A

Mini Project

On

SOCIAL MEDIA AND MISLEADING INFORMATION IN A DEMOCRACY A MECHANISM DESIGN APPROACH

(Submitted in partial fulfillment of the requirements for the award of Degree)

BACHELOR OF TECHNOLOGY

In

COMPUTER SCIENCE AND ENGINEERING

By

C. HEMA BINDU (207R1A05D4)

KAREMI. TEJA (207R1A05E7)

CH. NAVATEJ (207R1A05D2)

Under the Guidance of

K. SHILPA

(Assistant Professor)



DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING CMRTECHNICALCAMPUS

UGC AUTONOMOUS

(Accredited by NAAC, NBA, Permanently Affiliated to JNTUH, Approved by AICTE, New Delhi) Recognized Under Section 2(f) & 12(B) of the UGCAct.1956, Kandlakoya (V), Medchal Road, Hyderabad-501401.

2020-2024

DEPARTMENT OF COMPUTER SCIENCEAND ENGINEERING



CERTIFICATE

This is to certify that the project entitled "SOCIAL MEDIA ANDMISLEADING INFORMATION IN A DEMOCRACY A MECHANISM DESIGN APPROACH" being submitted by C. HEMA BINDU (207R1A05D4), KAREMI. TEJA (207R1A05E7), CH. NAVATEJ(2071RA05D2) in partial fulfillment of the requirements for the award of the degree of B.Tech in Computer Science and Engineering to the Jawaharlal Nehru Technological University Hyderabad, is a record of bonafide work carried out by them under our guidance and supervision during the year 2023-24.

The results embodied in this thesis have not been submitted to any otherUniversity or Institute for the award of any degree or diploma.

K. SHILPA	Dr.A.Raji Reddy	
(Assistant Professor)	DIRECTOR	
INTERNAL GUIDE		
Dr. K. Srujan Raju	EXTERNAL EXAMINER	
HOD		
Submitted for viva voice Examination held on		

ACKNOWLEDGEMENT

Apart from the efforts of us, the success of any project depends largely on the encouragement and guidelines of many others. We take this opportunity to express our gratitude to the people who have been instrumental in the successful completion of this project.

We take this opportunity to express our profound gratitude and deep regard to our guide **K. SHILPA**, Assistant Professor for her exemplary guidance, monitoring and constant encouragement throughout the project work. The blessing, help and guidance given by her shall carry us a long way in the journey of life on which we are about to embark.

We also take this opportunity to express a deep sense of gratitude to the Project Review Committee (PRC) Mr. G. Vinesh Shanker, Dr. J. Narasimha Rao, Ms. K. Shilpa, & Dr. K. Maheswari for their cordial support, valuable information and guidance, which helped us in completing this task through various stages.

We are also thankful to **Dr. K. Srujan Raju**, Head, Department of Computer Science and Engineering for providing encouragement and support for completing this project successfully.

We are obliged to **Dr. A. Raji Reddy**, Director for being cooperative throughout the course of this project. We also express our sincere gratitude to Sri. **Ch. Gopal Reddy**, Chairman for providing excellent infrastructure and a nice atmosphere throughout the course of this project.

The guidance and support received from all the members of **CMR Technical Campus** who contributed to the completion of the project. We are grateful for their constant support and help.

Finally, we would like to take this opportunity to thank our family for their constant encouragement, without which this assignment would not be completed. We sincerely acknowledge and thank all those who gave support directly and indirectly in the completion of this project.

C HEMA BINDU (207R1A05D4) KAREMI TEJA (207R1A05E7) CH NAVATEJ (2071RA05D2)

ABSTRACT

This project has a resource allocation mechanism for the study of the strategic behavior of social media interacting with citizens that form opinions in a democracy. In a world of information and the internet, it becomes imperative for social media to filter misleading opinions on their platforms. As this is too altruistic to expect from different social media to self-enforce, it proposes a mechanism design formulation that provides appropriate monetary incentives to social media leading to an efficient filter-wide system outcome. The proposed mechanism incentivizes strategic social media to efficiently filter misleading information and thus indirectly prevent the ever-emergent phenomenon of fake news. In particular, it consider an economically inspired mechanism that designs an implementable Nash equilibrium of efficient filtering of misleading information in a game of selfish social media platforms. It also show that the mechanism is individual rational and budget balance, two key characteristicsof ademocratic society.

Mechanism design was developed for the implementation of system-wide optimal solutions to problems involving multiple rational players with conflicting interests, each with private information about preferences. The fact that Mechanism design optimizes the behaviour of competing players has led to broad applications spanning different fields including economics, politics, wireless networks, social networks, internet advertising, spectrum and bandwidth trading, logistics, supply chain, management, grid computing, and resource allocation problems in decentralized systems.

CMRTC i

LIST OF FIGURES

FIGURE NO	FIGURE NAME	PAGE NO
Figure 3.1	Project Architecture	12
Figure 3.2	Use Case Diagram	14
Figure 3.3	Class Diagram	15
Figure 3.4	Sequence Diagram	16
Figure 3.5	Activity Diagram	17

CMRTC ii

LIST OF SCREENSHOTS

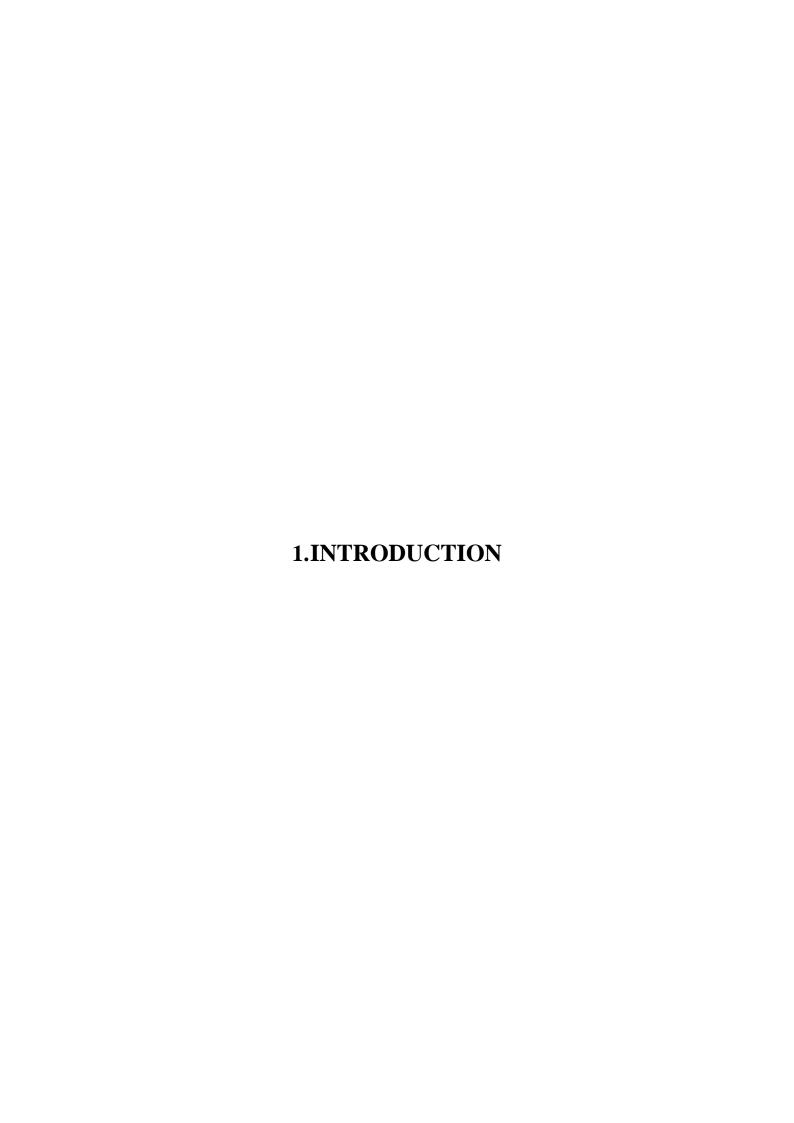
SCREENSHOT NO.	SCREENSHOT NAME	PAGE NO
Screenshot 6.1	True Dataset values	29
Screenshot 6.2	False Dataset values	29
Screenshot 6.3	Opening the CMD for running the program	30
Screenshot 6.4	Open the web-page for predicting data	30
Screenshot 6.5	Copy the link of Web-page	31
Screenshot 6.6	Paste the Web-page link in Search engine	31
Screenshot 6.7	After opening the web-page looks like this	32
Screenshot 6.8	Now login as Service Provider	32
Screenshot 6.9	After login the web-page looks like the	is 33
Screenshot 6.10	Register as Remote-user	33
Screenshot 6.11	Login as Remote-user	34
Screenshot 6.12	Giving true data for prediction	34
Screenshot 6.13	Predicting the News Type as True	35
Screenshot 6.14	Giving false data for prediction	35
Screenshot 6.15	Predicting the News Type as False	36
Screenshot 6.14	Pie-Chart comparison for algorithms	36

