. Visual Code Studio is a source code editor. It is a product of Microsoft which can be accessed easily in Windows, MacOS and Linux. VS code supports the development operations like version controlling, debugging and task running. As VS code supports many programming languages it is easy and efficient to use VS code as the IDE.

7.2.7. Summary of Technology Selection

Component	Tool/Technology	Version
Programming Language	Python	3.10.0

Deep learning libraries	Flask	2.0.3
	Tweepy	4.6.0
UI Framework	Express	4.17.3
	Mongoose	6.2.6
IDE	Vs code	1.65

Table 7.1 : Summary of Technology Selection

7.3. Implementation of Core Functionalities

Twitter API - Python code

```python
import tweepy
import configparser
import pandas as pd

#read configs
config = configparser.ConfigParser()
config.read('config.ini')

api_key= config['twitter']['api_key']
api_key_secret= config['twitter']['api_key_secret']

access_token = config['twitter']['access_token']
access_token_secret=config['twitter']['access_token_secret']

auth = tweepy.OAuthHandler(api_key,api_key_secret)
auth.set_access_token(access_token,access_token_secret)

api = tweepy.API(auth)

public_tweets = api.home_timeline()

print(public_tweets[0].text)

columns = ['User', 'Tweets']
data=[]
```

```
for tweet in public_tweets:
    data.append([tweet.user.screen_name, tweet.text])

df = pd.DataFrame(data,columns=columns)

print(df)

df.to_csv('tweets.csv')
```

Extension - HTML codes

```html
<!doctype html>
<html>
  <head>
    <title>Fake News Detector</title>
    <meta name="viewport" content="width=device-width, initial-scale=1.0">
    <meta charset="UTF-8">
    <link rel="stylesheet" href="CSS/Style1.css">
    <link rel="stylesheet" href="CSS/naviStyle.css">
    <body>
      <!DOCTYPE html>
<html lang="en">

    <!-- navigation bar -->
    <header class="navbarContainer" ><br>
      <img id="webname" src="Images/logo2.jpg" alt="Site logo">
      <h1 id="topic">FAKE-ND</h1>
    </header>
    <!-- <svg class="trapezium_1" height="100" width="300%"> -->
      <!-- <path d="M 1535 0 L 900 0 L 200 450 L 1535 450 L1535 0" fill="darkblue" opacity="0.9"></path> -->
      <!-- <path d="M 1535 0 L 450 0 L 700 450 L 1535 450 L1535 0" fill="gray" opacity="0.9"></path> -->
    <!-- </svg> -->
    <svg class="trapezium_1" height="900" width="100%">
      <path d="M 150 0 L 0 0 L 200 45000 L 1535 450 L1535 0" fill="darkblue"  opacity="0.8"></path>
      <path d="M 150 0 L 40 0 L 700 450 L 1535 450 L1535 0" fill="gray" opacity="0.7"></path>
    </svg>

    <div class="team_description">

      <div class="team_intro">
        <p><span>F</span>AKE-ND helps you to Detect Fake News on Twitter <br><br><br> Version 1.0<br><br><br><br><br><br><br>
        <div class="card-panel" id="info-card" style="visibility: visible; padding-top: 5px">
        <br>Select some text and then click ∞  to get started! <br><br></p></div>
      </div>
    </div>
```

```html
<!DOCTYPE html>
<html lang="en">
<head>
    <title>Fake News Detector</title>
    <meta charset="UTF-8">
    <link rel="stylesheet" href="CSS/Style1.css">
    <link rel="stylesheet" href="CSS/naviStyle.css">
</head>
<body>
    <!-- navigation bar -->
    <header class="navbarContainer" >
      <img id="webname" src="Images/logo2.jpg" alt="Site logo">
    </header>
    <svg class="trapezium_1" height="600" width="100%">
      <path d="M 1535 0 L 900 0 L 200 450 L 1535 450 L1535 0" fill="darkblue" opacity="0.3"></path>
      <path d="M 1535 0 L 450 0 L 700 450 L 1535 450 L1535 0" fill="lightblue" opacity="0.3"></path>
    </svg>

    <div class="team_description">
      <div>
        <h1>FAKE-ND</h1>
      </div>
      <div class="team_intro">
        <p><span>F</span>AKE-ND helps you to Detect Fake News on Twitter <br><br><br> Version 1.0<br><br>Select some text and then click
        <button type="button">∞</button> to get started!  </p>
      </div>
    </div>

</body>
<script src="popup.js"></script>
</html>
```
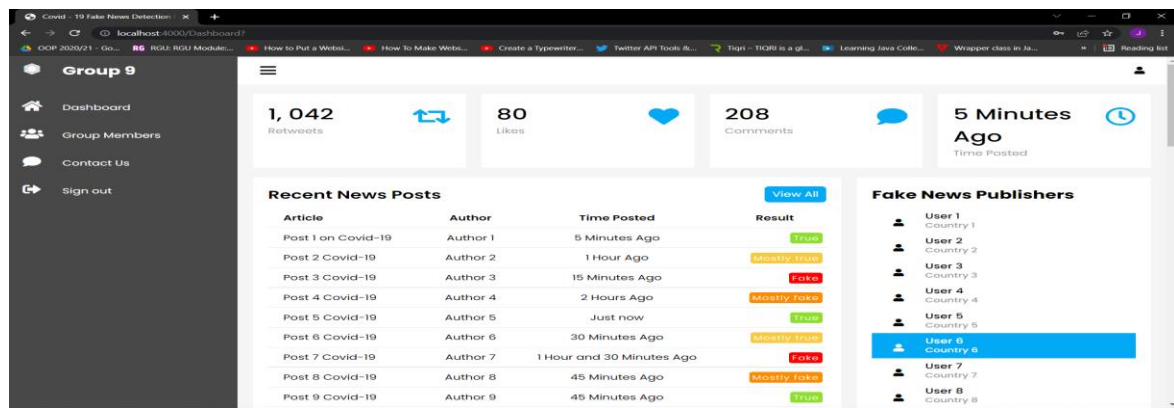
Extension - UI

