



**St. Thomas' College of Engineering and Technology**

**Department of Computer Science and Engineering**

# **Crime Data Analysis Using Data Mining**

**Prepared by**

<b><u>Student Name</u></b>	<b><u>University Roll</u></b>	<b><u>University Registration No.</u></b>
Rohan Bhownick	12200117030	171220110040 of 2017-2018
Ishika Chakrabarti	12200117045	171220110025 of 2017-2018
Anwesha Chakraborty	12200118011	181220120001 of 2018-2019

Under the guidance of

**Subhankar Malick**

Asst. Professor

# **Project Report**

Submitted in the partial fulfilment of the requirement for the degree of

B.Tech in Computer Science and Engineering

Affiliated to



Maulana Abul Kalam Azad University of Technology, West Bengal

June, 2021

**St. Thomas' College of Engineering and Technology**  
**Department of Computer Science and Engineering**

**CERTIFICATE OF APPROVAL**

This is to certify that the work in preparing the project titled "**Crime Data Analysis Using Data Mining**" has been carried out by **Rohan Bhowmick, Ishika Chakrabarti and Anwesha Chakraborty** under my guidance during the session **2020-2021** and accepted in partial fulfilment for the requirement of the degree of **B. Tech in Computer Science and Engineering**.

---

**Subhankar Malick**

Asst. Professor

Department of Computer Science and Engineering

---

**Mousumi Dutta**

Assoc. Prof, HOD

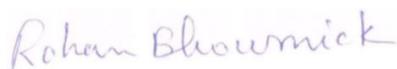
Department of Computer Science and Engineering

## St. Thomas' College of Engineering and Technology

### Department of Computer Science and Engineering

#### Acknowledgement

We are grateful to Almighty God for giving us the strength, knowledge and understanding to complete this project. Our profound gratitude goes to my wonderful mentor, Prof. Subhankar Malick for his invaluable support, patience, time and guidance in helping us to complete the project. Also our gratitude goes to our Head of Department, Dr. Mousumi Dutta who patiently saw us and reviewed our work regularly for correction and improvement to the completion of this work. I also extend gratitude and appreciation to our lecturers, Prof. Ghazala Yasmin, Prof. Samrat Sarkar, and Prof. Amit Paul of Computer Science and Engineering department who have taught us at one point or the other as well as reviewed our progress in this project from time to time. We also wish to acknowledge the great support of parents, siblings, friends and everyone else who have been a source of inspiration, help and guidance towards us during the course of this work. God bless you all as without your help this project would not have been a success.



---

Rohan Bhowmick, 12200117030



---

Ishika Chakrabarti, 12200117045



---

Anwesha Chakraborty, 12200118011

**St. Thomas' College of Engineering and Technology**  
**(Department of Computer Science and Engineering)**

**TABLE OF CONTENTS**

	<u>Content</u>	<u>Page Number</u>
<b>Pre-amble</b>		
1.1	<i>Vision of the Institute</i>	i
1.2	<i>Mission of the Institute</i>	i
1.3	<i>Vision of the Department</i>	i
1.4	<i>Mission of the Department</i>	i
1.5	<i>Course Outcome (CO)</i>	i
1.6	<i>Program Outcome (PO)</i>	i
1.7	<i>Program Educational Objective of Computer Science and Engineering</i>	ii
1.8	<i>Program Specific Outcome (PSO) of Computer Science and Engineering</i>	ii
1.9	<i>PO and PSO mapping with justification</i>	iii
<b>Chapter 1</b>	<b>Introduction</b>	
2.1	Problem Statement	1
2.2	Objective	1
2.3	Brief Discussion on Problem	2
2.4	Gantt Chart and Overall Planning	3
<b>Chapter 2</b>	<b>Literature Survey</b>	4
<b>Chapter 3</b>	<b>Concepts and Problem Analysis</b>	
3.1	Theory and Discussion	5
3.2	Basic Structure of Project	7
3.3	Functional and Non-Functional Requirements	8
<b>Chapter 4</b>	<b>Design and Methodology</b>	
4.1	Data Collection	10
4.2	Classification	11
4.3	Pattern Identification	12
4.4	Prediction	13
4.5	Visualization	13
4.6	System Architecture	13
4.7	Design Diagrams	14
<b>Chapter 5</b>	<b>Implementation and Sample Codes</b>	17
<b>Chapter 7</b>	<b>Future Work and Conclusion</b>	31
		32
	<b>References</b>	

## **Chapter 1: PEAMBLE**

### **1.1 Vision of the Institute:**

To evolve as an industry oriented, research-based Institution for creative solutions in various engineering domains, with an ultimate objective of meeting technological challenges faced by the Nation and the Society.

### **1.2 Mission of the Institute:**

1. To enhance the quality of engineering education and delivery through accessible, comprehensive and research-oriented teaching-learning-assessment processes in the state-of-art environment.
2. To create opportunities for students and faculty members to acquire professional knowledge and develop managerial, entrepreneurial and social attitudes with highly ethical and moral values.
3. To satisfy the ever-changing needs of the nation with respect to evolution and absorption of sustainable and environment friendly technologies for effective creation of knowledge-based society in the global era.

### **1.3 Vision of the Department:**

To continually improve upon the teaching-learning processes and research with a goal to develop quality technical manpower with sound academic and practical experience, who can respond to challenges and changes happening dynamically in Computer Science and Engineering.

### **1.4 Mission of the Department:**

1. To inspire the students to work with latest tools and to make them industry ready.
2. To impart research based technical knowledge.
3. To groom the department as a learning centre to inculcate advanced technologies in Computer Science and Engineering with social and environmental awareness.

### **1.5 Course Outcome(CO):**

1. To **Create and Test** an useful tool in the form of e-classroom for online learning.(**CO5**)
2. To **Apply** modern day software tools in frontend and backend along with suitable type of databases and construct an useful application beneficial to teacher and student community.(**CO1**)
3. To **Analyze** the Software Development Life Cycle and gain experience by the **Designing** and implementation of small scale project like this.(**CO2 & CO4**)

### **1.6 Program Outcome (PO):**

**PO1: Engineering Knowledge-** Apply the knowledge of mathematics, science, engineering fundamentals and engineering specialization to the solution of complex engineering problems.

**PO2: Problem Analysis-** Identify, formulate, review research literature, and analyse complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering science.

**PO3: Design & Development of Solutions-** Design solutions for complex engineering problems and design system components, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.

**PO4: Conduct Investigations of Complex Problems-** Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.

**PO5: Modern Tool Usage-** Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modelling to complex engineering activities with an understanding of the limitations.

**PO6: The Engineer and Society-** Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities relevant to the professional engineering practice.

**PO7: Environment and Sustainability-** Understand the impact of professional engineering solutions in social and environmental context and demonstrate the knowledge of, and need for sustainable development.

**PO8: Ethics-** Apply ethical principles and commit to professional ethics and responsibilities and norm of engineering practice.

**PO9: Individual and Team Work-** Function effectively as an individual and as a member or leader in diverse teams and in multi-disciplinary settings.

**PO10: Communication-** Communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations and give and receive clear instructions.

**PO11: Project Management and Finance-** Demonstrate knowledge and understanding of the engineering and management principles and apply these to one's own work, as a member and leader in a team to manage projects and in multi-disciplinary environments.

**PO12: Life-long Learning-** Recognize the need for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.

## 1.7 Program Educational Objective (PEO) of Computer Science & Engineering

1. Have skills to solve the problems by analysis, design, develop and implementation of algorithms leading to optimal solutions fulfilling the dynamic requirement of industry and society.
2. Have good understanding of Computer Science and Engineering concepts, making them practicing engineers with sound knowledge of logic and design, in Industries.
3. Undertake research in emerging fields in computer science and engineering so as to face the challenges of global challenges in their higher studies and by lifelong learning.

## 1.8 Program Specific Outcome (PSO) of Computer Science & Engineering

**PSO1: Programming skills-** Apply fundamental knowledge and programming aptitude to identify, design and solve real life problems.

**PSO2: Professional skills-** Students shall understand, analyse and develop software solutions to meet the requirements of industry and society.

**PSO3: Competency-** Students will be competent for competitive examinations for employment, higher studies and research.

### **1.9 PO and PSO mapping with justification:**

Proj ect	P O1	P O2	P O3	P O4	P O5	P O6	P O7	P O8	P O9	PO 10	PO 11	PO 12	PS O1	PS O2	PS O3
CS7 94	3	2	2	2	3	3	2	2	3	3	2	3	3	3	3

#### **Justification:**

**PO1:** Applied the knowledge of engineering fundamentals.

**PO2:** Identified and formulated a solution to problems using engineering science.

**PO3:** Analysis and interpretation of big quantity of data in a structured manner.

**PO4:** It may arise during implementation of the Project. We have to conduct some investigations on critical problems regarding the project.

**PO5:** Visual Studio Code is used for creating the apps. The basic environment for this project is VS Code where we use Python programming, Django libraries and PHP for backed and libraries and HTML, CSS, JavaScript with googlechartapi for frontend.

**PO6:** It can help law enforcement agencies deploy resources in a more effective manner, and assist detectives in identifying and apprehending suspects.

**PO7:** Crime data analysis is a law enforcement function that involves systematic analysis for identifying and analysing patterns and trends in crime and disorder.

**PO8:** We shall follow professional ethics and not submit a version of this application made by someone else.

**PO9:** Discussion in group and contribution at an individual level has helped to develop solutions for problems.

**PO10:** Communication has been an effective way for completing the pre-implementation stage. Team discussion during planning was immensely helpful to design this project.

**PO11:** The project has been designed keeping crime patterns in mind. This will be helpful for current system of the Supreme Security Committee.

**PO12:** Data Mining is a hot topic in the market now and implementation of this project will help us in the long run.

**PSO1:** Our programming skills in python will be tested in the implementation stage of this project. The basic fundamentals of system design will also be tested.

**PSO2:** The requirements have been analysed and a prototype has been designed for the same.

**PSO3:** This project will definitely help us in employment, higher studies and research. This will give us an idea on how to work in a project group.

