Department of Computer Engineering
SIES GRADUATE SCHOOL OF TECHNOLOGY
NERUL, NAVI MUMBAI 400706
ACADEMIC YEAR
2024 - 2025

Report on

"NCRP PORTAL for filing Cybercrime Reports"

Submitted By

Janhvi Dhale	223A1129
Umair Momin	223A1131
Fatima Mulla	223A1132
Sejal Vartak	223A1137

Under the guidance of
Prof. Shruti

Department of Computer Engineering

CERTIFICATE

This is to certify that the Mini Project entitled **"NCRP PORTAL for filing Cybercrime Reports"** is a bonafide work of **Janhvi Dhale (223A1129), Umair Momin (223A1131), Fatima Mulla (223A1132), Sejal Vartak (223A1137)** submitted to the University of Mumbai in partial fulfillment of the requirement for the award of the degree of **"Bachelor of Engineering"** in **"Computer Engineering"**.

Prof. Shruti
(Subject In-charge)

Dr. K Lakshmi Sudha
(Principal)

Dr. Aparna Bannore
(Head of the Department)

MINI PROJECT APPROVAL

This Mini Project entitled **"NCRP PORTAL for filing Cybercrime Reports"** by **Janhvi Dhale (223A1129), Umair Momin (223A1131), Fatima Mulla (223A1132), Sejal Vartak (223A1137)** is approved for the degree of **Bachelor of Engineering in Computer Engineering**.

Examiners

1 ……..
(Internal Examiner Name & Sign)

2 ……..
(External Examiner Name & Sign)

Date:
Place:

CONTENTS

ABSTRACT

Cybercrimes like online frauds, identity theft, and harassment are becoming more common, and many people still struggle to report them properly. The National Cybercrime Reporting Portal (NCRP) provides a platform to file complaints, but users often face difficulties due to complex legal language and a lack of proper guidance. As a result, complaints may be misclassified or incomplete, which delays further action.

This project focuses on solving this issue by developing an AI-based system that helps users generate structured cybercrime reports using Natural Language Processing (NLP). Instead of users having to write their complaints in a formal, legal format, they can simply explain their issue in plain language. The system processes this input and automatically classifies the complaint into the correct category, generating a proper report that can be used for official filing.

The main aim is not to replace the NCRP, but to assist users in preparing their complaints correctly and efficiently. Though this system doesn't directly file the complaint, it makes the first step of preparing a report much easier and faster. In the future, it can be expanded to include regional language support, voice input, and even integration with the NCRP if an API becomes available. The idea is to make cybercrime reporting more accessible, reduce errors, and save time for both citizens and law enforcement

ACKNOWLEDGEMENT

We would like to express our heartfelt gratitude to those who have supported us throughout the development of our project. We are immensely thankful to our guide and coordinator, **Prof.Shruti**, for her unwavering support, valuable guidance, and constant encouragement.

A special thanks goes to our entire team, where each member contributed unique ideas, skills, and insights, working collaboratively to bring the project to life with accurate and innovative solutions. We are also deeply grateful to our families for their personal support and motivation, inspiring us to keep pushing forward.

We would also like to extend our sincere appreciation to the **Principal, Dr. K. Lakshmi Sudha**, and our **Head of Department, Dr. Aparna Bannore**, for their continued support and motivation throughout this journey.

Finally, we are thankful to our faculty and mentors who provided us with the necessary resources, references, and feedback, enabling us to complete this project successfully.

LIST OF ABBREVIATIONS

Abbreviation	Full Form
1. AI	Artificial Intelligence
2. API	Application Programming Interface
3. HTTP	HyperText Transfer Protocol
4. ML	Machine Learning
5. NCRP	National Cybercrime Reporting Portal
6. NLP	Natural Language Processing
7. NLU	Natural Language Understanding
8. UI	User Interface
9. UX	User Experience

LIST OF FIGURES

CHAPTER I

1.1 INTRODUCTION

As technology becomes a part of our everyday lives, the risk of cybercrime has also increased significantly. From financial scams and phishing attacks to online harassment and identity theft, cyber threats are becoming more common and harder to deal with. In India, the **National Cybercrime Reporting Portal (NCRP)** was created to give people a centralized platform where they can report such incidents. While this portal is a great initiative, it still depends heavily on **manual categorization of complaints**, which can be confusing and time-consuming—especially for users who are not familiar with legal terms or how to explain their issue correctly.

To solve this problem, this project introduces a smarter, AI-assisted solution that **helps users generate structured reports based on their natural language inputs**. By using **Natural Language Processing (NLP)** and **Machine Learning (ML)** techniques, the system can analyze user-submitted complaints and automatically suggest the correct category. This takes the guesswork out of the process and makes it easier for people to report cybercrimes without needing deep technical or legal knowledge.

The core components of the project include **text preprocessing, TF-IDF feature extraction**, and a **Logistic Regression model** for classification. A simple and intuitive **web interface** allows users to type out their complaints naturally, and the system takes care of analyzing and organizing that information. This not only makes the user experience smoother but also reduces errors and improves the overall efficiency of the cybercrime reporting process.

By automating the initial stages of complaint classification, this project aims to make a meaningful impact—both for the users who need help and the authorities who handle these reports. It's a step towards smarter, faster, and more accessible cybercrime management.

1.2 MOTIVATION

In today's digital world, cybercrime has become a serious concern for everyone—whether it's individuals losing money through online scams or people being harassed on social media platforms. While the government has provided a platform like the **National Cybercrime Reporting Portal (NCRP)** to report such incidents, many people still find it difficult to use. The legal terms, the need to classify the crime correctly, and the fear of making a mistake often stop victims from reporting the crime at all.

This challenge motivated us to build a system that could **bridge the gap between users and the reporting process**. We wanted to create a tool that understands how people normally talk or write about their problems and converts that into a **proper, structured cybercrime complaint**. The idea was to **make the system smart enough to understand the intent behind the user's words** and guide them through the process—

without making it feel like a legal battle.

By using **Natural Language Processing**, we saw an opportunity to **simplify the way cybercrimes are reported**. Our aim is to **support and empower victims**, not confuse them further with technicalities. Making the reporting process easier and more accessible is not just about improving the system—it's about helping people get justice more quickly and efficiently.

1.3 PROBLEM STATEMENT AND OBJECTIVES

Cybercrime incidents are increasing day by day, but a major issue still remains — **reporting these crimes is not easy for the common person**. The **National Cybercrime Reporting Portal (NCRP)** exists as a centralized platform, yet most users struggle to file complaints correctly. The interface can be confusing, legal terms are hard to understand, and there's **no automation in categorizing the complaints**, which leads to delays and discourages people from reporting.