Web Site

Dashboard — Covid-19 Fake News Detection
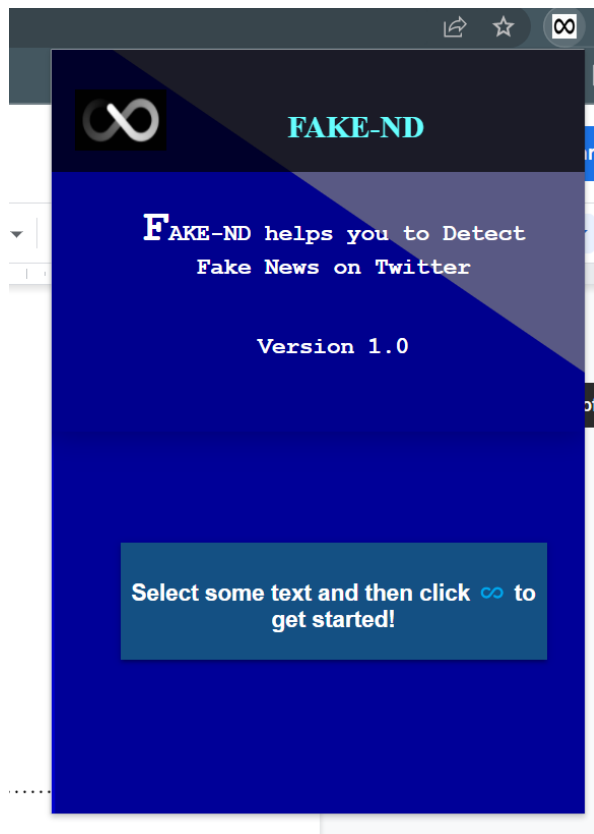
1, 042 Retweets	**80** Likes	**208** Comments	**5 Minutes Ago** Time Posted