CMRTC iii

TABLE OF CONTENTS

ABSTRACT	i
LIST OF FIGURES	ii
LIST OF SCREENSHOTS	ii
1. INTRODUCTION	
1.1 INTRODUCTION	1
1.2 PROJECT PURPOSE	1
1.3 PROJECT FEATURES	1
2 SYSTEM ANALYSIS	
2.2 SYSTEM ANALYSIS	3
2.3 PROBLEM DEFINITION	3
2.4 EXISTING SYSTEM	4
2.4.1 DISADVANTAGES OF THE EXISTING SYSTEM	5
2.5 PROPOSED SYSTEM	6
2.5.1 ADVANTAGES OF PROPOSED SYSTEM	8
2.6 FEASIBILITY STUDY	9
2.6.1 ECONOMIC FEASIBILITY	9
2.6.2 TECHNICAL FEASIBILITY	9
2.6.3 SOCIAL FEASIBILITY	9
2.6.4 OPERATIONAL FEASIBILITY	10
2.7 HARDWARE & SOFTWARE REQUIREMENTS	11
2.7.1 HARDWARE REQUIREMENTS	11
2.7.2 SOFTWARE REQUIREMENTS	11
3 ARCHITECTURE	
3.2 PROJECT ARCHITECTURE	12
3.3 DESCRIPTION	13
3.4 USE CASE DIAGRAM	14
3.5 CLASS DIAGRAM	15
3.6 SEQUENCE DIAGRAM	16
3.7 ACTIVITY DIAGRAM	17
4 IMPLEMENTATION	
4.2 PYTHON	18
4.3 ADVANTAGES OF PYTHON	18
4.4 DISADVANTAGES OF PYTHON	20

4.5 HISTORY PYTHON	20
4.6 MODULES USED	21
4.7 SAMPLE CODE	24
5 RESULTS	29
6 TESTING	
6.2 INTRODUCTION TO TESTING	37
6.3 TYPES OFTESTING	37
6.3.1 UNIT TESTING	37
6.3.2 INTEGRATION TESTING	38
6.3.3 FUNCTIONALTESTING	38
6.4 TEST CASES	39
7 CONCLUSION & FUTURE SCOPE	
7.2 PROJECT CONCLUSION	40
7.3 FUTURE SCOPE	40
8 REFERENCES	
8.2 REFERENCES	41
8.3 GITHUB LINK	43



1. INTRODUCTION

1.1 PROJECT SCOPE

The scope of the "Social Media and Misleading Information in a Democracy A Mechanism Design Approach" project encompasses the development of a robust and accurate decision making system. The primary aim is to create an intelligent model capable of identifying weather the news given is true or false related to the democracy of the country. This project would require interdisciplinary collaboration among researchers, data scientists, social media experts, policymakers, and stakeholders to address the complex challenges posed by misleading information on social media in the context of a democracy. The ultimate goal is to strengthen democratic processes and ensure the responsible use of social media platforms. The project will start with the collection of a diverse and representative dataset of misleading information, including both true news and fake news. Data pre-processing will involve removing stop words and all the unwanted data is removed and splitting it into training data and testing data.

1.2 PROJECT PURPOSE

The purpose of the "Social Media and Misleading Information in a Democracy A Mechanism Design Approach" is to investigate, analyze, and address the pervasive issue of misleading information on social media platforms and its profound implications for democratic processes. By employing principles of mechanism design, the project aims to develop innovative strategies, policies, and interventions that can mitigate the spread of misinformation while preserving democratic values, fostering informed civic participation, and enhancing the integrity of public discourse in the digital age. Through rigorous research, modeling, stakeholder collaboration, and ethical considerations, this project seeks to contribute to the advancement of responsible information sharing and the protection of democratic institutions. The objective of this project is to explore the impact of media on the dissemination of misleading information, within a system using a mechanism design approach.

1.3 PROJECT FEATURES

The project "Social Media and Misleading Information in a Democracy: A Mechanism Design Approach" incorporates several key features to comprehensively address the complex issue of misleading information in the context of democracy:

Data-Driven Analysis:

It utilizes extensive data collection and analysis from social media platforms to gain insights into the dissemination patterns and impact of misinformation, enabling evidence-based decision-making.

Public Awareness and Education:

The project emphasizes the importance of educating the public to recognize and resist misinformation, leading to greater resilience against deceptive content.

Interdisciplinary Collaboration:

The project promotes collaboration among researchers, data scientists, social media experts, economists, and policymakers to leverage diverse expertise for a holistic understanding of the problem and the design of effective mechanisms.

Policy Formulation:

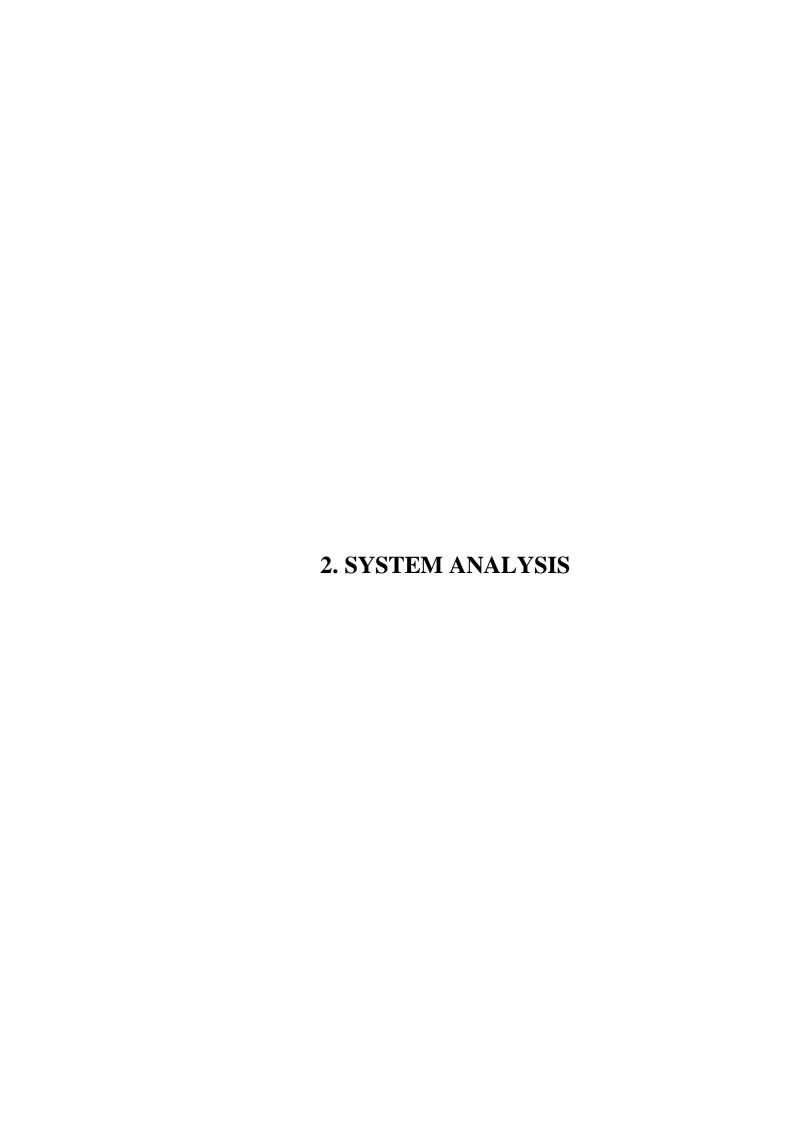
It aims to propose well-informed policies and interventions that can be implemented by social media platforms, government agencies, and civil society organizations to combat misinformation while respecting democratic values.

Budget and Resource Management:

Detailed budget planning and resource allocation ensure the efficient execution of the project's research, development, and outreach activities.

Ethical Considerations:

Ethical issues related to privacy, freedom of speech, and the balance between combatting misinformation and safeguarding democratic principles are carefully addressed in the design of mechanisms and policies.



2. SYSTEM ANALYSIS

2.1 SYSTEM ANALYSIS

System Analysis is the important phase in the system development process. The System is studied to the minute details and analyzed. The system analyst plays an important role of an interrogator and dwells deep into the working of the present system. In analysis, a detailed study of these operations performed by the system and their relationships within and outside the system is done. A key question considered here is, "what must be done to solve the problem?" The system is viewed as a whole and the inputs to the system are identified. Once analysis is completed the analyst has a firm understanding of what is to be done.

2.2 PROBLEM DEFINITION

The problem of social media and misleading information in a democracy, from a mechanism design perspective, can be defined as follows:

In a democratic society, information plays a crucial role in shaping public opinion and decision-making. Social media platforms have become a primary source of information dissemination. However, these platforms also face challenges related to the spread of misleading or false information, often referred to as "fake news" or misinformation. This poses a significant threat to the functioning of a democracy, as it can influence public opinion, electoral outcomes, and policy decisions.

A mechanism design approach to this problem involves designing incentives and rules within social media platforms to mitigate the spread of misleading information while preserving free speech and open discourse. This may include the design of algorithms, content moderation policies, and user incentives to encourage the sharing of accurate and reliable information while discouraging the dissemination of false or misleading content. The challenge lies in striking a balance between combating misinformation and upholding principles of freedom of expression and diversity of perspectives, which are essential in a democratic society.

2.3 EXISTING SYSTEM

Social media in particular, has generated extraordinary concern, in large part because of its potential effects on public opinion, political polarization, and ultimately democratic decision making. Recently, however, a handful of papers have argued that both the prevalence and consumption of "fake news" per se is extremely low compared with other types of news and news- relevant content. Although neither prevalence nor consumption is a direct measure of influence, this work suggests that proper understanding of misinformation and its effects requires a much broader view of the problem, encompassing biased and misleading—but not necessarily factually incorrect— information that is routinely produced or amplified by mainstream news organizations.

In the existing system we use Naive Bayes Classifier and Support Vector Machine.