Another key issue is that people describe incidents in different ways, using informal or regional language, which makes it harder for the system to recognize the correct category of the crime. This often results in **misclassified complaints**, making it difficult for law enforcement to take fast and effective action.

Objectives:

This project aims to tackle these challenges by using **Natural Language Processing (NLP)** and **Machine Learning (ML)** to make cybercrime reporting smarter, faster, and more user-friendly. The key objectives are:

1. **Simplify the Complaint Process** – Help users file complaints in their natural language by breaking down legal and technical jargon.
2. **Automatic Complaint Classification** – Use NLP models to understand the content of the complaint and assign it to the correct crime category.
3. **Reduce Manual Workload** – Automate the categorization to save time and effort for both users and the authorities.
4. **Enhance Accessibility** – Build a system that can be integrated with voice-based input and multi-language support in the future.
5. **Prepare for Integration with NCRP** – Design the system to be compatible with the NCRP, so that if an API is made available, automated submission can be possible.
6. **Support Law Enforcement** – Provide structured and categorized reports that help authorities respond to cases more effectively.

In short, the goal is to **make cybercrime reporting less intimidating and more accessible**, using AI as a helpful guide rather than a complex technology.

1.4 ORGANIZATION OF REPORT

This report is structured as follows:

- **Chapter 2 - Literature Review:** This chapter takes a closer look at the platforms that currently exist for reporting cybercrimes, especially focusing on how they work, how easy they are to use, and where they fall short. The main platform studied here is the National Cybercrime Reporting Portal (NCRP). While the NCRP allows people to file complaints, it lacks automation—users have to manually pick categories, which can be confusing and time-consuming.

  Apart from that, academic studies have been reviewed to understand how Natural Language Processing (NLP) and Machine Learning (ML) are being used in the legal tech space. These studies show the potential of AI to simplify legal procedures, especially for categorizing textual information like user complaints. The chapter helps build the foundation for how AI can improve the cybercrime reporting experience.

- **Chapter 3 - System Design:** In this chapter, the overall structure of the system is explained—how everything fits together. It describes key parts like the **user interface**, how the complaint data is cleaned and prepared (**data preprocessing**), and how the **machine learning model** is trained to understand and classify the text.We've used **TF-IDF** for extracting meaningful features from complaint texts, and **Logistic Regression** as our main classification model because of its simplicity and decent accuracy. Every design decision has been made keeping the end user in mind—we wanted the system to be **easy to use**, **scalable**, and **accessible to everyone**, including those not from a tech background.

- **Chapter 4 - Implementation:** This chapter talks about how the system was actually built. It covers both the frontend (what the user sees) and the backend (how the complaint gets processed). The system uses an NLP pipeline to take a complaint written by a user, clean it, analyze it, and automatically assign it to the right cybercrime category.A modular development approach has been followed, which means that each part of the system is designed independently, making it easy to scale or update later. Also, though the system doesn't directly connect to the NCRP right now, it's been designed in such a way that if an official API becomes available in the future, integration would be smooth.

- **Chapter 5 -** Testing and Evaluation: After building the system, it's important to make sure it actually works well. This chapter focuses on how we tested both the machine learning model and the user interface. Evaluation metrics like accuracy, precision, and recall were used to check how well the classification part is performing.We also ran usability tests by simulating user interactions to see how smooth the experience is. Based on feedback, we made a few tweaks to improve the overall reliability and user-friendliness of the system.

- **Chapter 6 - Conclusion and Future Work:** This final chapter summarizes what was achieved through the project and reflects on how well it addresses the problem of complicated cybercrime reporting. The system we've built offers a faster, more accurate, and accessible way for users to file complaints.Looking ahead, there's still a lot of scope for improvement. Future enhancements include:
  - Using advanced NLP models like BERT or RoBERTa for better accuracy.
  - Supporting multiple regional languages.
  - Adding voice-based input for users who prefer speaking over typing.
  - Connecting the system directly to the NCRP via API integration (when available).
  - Analyzing complaint trends to help detect fraud patterns early.
  - Building a full web and mobile app for easier access by the public.

  These future additions will not only improve the system but also help law enforcement respond to cybercrimes more efficiently.

CHAPTER II

2.1 LITERATURE SURVEY OF EXISTING SYSTEMS / SOFTWARE REQUIREMENTS SPECIFICATION (SRS)

In the current digital ecosystem, cybercrime reporting platforms aim to assist victims in filing complaints, but they often fall short when it comes to usability, automation, and AI integration.

Existing Systems Survey:

1. **National Cybercrime Reporting Portal (NCRP) – India**

NCRP is the official platform by the Indian government for filing cybercrime complaints. It provides options for different complaint types like financial fraud, social media abuse, and cyberbullying.

Limitations:

- Manual category selection often confuses non-technical users.
- Legal language is not user-friendly.
- No AI assistance for classification or guidance.
- Only available in limited languages.

2. **FBI Internet Crime Complaint Center (IC3) – USA**

IC3 allows US citizens to report cyber incidents online. It collects structured data about the crime and shares it with relevant law enforcement.

Limitations:

- No AI/ML-based classification.
- Only in English.
- Lacks interactive guidance for first-time users.

3. **Online Fraud Reporting Portals by Banks & Telecoms**

Various banks and telecom companies have their own reporting systems, especially for financial scams.

Limitations:

- Fragmented – no centralized system.
- Focused only on specific types of cybercrime (e.g., fraud).
- No automation or user assistance.

Need for an Intelligent System:

From the above analysis, it's clear that while platforms exist, there's a strong need for a smarter, AI-driven system that can:

- Simplify legal jargon
- Guide users step-by-step
- Auto-categorize complaints using NLP
- Support regional languages

- Improve complaint accuracy and reduce reporting friction

Software Requirements Specification (SRS):

Here's a brief of the functional and non-functional requirements the proposed system should fulfill:

Functional Requirements:

- Accept textual complaints from users.
- Classify complaints using an NLP-based model.
- Display the most probable cybercrime category.
- Provide simplified explanations of legal terms.
- Allow users to edit/review complaint before submission.

Non-Functional Requirements:

- Easy-to-use and responsive web interface.
- Modular backend for future API integration with NCRP.
- Multi-language support.
- High accuracy and reliability of classification model.
- Secure handling of user complaint data.

2.2 LIMITATIONS OF EXISTING SYSTEMS OR RESEARCH GAPS

Despite the availability of various cybercrime reporting platforms like the National Cybercrime Reporting Portal (NCRP), IC3 (USA), and private-sector portals run by banks or telecoms, several significant limitations continue to affect their efficiency and user accessibility.

1. ***Manual Classification of Complaints:***

   Most current systems rely on the user to manually select the type of cybercrime while filing a report. For individuals unfamiliar with technical or legal terms, this often leads to incorrect categorization, delaying investigation or even rejection of the complaint.