Recent News Posts — View All

Article	Author	Time Posted	Result
Post 1 on Covid-19	Author 1	5 Minutes Ago	True
Post 2 Covid-19	Author 2	1 Hour Ago	Mostly true
Post 3 Covid-19	Author 3	15 Minutes Ago	Fake
Post 4 Covid-19	Author 4	2 Hours Ago	Mostly fake
Post 5 Covid-19	Author 5	Just now	True
Post 6 Covid-19	Author 6	30 Minutes Ago	Mostly true
Post 7 Covid-19	Author 7	1 Hour and 30 Minutes Ago	Fake
Post 8 Covid-19	Author 8	45 Minutes Ago	Mostly fake
Post 9 Covid-19	Author 9	45 Minutes Ago	True
Post 10 Covid-19	Author 10	15 Minutes Ago	Fake

Fake News Publishers

- User 1 — Country 1
- User 2 — Country 2
- User 3 — Country 3
- User 4 — Country 4
- User 5 — Country 5
- User 6 — Country 6
- User 7 — Country 7
- User 8 — Country 8

OUR TEAM

MR. JENUK GILES
SMALL DESCRIPTION

CONTACT US
Description

Address
1234 Road 1,
Town,
City,
12345

Phone
+94 77 123 4567

Email
someone@email.com

Send Message
Please Enter Your Name
Please Enter Your Email
What would you like to say to us?
Send

Create Account

Username
Email Address
Password
Confirm Password

Sign up

Already have an account? Sign in
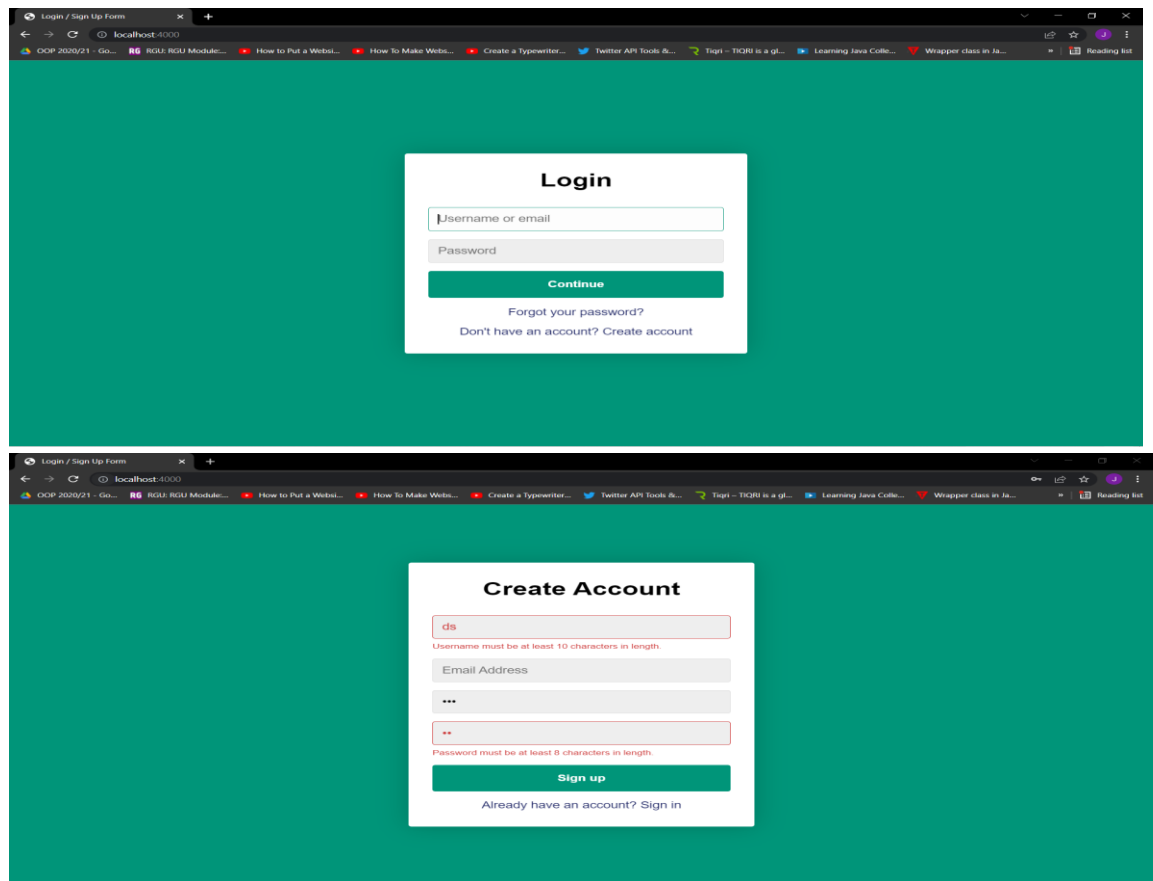
7.4. Chapter Summary