# Chapter 1: INTRODUCTION

## 1.1 Problem Statement

In our country India the current system of the Supreme Security Committee (SSC) is a manual system. We aim to explore in this work the applicability of data mining technique in the efforts of crime prevention with particular emphasis to the dataset we collected from web sites (Kaggle), news sites, blogs, social media, RSS feeds etc. We propose to implement a model that could help us to extract crime patterns. These patterns will be applied to some data mining algorithms such as association rules mining and clustering to classify crime records on the basis of the values of attributes crime. Applying such algorithms will illustrate the overall results of using both algorithms to perform better results rather in association rule mining or in clustering [1]. The rules generated by association rule mining could be easily presented in human language which might be used by SSC officers to help them decided a crime prevention strategy. Also, from this data analysis technic we will come to know:

- How crime data analysis is a law enforcement function that involves systematic analysis for identifying and analysing patterns and trends in crime and disorder?
- How information on patterns can help law enforcement agencies deploy resources in a more effective manner, and assist detectives in identifying and apprehending suspects?
- How crime analysis also plays a role in devising solutions to crime problems, and formulating crime prevention strategies?

## 1.2 Objective

The objective of this project is:

- To identify the nature of crime and the crime prevention process to extracting named entities from narrative reports.
- To explore and choose among the various data mining software that support clustering and association rule mining technique to experiment with crime records.
- To build and train as well as test the performance of the model.
- To interpret and analyse the results of the model that how strong is the model to extract crime data patterns.
- To compare the clustering and association rules data mining techniques and select the one which performs the best results.
- To compare our proposed model with some recent working model.
- Predicting crime before it takes place.
- Predicting hotspots of crime.
- Understanding crime pattern.
- Classify crime based on location.
- Analysis of crime in India.
- Finally, to forward recommendations based on the findings of the study.

### 1.3 Brief Discussion of Project

Day by day the crime rate is increasing considerably. Crime cannot be predicted since it is neither systematic nor random. Also, the modern technologies and hi-tech methods help criminals in achieving their misdeeds. According to Crime Records Bureau crimes like burglary, arson, etc have been decreased while crimes like murder, sex abuse, gang rape, etc have been increased. Even though we cannot predict who all may be the victims of crime but can predict the place that has probability for its occurrence. The predicted results cannot be assured of 100% accuracy but the results shows that our application helps in reducing crime rate to a certain extent by providing security in crime- sensitive areas [2]. So, for building such a powerful crime analytics tool we have to collect crime records and evaluate it. Crime analysis and prevention is a systematic approach for identifying and analysing patterns and trends in crime.

The crime analysis can be performed procedure which is similar to figure (Figure1) which specifies each module which is used for data mining and machine learning algorithm to predict the crime or form group of clusters of criminals according to crime records. The criminals can hold certain properties and their crime characteristics and crime careers may vary from one criminal to another. Such a type of information can be taken as the input dataset [1]. The input dataset is given to a pre-processor which performs the pre-processing based on the requirements. Our system can predict regions which have high probability for crime occurrence and can visualize crime prone areas. With the increasing advent of computerized systems, crime data analysts can help the Law enforcement officers to speed up the process of solving crimes. It will also help to create a somewhat safer, crime-free neighbourhood for the citizens as they will be more aware of different types of crimes occurring and the time in which it is more likely to happen [7].

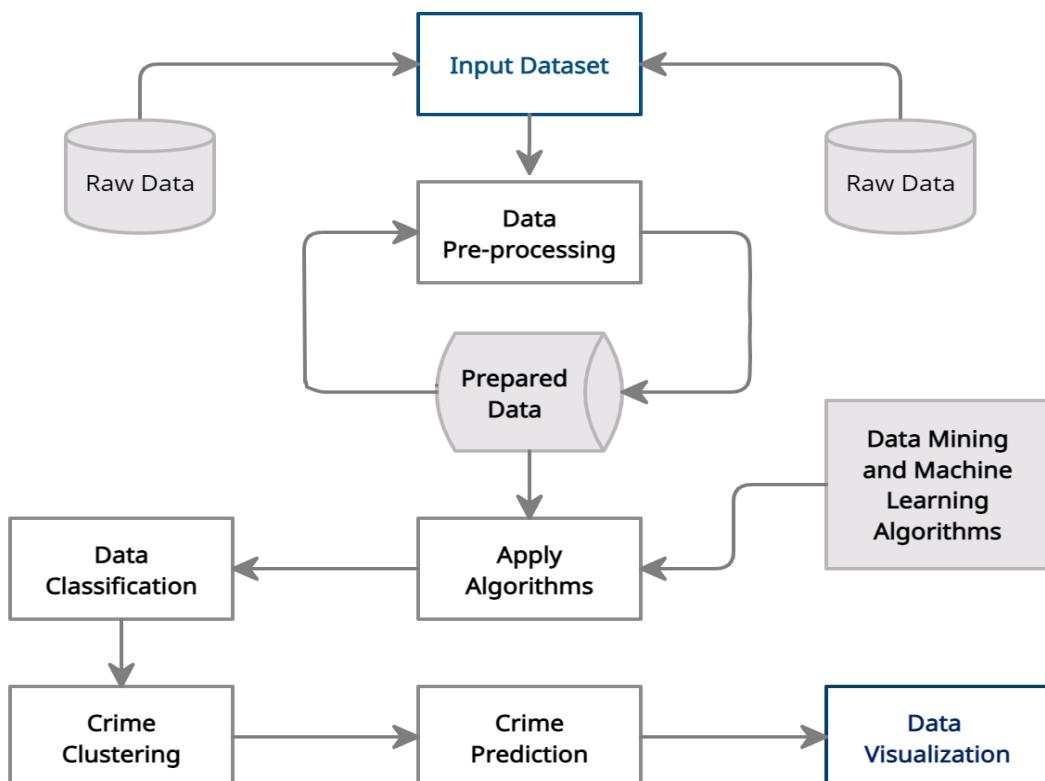


Figure 1: Crime Prediction and Crime clustering based on the input dataset

## 1.4 Gantt Chart and Overall Planning

The overall planning for this project is to analyse the available crime data of a particular region or area using data mining and process the data to predict the crime rate pertaining in that area. For the data analysis we are considering factors like the types of crime committed in an area, the intensity of different crimes and rank them from high to low intensity. We are identifying the areas which are more crime prone than others, also the time of the day, day of the week and months during which it is most likely to occur. We are going to implement this obtained data with the help of a website application which will give the users of the application the crime rate of a particular area within a city and when to avoid that area [8]. Below figure (Figure 2) shows the Gantt Chart for this crime analysis project.

Much of the current work is focused in two major directions:

- Predicting surges and hotspots of crime, and
- Understanding patterns of criminal behaviour that could help in solving criminal investigations.

**Crime Data Analysis using Data Mining**

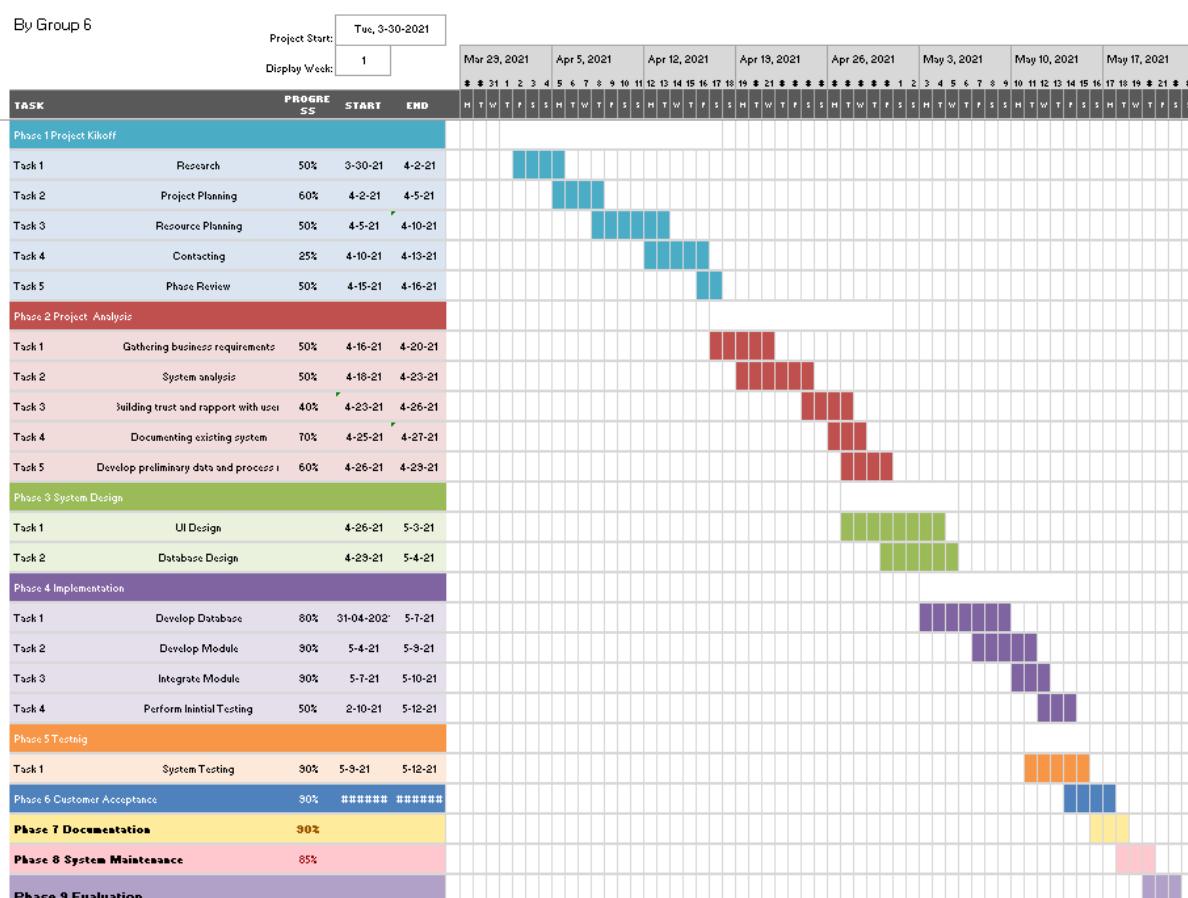


Figure 2: Gantt Chart for Crime Data Analysis Project

## Chapter 2: LITERATURE SURVEY

Historically solving crimes has been the right of the criminal justice and law enforcement specialists. With the increase in the use of the computerized systems to track crimes and trace criminals, computer data analysts have started lending their hands in helping the law enforcement officers and detectives to speed up the process of solving crimes [1]. Criminology is process that is used to identify crime and criminal characteristics. The criminals and the crime occurrence possibility can be assessed with the help of criminology techniques [4].