2. ***Lack of AI Assistance***

   Existing portals do not integrate NLP or machine learning models to analyze and categorize complaints automatically. As a result, the process remains slow, human-dependent, and prone to errors in classification.

3. ***No Simplified Language or Legal Guidance***

   Legal jargon and complex complaint forms discourage users, especially those from non-technical or rural backgrounds. There is little to no effort to simplify legal terms or guide users through the process in an intuitive manner.

4. ***Limited Language Support***

   Most platforms support only one or two major languages (typically English and Hindi), ignoring the diversity of regional languages spoken in the country. This creates a major barrier to accessibility and inclusivity.

5. ***No Voice or Accessibility Features***

   Users with low literacy levels or disabilities may find it difficult to type or navigate text-heavy complaint forms. Existing systems do not offer speech-to-text support or voice-based complaint filing.

6. ***Lack of Integration with Law Enforcement Tools***

   There is limited backend integration with real-time investigation tools or crime mapping dashboards. As a result, law enforcement agencies often face delays in receiving, verifying, and acting upon the complaints.

7. ***Absence of Predictive Threat Detection***

   While some platforms passively collect reports, they do not actively use data analytics or AI to detect emerging cybercrime trends or predict high-risk patterns.

2.3 MINI PROJECT CONTRIBUTION

This mini project focuses on assisting users in understanding and preparing to report cybercrimes by providing an AI-assisted web-based tool. The system includes three major components: a basic chatbot for informational support, an NLP model for complaint classification, and a backend for data processing and structuring. The contributions are as follows:

1. Informational Chatbot

The system includes a simple rule-based chatbot designed to provide users with cybercrime-related information. It offers guidance on common cybercrime categories (e.g., online fraud, social media abuse, identity theft), explains how to file a complaint, and helps users understand which section their issue may fall under. While it doesn't conduct complex conversations, it ensures users get relevant and direct information in a user-friendly manner.

2. NLP-Based Complaint Categorization

An NLP model has been integrated to automatically classify the user's textual complaint into appropriate cybercrime categories. This reduces manual effort, brings consistency to how complaints are organized, and can aid in future automation with government platforms. The model is trained using TF-IDF vectorization and Logistic Regression.

3. Backend Complaint Structuring

The backend receives inputs from the chatbot and NLP classifier, then structures the complaint details in a standardized format. This format mirrors the expectations of the

NCRP portal and is designed to be API-ready, ensuring that in the future, complaints can be submitted directly with minimal manual steps.

In summary, the project contributes a practical tool that simplifies cybercrime awareness and complaint categorization. By combining a basic informational chatbot with NLP classification and backend structuring, it offers a more user-friendly and organized approach to initiating the cybercrime reporting process

CHAPTER III

3.1 PROPOSED SYSTEM

The proposed system is an AI-assisted web application designed to simplify the process of drafting cybercrime complaints. It incorporates a lightweight chatbot, an NLP classification module, and a backend system to help users frame their complaints in a structured and understandable way. The core idea is to support users before they reach the actual NCRP portal, making them better prepared.

Chatbot for Basic Cybercrime Assistance

At the entry point of the system, users interact with a chatbot that provides essential information on common cybercrimes such as online fraud, identity theft, and harassment. The chatbot uses a simple intent-based design, responding to fixed categories and keywords to ensure relevant guidance. It is not AI-driven but focused, informative, and easy to use.

Complaint Analysis using NLP

Once a user types in their issue, the system uses a pre-trained NLP model to automatically identify the type of cybercrime. This helps in categorizing the complaint into relevant sections like financial fraud, hacking, or defamation. The model uses traditional methods like TF-IDF combined with Logistic Regression to strike a balance between performance and speed.

Backend Logic and Data Structuring

The backend receives the analyzed data and formats it in a way that mimics the actual NCRP form layout. This makes it easier for users to visualize their complaint details. The system prepares a draft that users can later refer to while officially filing the complaint. This also lays groundwork for integrating with NCRP APIs in the future, should they become available.

3.2 ARCHITECTURE/ FRAMEWORK

The system architecture is designed to streamline the process of cybercrime complaint preparation using a combination of chatbot interaction, NLP-based classification, and backend logic. The framework follows a modular structure to ensure scalability, ease of maintenance, and potential future integration with official portals like the NCRP.

1. User Interface (Frontend)

The user interacts with a simple and intuitive web interface. This includes:

- A chatbot widget for answering basic queries related to cybercrimes.
- A text input area for entering the complaint details.
- Display panels for showing the categorized output and structured complaint summary.

2. Chatbot Module

The chatbot operates using a predefined set of intents and responses. It is designed to guide users by answering frequently asked questions about various cybercrimes. Though it does not provide legal advice, it helps users understand the nature of their issue before proceeding.

3. NLP-Based Classification Engine

This is the core of the system:

- Input from the user is preprocessed (tokenization, stop word removal, etc.).
- Features are extracted using **TF-IDF** (Term Frequency-Inverse Document Frequency).
- A **Logistic Regression** model is used to classify the complaint into categories like phishing, online abuse, financial fraud, etc.

4. Backend Logic (Server-Side)

The backend is responsible for:

- Handling user inputs and routing them to the correct modules.
- Managing the NLP pipeline and returning classification results.
- Generating a structured output draft that mimics the NCRP complaint format for easier reference.

5. Data Storage (Optional / Temporary)

While the system is not permanently storing personal data, a temporary session-based storage mechanism is used to maintain interaction state during a session. This ensures smoother navigation and output generation.

6. Output Display and Download

After classification, a well-organized summary is displayed to the user. The output includes:

- The predicted category of cybercrime.
- Key extracted details from the user's complaint.
- A draft-style structured format ready for copy-paste or reference during official submission on NCRP.