This chapter has started with the technology selection. Technology stack, dataset selection, frontend and backend development, programming languages, libraries and IDE used are clearly mentioned in this chapter. Finally the Implementation of Core Functionalities are described through pseudo codes.

CHAPTER 8 - TESTING

8.1 Chapter Overview

In this chapter cover the testing methods of the project. First part explains objectives and goals of Testing. Then clearly mention functional and noin-functional testing methods. The last part of the chapter describes limitations of testing.

8.2 Objectives and Goals of Testing

The object of the test is to maintain the quality of the system. And ensure the requirements are completed or not.

Goals of the testing are mentioned below.

1. Check the functional and non-functional requirements are satisfied by the Fake News Detection system.
2. Check code quality of implemented code.
3. Find bugs not found in development.

8.3 Testing Criteria

As previously said, active and non-active tests are mostly used in the testing process but assessing the quality of the run code is also an important component of the testing section.

Software quality of the system can be divided into following categories ,

● **Functional quality**

Functional and technical requirement quality are mainly focused on the final code implementation.

● **Structural quality**

Performance of functional and non-functional requirement are mainly focus of the system.

● **Process quality**

Business process of the system is discussed in this section.

8.4 Model Evaluation

8.4.1 Confusion Matrix (Accuracy/F1Score/Precision/Recall)

8.4.1.1.Accuracy

```python
accuracy = np.trace(cm) / float(np.sum(cm))
misclass = 1 - accuracy
```

8.4.1.2.F1Score/Precision/Recall

```
SVM
test:
['real' 'fake' 'fake' ... 'real' 'real' 'real']
Our daily update is published. States reported 734k tests 39k new cases and 532
deaths. Current hospitalizations fell below 30k for the first time since June 2
2. https://t.co/wzSYMe0Sht
['fake']
[[ 966   90]
 [  46 1125]]
              precision    recall  f1-score   support

        fake       0.95      0.91      0.93      1056
        real       0.93      0.96      0.94      1171

    accuracy                           0.94      2227
   macro avg       0.94      0.94      0.94      2227
weighted avg       0.94      0.94      0.94      2227

Accuracy :  0.9389312977099237
Precison :  0.939839049289118
Recall :  0.9389312977099237
F1 :  0.9390178836782864
```

Confusion Matrix of SVM on Test Dataset
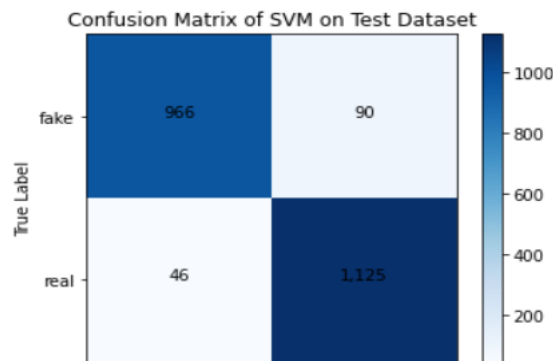
8.5 Benchmarking

Benchmarking is comparing current system with existing systems. Competitive benchmarking is the one main category of benchmarking. There are so many systems available for detecting fake news. But this is the system to detect fake news on Twitter about Covid.