The criminology aids the police department, the detective agencies and crime branches in identifying the true characteristics of a criminal. The criminology department has been used in the proceedings of crime tracking ever since 1800 [6]. Crimes are a social nuisance and cost our society dearly in several ways. Even, the Indian Government has taken steps to develop applications and software for the use of State and Central Police in relation with the National Crime Records Bureau (NCRB) [7]. Any research that can help in solving crimes faster will pay for itself. About 10% of the criminals commit about 50% of the crimes [15]. People who study criminology will be able to identify the criminals based on the traces, characteristics and methods of crime which can be collected from the crime scene. In the middle of 1990s, data mining came into existence as a strong tool to extract useful information from large datasets and find the relationship between the attributes of the data [11].

Data mining originally came from statistics and machine learning as an interdisciplinary field, but then it was grown a lot that in 2001 it was considered as one of the top 10 leading technologies which will change the world [12]. According to many researchers such as Nath [3], solving crimes is a difficult and time-consuming task that requires human intelligence and experience and data mining is one technique that can help us with crime detection problems. For solving crimes faster, we have to develop a data mining paradigm that performs an interdisciplinary approach between computer science and criminal justice. As said earlier, the Criminology is a process that aims to identify crime characteristics and it is one of the most important fields for applying data mining. By using this, data mining algorithms will be able to produce crime reports and help in the identification of criminals much faster than any human could. Because of this remarkable feature, there is a growing demand for data mining in criminology [13]. Actually, Crime analysis is a process which includes exploring the behaviour of the crimes, detecting crimes and their relationships with criminals [10]. The huge volume of crime and criminal datasets and the complexity of relationships between these kinds of information have made criminology an appropriate field for applying data mining techniques. Identifying crime characteristics is the first step for proceeding with any further analysis.

The quality of data analysis depends greatly on background knowledge of analyst. A criminal can range from civil infractions such as illegal driving to terrorism mass murder such as the 9/11 attacks, therefore it is difficult to model the perfect algorithm to cover all of them [12]. The knowledge that is gained from Data Mining approaches is a very useful and this can help and support, the police. More specifically, we can use classification and clustering based models to help in identification of crime patterns and criminals [7]. The wide range of data mining applications in the criminology has made it an important field of research. Data mining systems have played as a key role in assisting humans in this forensic domain and criminology domain. This makes it one of the most challenging decision-making environments for research.

Data mining is one of the most useful techniques that help entrepreneurs, researchers, and individuals to extract valuable information from huge sets of data. Data mining is also called Knowledge Discovery in Database (KDD) [16]. The knowledge discovery process includes Data cleaning, Data integration, Data selection, Data transformation, Data mining, Pattern evaluation, and Knowledge presentation [3]. The process of extracting information to identify patterns, trends, and useful data that would allow the

business to take the data- driven decision from huge sets of data is called Data Mining [12]. In other words, we can say that Data Mining is the process of investigating hidden patterns of information to various perspectives for categorization into useful data, which is collected and assembled in particular areas such as data warehouses, efficient analysis, data mining algorithm, helping decision making and other data requirement to eventually cost-cutting and generating revenue [14]. Types of data mining are Relational Database, Data Warehouse, Data Repositories, Object-Relational Database, Transactional Database etc [2],[5]. Data science is an inter-disciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from many structural and unstructured data. Data science is related to data mining, machine learning and big data [8],[9].

## **Chapter 3: CONCEPTS AND PROBLEM ANALYSIS**

### **3.1 Theory and Discussion**

#### **3.1.1 Why Analyse Crime?**

Crime usually tends to justify their existence as crime analysis in what is known as law enforcement agency, it is important to articulate some of the reasons it makes sense to analyse crime. Some good reasons are listed below [3]. There may be more other reasons depending on the community culture, geographic effects, and others.

- Analyse crime to inform law enforcers about general and specific crime trends, patterns, and series in an ongoing, timely manner.
- Analyse crime to take advantage abundance of information existing in law enforcement agencies the criminal justice system, and the public domain.
- Analyse crime to maximize the use of limited law enforcement resources.

#### **3.1.2 Types of Crime Analysis:**

##### **A. Tactical Crime Analysis**

The tactical crime analysis involves analysing data to develop information on where, when, and how crimes happen in order to assist officers and investigators in identifying and understanding specific and immediate crime problems [5],[1]. Tactical crime analysis units will work closely with patrol officers and investigators. The goal of tactical analysis is to promote a rapid response to a crime problem happening currently. One of the roles as a tactical crime analysis is to detect current patterns of criminal's activity to predict possible future crime events.

##### **B. Strategic Crime Analysis**

Strategic crime analysis is concerned with long-range problems and planning for long-term projects. Strategic analysis examine long term increases or decreases in crime, known as “crime trends”, A crime trend is the direction of movement of crime and reflects either no change or increases/decreases in crime frequencies within a specific jurisdiction or area. For instance, strategic analysts might study increased car thefts during the winter months when citizens warm up their cars, leaving them unlocked and unattended in various locations [14].

### **C. Administrative Crime Analysis**

Administrative crime analysis focuses on providing summary data, statistics, and general trend information to police managers. This type of analysis involves providing descriptive information about crime to department administrators, command staff, and officers, as well as to other city government personnel and the public. Such reports provide support to administrators as they determine and allocate resources or help citizens to have a better understanding of the community crime and disorder problems [13].

### **D. Investigative Crime Analysis**

Investigative crime analysis involves profiling suspect and victims for investigators based on analysis of available information. It is sometimes called “criminal investigative analysis”. Generally; it focuses on hypothesizing about what type of person is committing a particular crime series [5].

### **E. Intelligence Analysis**

Intelligence analysis focuses on organized crime, terrorism, and supporting specific investigations with information analysis and presentation. Analysts can support investigations by becoming the “processor” of information for officers [5]. In a homicide investigation, the tools of analysis can be used to organize investigative information and display it in the form of time lines and association link charts.

### **F. Operations Analysis**

Operations analysis examines how a law enforcement agency is using its resources. It focuses on such topics as deployment, use of grant funds, redistricting assignments, and budget issues. In many agencies’ crime analysts are asked to assist on special projects for the department that fall into the category of operations analysis [12].

#### **3.1.3 Data Mining and Crime Pattern**

We will look at how to convert crime information into a data-mining problem. In this case it can help the analysts to identify crimes faster and help to make faster decisions. We have seen that in crime terminology a cluster is a group of crimes in a geographical region or a hot spot of crime. Whereas, in data mining terminology a cluster is group of similar data points which can be a possible crime pattern [11]. Thus, appropriate clusters or a subset of the cluster will have a one-to-one correspondence to crime patterns. Thus, clustering algorithms in data mining are equivalent to the task of identifying groups of records that are similar between themselves but different from the rest of the data. In our case some of these clusters will be useful for identifying a crime spree committed by one or same group of suspects. Given this information, the next challenge is to find the variables providing the best clustering. These clusters will then be presented to the detectives to drill down using their domain expertise [10],[13].

It is only within the last few decades that the technology made spatial data mining a practical solution for wide audiences of Law enforcement officials which is affordable and available. Since the availability of criminal data or records is limited, we are collecting crime data from various sources like web sites, news sites, blogs, social media, RSS feeds etc [7]. This huge data is used as a record for creating a crime record database. So, the main challenge in front of us is developing a better, efficient crime pattern detection tool to identify crime patterns effectively. The main challenges we are facing are:

- Increase in crime information that has to be stored annualized.
- Analysis of data is difficult since data is incomplete and inconsistent.
- Accuracy of the program depends on accuracy of the training set.
- Limitation in getting crime data records from Law Enforcement department.

Finding the patterns and trends in crime is a challenging factor. To identify a pattern, crime analysts take a lot of time, scanning through data to find whether a particular crime fits into a known pattern. If it does not fit into an existing pattern then the data must be classified as a new pattern. After detecting a pattern, it can be used to predict, anticipate and prevent crime [14].