According to the Bayes' theorem, "conditional probability" is the "probability that something will happen given that something else has already occurred." Naive Bayes refers to this kind of calculation. Since we already know how likely it is for something to happen, we can estimate the likelihood of it. A supervised learning method, naive Bayes, is a sort of classifier. For each class, it predicts "membership probability" and hence belongs to the machine language class. Among other things, it estimates the chance that a given piece of evidence or record belongs to a certain class. If one is looking to identify the "most likely class," they must first identify the class with the highest probability. MAP categorization is another name for this technique. Alternative interpretation is as follows: The "naïve" premise that all features are unrelated underlies the naïve Bayes classifier. There is no evidence to support this claim of independence in most cases. Take a look at this scenario: During a scan of an article, the naive Bayes classifier finds the term "Barack." It is not uncommon for the same story to also mention Barack Obama. This strategy will overestimate "the possibility that an object belongs to a certain class," as previously mentioned, despite the fact that these two attributes are interrelated. The research supports the idea that the naive Bayes classifier is unsuited for text classification because it overestimates the likelihood of dependency. "Strong dependencies" are no problem for the naive Bayes classifier since the dependencies will almost always cancel each other out. The popularity of the naive Bayes classifier may be attributed in part to the fact that it is both quick and readily accessible. For "text classification issues," it is a superb option because of its adaptability. It may be used for binary or multi class classifications. Because the naive Bayes classifier is so straightforward to use, it does not need a big amount of counts to provide its findings. Therefore, as the name says, it is easy to train on a little data set.

One of the most used supervised learning algorithms is the SVM, which may be used interchangeably with the support vector network (SVN). SVM are trained using data that has previously been divided into two groups. As a result, the model is built only after it has been trained before. Furthermore, the SVM method's purpose is to identify which group fresh data fit into, as well as to optimize the margin between the two classes. The SVM ideal objective is to discover a hyper-plane that splits the data set into two distinct groups. "The data points closest to the hyper-plane" are "support vectors," and removing them would change the placement of their dividing hyper-plane. This is why the support vectors are so important. A hyper-plane may be described as "a line that linearly divides and classifies a group of data" and "the farther from the hyper-plane our data points reside, the greater the possibility that our data points have been properly categorized". Because of this, it is a good idea to use the SVM approach because it is incredibly accurate and does well on data sets that are tiny and succinct.

2.3.1 DISADVANTAGES OF EXISTING SYSTEM

Algorithmic Amplification:

Social media algorithms often prioritize sensational or emotionally charged content, which can incentivize the creation and spread of misleading information for increased engagement and profit.

Incentive Misalignment:

The economic interests of social media platforms may conflict with the need to combat misleading information. Platforms benefit from user engagement, which may be driven by sensational or false content.

Lack of Algorithmic Transparency:

Many social media algorithms are proprietary and lack transparency. Users often do not understand how content is ranked or recommended, making it difficult to hold platforms accountable. The system doesn't have facility to train and test on large number of numbers.

2.4 PROPOSED SYSTEM

To tackle this growing phenomenon of misinformation, in this paper, it consider a finite group of social media platforms, whose users represent the citizens in a democracy, and a democratic government. Every post in the platforms is associated with a parameter that captures its informativeness, which can take values between two extremes: (i) completely factual and (ii) complete misinformation. This framework, posts that exhibit misinformation can lead to a decrease in trust on common knowledge among the users. In addition, social media platforms are considered to have the technologies to *filter*, or label, posts that intend to sacrifice trust on common knowledge. Thus, the government seeks to incentivize the social media platforms to use these technologies and filter any misinformation included in the posts.

In this framework, the government is also a strategic player, whose utility increases as the trust of the users of social media platforms on common knowledge increases. Consequently, increasing filtering of misinformation by the social media platforms increases the utility of the government. Thus the government is willing to make an investment to incentivize the social media platforms to filter misinformation.

In this approach, it use mechanism design to distribute this investment among the platforms optimally, and in return, implement an optimal level of filtering. Mechanism design was developed for the implementation of system-wide optimal solutions to problems involving multiple rational players with conflicting interests, each with private information about preferences. Note that this approach is different from traditional approaches to decentralized control with private information because the players are not a part of the same time, but in fact, have private and competitive utilities. The fact that Mechanism design optimizes the behaviour of competing players has led to broad applications spanning different fields including economics, politics, wireless networks, social networks, internet advertising, spectrum and bandwidth trading, logistics, supply chain, management, grid computing, and resource allocation problems in decentralized systems. Here in proposed system we use Logistic Regression, Random Forest algorithm and Decision Tree Algorithm.

In this approach, it use mechanism design to distribute this investment among the platforms optimally, and in return, implement an optimal level of filtering. Mechanism design was developed

for the implementation of system-wide optimal solutions to problems involving multiple rational players with conflicting interests, each with private information about preferences. Note that this approach is different from traditional approaches to decentralized control with private information because the players are not a part of the same time, but in fact, have private and competitive utilities. The fact that Mechanism design optimizes the behaviour of competing players has led to broad applications spanning different fields including economics, politics, wireless networks, social networks, internet advertising, spectrum and bandwidth trading, logistics, supply chain, management, grid computing, and resource allocation problems in decentralized systems. Here in proposed system we use Logistic Regression, Random Forest algorithm and Decision Tree Algorithm.

The Logistic Regression is quite good in solving binary classifications due to its predictive power in-probability values are taken. Logistic Regression detection model works well in dealing and also short input text and the range of accuracy can be achieved is within 79.0% to 89.0% based on the data on the table. The algorithm that used to predict is depends on logical regression, and the binary variable that contains the code yes, success etc or no, failure, etc for yes it takes 0 and for no it takes 1, and in other words, the logistic regression model predicts as P(y=1) as a function of x.

The Decision Tree algorithm is a commonly used machine learning algorithm that builds a tree-like model to make predictions based on the input features. The tree-like model is constructed by recursively partitioning the feature space into regions that correspond to specific classifications. The partitions are determined by selecting the feature that provides the highest information gain or the most significant reduction in entropy at each node of the tree. In the proposed method, Decision Tree was used to classify articles based on their content features extracted using NLP techniques. The features extracted include the frequency of occurrence of specific words or phrases, and other linguistic features. By using these features, Decision Tree was able to classify articles as real or fake with a high degree of accuracy.

Random Forest can be defined as a procedure that reduces the variance of an estimated function of prediction". Random Forest works efficiently with high variance and low bias techniques like trees in classification. Random forests are a significant innovation of the bagging in which it forms a large group of correlated trees, and after that, take an average for them. Random Forest

enhanced on bagging through decreasing correlation between trees with no increase in the variance. In many situations, the random forest performance is like boosting in which they are simpler to be trained and tuned. As a result, random forests are widespread algorithms that are applied to various packages.

2.4.1 ADVANTAGES OF THE PROPOSED SYSTEM

Improved Information Accuracy:

The mechanism design approach can incentivize social media platforms to prioritize accurate and reliable information over misleading content. Algorithms can be designed to reward quality and penalize misinformation.

Fact-Checking Mechanisms:

The mechanism design approach can facilitate the development of effective fact-checking mechanisms, involving both user-generated efforts and independent experts. This can help verify the accuracy of information circulating on social media.

Data Privacy Protection:

Mechanisms can be designed to protect user data and limit its use for targeted advertising, reducing privacy concerns associated with social media platforms.

Enhanced Algorithmic Transparency:

The proposed system can make social media algorithms more transparent, allowing users to understand how content is ranked and recommended. This transparency can help users make more informed choices about the information they consume.

Efficiency:

The proposed system aims to provide an automated, non-invasive method for diagnosing cardiac arrhythmia, which can be more efficient than traditional manual methods. It gives high efficiency.

2.5 FEASIBILITY STUDY

The feasibility of the project is analyzed in this phase and business proposal is put forth with a very general plan for the project and some cost estimates. During system analysis the feasibility study of the proposed system is to be carried out. This is to ensure that the proposed system is not a burden to the company. For feasibility analysis, some understanding of the major requirements for the system is essential.

Three key considerations involved in the feasibility analysis are

- ECONOMICAL FEASIBILITY
- TECHNICAL FEASIBILITY
- SOCIAL FEASIBILITY
- OPERATIONAL FEASIBILITY

2.5.1 ECONOMICAL FEASIBILITY

This study is carried out to check the economic impact that the system will have on the organization. The amount of fund that the company can pour into the research and development of the system is limited. The expenditures must be justified. Thus the developed system as well within the budget and this was achieved because most of the technologies used are freely available. Only the customized products had to be purchased.

2.5.2 TECHNICAL FEASIBILY

This study is carried out to check the technical feasibility, that is, the technical requirements of the system. Any system developed must not have a high demand on the available technical resources. This will lead to high demands on the available technical resources. This will lead to high demands being placed on the client. The developed system must have a modest requirement, as only minimal or null changes are required for implementing this system.

2.5.3 SOCIAL FEASIBILITY

The aspect of study is to check the level of acceptance of the system by the user. This includes the process of training the user to use the system efficiently. The user must not feel threatened by the system, instead must accept it as a necessity. The level of acceptance by the users solely depends on the methods that are employed to educate the user about the system and to make him familiar

with it. His level of confidence must be raised so that he is also able to make some constructive criticism, which is welcomed, as he is the final user of the system.

2.5.4 OPERATIONAL FEASIBILITY

Operational Feasibility deals with the study of prospects of the system to be developed. This system operationally eliminates all the tensions of the Admin and helps him in effectively tracking the project progress. This kind of automation will surely reduce the time and energy, which previously consumed in manual work. Based on the study, the system is proved to be operationally feasible.

2.6 HARDWARE & SOFTWARE REQUIREMENTS

2.6.1 HARDWARE REQUIREMENTS:

Hardware interfaces specify the logical characteristics of each interface between the software product and the hardware components of the system. The following are some hardware requirements.

Processor: Intel Dual Core I5 and above

• Hard disk: 10GB and above

• RAM: 2GB and above

• Input devices: Keyboard, mouse.