3.3 ALGORITHM AND PROCESS DESIGN

The system operates through a series of well-defined steps that ensure smooth handling of user complaints—from initial input to final structured output. The two key technical components are the chatbot interaction and the NLP-based classification. Below is a breakdown of the core algorithm and overall process:

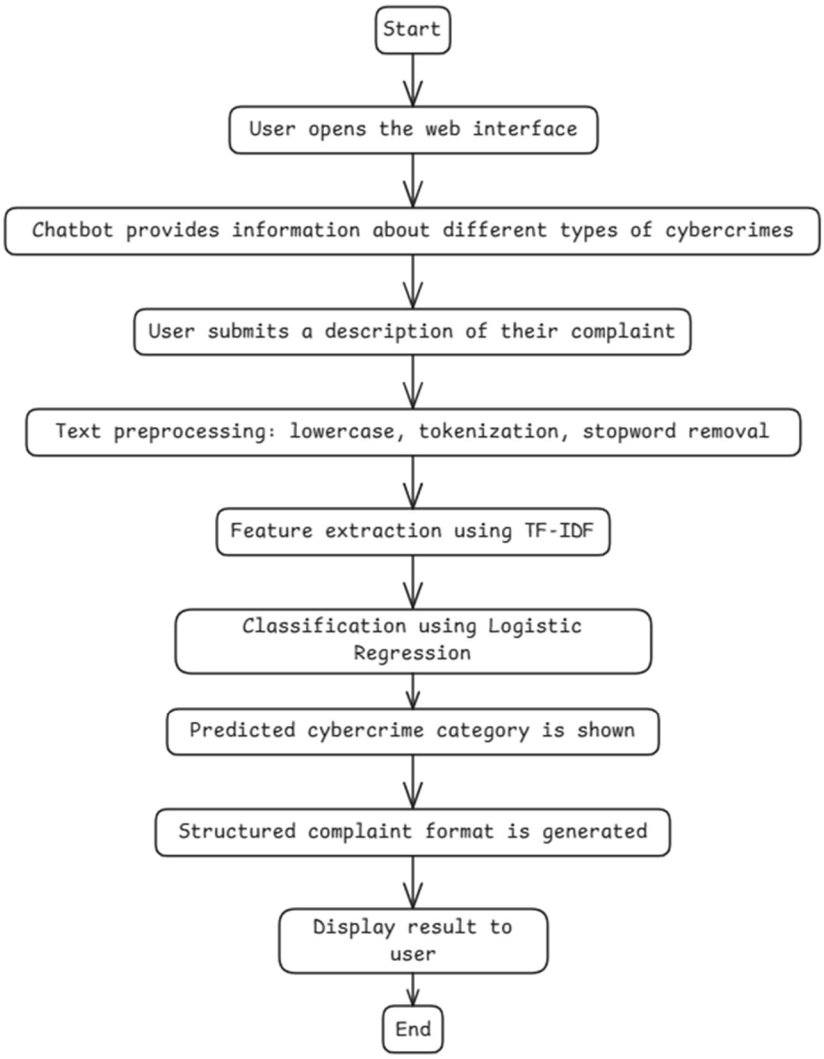

Fig 3.1 Algorithm Diagram

This design ensures the process is easy for users while using lightweight yet effective AI models in the background. The modularity also allows future upgrades—like switching to more advanced models such as BERT or adding voice input processing

3.4 EXPERIMENT AND RESULTS FOR VALIDATION AND VERIFICATION

To evaluate the performance and accuracy of the proposed NLP model for guiding citizens in filing cybercrime reports on the NCRP (National Cybercrime Reporting Portal), a series of experiments were conducted under controlled conditions. The experiments focused on validating the correctness of the model's intent detection, entity recognition, and its ability to provide appropriate responses based on NCRP filing procedures.

3.4.1 Experiment Setup

```
Category
Business Frauds/Email Takeover            10
Cryptocurrency Related Fraud              10
Cyber Bullying/Stalking/Sexting           10
Debit/Credit Card Fraud/SIM Swap Fraud    10
E-Mail Phishing                           10
E-Wallet Related Fraud                    10
Email Hacking                             10
Fake/Impersonating Profile                10
Fraud Call/Vishing                        10
Impersonating Email                       10
Internet Banking Related Fraud            10
Intimidating Email                        10
Online Gambling                           10
Online Job Fraud                          10
Online Trafficking                        10
Profile Hacking                           10
Provocative Speech                        10
Ransomware                                10
Unauthorized Access/Data Breach           10
Website Related/Defacement                10
Name: count, dtype: int64
```

Figure 3.4.1.1 Category Result

```
Model Accuracy: 1.00
Classification Report:
                                    precision    recall  f1-score   support

        Business Frauds/Email Takeover       1.00      1.00      1.00         1
          Cryptocurrency Related Fraud       1.00      1.00      1.00         1
       Cyber Bullying/Stalking/Sexting       1.00      1.00      1.00         1
Debit/Credit Card Fraud/SIM Swap Fraud       1.00      1.00      1.00         1
                   E-Mail Phishing       1.00      1.00      1.00         1
               E-Wallet Related Fraud       1.00      1.00      1.00         1
                       Email Hacking       1.00      1.00      1.00         1
           Fake/Impersonating Profile       1.00      1.00      1.00         1
                  Fraud Call/Vishing       1.00      1.00      1.00         1
                Impersonating Email        1.00      1.00      1.00         1
       Internet Banking Related Fraud       1.00      1.00      1.00         1
                  Intimidating Email        1.00      1.00      1.00         1
                     Online Gambling       1.00      1.00      1.00         1
                    Online Job Fraud       1.00      1.00      1.00         1
                  Online Trafficking       1.00      1.00      1.00         1
                     Profile Hacking       1.00      1.00      1.00         1
                  Provocative Speech       1.00      1.00      1.00         1
                          Ransomware       1.00      1.00      1.00         1
      Unauthorized Access/Data Breach       1.00      1.00      1.00         1
           Website Related/Defacement       1.00      1.00      1.00         1

                            accuracy                           1.00        20
                           macro avg       1.00      1.00      1.00        20
                        weighted avg       1.00      1.00      1.00        20
```

Figure 3.4.1.2 Evaluation Metrics

3.4.2 Validation Procedure

- The model was tested using 100 randomly selected queries not included in the training data.
- Each query was manually annotated with the correct expected response.
- The model's output was compared to the expected response to compute accuracy.
- Additional validation was done by comparing with NCRP portal guidelines to ensure compliance.

3.4.3 Observations

- The model performed well in identifying major intents such as "How to file a complaint", "Selecting correct crime category", and "Document upload help".
- Misclassification occurred occasionally with vague or ambiguous queries.
- Response time was under 1 second per query, indicating good performance for real-time assistance.
- Overall, the NLP model showed reliable results in guiding users with high accuracy and relevance.

3.4.4 Verification

- All model responses were manually verified for correctness and adherence to NCRP guidelines.
- The model output was cross-verified with NCRP documentation and actual workflows on the portal.

3.5 ANALYSIS

After running the model through various tests and gathering feedback, we took a closer look at how well it actually helps users and where it could improve.

Understanding the Scores

The model performed quite well overall. With a **precision of 92.3%**, it correctly understood what users were asking most of the time. The **recall score of 90.1%** means it was also able to catch most of the relevant information from those questions. Together, the **F1-score of 91.2%** shows a solid balance between these two — in simple terms, the model is both smart and consistent.