Citation	Limitation	Technology/Algorithm	Advantages
Sing, Dasgupta, Sonagra, Raman, & Ghosh	This model does not consider social media account or specific personal posts to compare. It focuses mainly on a broader	Logistic Regression Support Vector Machine Random Forest Decision Tree K-Neighbor Classifier.	Out of the algorithms used , Support Vector Machine method gave the best prediction results.

	picture when considering whether the article is real news or fake news.		
Mazzeo, Rapisarda, & Giuffrida	URL pattern analysis via phishing detection techniques can help ML algorithms' ability to detect the spread of fake news beyond the World Wide Web.	Machine learning workflow for training algorithms and classification of results from web search engines.	News has some specific textual properties that can be used to classify it as fake or real.
Ahmad, Yousaf, Yousaf, & Ahmad, 2020	Automated classification of a text article as misinformation or disinformation is a challenging task. Even an expert in a particular domain has to explore multiple aspects before giving a verdict on the truthfulness of an article.	Logistics Regression Support Vector Machine Multilayer Perceptron K-Nearest Neighbors Random Forest	The advent of the World Wide Web and the rapid adoption of social media platforms (such as Facebook and Twitter) paved the way for information dissemination that has never been witnessed in the human history before.

8.6 Functional Testing

Functional testing is where you test the functions of a software or application to see if it caters to the requirements of the user. Functional testing is also called black box testing, and it always happens before white box testing.

As our project is based on fake news detection, we tested the data against the accuracy of the prediction (if the news is properly classified as true or fake), we tested the functionality of the web page, and we also tested the extension. For the web page we used Selenium IDE.
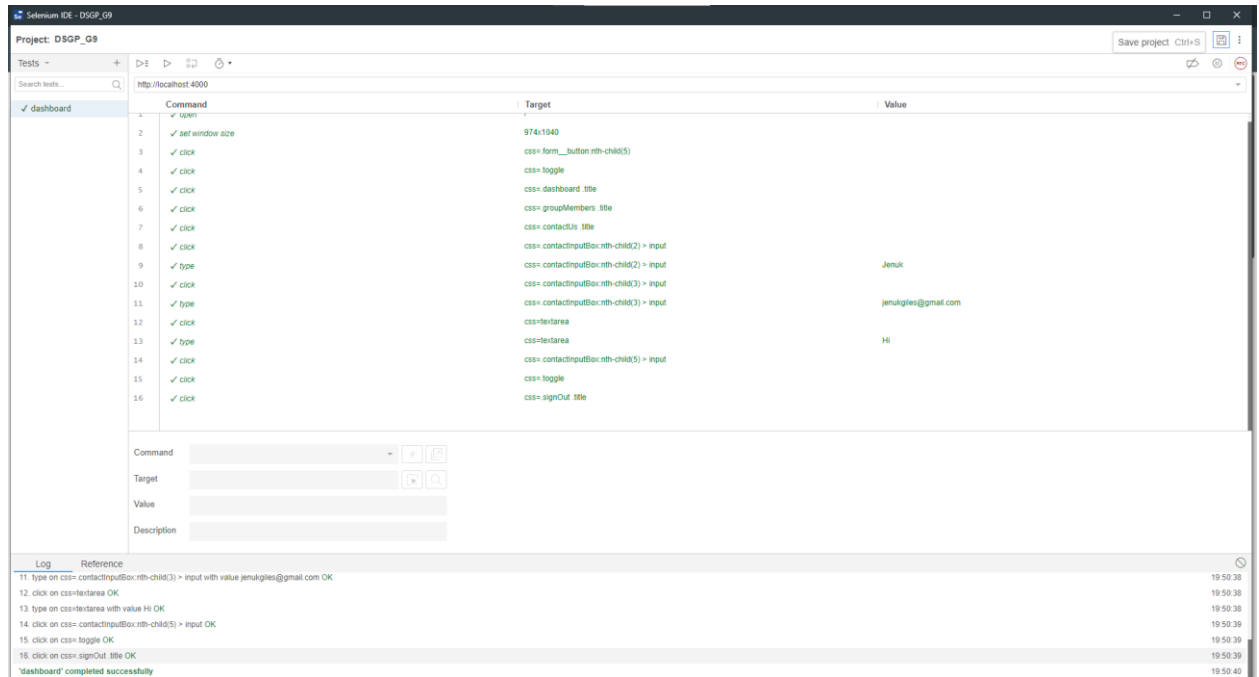
Test Case #	Description	Input Data/User action	Expected Outcome	Actual Outcome	Status
1	System categorizes the news feed	Activate AI tool.	The system should get the nes feed, categorize each individual news post and save it in a file	The system should get the nes feed, categorize each individual news post and save it in a file	PASS
2	System shows the result through a web extension	-	The system should display the result through the web extension.	The system displays the result through the web extension.	PASS
3	The website shows a summary of past news posts	Login to the web page	The system should validate the username and password(s) of the user and show a dashboard in the main page	The system validates the username and password(s) of the user and show a dashboard in the main page	PASS