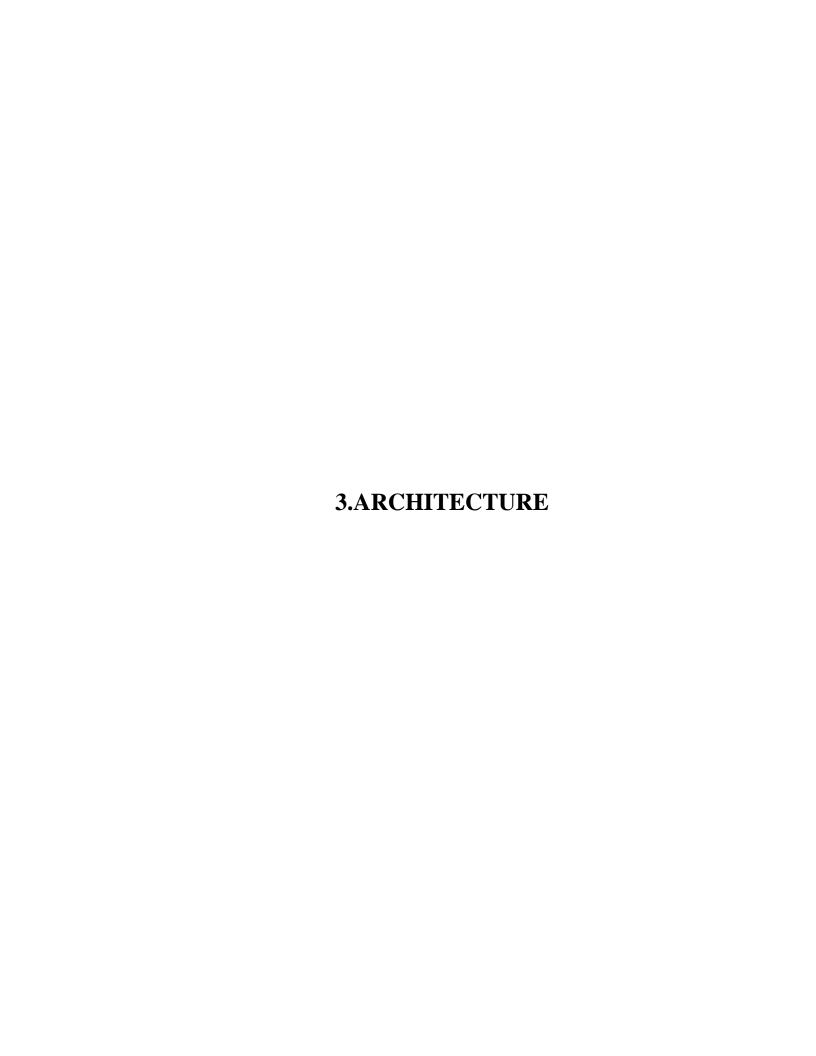
2.6.2 SOFTWARE REQUIREMENTS:

Software Requirements specifies the logical characteristics of each interface and software components of the system. The following are some software requirements.

Operating system: Windows 8 and above

Languages: Python

Tools: VisualStudio Code



3. ARCHITECTURE

3.1 PROJECT ARCHITECTURE

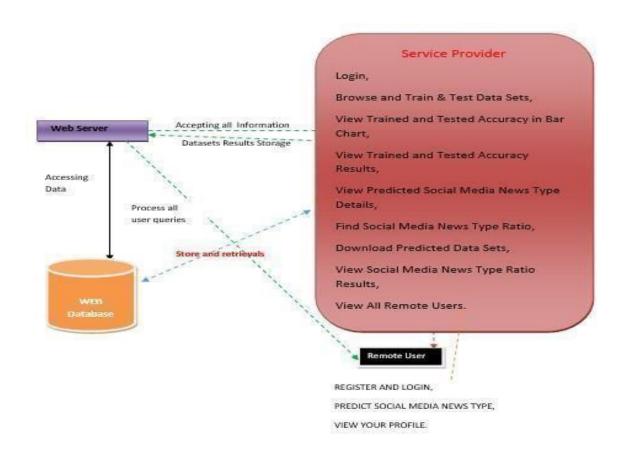


Figure 3.1: Architecture for Social Media and Misleading Information in a Democracy A Mechanism Design Approach.

3.2 DESCRIPTION

The data set is considered for evaluating the performance of overall system. Initially, the data set is split-ted into 80% for training and 20% for testing. Then, the entire operations are going to be perform on both training and testing data sets. Further, pre-processing operation is carried out to remove the missing symbols and unknown characters, special characters. The pre-processing operation also normalizes the number of rows and columns presented in the data set. Further, both logical regression, decision tree and random forest algorithm models are applied to evaluate the prediction of data set. So, through this prediction it is going to identify the cardiac arrhythmia presented in overall data set. Finally, performance comparison is takes place between logical regression, decision tree and random forest algorithm.

3.3 USE CASE DIAGRAM

In the use case diagram, we have basically one actor who is the user in the trained model. A use case diagram is a graphical depiction of a user's possible interactions with a system. A use case diagram shows various use cases and different types of usersthe system has. The use cases are represented by either circles or ellipses. The actors are often shown as stick figures.

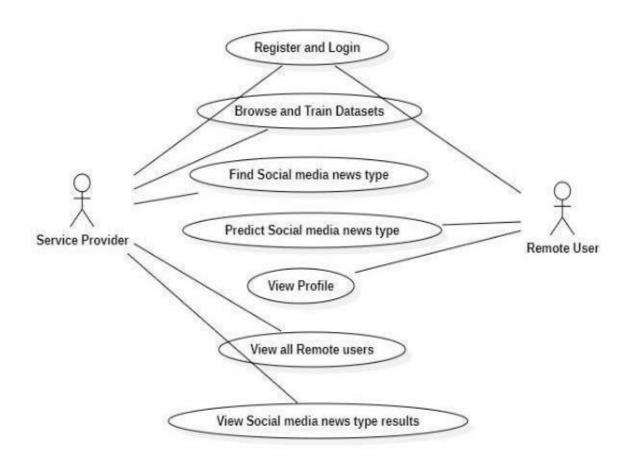


Figure 3.3: Use Case Diagram for Social Media and Misleading Information in a Democracy A

Mechanism Design Approach

3.4 CLASS DIAGRAM

Class diagram is a type of static structure diagram that describes the structure of a system by showing the system's classes, their attributes, operations (or methods), and the relationships among objects.

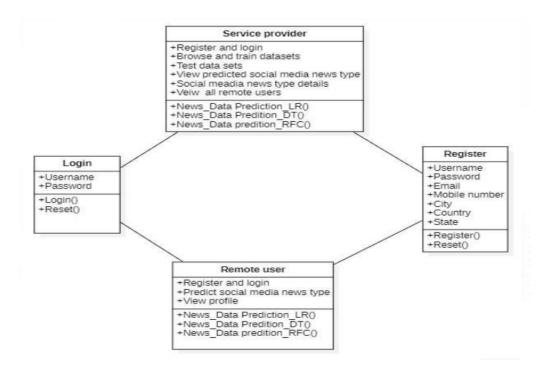


Figure 3.4: Class Diagram for Social Media and Misleading Information in a Democracy A

Mechanism Design Approach

3.5 SEQUENCE DIAGRAM

A sequence diagram shows object interactions arranged in time sequence. It depicts the objects involved in the scenario and the sequence of messages exchanged between the objects needed to carryout the functionality of the scenario. Sequence diagrams are typically associated with use case realizations in the logical view of the system under development.

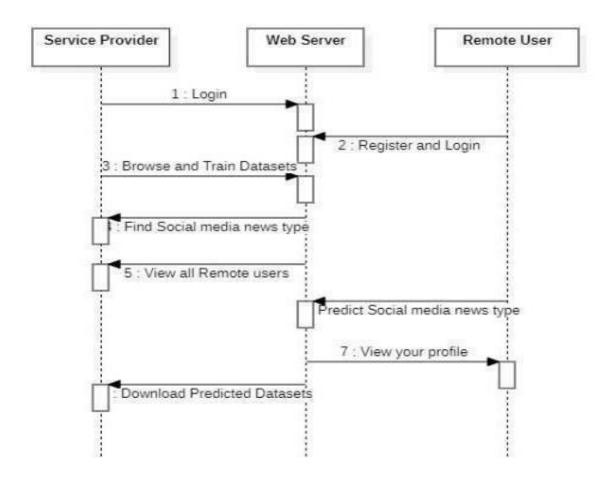


Figure 3.5: Sequence Diagram for Social Media and Misleading Information in a Democracy

A Mechanism Design Approach

3.6 ACTIVITY DIAGRAM

Activity diagrams are graphical representations of workflows of stepwise activities and actions with support for choice, iteration and concurrency. They can also include elements showing the flow of data between activities through one or more datastores.