### 3.2 Basic Structure of Project

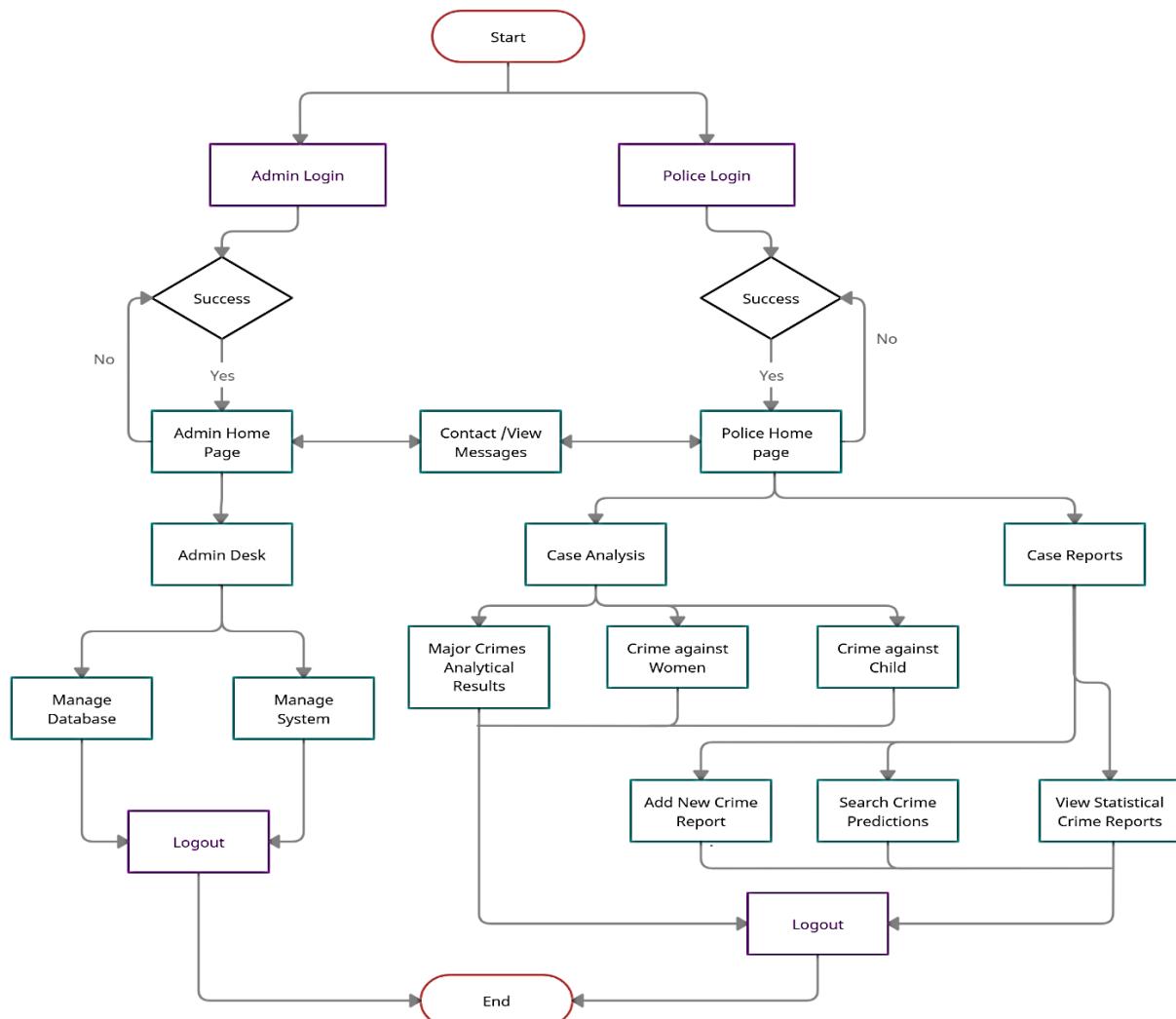


Figure 3: Basic Structure of AIR-Crime Data Analysis System.

AIR Crime Analyzer is a Crime Data Analysis Project using Data Mining Techniques (shown in Figure 3). Crime analyses is one among the important application of knowledge mining. Crime Analyzer is a law enforcement function that involves systematic analysis for identifying and analyzing patterns and trends in crime and disorder. Data processing contains many tasks and techniques including Classification, Association, Clustering, Prediction each of them has its own importance and applications. It can help the analysts to spot crimes faster and help to form faster decisions.

The main objective of crime analysis is to seek out the meaningful information from great deal of knowledge and disseminates this information to officers and investigators within the field to help in their efforts to apprehend criminals and suppress criminal activity. In this project, K-means Clustering is employed for crime data analysis. With the advent of the Big Data era and the availability of fast, efficient algorithms for data analysis, understanding patterns in crime from data is an active and growing field of research.

The inputs to our algorithms are time (hour, day, month, and year), place (latitude and longitude), and class of crime:

- Act 379 - Robbery
- Act 13 - Gambling
- Act 279 - Accident
- Act 323 - Violence
- Act 302 - Murder
- Act 363 - Kidnapping

The output is the class of crime that is likely to have occurred. We try out multiple classification algorithms, such as KNN (K-Nearest Neighbours), Random Forests. We also perform multiple classification tasks, we first try to predict which of 6 classes of crimes are likely to have occurred, and later try to differentiate between violent and non-violent crimes.

### 3.3 Functional and Non-functional Requirements

#### 3.3.1 Functional Requirements

- a. **Login:** Admin and Police both users will have their login details in confidential mail. They can easily login in their specific profiles using those login details.
- b. **Admin:**
  - **Manage Database:** Admin can manage datasets by updating, modifying, deleting data from datasets and also can change any dataset using queries.
  - **Manage System:** Admin can also modify system software, and interfaces by coding.
- c. **Police:**
  - **Add New Crime Report:** Police can add newly filed crime report for crime investigation purposes.
  - **Search Crime Prediction Results:** On basis of reported crimes for last 5 years we can predict crime rate of any location.
  - **View Analytical Results:** Police can view different crime analytical results.
- d. **Contact / View Messages:** If some issues were generated from user side, then user can send messages to admin and admin can reply to it vice-versa.
- e. **Logout:** Both admin and police can logout when ever they by clicking logout button.

### 3.3.2 Interface Requirements

- a. **GUI:** Various interfaces for the project could be
  - Index Page
  - Login Page (for Admin)
  - Login Page (for Police)
  - Admin Home Page
  - Admin Desk
  - Police Home Page
  - Crime Analysis Page
  - Crime Reports page
- b. **Hardware Requirements:** Preferred hardware set for our project is
  - 4GB RAM
  - Pentium IV Processor
  - Input device: Keyboard, Mouse
  - Output device: Monitor, Printer
- c. **Software Requirements**
  - Operating System: Windows 10, Linux, Unix.
  - Browser: Google Chrome, Mozilla Firefox, Microsoft Edge.
- d. **Technology Requirements**
  - Frontend: HTML5, CSS3, JavaScript, Bootstrap
  - Backend: Python (NumPy, SciPy, Matplotlib, Seaborn), Django
  - Database: SQLite, Excel

### 3.3.3 Other Non-functional Requirements

- a. **Performance Requirements:** The software provides good graphical interface for the user and the admin can operate on it. So, the system should have updated GUI. It also needs good broadband internet access (cable, DSL etc.), Updated version of used web browsers and excel.
- b. **Safety Requirements:** Both admin and police login details should not be sheared, maintain sincerity.
- c. **Reliability:** The system provides secured and accurate record of data about reported crimes in India. The main pillar of reliability of the system is the backup of the database, which is continuously maintained and updated to reflect the most recent changes. The overall stability of the system depends on the stability of the application and the databases.
- d. **Availability:** The system should be available at all times, mean the user can access it using their registered accounts anytime and anywhere. In case of a hardware failure or database corruption, a replacement interface will be shown. Also, in case of a hardware failure or database corruption, backups of the database should be retrieved from the server and saved by the administrator. Then the service will be restarted. It means 24 X 7 availability.
- e. **Maintainability:** The system is very easy to maintain and all cached data will be rebuilt during every start-up. There is no recovery of user data if it is lost, Default values of system data will be assigned when necessary.

## Chapter 4: DESIGN AND METHODOLOGY

### 4.1 Methodology

There are 5 steps in doing Crime Analysis, Figure 4 shows these steps:

- 1) Data Collection
- 2) Classification
- 3) Pattern Identification
- 4) Prediction
- 5) Visualization

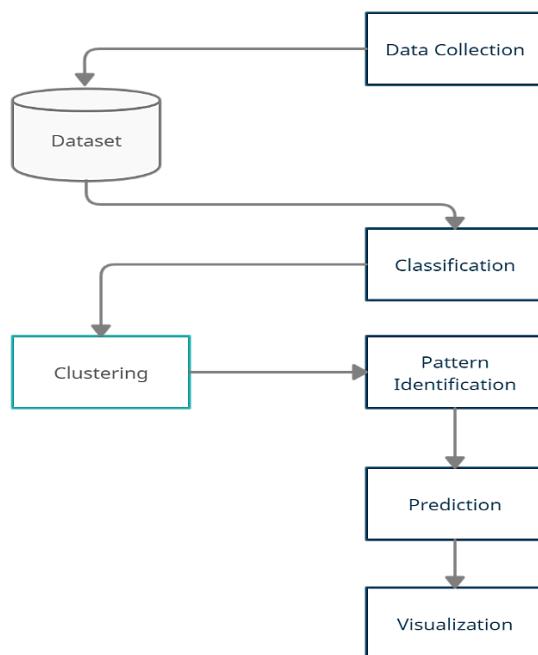


Figure 4: Steps in Crime Data Analysis

#### 4.1.1 Data Collection:

In data collection step we are collecting data from different web sites like news sites, blogs, social media, Kaggle datasets etc. The collected data is stored into a datasheet for further process. Since the collected data is unstructured data, we use MS Excel and Google Sheets. Crime data is an unstructured data since the no of field, content, and size of the document can differ from one document to another the better option is to have a schema less datasheet. Also, the absence of joins reduces the complexity [12]. Other benefits of using an unstructured datasheet are that:

- Large volumes of structured, semi-structured, and unstructured data.
- Object-oriented programming that is easy to use and flexible. The advantage of NoSQL database over SQL database is that it allows insertion of data without a predefined schema. Unlike SQL database it not need to know what we are storing in advance, specify its size etc.

Below figure (Figure 5) shows the crime dataset which we prepared after data collection process.