Functional test pass rate $= \frac{Number\ of\ tests\ passed}{Number\ of\ tests\ performed} \times 100\% = \frac{3}{3} \times 100\% = 100\%$

8.7 Module and Integration Testing

Integration testing is the type of testing where software modules are integrated logically and tested as a group. Usually a software project consists of multiple software modules, each coded by different programmers.

Test Case #	Description	Input Data/User action	Expected Outcome	Actual Outcome	Status
1	System scrapes data from twitter feeds.	Activate AI tool.	The twitter API should get the data/news feed from the twitter page.	The twitter API got the data/news feed from the twitter page.	PASS
2	System verifies the news feed.	-	The AI tool should be able to identify fake news and true news.	The AI identified fake news and true news.	PASS
3	Saves the data.	-	The AI tool should be able to save the identified news.	The AI tool saves the identified news in a csv file.	PASS
4	System shows the saved data.	-	The web extension should show the fake news.	The web extension shows fake news.	PASS
5	User can enter username in the login page of the dashboard.	Input user's username.	System should be able to verify the username of the user correctly.	System can verify the username of the user correctly.	PASS
6	User can enter password in the login page of the system.	Input user's password.	System should be able to verify the password of the user correctly.	System can verify the password of the user correctly.	PASS
7	Acknowledgement of invalid input.	-	System should be able to identify invalid inputs for username and password.	System shows when there are invalid inputs.	PASS
8 - 11	New users can register using the login page, can enter a new	Input user's Username and	System should be able to save the relevant details of the new user	System saves the relevant details of the new user correctly and the	PASS

	password and the system acknowledgement of invalid inputs.	password.	correctly and the system should be able to identify invalid inputs from the user	system shows when there are invalid inputs.	
12	System has a login button in the login page that the user can click after putting the correct username and password.	Click Login button in the login page	System should navigate to the home page after clicking login button	System navigates to the home page after clicking login button	PASS
13	System displays a summary as a dashboard in the main page	-	System should display a summary as a dashboard.	System displays a summary as a dashboard.	PASS
14	Users can scroll to view other pages in the web page	Scroll up and down in the page	System should be able to navigate between pages.	System navigates between pages..	PASS
15	Users can contact the developers through the "Our Team" page and the "Contact us" page	View the developers twitter accounts and other social media accounts.	System should be able to navigate to the relevant profile pages.	System navigates to the relevant profile pages.	PASS
16	Users can logout	Click the "Sign out" button	System should return to the login page	System returns to the login page	PASS

Functional test pass rate $= \dfrac{Number\ of\ tests\ passed}{Number\ of\ tests\ performed} \times 100\% = \dfrac{16}{16} \times 100\% = 100\%$

8.8 Non-Functional Testing

8.8.1 Accuracy Testing

Accuracy is the measurement used to determine which model is best at identifying relationships and patterns between variables in a dataset based on the input, or training, data.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN}$$

TP = True Positives
TN = True Negatives
FP = False Positives
FN = False Negatives

8.8.2 Performance Testing

The accuracy and performance of the activated system are critical. The graph below depicts the active system's performance during model training. The model's performance is great, because the training loss is minimal.