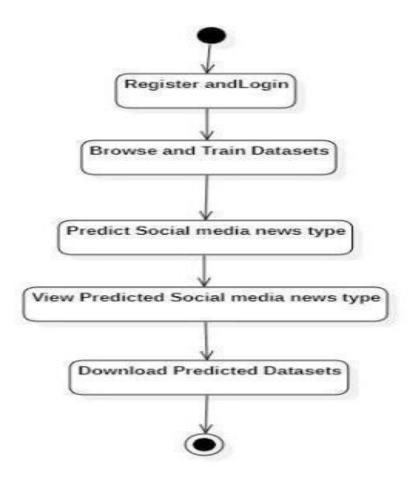
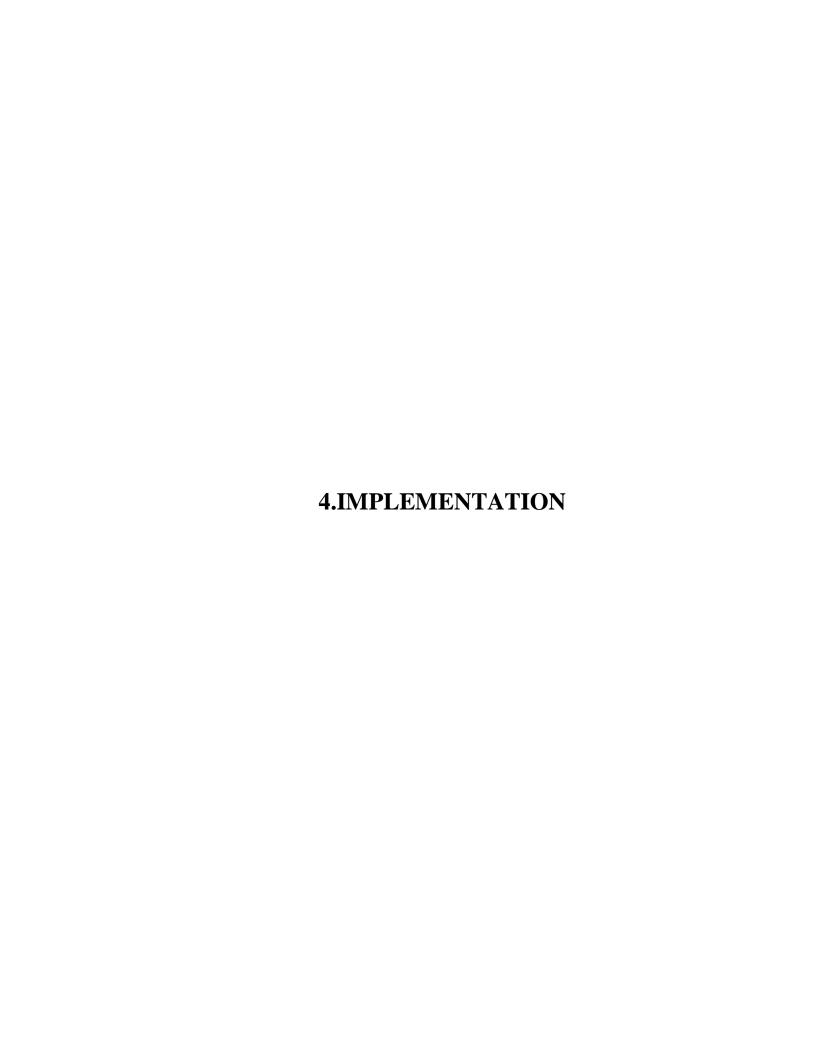


Figure 3.6: Activity Diagram for Social Media and Misleading Information in a Democracy A

Mechanism Design Approach



4. IMPLEMENTATION

4.1 Python

- Below are some facts about Python.
- Python is currently the most widely used multi-purpose, high-level programming language.
- Python allows programming in Object-Oriented and Procedural paradigms. Python programs generally are smaller than other programming languages like Java.
- Programmers have to type relatively less and indentation requirement of the language, makes them readable all the time.
- Python language is being used by almost all tech-giant companies like Google, Amazon,
 Facebook, Instagram, Dropbox, Uber... etc.
- The biggest strength of Python is huge collection of standard library which can be used for the following –
 - o GUI Applications (like Kivy, Tkinter, PyQt etc.)
 - o Web frameworks like Django (used by YouTube, Instagram, Dropbox)
 - o Image processing (like Opency, Pillow)
 - o Web scraping (like Scrapy, BeautifulSoup, Selenium)
 - Test frameworks
 - o Multimedia

4.2 Advantages of Python

Let's see how Python dominates over other languages.

1. Extensive Libraries

Python downloads with an extensive library and it contain code for various purposes like regular expressions, documentation-generation, unit-testing, web browsers, threading, *databases*, *CGI*, *email*, *image manipulation*, *and more*. So, we don't have to write the complete code for that manually.

2. Extensible

As we have seen earlier, Python can be extended to other languages. You can write some of your code in languages like C++ or C. This comes in handy, especially in projects.

3. Embeddable

Complimentary to extensibility, Python is embeddable as well. You can put your Python code in your source code of a different language, like C++. This lets us add scripting capabilities to our code in the other language.

4. Improved Productivity

The language's simplicity and extensive libraries render programmers more productive than languages like Java and C++ do. Also, the fact that you need to write less and get more things done.

5. IOT Opportunities

Since Python forms the basis of new platforms like Raspberry Pi, it finds the future bright for the Internet Of Things. This is a way to connect the language with the real world.

6. Readable

Because it is not such a verbose language, reading Python is much like reading English. This is the reason why it is so easy to learn, understand, and code. It also does not need curly braces to define blocks, and indentation is mandatory. These further aids the readability of the code.

7. Free and Open-Source

Like we said earlier, Python is freely available. But not only can you **download Python** for free, but you can also download its source code, make changes to it, and even distribute it. It downloads with an extensive collection of libraries to help you with your tasks.

8. Portable

When you code your project in a language like C++, you may need to make some changes to it if you want to run it on another platform. But it isn't the same with Python. Here, you need to code only once, and you can run it anywhere. This is called Write Once Run Anywhere (WORA). However, you need to be careful enough not to include any system-dependent features.

9. Interpreted

Lastly, we will say that it is an interpreted language. Since statements are executed

one by one, debugging is easier than in compiled languages.

4.3 Disadvantages of Python

So far, we've seen why Python is a great choice for your project. But if you choose it, you should be aware of its consequences as well. Let's now see the downsides of choosing Python over another language.

1. Speed Limitations

We have seen that Python code is executed line by line. But since <u>Python</u> is interpreted, it often results in **slow execution**. This, however, isn't a problem unless speed is a focal point for the project. In other words, unless high speed is a requirement, the benefits offered by Python are enough to distract us from its speed limitations.

2. Weak in Mobile Computing and Browsers

While it serves as an excellent server-side language, Python is much rarely seen on the **client-side**. Besides that, it is rarely ever used to implement smartphone-based applications. One such application is called **Carbonnelle**.

3. Design Restrictions

As you know, Python is dynamically-typed. This means that you don't need to declare the type of variable while writing the code. It uses duck-typing. But wait, what's that? Well, it just means that if it looks like a duck, it must be a duck. While this is easy on the programmers during coding, it can raise run-time errors.

4. Underdeveloped Database Access Layers

Compared to more widely used technologies like JDBC (Java DataBase Connectivity) and ODBC (Open DataBase Connectivity), Python's database access layers are a bit underdeveloped. Consequently, it is less often applied in huge enterprises.

4.4 History of Python

Python laid its foundation in the late 1980s.

- In February 1991, Guido Van Rossum published the code (labeled version 0.9.0) to alt.sources.
- In 1994, Python 1.0 was released with new features like lambda, map, filter, and reduce.
- Python 2.0 added new features such as list comprehensions, garbage collection systems.
- On December 3, 2008, Python 3.0 (also called "Py3K") was released. It was designed to rectify the fundamental flaw of the language.
- The implementation of Python was started in December 1989 by **Guido Van Rossum** at CWI in Netherland.
 - The following programming languages influence Python:
 - ABC language.
 - Modula-3

4.5 Modules used

Pandas

Pandas is an open source library in Python. It provides ready to use high-performance data structures and data analysis tools. Pandas module runs on top of NumPy and it is popularly used for data science and data analytics. Pandas is a Python library used for working with data sets. It has functions for analyzing, cleaning, exploring, and manipulating data. To install Pandas using pip, enter pip install pandas or pip3 install pandas in the terminal or command line. To install Pandas using conda, execute the following command on a terminal or command line: conda install pandas. ImportError usually arises when the Python working environment doesn't recognize pandas. Pandas has been one of the most commonly used tools for Data Science and Machine learning, which is used for data cleaning and analysis. Here, Pandas is the best tool for handling this real-world messy data. And pandas is one of the open-source python packages built on top of NumPy.

Seaborn

Seaborn is a Python data visualization library based on matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics. In order to install the Seaborn library in Python, you can use either "pip install seaborn" or "conda install seaborn", depending on which package manager you use.

OpenCV

OpenCV is a huge open-source library for computer vision, machine learning, and image processing. OpenCV supports a wide variety of programming languages like Python, C++, Java, etc. It can process images and videos to identify objects, faces, or even the handwriting of a human. When it is integrated with various libraries, such as <u>Numpy</u> which is a highly optimized library for numerical operations, then the number of weapons increases in your Arsenal i.e whatever operations one can do in Numpy can be combined with OpenCV.

Numpy

Numpy is a general-purpose array-processing package. It provides a high- performance multidimensional array object, and tools for working with these arrays.

It is the fundamental package for scientific computing with Python. It contains various features including these important ones:

- A powerful N-dimensional array object.
- Sophisticated (broadcasting) functions.
- Tools for integrating C/C++ and Fortran code.

Sklearn

Scikit-Learn, also known as sklearn is a python library to implement machine learning models and statistical modelling. Through scikit-learn, we can implement various machine learning models for regression, classification, clustering, and statistical tools for analyzing these models. Sklearn model implements a range of machine learning, pre-processing, cross-validation, and visualization algorithms using a unified interface. Important features of scikit-learn: Simple and efficient tools for data mining and data analysis.

How to Install Scikit-Learn:

- Download SciPy installer using the link SciPy: Scientific Library for Python Browse /scipy/0.16. 1 at SourceForge.net.
- Install Pip by typing python get_pip.py in the command line terminal.
- Install scikit-learn by typing pip install scikit-learn in the command line.

Sklearn model

Scikit-learn (Sklearn) is the most useful and robust library for machine learning in Python. It provides a selection of efficient tools for machine learning and statistical modeling including classification, regression, clustering and dimensionality reduction via a consistence interface in Python.