A	B	C	D	E	F	G	H	I	J	K	L	M
IncidentNum	Category	Descript	DayOfWeek	Date	Time	PdDistrict	Resolution	Address	X	Y	Location	PdId
1	120058272 WEAPON LAWS	POSS OF PROHIBITED WEAPON	Friday	01/29/2016	11:00	SOUTHERN ARREST, BOOKED	800 Block of BRYAN	-122.403	37.77542	37.775420706711	-122.40	1.2E+13
2	120058272 WEAPON LAWS	FIREARM, LOADED, IN VEHICLE, POSSESSION OR USE	Friday	01/29/2016	11:00	SOUTHERN ARREST, BOOKED	800 Block of BRYAN	-122.403	37.77542	37.775420706711	-122.40	1.2E+13
3	141059263 WARRANTS	WARRANT ARREST	Monday	04/25/2016	14:59	BAYVIEW ARREST, BOOKED	KEITH ST / SHAFTER	-122.389	37.72998	37.729980967996	-122.3	1.4E+13
5	160013662 NON-CRIMINAL	LOST PROPERTY	Tuesday	*****	23:50	TENDERLC NONE	JONES ST / OFARREI	-122.413	37.78579	37.785783766888	-122.4	1.6E+13
6	160012749 NON-CRIMINAL	LOST PROPERTY	Friday	*****	00:30	MISSION NONE	16TH ST / MISSION	-122.42	37.76505	37.765050124668	-122.4	1.6E+13
7	160002869 ASSAULT	BATTERY	Friday	*****	21:35	NORTHERN NONE	1700 Block of BUSH	-122.426	37.78802	37.78801855829	-122.42	1.6E+13
8	160003138 OTHER OFFENSES	PAROLE VIOLATION	Saturday	*****	00:04	SOUTHERN ARREST, BOOKED	MARY ST / HOWARD	-122.406	37.78088	37.7808789360214	-122.4	1.6E+13
9	160003259 NON-CRIMINAL	FIRE REPORT	Saturday	*****	01:02	TENDERLC NONE	200 Block of EDDY S	-122.412	37.78398	37.783980559263	-122.4	1.6E+13
10	160003972 WARRANTS	WARRANT ARREST	Saturday	*****	12:21	SOUTHERN ARREST, BOOKED	4TH ST / BERRY	-122.393	37.77579	37.7757876218293	-122.3	1.6E+13
11	160003641 MISSING PERSON	FOUND PERSON	Friday	*****	10:06	BAYVIEW NONE	100 Block of CAMEF	-122.387	37.72097	37.7209669615499	-122.3	1.6E+13
12	160086863 LARCENY/THEFT	ATTEMPTED THEFT FROM LOCKED VEHICLE	Friday	01/29/2016	22:30	TARAVAI NONE	1200 Block of 19TH	-122.477	37.76448	37.7644781576899	-122.4	1.6E+13
13	160004053 NON-CRIMINAL	AIDED CASE, MENTAL DISTURBED	Saturday	*****	13:30	TARAVAI NONE	2200 Block of 21ST	-122.478	37.74574	37.7457389429655	-122.4	1.6E+13
14	160073014 OTHER OFFENSES	RESISTING ARREST	Monday	01/25/2016	23:20	BAYVIEW ARREST, BOOKED	200 Block of HUNTE	-122.377	37.7357	37.7356970275482	-122.3	1.6E+13
15	140776777 ASSAULT	AGGRAVATED ASSAULT WITH A GUN	Thursday	09/15/2016	07:40	INGLESIDE ARREST, BOOKED	SILVER AV / CAMELI	-122.432	37.72927	37.7292705199592	-122.4	1.4E+13
16	160004062 BURGLARY	BURGLARY/STORE UNDER CONSTRUCTION, FORCIBLE ENTRY	Saturday	*****	01:43	CENTRAL NONE	100 Block of SANOS	-122.401	37.79164	37.791642982384	-122.40	1.6E+13
17	160004156 STOLEN PROPERTY	STOLEN CHECKS, POSSESSION	Saturday	*****	11:54	SOUTHERN ARREST, BOOKED	900 Block of MARKE	-122.409	37.78371	37.7837069301545	-122.4	1.6E+13
18	160004241 ROBBERY	ROBBERY, ARMED WITH A KNIFE	Saturday	*****	14:11	MISSION NONE	900 Block of POTRE	-122.407	37.75729	37.7572895904579	-122.4	1.6E+13
19	160004556 ASSAULT	BATTERY WITH SERIOUS INJURIES	Saturday	*****	16:40	MISSION NONE	VALENCIA ST / 26TH	-122.42	37.74891	37.7489063051829	-122.4	1.6E+13
20	160004655 ASSAULT	BATTERY	Saturday	*****	17:05	INGLESIDE NONE	900 Block of GENEV	-122.44	37.71577	37.715765426995	-122.43	1.6E+13
21	160004837 LARCENY/THEFT	PETTY THEFT SHOPLIFTING	Saturday	*****	17:39	SOUTHERN JUVENILE BOOKED	900 Block of MARKE	-122.408	37.78357	37.7835699386914	-122.4	1.6E+13
22	160005421 FRAUD	FRAUDULENT USE OF AUTOMATED TELLER CARD	Saturday	*****	21:00	NORTHERN NONE	GOUGH ST / ROSE S	-122.422	37.77362	37.7736186276456	-122.4	1.6E+13
23	160005750 WARRANTS	WARRANT ARREST	Sunday	*****	04:03	NORTHERN ARREST, BOOKED	FRANKLIN ST / WAS	-122.425	37.79284	37.7928412840447	-122.4	1.6E+13
24	160005841 DRUG/NARCOTIC	POSSESSION OF NARCOTICS/PARAPHERNALIA	Sunday	*****	06:50	MISSION ARREST, BOOKED	23RD ST / FOLSOM	-122.414	37.75451	37.754510986882068	-122.4	1.6E+13
25	160005841 OTHER OFFENSES	PROBATION VIOLATION	Sunday	*****	06:50	MISSION ARREST, BOOKED	23RD ST / FOLSOM	-122.414	37.75451	37.75450986882068	-122.4	1.6E+13
26	160006071 ASSAULT	THREATS AGAINST LIFE	Sunday	*****	09:38	RICHMOND ARREST, BOOKED	FULTON ST / 47TH	-122.508	37.77149	37.771493996416	-122.5	1.6E+13
27	160006598 OTHER OFFENSES	VIOLATION OF STAY AWAY ORDER	Sunday	*****	15:09	TARAVAI ARREST, BOOKED	0 Block of CAMBON	-122.474	37.7183	37.718302204766	-122.47	1.6E+13
28	160006770 OTHER OFFENSES	LOST/STOLEN LICENSE PLATE	Sunday	*****	10:30	MISSION NONE	16TH ST / LANDERS	-122.428	37.76458	37.7645752317615	-122.4	1.6E+13
29	160006789 BURGLARY	BURGLARY, VEHICLE (ARREST MADE)	Sunday	*****	15:45	NORTHERN ARREST, BOOKED	0 Block of HELMICK	-122.419	37.78744	37.7874478309112	-122.4	1.6E+13
30	160006992 OTHER OFFENSES	TRAFFIC VIOLATION ARREST	Sunday	*****	17:02	MISSION ARREST, BOOKED	26TH ST / TRENT AV	-122.413	37.74937	37.74937884539	-122.4	1.6E+13

Figure 5: Collecting data from different datasets

#### 4.1.2 Classification:

For classification we are using an algorithm called Naïve Bayes which is a supervised learning method as well as a statistical method for classification. Naive Bayes classifier is a probabilistic classifier which when given an input gives a probability distribution of set of all classes rather than providing a single output [15]. The algorithm classifies a news article into a crime type to which it fits the best. From figure 6, what we get is “What is the probability that a crime documents belongs to a given class?”.

```

OUTPUT TERMINAL DEBUG CONSOLE PROBLEMS
[Running] python -u "f:\A FINAL YEAR PROJECT\main.py"
(150500, 13)
| IncidntNum ... PdId
0 120058272 ... 12005827212120
1 120058272 ... 12005827212168
2 141059263 ... 14105926363010
3 160013662 ... 16001366271000
4 160002740 ... 16000274071000

[5 rows x 13 columns]
| | IncidntNum X Y PdId
count 1.505000e+05 150500.000000 150500.000000 1.505000e+05
mean 1.616440e+08 -122.423599 37.768921 1.616440e+13
std 5.535976e+06 0.026210 0.023637 5.535976e+11
min 1.135121e+07 -122.513642 37.707922 1.135121e+12
25% 1.603283e+08 -122.434036 37.756486 1.603283e+13
50% 1.606541e+08 -122.416903 37.775421 1.606541e+13
75% 1.609764e+08 -122.406605 37.785063 1.609764e+13
max 9.910090e+08 -122.365565 37.819975 9.910090e+13
IncidntNum 0
Category 0
Descript 0
DayOfWeek 0
Date 0
Time 0
PdDistrict 1
Resolution 0
Address 0
X 0
Y 0
Location 0
PdId 0
dtype: int64
False

```

Figure 6: Classifying datasets.

The advantage of using Naive Bayes Classifier is that it is simple, and converges quicker than logistic regression. Compared to other algorithms like SVM (Support Vector Machine) which takes lot of memory the easiness for implementation and high performance makes it different from other algorithms. Also, in case of SVM as size of training set increases the speed of execution decreases [16]. Using Naive Bayes algorithm, we create a model by training crime data related to vandalism, murder, robbery, burglary, sex abuse, gang rape, arson, armed robbery, highway robbery, snatching etc. By training means we have to teach them on particular inputs such that we can test them for unknown inputs. For testing the accuracy of the model, we apply test data. Unlike SVM as the size of training data increases accuracy of test set also increases. Another advantage of Naïve Bayes is that it works well for small amount of training to calculate the classification parameters. Also, it fixes the Zero-frequency problem i.e. while estimating probability sometimes while checking a probability  $P(A) * P(B/D) * P(C/D) * P(E/D)$  where  $P(C/D) = 0$ . So, the estimated probability results always give zero which leads to uncertainty in results [2]. To avoid this condition, we add +1 to the count of every zero value classes to achieve uniform distribution. Test results shows that Naive Bayes shows more than 90% accuracy since it categorises each word as tokens and removing frequent words like “the”, “and”, “of” etc which improves accuracy [1]. Below figure 7 shows a sample pseudo code for data classification.

```
# check the shape of the data
print(data.shape)

#To see the first 5 rows in the data set
print(data.head())

#To Describe the data set
print(data.describe())

#To check if there are any null values
print(data.isnull().sum())

#Filling the missing value in PdDistrict using the mode values
data['PdDistrict'].fillna(data['PdDistrict'].mode()[0], inplace = True)
print(data.isnull().any().any())
```

Figure 7: Pseudo Code for Data Classification

#### 4.1.3 Pattern Identification:

Third phase is the pattern identification phase where we have to identify trends and patterns in crime. For finding crime pattern that occurs frequently we are using Apriori algorithm. Apriori can be used to determine association rules which highlight general trends in the database [3],[5]. The result of this phase is the crime pattern for a particular place. Here corresponding to each location, we take the attributes of that place like VIP presence, weather attributes, area sensitivity, notable event, presence of criminal groups etc. After getting a general crime pattern for a place, when a new case arrives and if it follows the same crime pattern then we can say that the area has a chance for crime occurrence [6],[2]. Information regarding patterns helps police officials to facilitate resources in an effective manner. They can avoid crime occurrence by providing security/ patrolling in crime prone areas, fixing burglar alarms / CCTV etc.