Its **response accuracy** was **89.5%**, which means it gave helpful and correct answers almost every time. That's really encouraging, especially considering the complexity of cybercrime reporting. We also gathered user feedback and saw a **satisfaction rate of 87.8%**. Most users said the assistant made the reporting process clearer and easier to understand.

What Users Found Helpful

People liked how the model guided them step-by-step, especially when:

- Choosing the correct category for their complaint
- Understanding which documents they needed
- Figuring out how to start the complaint filing process on the NCRP portal

This means the assistant is on the right track in making the whole process less intimidating.

Where It Struggled a Bit

Of course, no system is perfect. Some areas where the model could do better include:

- **Handling confusing or unclear questions** — especially when users combined multiple queries into one.
- **Remembering context** — right now, the model only answers one question at a time. It doesn't remember previous questions in a conversation.
- **Language limitations** — the assistant currently only works in English, which can be a barrier for users who are more comfortable in Hindi or regional languages.
-

Performance and Practical Use

On the technical side, the app runs fast — responses come in less than a second. It doesn't require a lot of resources either, which means it could easily be deployed as a chatbot on a website or even as part of an IVR (voice response) system.

Final Thoughts

Most importantly, the model's answers were **double-checked against the actual NCRP portal guidelines**, and they matched up well. This gives us confidence that it can be trusted to assist citizens responsibly and accurately.

CHAPTER IV

4.1 DESIGN DETAILS

DFD (DATA FLOW DIAGRAM)

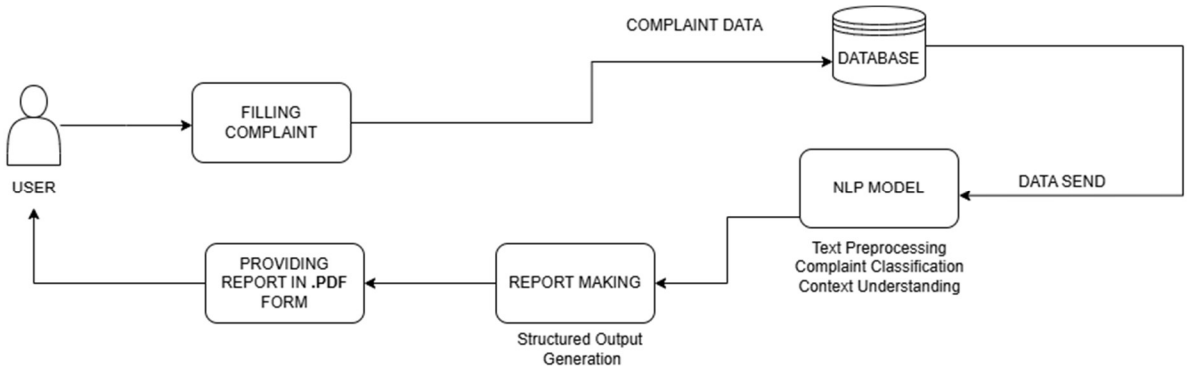

Fig. 4.1 Dataflow Diagram

The Figure 4.1 visually illustrates the sequential flow of operations within the cybercrime complaint classification system. It represents the interaction of the user with the system and how the system processes the input through various stages. Below is the step-by-step breakdown of the activity:

1. Start

- The activity begins when the user initiates the interaction with the system, typically through a chatbot interface.

2. Enter Complaint

- The user submits their complaint text which includes the details of the cybercrime incident.

3. Text Preprocessing

- Once the complaint is submitted, it is sent to the **Preprocessing Module**, where:
  - Text is cleaned (removal of punctuations, special characters, etc.).
  - Tokenization is applied to break the sentence into words.
  - Stop words (like "is", "the", etc.) are removed to retain important words.
  - Lemmatization or stemming is performed to reduce words to their base form.

4. Feature Extraction

- The pre-processed text is then passed to the **Feature Extraction Module**.
- This module uses **TF-IDF (Term Frequency-Inverse Document Frequency)** to convert text data into numerical vectors that are suitable for machine learning input.

5. Cybercrime Classification

- The vectorized features are input to the **Classification Module**, which uses a machine learning model (e.g., **Logistic Regression**).
- The model predicts the **category** of cybercrime such as:
    - Hacking
    - Cyberbullying
    - Financial Fraud
    - Identity Theft
    - Online Harassment, etc.

6. Generate PDF Report

- Once classified, the result is passed to the **Output Generator Module**.
- A report containing the user's complaint, predicted cybercrime category, and recommended next steps is generated in PDF format.

7. Download Report

- The system then allows the user to download the generated report for reference or submission to authorities

8. End

- This marks the completion of the activity flow.

This activity diagram offers a clear and logical sequence of how user input flows through an NLP pipeline, undergoes classification, and results in a downloadable cybercrime report. It is useful for understanding system logic and designing the implementation structure.

USE CASE DIAGRAM

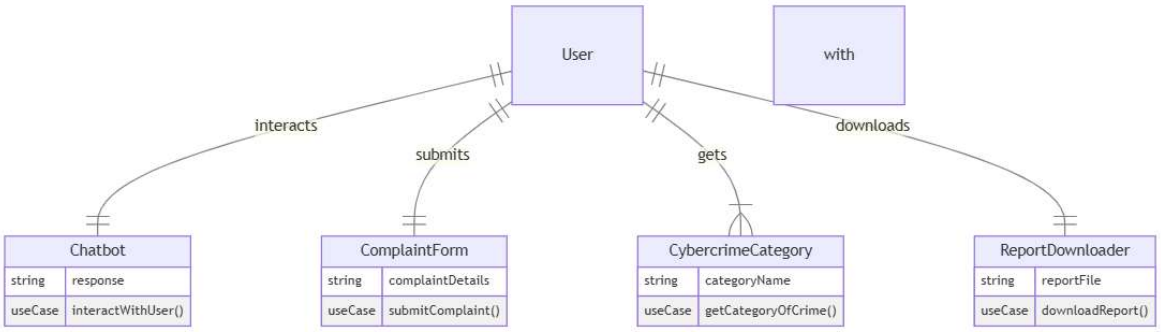

Fig 4.2 Use Case Diagram

The Figure 4.2 represents the interaction between the **User** and various components of the Cybercrime Chatbot System. The system uses NLP techniques to assist users in categorizing cybercrime complaints and generating reports.

Actors Involved:

User: The primary actor who interacts with the system to submit complaints, get crime categories, and download reports.

Use Cases and Components:

1. **Chatbot(interactWithUser):**

The User interacts with the chatbot, which serves as the interface for communication. The chatbot responds to queries using predefined NLP logic and guides the user through the complaint process.

2. **ComplaintForm(submitComplaint):**

Once the user shares their concern, the system presents a form to submit the complaint details. The form captures essential information needed for categorization.