Sklearn metrics

The sklearn. metrics module implements several loss, score, and utility functions to measure classification performance. Some metrics might require probability estimates of the positive class, confidence values, or binary decisions values.

Sklearn accuracy

Sklearn provides a simple function accuracy_score to compute the Accuracy score. The function takes two arrays as input: y_true and y_pred . y_true is an array of true labels, and y_pred is an array of predicted labels. The function returns the accuracy score, which is the proportion of correctly classified examples. It takes two arguments: the true labels and the predicted labels. It returns a value between 0 and 1, where 1 represents a perfect prediction and 0 represents a complete mismatch.

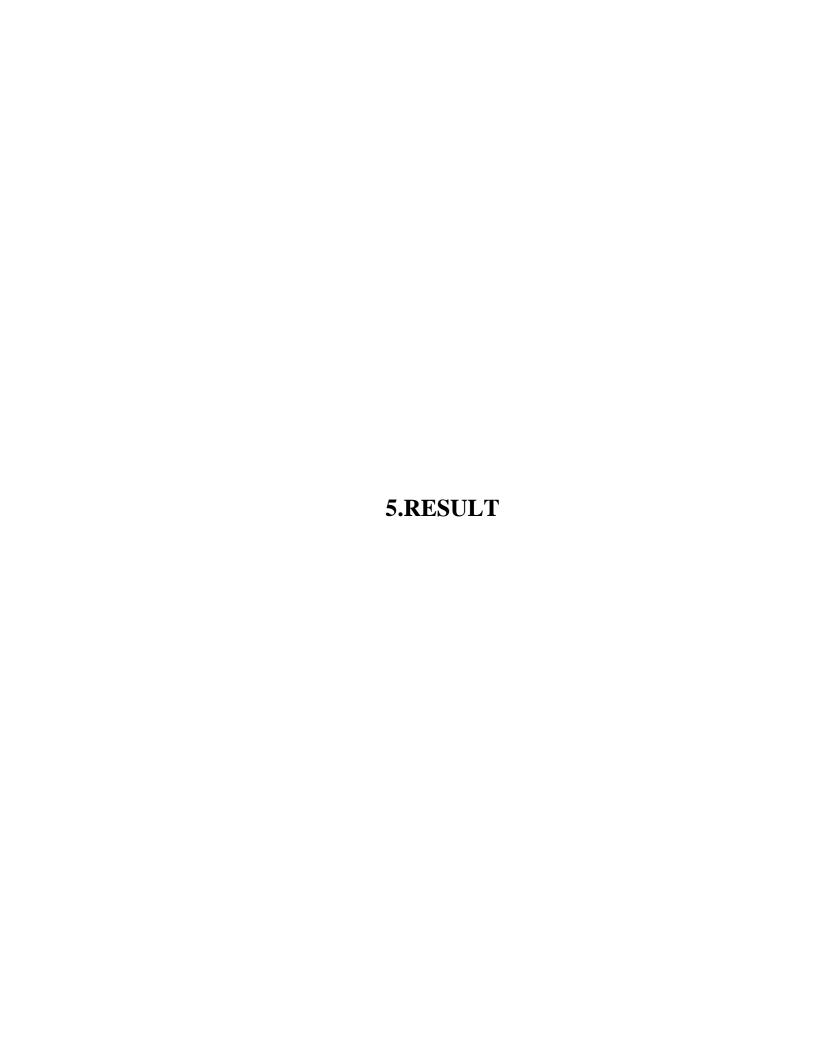
4.6 SAMPLE CODE

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score
from sklearn.metrics import classification_report
import re
import string
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.metrics import accuracy_score, confusion_matrix, classification_report
from sklearn.ensemble import VotingClassifier
def Train_Test_DataSets(request):
  detection_accuracy.objects.all().delete()
  df_fake = pd.read_csv("Fake.csv")
  df_true = pd.read_csv("True.csv")
  df_fake.head()
  df_true.head(5)
  df_fake["class"] = 0
  df_true["class"] = 1
  df_fake.shape, df_true.shape
  # Removing last 10 rows for manual testing
  df_fake_manual_testing = df_fake.tail(10)
  for i in range(23480, 23470, -1):
```

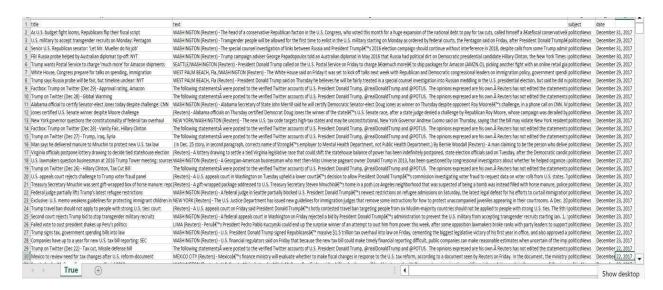
```
df_fake.drop([i], axis=0, inplace=True)
  df_true_manual_testing = df_true.tail(10)
  for i in range(21416, 21406, -1):
     df_true.drop([i], axis=0, inplace=True)
     df fake.shape, df true.shape
     df_fake_manual_testing["class"] = 0
     df_true_manual_testing["class"] = 1
     df_fake_manual_testing.head(10)
     df_true_manual_testing.head(10)
df_manual_testing=pd.concat([df_fake_manual_testing,df_true_manual_testing],axis=0) x
     df_manual_testing.to_csv("manual_testing.csv")
     df_merge = pd.concat([df_fake, df_true], axis=0)
     df_merge.head(10)
     df_merge.columns
     df = df_merge.drop(["title", "subject", "date"], axis=1)
     df.isnull().sum()
     df = df.sample(frac=1)
    df.head()
    df.reset_index(inplace=True)
    df.drop(["index"], axis=1, inplace=True)
     df.columns
     df.head()
     def wordopt(text):
       text = text.lower()
       text = re.sub('\[.*?\]', '', text)
       text = re.sub("\W", "", text)
       text = re.sub('https?://\S+|www\.\S+', ", text)
       text = re.sub('<.*?>+', ", text)
       text = re.sub('[%s]' % re.escape(string.punctuation), ", text)
       text = re.sub('\n', '', text)
```

```
text = re.sub('\w^*\d\w^*', '', text)
       return text
  cv = CountVectorizer()
  df["text"] = df["text"].apply(wordopt)
  x = df["text"]
  y = df["class"]
  x = cv.fit_transform(x)
  x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.25)
  from sklearn.linear_model import LogisticRegression
  print("Logistic Regression")
  LR = LogisticRegression()
  LR.fit(x_train, y_train)
  pred_lr = LR.predict(x_test)
  LR.score(x_test, y_test)
  print("ACCURACY")
  print(accuracy_score(y_test, pred_lr) * 100)
  print("CLASSIFICATION REPORT")
  print(classification_report(y_test, pred_lr))
  print("CONFUSION MATRIX")
  print(confusion_matrix(y_test, pred_lr))
  detection_accuracy.objects.create(names="Logistic Regression", ratio=accuracy_score(y_test,
pred_lr) * 100)
  print("Decision Tree Classifier")
  from sklearn.tree import DecisionTreeClassifier
  DT = DecisionTreeClassifier()
  DT.fit(x_train, y_train)
  pred_dt = DT.predict(x_test)
  DT.score(x_test, y_test)
  print("ACCURACY")
```