#### 4.1.4 Prediction:

For prediction we are using the decision tree concept. A decision tree is similar to a graph in which internal node represents test on an attribute, and each branch represents outcome of a test. The main advantage of using decision tree is that it is simple to understand and interpret [8]. The other advantages include its robust nature and also it works well with large data sets. This feature helps the algorithms to make better decisions about variables [13].

#### 4.1.5 Visualization:

The crime prone areas can be graphically represented using a statistical representation (bar chart, pi chart) which indicates level of activity, usually darker colours to indicate low activity and brighter colours to indicate high activity. Some statistically analysed figures are shown below [9].

## 4.2 System Architecture

The system architectural design is the design process for identifying the subsystems making up the system and framework for subsystem control and communication. The goal of the architectural design is to establish the overall structure of software system which is shown in figure 8.

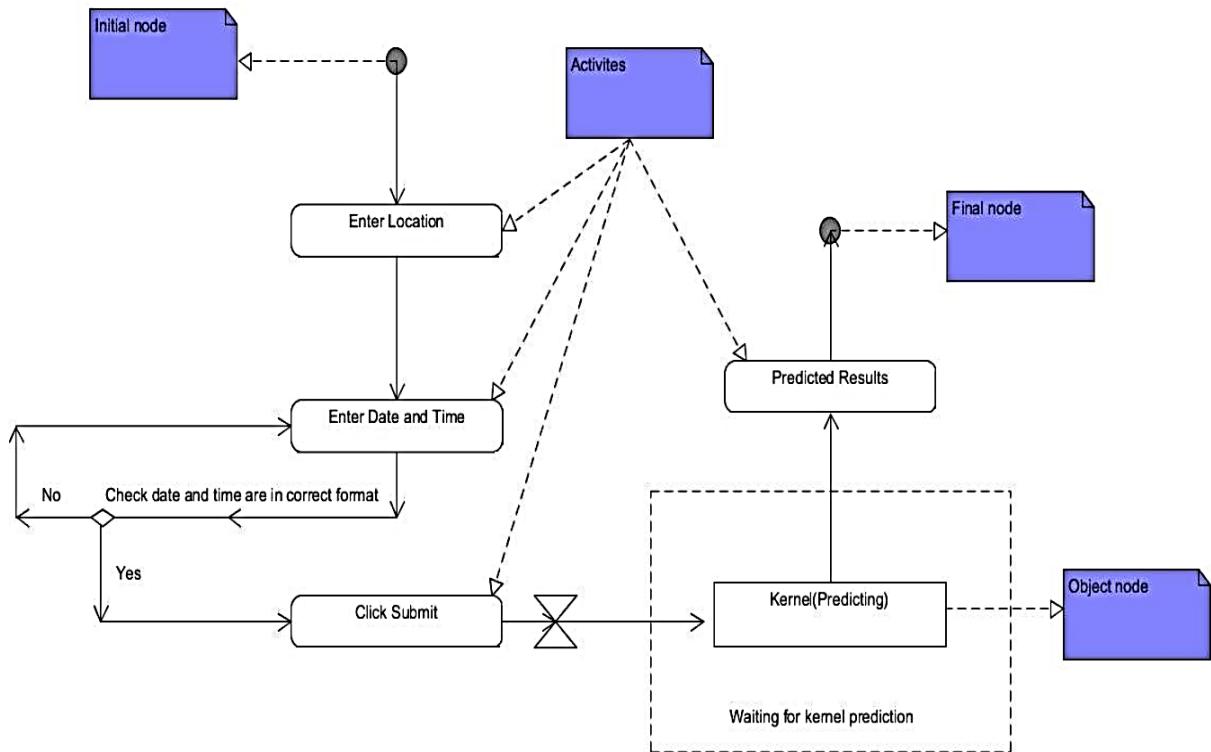


Figure 8: System Architecture of AIR-Crime Analyzer

## 4.3 Design Diagrams

### 4.3.1 Use case diagram

Use case diagram represent the overall scenario of the system. A scenario is nothing but a sequence of steps describing an interaction between a user and a system. Thus, use case is a set of scenarios tied together by some goal. The use case diagram (Figure 9) is drawn for exposing the functionalities of the system.

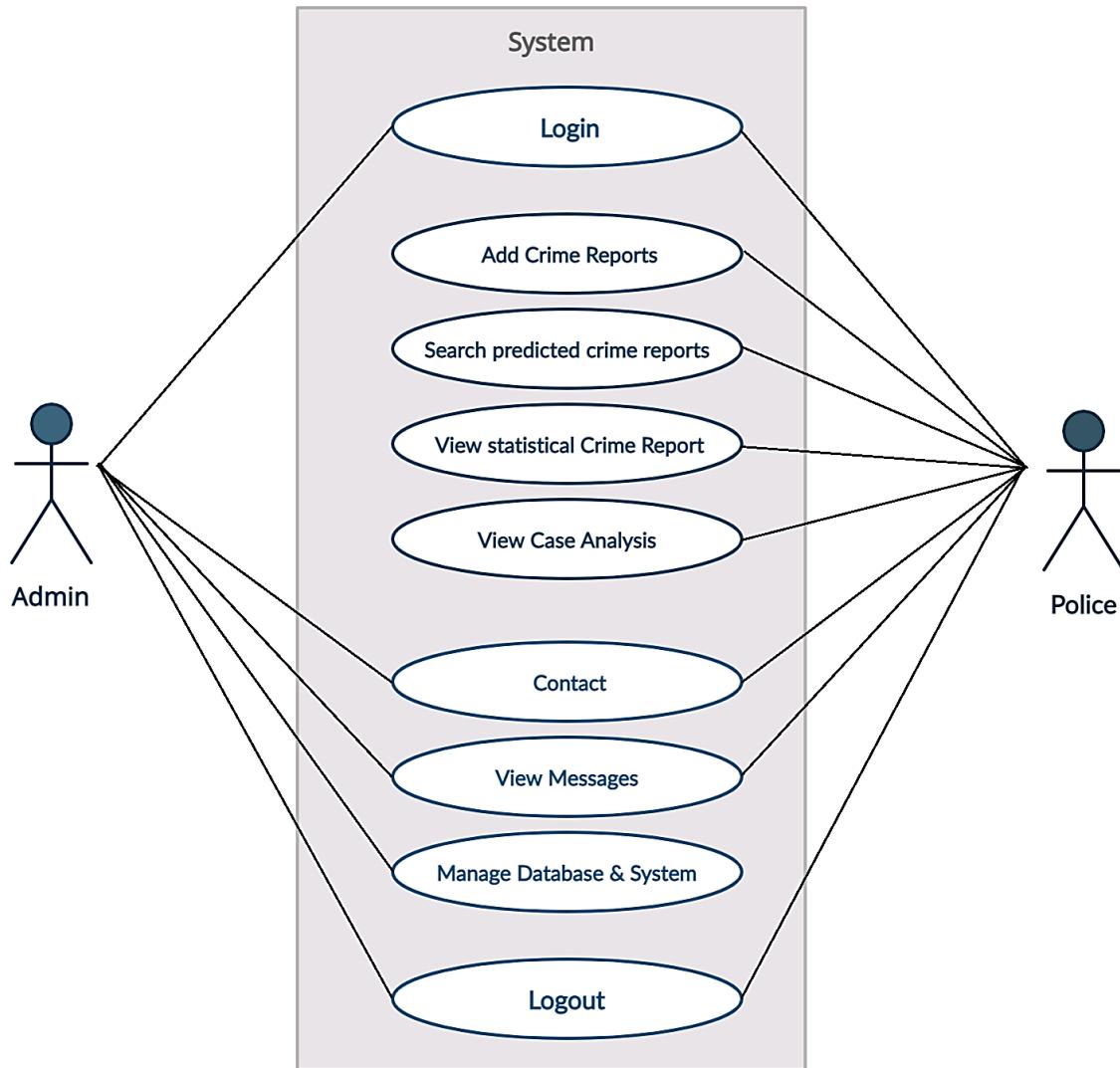


Figure 9: Use-case Diagram of AIR-Crime Analyzer

### 4.3.2 Activity diagram

The activity diagram is a graphical representation for representing the flow of interaction within specific scenarios. It is similar to a flowchart in which various activities that can be performed in the system are represented. Figure 10, represents the activity of the admin and Figure 11, represents the activity of police.