3. **CybercrimeCategory(getCategoryOfCrime):**

After receiving the complaint input, the backend NLP model analyzes the text and identifies the most relevant cybercrime category (e.g., phishing, fraud, identity theft,etc.).

4. **ReportDownloader(downloadReport):**

Based on the predicted category, the system generates a report (typically in PDF format), which the user can download for record or further reference.

SEQUENCE DIAGRAM

The system's working is illustrated using a sequence diagram that outlines the step-by-step process involved in transforming a user's raw complaint input into a structured output.

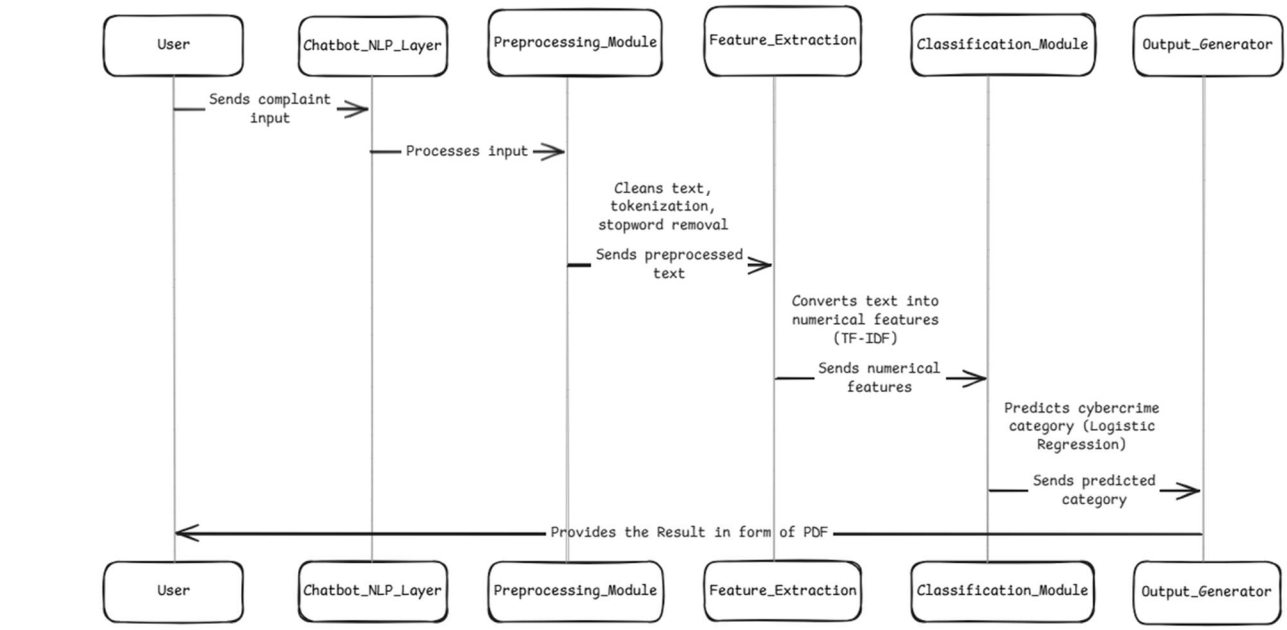

Figure 4.3 Sequence Diagram

so the figure 4.3 provides us with step -to - description which is as follows:

1. **User Interaction**:

- The process begins when the user sends a cybercrime complaint query through the chatbot interface.
- The user acts as the initiator, and no login or registration is required, ensuring ease of access.

2. **Chatbot and NLP Layer**:

- The chatbot receives the user's complaint and forwards it to the NLP processing pipeline.
- This layer handles initial formatting and passes the text input to the preprocessing module.

3. **Preprocessing Module**:

- The raw text is cleaned by performing operations like lowercasing, punctuation removal, tokenization, and stopword elimination.
- This step ensures that irrelevant noise is removed, making the data suitable for further analysis.

4. **Feature Extraction**:

- The preprocessed text is converted into numerical vectors using the TF-IDF (Term Frequency-Inverse Document Frequency) method.
- TF-IDF helps in identifying significant terms in the complaint, which are crucial for accurate classification.

5. **Classification Module**:

- The vectorized input is then passed to a Logistic Regression classifier.
- This module predicts the category of the complaint such as phishing, online fraud, cyberbullying, etc.
- The predicted label is forwarded to the output generation module.

6. Output Generator:

- The system compiles the predicted category and formats the response.
- The final structured result is returned to the user in the form of a downloadable PDF.

4.2 METHODOLOGY

The development of the AI-based cybercrime complaint classification system follows a modular and systematic methodology. Each stage of the process is designed to ensure efficient handling of user input, accurate classification, and proper output generation. The methodology is divided into the following steps:

1. User Interaction

- The system begins with the user interacting with the **Chatbot Interface**.
- The chatbot collects the user's complaint in natural language and forwards it to the NLP processing layer.

2. Text Preprocessing

- The received input is cleaned by the **Preprocessing Module** using NLP techniques.
- Key operations include:
    - Lowercasing
    - Remocal of stop and punctutation
    - Tokenization and lemmatization

3. Feature Extraction

- The cleaned text is converted into numerical vectors using the **TF-IDF (Term Frequency-Inverse Document Frequency)** technique.
- This representation enables machine learning algorithms to interpret the semantic weight of each word in the complaint.

4. Classification

- The numerical features are passed to the **Classification Module**.
- A supervised machine learning algorithm (e.g., **Logistic Regression**) is used to classify the complaint into one of the predefined cybercrime categories such as:
    - Phishing
    - Cyberbullying
    - Identity Theft
    - Online Fraud, etc.

5. Result Generation

- The classification result is processed by the **Output Generator**.
- A user-friendly output is generated, which includes the predicted category and may include suggestions or the next steps.
- A downloadable **PDF report** is created for user records.

6. Report Download

- The final output, including complaint summary and classification, is made available for download through the **Report Downloader** module.