```
print(accuracy_score(y_test, pred_dt) * 100)
  print("CLASSIFICATION REPORT")
  print(classification_report(y_test, pred_dt))
  print("CONFUSION MATRIX")
  print(confusion_matrix(y_test, pred_dt))
  detection_accuracy.objects.create(names="Decision Tree Classifier", accuracy_score(y_test,
pred_dt) * 100)
  print("Random Forest Classifier")
  from sklearn.ensemble import RandomForestClassifier
  RFC = RandomForestClassifier(random_state=0)
  RFC.fit(x_train, y_train)
  pred_rfc = RFC.predict(x_test)
  RFC.score(x_test, y_test)
  print("ACCURACY")
  print(accuracy_score(y_test, pred_rfc) * 100)
  print("CLASSIFICATION REPORT")
  print(classification_report(y_test, pred_rfc))
  print("CONFUSION MATRIX")
  print(confusion_matrix(y_test, pred_rfc))
  detection accuracy.objects.create(names="Random Forest Classifier",
ratio=accuracy_score(y_test, pred_rfc) * 100)
  obj = detection_accuracy.objects.all()
  return render(request, 'SProvider/Train_Test_DataSets.html', {'objs': obj})
```



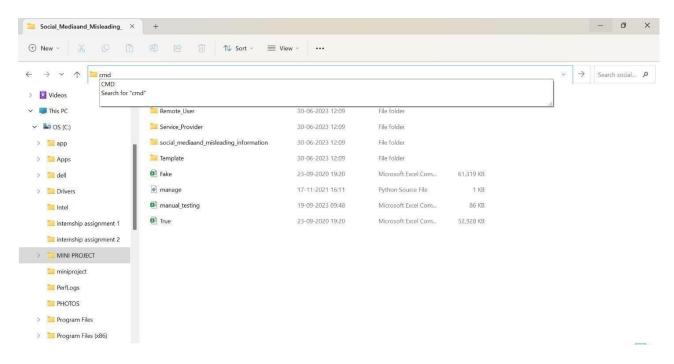
5. Results



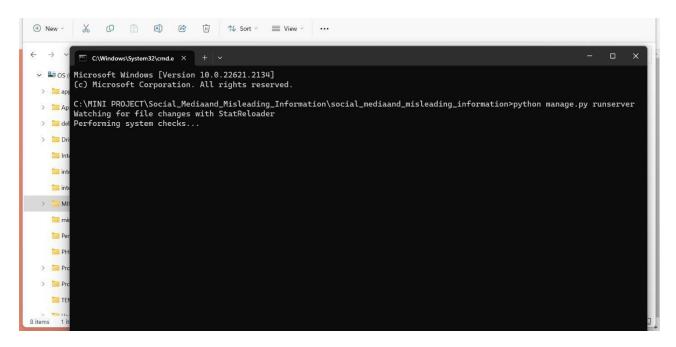
Screenshot 5.1: True Data Set Values



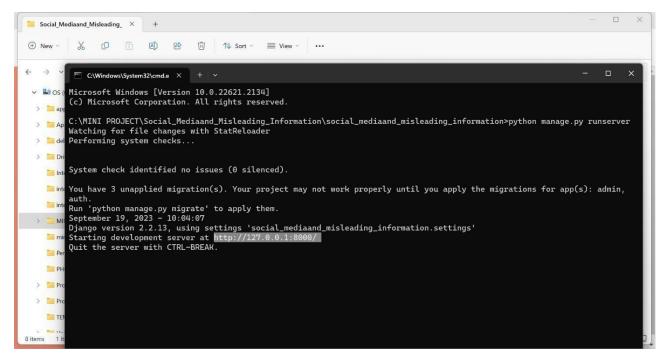
Screenshot 5.2: False Data set Values



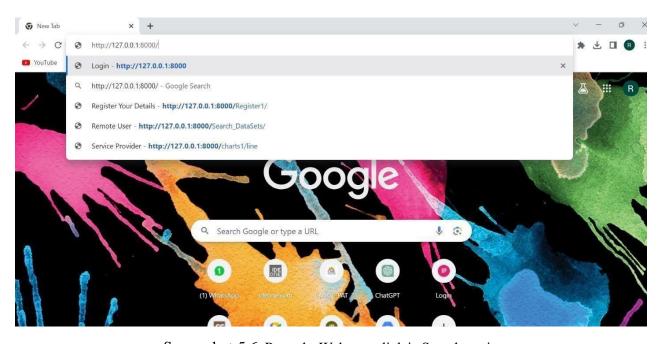
Screenshot 5.3: Opening the CMD for running the program



Screenshot 5.4: Open the web-page for predicting data



Screenshot 5.5: Copy the link of Web-page



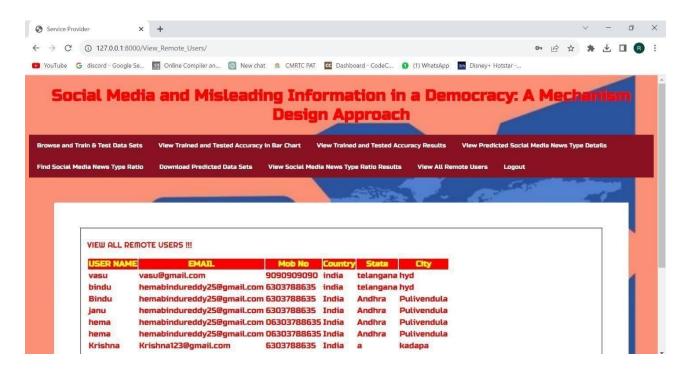
Screenshot 5.6: Paste the Web-page link in Search engine



Screenshot 5.7: After opening the web-page looks like this



Screenshot 5.8: Now login as Service Provider



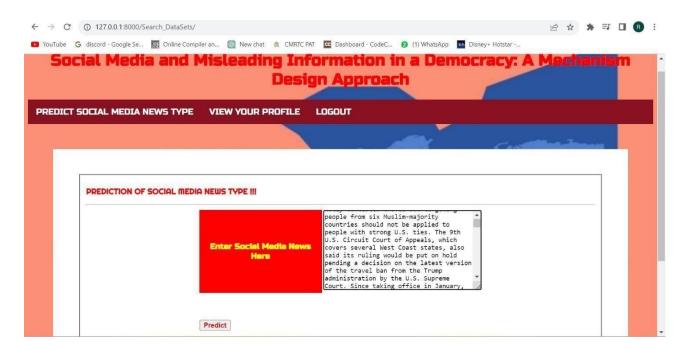
Screenshot 5.9: After login the web-page looks like this



Screenshot 5.10: Register as Remote-user



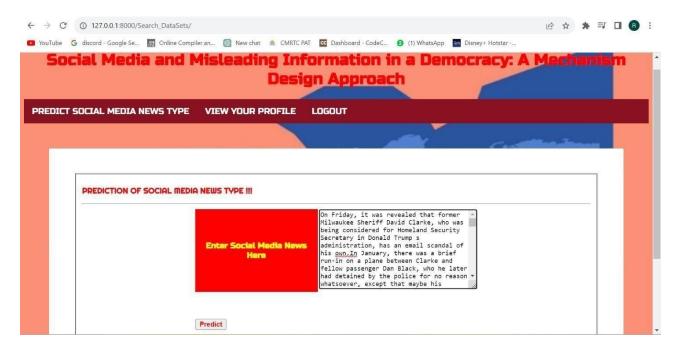
Screenshot 5.11: Login as Remote-user



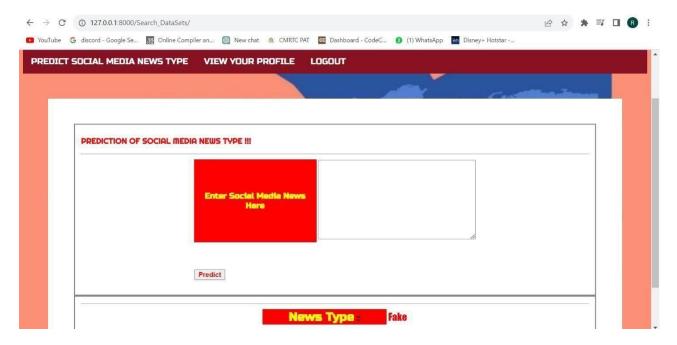
Screenshot 5.12: Giving True Data-set values for prediction



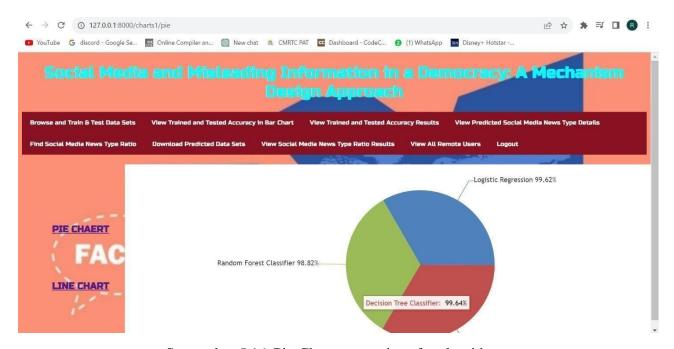
Screenshot 5.13: Predicting the news type as True



Screenshot 5.14: Giving False Data-set values for prediction



Screenshot 5.15: Predicting the news type as False



Screenshot 5.16: Pie-Chart comparison for algorithms



6. TESTING

6.1 INTRODUCTION TO TESTING

The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the functionality of components, subassemblies, assemblies and/or a finished product. It is the process of exercising software with the intent of ensuring that the Software system meets its requirements and user expectations and does not fail in an unacceptable manner. There are various types of tests. Each test type addresses a specific testing requirement.

Software testing is one of the main stages of project development life cycle to provide our cessation utilizer with information about the quality of the application and ours, in our Project we have under gone some stages of testing like unit testing where it's done in development stage of the project when we are in implementation of the application after the Project is yare we have done manual testing with different Case of all the different modules in the application we have even done browser compatibility testing in different web browsers in market, even we have done Client side validation testing on our application.

6.2 TYPES OF TESTING

6.2.1 Unit testing

Unit testing involves the design of test cases that validate that the internal program logic is functioning properly, and that program inputs produce valid outputs. All decision branches and internal code flow should be validated. It is the testing of individual software units of the application it is done after the completion of an individual unit before integration. This is a structural testing, that relies on knowledge of its construction and is invasive. Unit tests perform basic tests at component level and test a specific business process, application, and/or system configuration. Unit tests ensure that each unique path of a business process performs accurately to the documented specifications and contains clearly defined inputs and expected results.

6.2.2 Integration testing

Integration tests are designed to test integrated software components to determine if they actually run as one program. Testing is event driven and is more concerned with the basic outcome of screens or fields. Integration tests demonstrate that although the components were individually satisfaction, as shown by successfully unit testing, the combination of components is correct and consistent. Integration testing is specifically aimed at exposing the problems that arise from the combination of components.

6.2.3 Functional testing

Functional tests provide systematic demonstrations that functions tested are available as specified by the business and technical requirements, system documentation, and user manuals. Functional testing is centered on the following items:

Valid Input : identified classes of valid input must be accepted.

Invalid Input : identified classes of invalid input must be rejected.

Functions : identified functions must be exercised.

Output : identified classes of application outputs must be exercised.

Systems/Procedures: interfacing systems or procedures must be invoked.

6.3 TEST CASES

S.no	Test case	Purpose	Input	Output
	name			
1	Identifying	The user identifies the news type as true or fake	On Christmas day, Donald Trump	Fake
	News Type		announced that he would be back to	
			work the following day, but he is	
			golfing for the fourth day in a row.	
			The former reality show star blasted	
			former	
2	Identifying News Type	The user identifies the news type as true or fake	ABUJA (Reuters) - Nigerian President Muhammadu Buhari will pass through London, where he spent five months on medical leave this year, on his way home from the United Nations	True
3	Identifying News Type	The user identifies the news type as true or fake	Donald Trump has a white supremacy problem, and now it is causing Great Britain, America s closest ally, to rebuke him in an historically public way	Fake
4	Identifying News Type	The user identifies the news type as true or fake	LONDON (Reuters) - British police are chasing down the people they suspect of being behind a bomb which injured 29 people on a packed commuter train in west London on Friday, the country s most senior counter-terrorism officer said	True



7. CONCLUSION

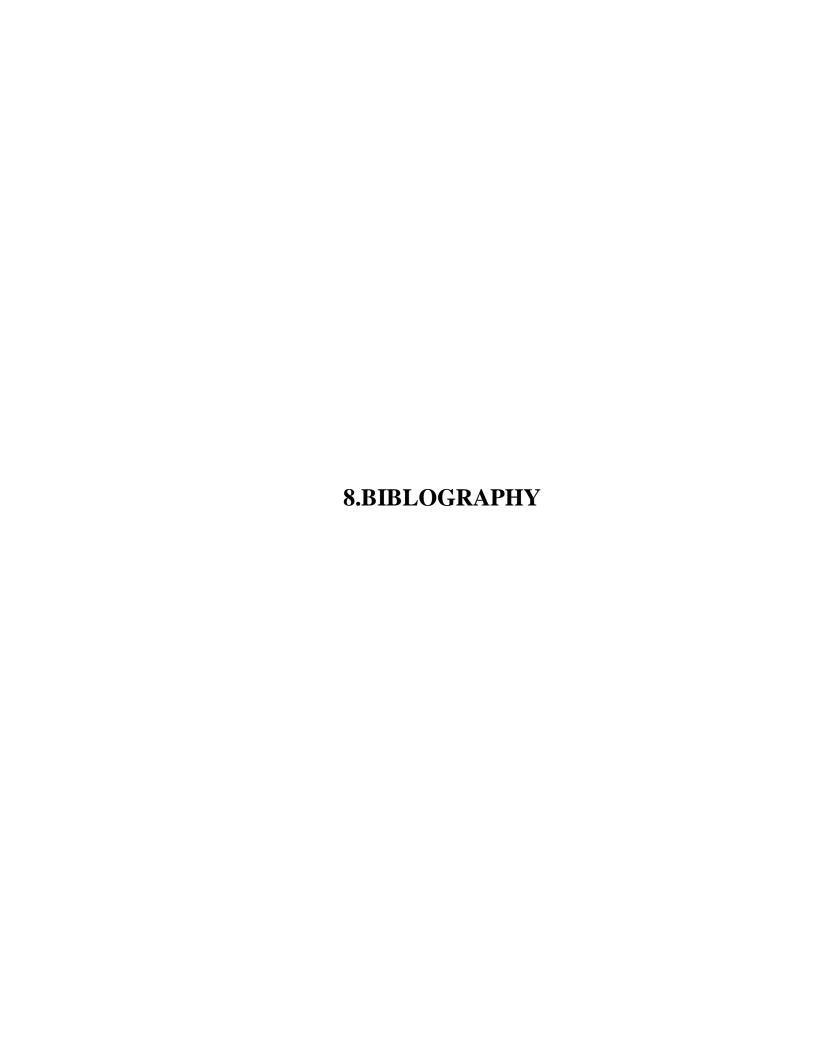
7.1 PROJECT CONCLUSION