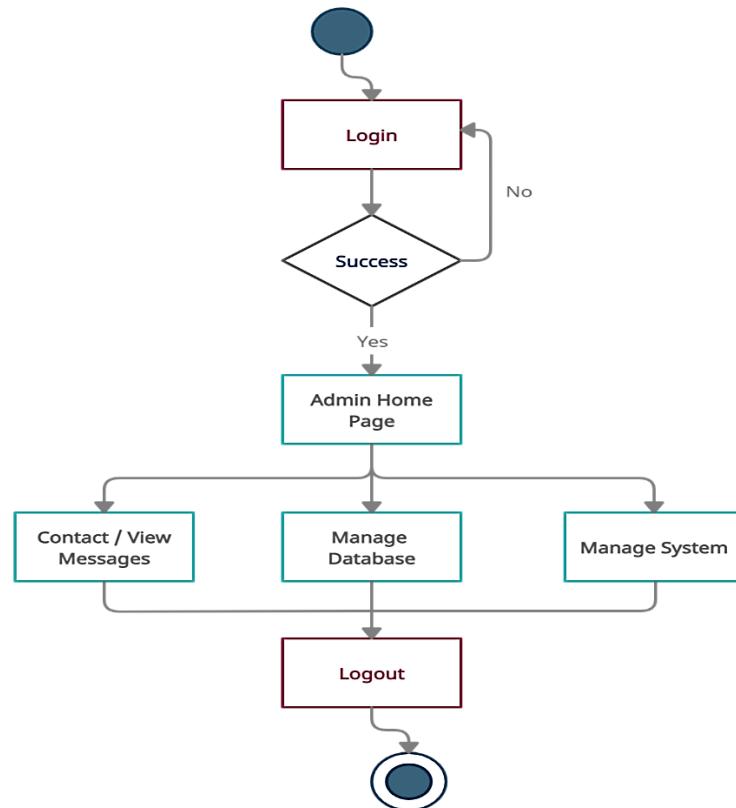


Figure 10: Admin Activity Diagram

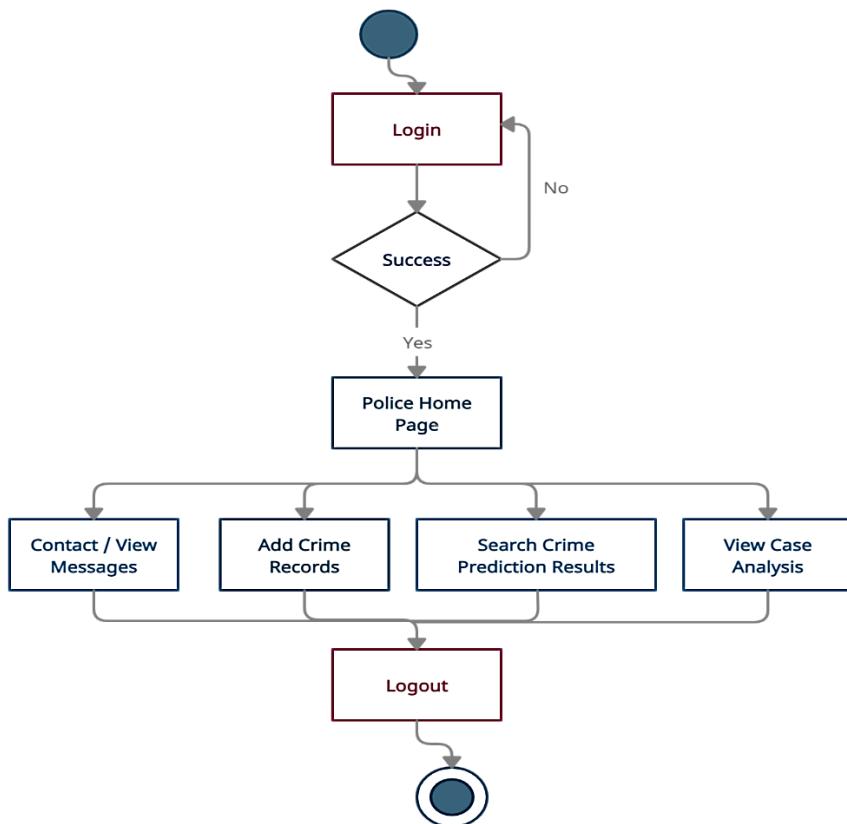


Figure 11: Police Activity Diagram

### 4.3.3 Sequence Diagram

In the sequence diagram how, the object interacts with the other object is shown. There is sequence of events that are represented by a sequence diagram. It is a time-oriented view of the interaction between objects to accomplish a behavioural goal of the system. Figure 12, represents the job sequence of police and Figure 13, represents the job sequence of the admin.

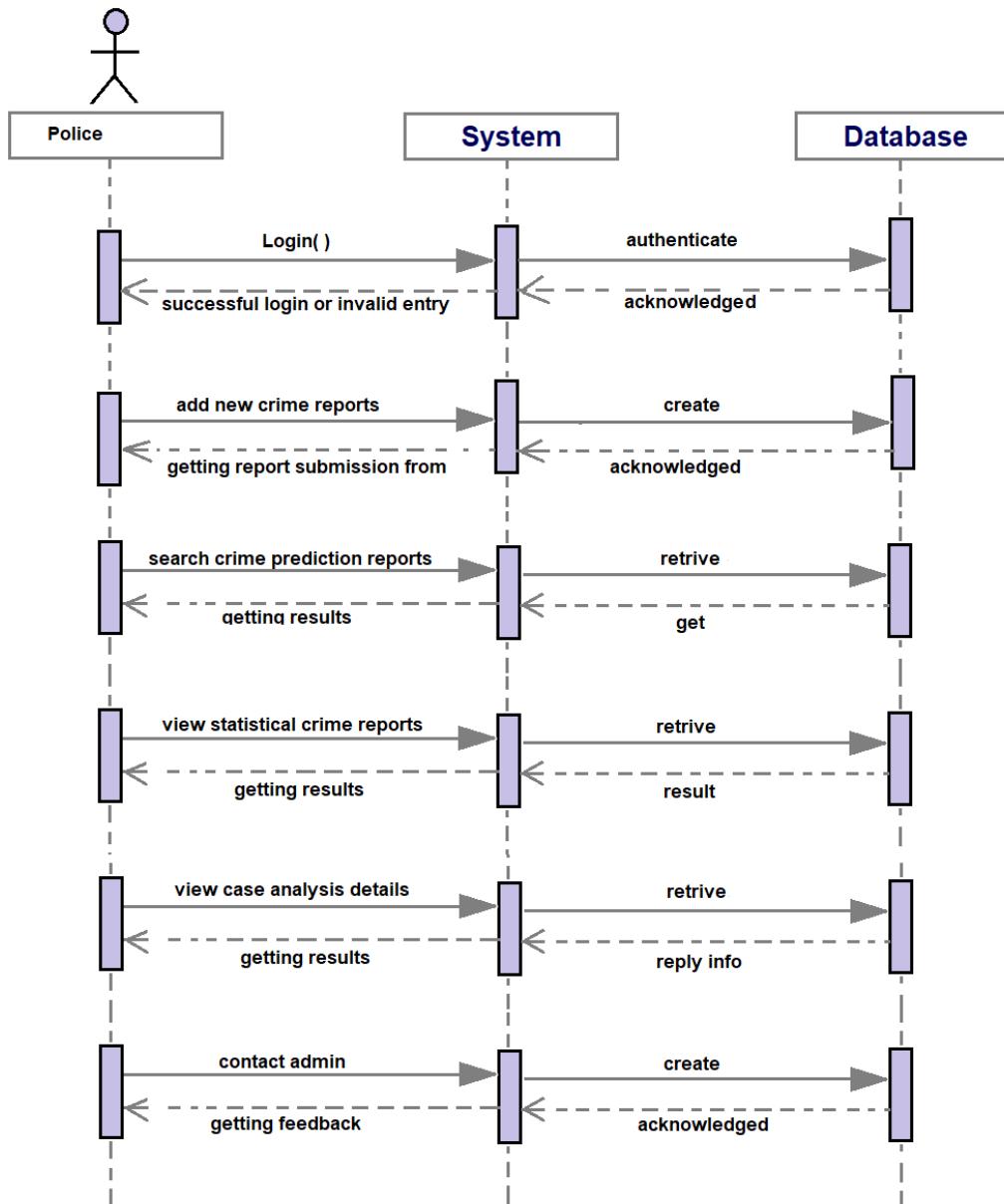


Figure 12: Sequence diagram for police user.

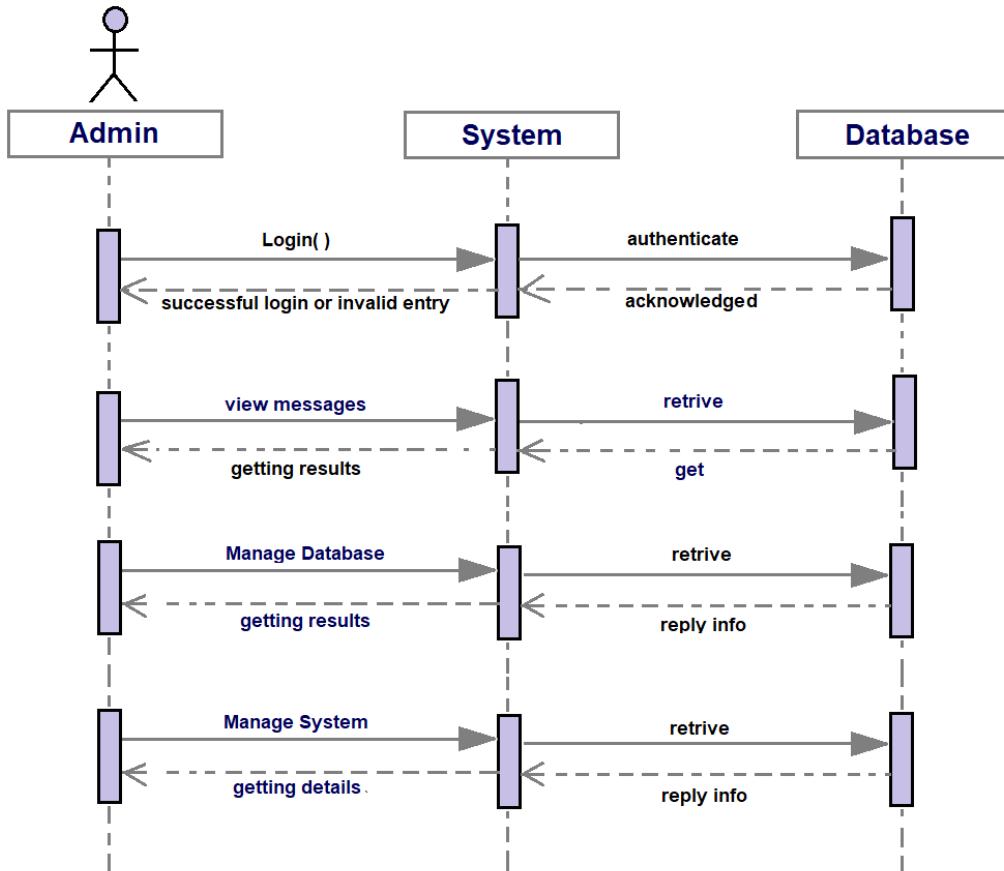


Figure 13: Sequence diagram for admin.