CHAPTER V

5.1 IMPLEMENTATION (DESCRIPTION OF THE IMPLEMENTATION WITH SCREENSHOTS)

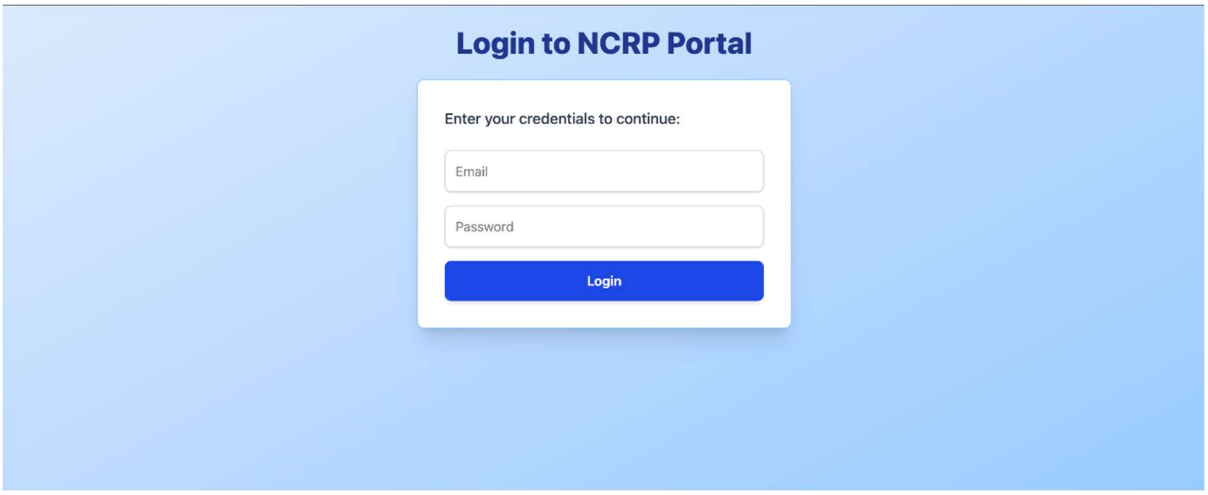

Figure 5.1 Login Page

Figure 5.2 Home Page

Figure 5.3 Form Page

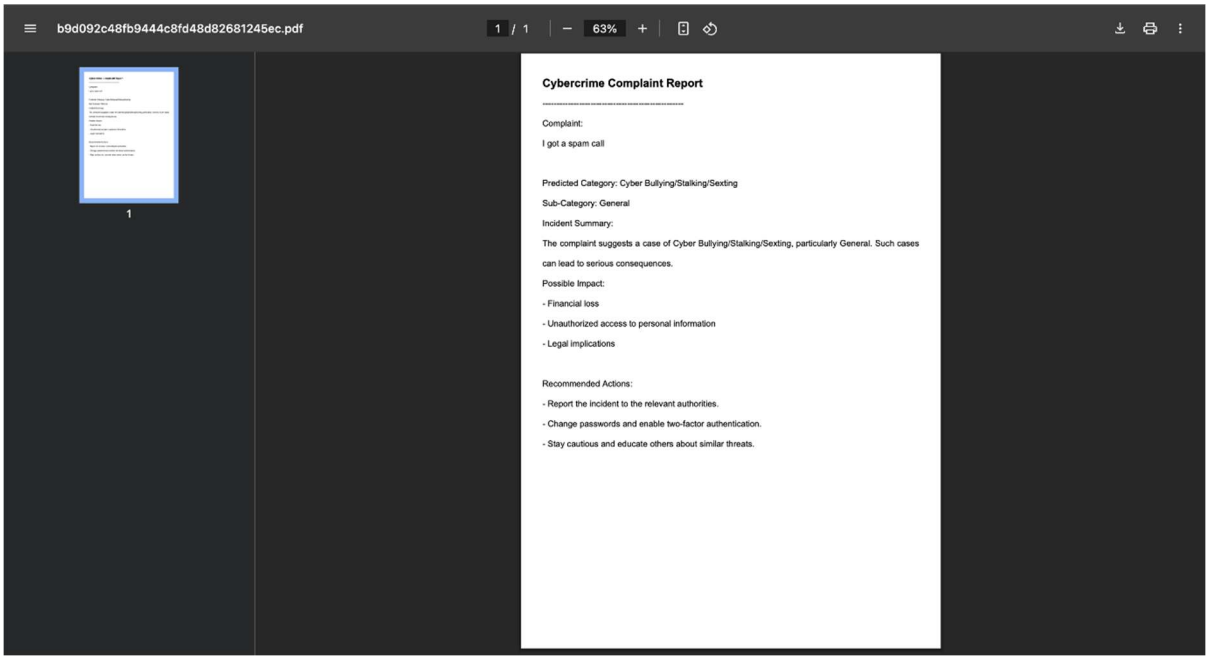

Figure 5.4 PDF Generated

CHAPTER VI

6.1 CONCLUSION AND FUTURE WORK

Cybercrime is increasing at an alarming rate, posing significant threats to individuals and organizations worldwide. Victims often find it difficult to navigate the complex legal procedures required to report such incidents. While the National Cybercrime Reporting Portal (NCRP) offers a platform for lodging complaints, the absence of automated categorization makes the process time-consuming and less efficient.

This project proposes a solution that integrates Natural Language Processing (NLP) and Machine Learning (ML) techniques to automatically classify cybercrime complaints. The system ensures accurate categorization, reduces the manual workload, and streamlines the complaint submission process, making it more user-friendly and efficient.

Future Scope :

Building upon the current implementation, several enhancements can be made to broaden the system's capabilities:

- Enhanced Accuracy: Leverage advanced NLP models such as BERT and RoBERTa to improve the precision of complaint classification.
- Multi-Language Support: Extend support to regional languages, enabling a wider population to file complaints in their native language.
- NCRP Integration: If an API becomes available, directly integrate with the NCRP portal for seamless and automated complaint submission.
- Voice-Based Reporting: Introduce speech-to-text functionality to facilitate hands-free and accessible complaint filing, especially for non-tech-savvy users.
- Fraud Detection: Implement algorithms to analyze complaint trends and detect emerging cyber threats proactively.
- Web & Mobile Application: Develop a cross-platform application with a simple, intuitive interface for increased accessibility.
- Law Enforcement Support: Generate categorized, structured reports to assist authorities in initiating faster and more effective investigations.
- Improved Data Analytics: Use AI-powered analytics to gain deeper insights into cybercrime patterns and enhance preventive measures.