The primary goal in this project is to design a mechanism to induce a GNE solution in the misinformation filtering game, where (i) each platform agrees to participate voluntarily, and (ii) the collective utility of the government and the platforms is maximized. Here the mechanism is designed and proved that it satisfies these properties along with budget balance. It also presented an extension of the mechanism with weaker technical assumptions.

In this it focuses on improving the valuation and average trust functions of the social media platforms based on data. Here it also consider the incorporating uncertainty in a platform's estimates of the impact of their filter. These refinements of the modeling framework will allows to make our mechanism more practical for use in the real world.

7.2 Future Scope

Though deep learning networks produces excellent results, the disadvantage lies in the insufficient understanding of the complex inner mechanisms of the deep learning networks. This could be overcome by remodelling the nonlinear deep networks to a linear form by computing eigenvalues and eigenvectors in different time steps. The future work can be the collection of realworld datasets from hospitals having cardiac care units and the application of the same methodologies to the real datasets.



8. BIBLIOGRAPHY

8.1 REFERENCES

- [1] W. Davies, "The age of post-truth politics," *The New York Times*, vol. 24, p. 2016, 2016.
- [2] J. Cone, K. Flaharty, and M. J. Ferguson, "Believability of evidence matters for correcting social impressions," *Proceedings of the National Academy of Sciences*, vol. 116, no. 20, pp. 9802–9807, 2019.
- [3] Z. Tufekci, "Youtube, the great radicalizer," *The New York Times*, vol. 10,2018.
- [4] A. D. Kramer, J. E. Guillory, and J. T. Hancock, "Experimental evidence of massive-scale emotional contagion through social networks," *Proceedings of the National Academy of Sciences*, vol. 111, no. 24, pp.8788–8790, 2014.
- [5] J. Weedon, W. Nuland, and A. Stamos, "Information operations and facebook," *Retrieved from: https://fbnewsroomus. files. wordpress.com/2017/04/facebook-and-information-operations-v1. pdf*, 2017.
- [6] H. Farrell and B. Schneier, "Common-knowledge attacks on democracy," *Berkman Klein Center Research Publication*, no. 2018-7, 2018.
- [7] H. Allcott and M. Gentzkow, "Social media and fake news in the 2016election," *Journal of economic perspectives*, vol. 31, no. 2, pp. 211–36,2017.
- [8] O. Analytica, "Russia will deny cyberattacks despite more us evidence," *Emerald Expert Briefings*, no. oxan-db, 2018.
- [9] A. Bessi, M. Coletto, G. A. Davidescu, A. Scala, G. Caldarelli, and W. Quattrociocchi, "Science vs conspiracy: Collective narratives in theage of misinformation," *PloS one*, vol. 10, no. 2, p. e0118093, 2015.
- [10] E. Brown, "Propaganda, misinformation, and the epistemic value ofdemocracy," *Critical Review*, vol. 30(3–4), pp. 194–218, 2018.
- [11] J. A. Tucker, Y. Theocharis, M. E. Roberts, and P. Barber'a, "From liberation turmoil: Social media and democracy," *Journal of Democracy*, vol. 28(4), pp. 46–59, 2017.
- [12] A. Sternisko, A. Cichocka, and J. J. Van Bavel, "The dark side of socialmovements: Social identity, non-conformity, and the lure of conspiracytheories," *Current opinion in psychology*, vol. 35, pp. 1–6, 2020.

- [13] R. Jaakonm aki, O. Muller, and J. Vom Brocke, "The impact of content, context, and creator on user engagement in social media marketing," *Proceedings of the 50th Hawaii international*.
- [14] O. Candogan and K. Drakopoulos, "Optimal signaling of content accuracy: Engagement vs. misinformation," *Operations Research*, vol. 68,no. 2, pp. 497–515, 2020.
- [15] E. A. Vogels, A. Perrin, and M. Anderson. (2020) Most americans think social media sites censor political viewpoints. [Online].
- [16] S. Vosoughi, D. Roy, and S. Aral, "The spread of true and false newsonline," *Science*, vol. 359, no. 6380, pp. 1146–1151, 2018. A. Mas-Colell, M. D. Whinston, and J. R. Green, *Microeconomic theory*.Oxford University Press, 1995. on user engagement in social media marketing," *Proceedings of the 50th Hawaii international conference on system sciences*, 2017.
- [17] O. Candogan and K. Drakopoulos, "Optimal signaling of content accuracy: Engagement vs. misinformation," *Operations Research*, vol. 68,no. 2, pp. 497–515, 2020.
- [18] E. A. Vogels, A. Perrin, and M. Anderson. (2020) Most americans think social media sites censor political viewpoints. [Online].
- [19] S. Vosoughi, D. Roy, and S. Aral, "The spread of true and false newsonline," *Science*, vol. 359, no. 6380, pp. 1146–1151, 2018.
- [20] A. Mas-Colell, M. D. Whinston, and J. R. Green, *Microeconomic theory*.Oxford University Press, 1995.
- [21] A. Dave and A. Malikopoulos, "The prescription approach to decentralized stochastic control with word-of-mouth communication," *arXive-prints*, p. arXiv:1907.12125, Sep 2019.
- [22] A. Mahajan, N. C. Martins, M. C. Rotkowitz, and S. Y"uksel, "Information structures in optimal decentralized control," in 2012 IEEE 51st IEEE Conference on Decision and Control (CDC). IEEE, 2012, pp.1291–1306.
- [23] A. Nayyar, A. Mahajan, and D. Teneketzis, "Decentralized stochastic control with partia history sharing: A common information approach," *IEEE Transactions on Automatic Control*, vol. 58, no. 7, pp. 1644–1658, 2013.
- [24] A. A. Malikopoulos, C. G. Cassandras, and Y. J. Zhang, "A decentralized energy-optimal control framework for connected automated vehicles at signal-free intersections," *Automatica*, vol. 93, no. April, pp. 244–256, 2018.
- [25] S. Sharma and D. Teneketzis, "Local public good provisioning in networks: A Nash implementation mechanism," *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 11, pp. 2105–2116, 2012.

- [26] A. Sinha and A. Anastasopoulos, "Generalized proportional allocation mechanism design for multi-rate multicast service on the internet," *51st Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pp. 146–153, 2013.
- [27] A. Kakhbod and D. Teneketzis, "An efficient game form for unicast service provisioning," *IEEE Transactions on Automatic Control*, vol. 57, no. 2, pp. 392–404, 2011.
- [28] R. Jain and J. Walrand, "An efficient nash-implementation mechanism for network resource allocation," *Automatica*, vol. 46(8), pp. 1276–1283, 2010.

8.2 GITHUB LINK

LINK: https://github.com/navatej99/social-media-and-misleading-information-in-a-democracy-a-mechanism-design-approach