8.9 Limitations

Several test procedure constraints have been identified, and the limits are listed separately below.

● The system's accuracy can only be calculated using the code; no other methods can be used to calculate the accuracy.

8.10 Chapter Summary

This chapter covers the testing of the proposed system as well as the testing technique. The test objectives are stated at the outset, emphasizing the significance of evaluating the system. Following that, the functional and non-active requirements are listed in order of implementation. Finally, the chapter discusses the testing procedure's limitations.

CHAPTER 9 - EVALUATION

9.1 Chapter Overview

This chapter discusses how domain experts and target audiences rated the system's importance, development, and usability, among other things. Authors' self-assessment is also mentioned in their evaluation.

9.2 Evaluation Methodology and Approach

Deep learning is used in the detection of fake news.There are just a few approaches for assessing this system. One is for a developer to contact that system's users in person or online and obtain information in the categories listed below.

- Model performance
- Usability of the system
- User friendly of the system

9.3 Evaluation Criteria

As a quantitative technique, evaluation criteria are used to evaluate implemented systems. The following assessment criteria will be used to identify and critically evaluate evaluators.

Criteria	Purpose to evaluate
The overall concept	Obtain the understanding and gain target audience perceptions of the system domain. And check whether they have understood the concept.
Project Scope	
System design, architecture and implementation	Verify that system and architectural design standards are conformed.
Solution and Prototype	The activated prototype acts as evidence of the concept for

	the proposed solution to be evaluated.
Application difficulties	Fulfill the non-functional requirements in the system
User interface and user application experience	Fulfill the non-functional and functional requirements in the system

9.4 Self-Evaluation

Criteria	Author's Evaluation
Overall Project Concept	Implemented a system for detecting fake news about Covid on twitter. Train the python model and it detects fake news and selects the fake parts of news using chrome extension.
Project's Scope	The scope specified at the start of the project is well-suited to the project's timeline. In terms of the scope of the project, detect text formatted news in Twitter. This can be expands in the future to image data also.
System Architecture, Design and Implementation	The design and architecture were both excellently executed. The system is designed with a high-level architecture for future enhancements. The implementation was perfectly done.
Final Prototype and the Solution	The prototype has been developed with great architecture. Anyone can easily access the system. The prototype gives a solution to the problem that has been defined.
User Interface and Accessibility	The web page is designed in an attractive way for the audience. It can be accessible easily without any problems.
Limitation of the System	Only text formatted news can be detected as fake or real.

9.5 Evaluation on Functional Requirements

	Requirement and Description	Evaluation	Priority Level
FR01	**Detect the posts that are related to covid** This is the input for the model to clarify whether the news is fake or true.	**Implemented**	Critical
FR02	**Filter out the unnecessary parts of the post** This is done to ensure that only Covid -19 related text within a post will be used in the model.	**Implemented**	Critical
FR03	**Detect the intent of the post** This is done to check whether the posted details are a question or whether an individual is making a statement regarding Covid.	**Implemented**	Critical
FR04	**Separate depending on the variant of Covid** The system will detect which variant of Covid the news belongs to.	**Implemented**	Not important
FR05	**Use trusted sources to compare** The system will use trusted sources like the WHO web page to cross reference the facts on COVID.	**Implemented**	Important
FR06	**Present the news**	**Implemented**	Not important