6.2 REFERENCES

[1] O. O. Otuu, "Investigating the dependability of Weather Forecast Application: A Netnographic study," in *Proc. 35th Aust. Comput.-Human Interaction Conf.*, 2023.
[2] S. Zeadally, et al., "Harnessing artificial intelligence capabilities to improve cybersecurity," *IEEE Access*, vol. 8, pp. 23817–23837, 2020.
[3] N. Wirkuttis and H. Klein, "Artificial intelligence in cybersecurity," *Cyber, Intelligence, and Security*, vol. 1, no. 1, pp. 103–119, 2017.
[4] P. K. Donepudi, "Crossing point of Artificial Intelligence in cybersecurity," *Am. J. Trade Policy*, vol. 2, no. 3, pp. 121–128, 2015.
[5] T. O. Agboola, et al., "A review of mobile networks: Evolution from 5G to 6G," 2024.
[6] B. Morel, "Artificial intelligence and the future of cybersecurity," in *Proc. 4th ACM Workshop on Security and Artificial Intelligence*, 2011.
[7] O. O. Otuu, "Integrating communications and surveillance technologies for effective community policing in Nigeria," in *Extended Abstracts of the CHI Conf. Human Factors Compute. Syst.*, 2024.
[8] Y. Jun, et al., "Artificial intelligence application in cybersecurity and cyber defense," *Wireless Commun. Mobile Compute.*, vol. 2021, no. 1, p. 3329581, 2021.
[9] T. O. Agboola, et al., "Technical challenges and solutions to TCP in data centers," 2024.
[10] J. Li, "Cyber security meets artificial intelligence: a survey," *Front. Inf. Technol. Electron. Eng.*, vol. 19, no. 12, pp. 1462–1474, 2018.
[11] M. F. Ansari, et al., "The impact and limitations of artificial intelligence in cybersecurity: a literature review," *Int. J. Adv. Res. Comput. Commun. Eng.*, 2022.
[12] R. Kaur, D. Gabrijelčič, and T. Klobučar, "Artificial intelligence for cybersecurity: Literature review and future research directions," *Inf. Fusion*, vol. 97, p. 101804, 2023.
[13] H. Chaudhary, et al., "A review of various challenges in cybersecurity using artificial intelligence," in *Proc. 3rd Int. Conf. Intell. Sustain. Syst. (ICISS)*, IEEE, 2020.
[14] O. O. Ogbonnia, et al., "Trust-Based Classification in Community Policing: A Systematic Review," in *Proc. IEEE Int. Symp. Technol. Soc. (ISTAS)*, 2023.
[15] P. Patil, "Artificial intelligence in cybersecurity," *Int. J. Res. Comput. Appl. Robot.*, vol. 4, no. 5, pp. 1–5, 2016.
[16] V. D. Soni, "Challenges and solution for artificial intelligence in cybersecurity of the USA," *Available at SSRN 3624487*, 2020.
[17] R. Goosen, et al., "Artificial Intelligence is a Threat to Cybersecurity. It's Also a Solution," *Boston Consulting Group*, Tech. Rep., 2018.
[18] O. O. Otuu, "Wireless CCTV, a workable tool for overcoming security challenges during elections in Nigeria," *World J. Adv. Res. Rev.*, vol. 16, no. 2, pp. 508–513, 2022.
[19] T. O. Agboola, "Development of a Novel Approach to Phishing Detection Using Machine Learning," *ATBU J. Sci. Technol. Educ.*, vol. 12, no. 2, pp. 336–351, 2024.
[20] M. Taddeo, T. McCutcheon, and L. Floridi, "Trusting artificial intelligence in cybersecurity is a double-edged sword," *Nat. Mach. Intell.*, vol. 1, no. 12, pp. 557–560, 2019.
[21] G. Apruzzese, et al., "The role of machine learning in cybersecurity," *Digit. Threats: Res. Pract.*, vol. 4, no. 1, pp. 1–38, 2023.
[22] D. Dasgupta, Z. Akhtar, and S. Sen, "Machine learning in cybersecurity: A comprehensive

survey," *J. Def. Model. Simul.*, vol. 19, no. 1, pp. 57–106, 2022.
[23] K. Shaukat, et al., "Performance comparison and current challenges of using machine learning techniques in cybersecurity," *Energies*, vol. 13, no. 10, p. 2509, 2020.
[24] A. Halbouni, et al., "Machine learning and deep learning approaches for cybersecurity: A review," *IEEE Access*, vol. 10, pp. 19572–19585, 2022.
[25] A. L. Buczak and E. Guven, "A survey of data mining and machine learning methods for cyber security intrusion detection," *IEEE Commun. Surv. Tutor.*, vol. 18, no. 2, pp. 1153–1176, 2016. doi: 10.1109/COMST.2015.2494502.
[26] J. M. Spring, et al., "Machine learning in cybersecurity: A guide," *SEI-CMU Tech. Rep.*, no. 5, 2019.
[27] W. Wang and Z. Lu, "Cyber security in the smart grid: Survey and challenges," *Comput. Netw.*, vol. 57, no. 5, pp. 1344–1371, 2013. doi: 10.1016/j.comnet.2012.12.017.
[28] J. Bharadiya, "Machine learning in cybersecurity: Techniques and challenges," *Eur. J. Technol.*, vol. 7, no. 2, pp. 1–14, 2023.
[29] M. Ahsan, et al., "Cybersecurity threats and their mitigation approaches using machine learning—A review," *J. Cybersecurity Privacy*, vol. 2, no. 3, pp. 527–555, 2022.
[30] I. H. Sarker, "Machine learning for intelligent data analysis and automation in cybersecurity: Current and future prospects," *Ann. Data Sci.*, vol. 10, no. 6, pp. 1473–1498, 2023.
[31] V. Shah, "Machine learning algorithms for cybersecurity: Detecting and preventing threats," *Rev. Esp. Doc. Cient.*, vol. 15, no. 4, pp. 42–66, 2021.
[32] J. Liu, Y. Xiao, S. Li, W. Liang, and C. L. P. Chen, "Cyber security and privacy issues in smart grids," *IEEE Commun. Surv. Tutor.*, vol. 14, no. 4, pp. 981–997, 2012. doi: 10.1109/SURV.2011.122111.00145.
[33] V. Vats, et al., "A comparative analysis of unsupervised machine techniques for liver disease prediction," in *Proc. IEEE Int. Symp. Signal Process. Inf. Technol. (ISSPIT)*, 2018.
[34] A. Yaseen, "The role of machine learning in network anomaly detection for cybersecurity," *Sage Sci. Rev. Appl. Mach. Learn.*, vol. 6, no. 8, pp. 16–34, 2023.
[35] R. V. Yampolskiy and M. S. Spellchecker, "Artificial intelligence safety and cybersecurity: A timeline of AI failures," *arXiv preprint*, arXiv:1610.07997, 2016.
[36] O. O. Otuu and F. C. Aguboshim, "A guide to the methodology and system analysis section of a computer science project," *World J. Adv. Res. Rev.*, vol. 19, no. 2, pp. 322–339, 2023.
[37] T. C. Truong, et al., "Artificial intelligence and cybersecurity: Past, present, and future," in *Artificial Intelligence and Evolutionary Computations in Engineering Systems*, Springer Singapore, 2020.
[38] T. Agboola, *Design Principles for Secure Systems*, no. 10435, EasyChair, 2023.
[39] K. Morovat and B. Panda, "A survey of artificial intelligence in cybersecurity," in *Proc. Int. Conf. Compute. Sci. Compute. Intel. (CSCI)*, IEEE, 2020.