9.6 Evaluation on Non-Functional Requirements

	Requirement and Description	Specification	Evaluation	Priority Level
NFR01	System should be able to handle multiple simultaneous requests.	Performance	**Implemented**	Important
NFR02	The system should respond to a user query and return a result in less than 10 seconds.	Performance	**Implemented**	Not important
NFR03	Time taken by Machine Learning algorithms should be in a few seconds.	Performance	**Implemented**	Important
NFR04	The system should be user friendly and easy to use.	Usability	**Implemented**	Not important
NFR05	The user should be able to learn the system in less than 5 minutes.	Usability	**Implemented**	Not important
NFR06	System's accuracy level should be greater than 95%.	Accuracy	**Implemented**	Impotant

9.7 Chapter Summary

This chapter explained different ways of evaluation with different site's experts. Firstly, briefly explain about a variety of evaluation methods and its approaches. Then evaluation criteria were documented. Next author's self-evaluation is also mentioned. Finally, Evaluation results were evaluated according to concept, scope, design, implementation, prototypes and limitations.

CHAPTER 10 - CONCLUSION

10.1 Chapter Overview

This chapter concludes with all the details related to the project by analysing the project completion status. Initially, the objectives of the project will be discussed, and the skills and knowledge gained from the various course modules for completing the project will be discussed. The knowledge gained after the completion of the project will discuss the challenges faced in the project and the limitations of the project. The final section discusses future improvements to the chapter, and the project documentation concludes with a final conclusion about the project.

10.2 Achievements of Research Aims & Objectives

Description	Status
Literature Review	
Critical evaluation of existing systems.	**Completed**
Requirement gathering	
Critical evaluation about requirement collections.	**Completed**
Design	
Designed Fake News Detection System.	**Completed**
Development	
Develop proper protype with efficient solution using relevant tools to meet requirements and design.	**Completed**
Testing	
Model testing and whole project testing.	**Not-completed**

10.3 Utilization of Knowledge from the Course

Module	Description
Machine Learning	Python model of the system implemented using machine learning.
Web Technology	Hypertext markup language (HTML), Cascading Style Sheets (CSS) and JavaScripts are used to create Web Site and Chrome Extension.
Advanced Mathematics for Data Science	Advanced mathematics formulas are used to implement the system.

10.4 Use of Existing Skills

Throughout the Artificial Intelligence and Data Science degree program, several skills are helpful to develop the system. Python, HTML, CSS and machine learning skills are highly required to implement this system.

10.5 Achievement of Learning Outcomes

- Identify and compare contemporary software development methodologies, project management methodologies, quality assurance standards and methodologies related to innovation and marketing for producing Data Science software.
- Apply a suitable software development methodology to meet a designated set of requirements using appropriate languages and tools.
- Develop skills necessary to collaborate within a team on a software development project.
- Present an analysis of the user centered design process, cognitive aspects, research methods, modeling and prototyping used to produce applications related to Data Science with a reflection on legal, ethical, professional and social issues.

10.6 Problems and Challenges Faced

In the implementation of any research project the researcher faces many problems and challenges and in this research project also few challenges have to be faced and details regarding that are described in the table below.

Challenge	Solution/ Description
Lack of knowledge on the Twitter	The knowledge on twitter was very low, and for the implementation of the system knowledge on that domain was necessary.
Time management	Due to continuous change in the functional requirements in the project time wasted without any use. But with the use of project management methodologies overcame the challenge.
Data loss	All the data related to the implemented system and the documentation of the system is vulnerable, so all the data is stored in platforms like GitHub and Google Drive.

10.7 Individual Contribution

Name	Contribution
Dilan Perera	Model and Twitter API.
Chathuni Manage	Chrome Extension
Chamathi Gunawardena	Rest API
Jenuk Giles	Database and web site

10.8 Chapter Summary

This chapter provided a quick overview of the research goals and objectives, as well as the usage of new and current abilities, as well as the difficulties encountered throughout the project. Also discussed is the impact of previously learned modules on the deployment of this system. Finally, the contribution of each individual is mentioned.