## Chapter 5: IMPLEMENTATION

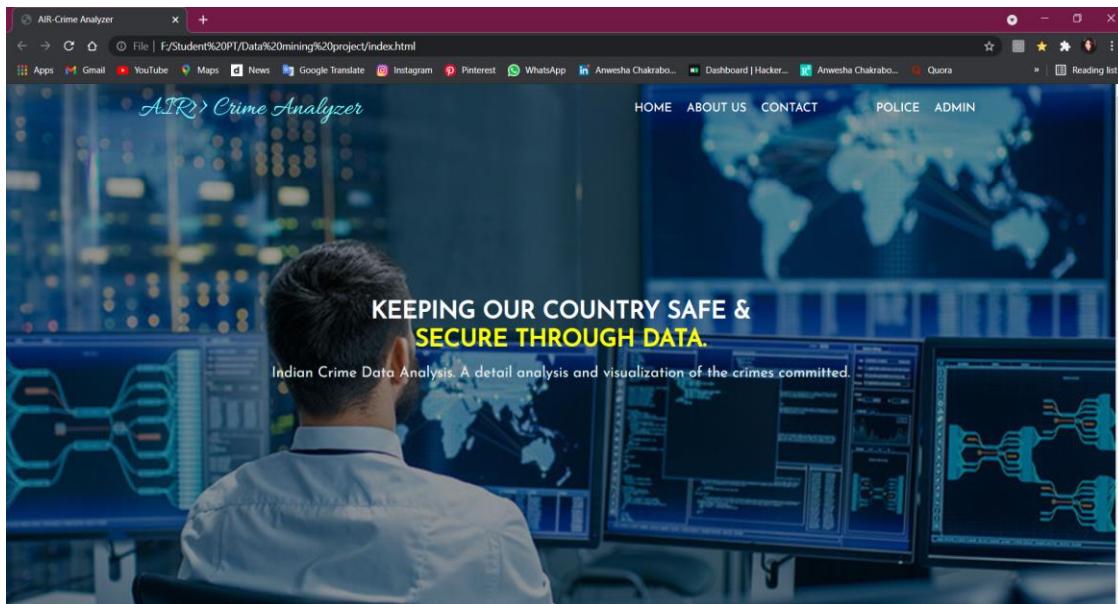
The implementation of the project is done with the help of python language. To be particular, for the purpose of machine learning Anaconda is being used. Anaconda is one of several Python distributions. Anaconda is a new distribution of the Python. It was formerly known as Continuum Analytics. Anaconda has more than 100 new packages. Anaconda is used for scientific computing, data science, statistical analysis, and machine learning. On Python technology, we found out Anaconda to be easier.

Since it helps with the following problems:

- Installing Python on multiple platforms.
- Separating out different environments.
- Dealing with not having correct privileges.
- Getting up and running with specific packages and libraries.

This data was scraped from the publicly available data from Indore police website which had been made by people in police station of different areas. Implementation of the idea started from the Indore city itself so as to limit an area for the prediction and making it less complex. The data was sorted and converted into a new format of timestamp, longitude, latitude, which was the input that machine would be taking so as to predict the crime rate in particular location or city.

## 5.1 Prototype Model



### *ABOUT US*

AIR Crime Analyzer is a Crime Data Analysis Project using Data Mining Techniques. Crime analyses is one among the important application of knowledge mining. Crime Analyzer is a law enforcement function that involves systematic analysis for identifying and analyzing patterns and trends in crime and disorder. Data processing contains many tasks and techniques including Classification, Association, Clustering, Prediction each of them has its own importance and applications. It can help the analysts to spot crimes faster and help to form faster decisions. The main objective of crime analysis is to seek out the meaningful information from great deal of knowledge and disseminates this information to officers and investigators within the field to help in their efforts to apprehend criminals and suppress criminal activity. In this project, Kmeans Clustering is employed for crime data analysis.

### *OUR SERVICES*



#### Data Mining

Officers can view all the latest case Updates and add any available current cases to keep updating the registry using Data Mining Techniques.



#### Data Analysis

Officers can make use of Machine learning algorithm for data analysis to make accurate predictions that can very much help them to solve cases.



#### Data Visualizations

Officers can view interactive plots and visualizations to help them understand the crimes and take actions accordingly.

### *CONTACT US*

Email
Subject
Message
Submit

Figure 14: AIR-Crime Analyzer Home Page

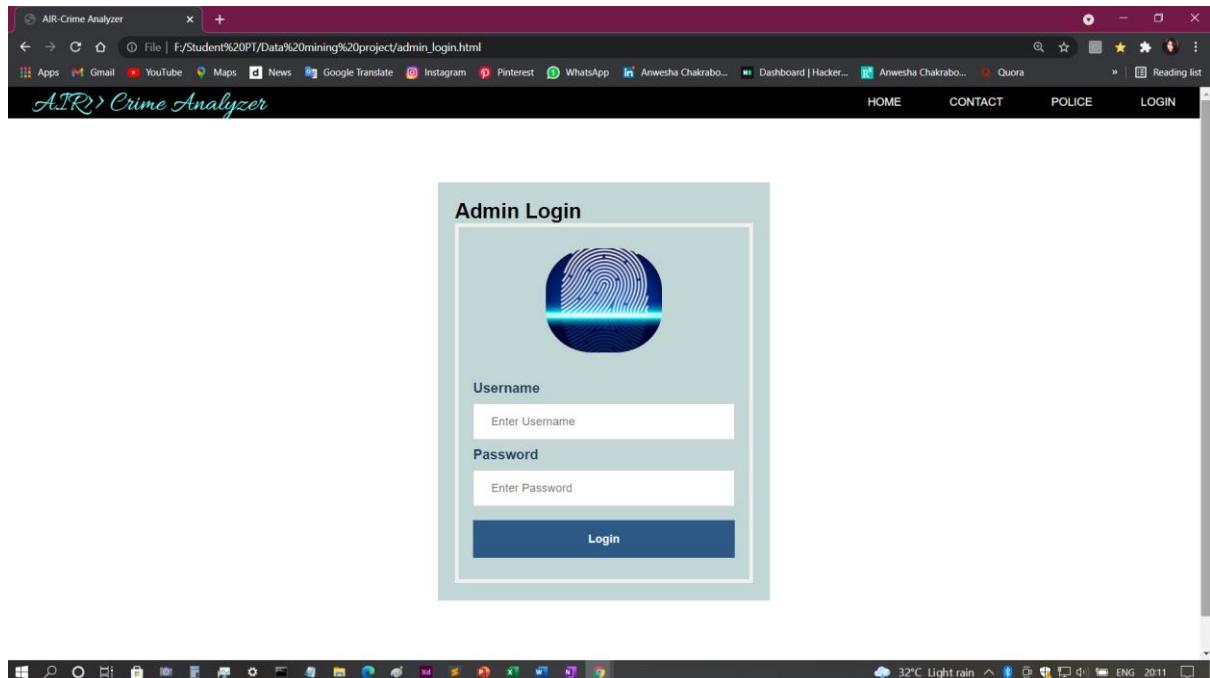


Figure 15: Admin Login Page

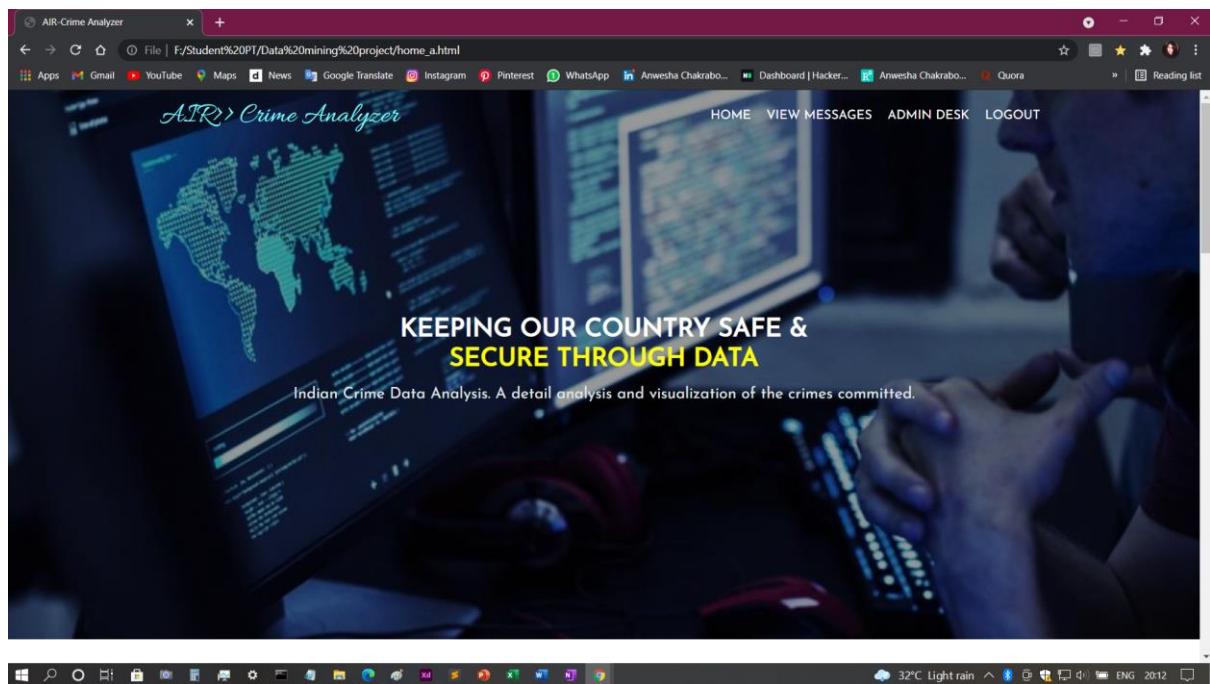


Figure 16: Admin Home Page

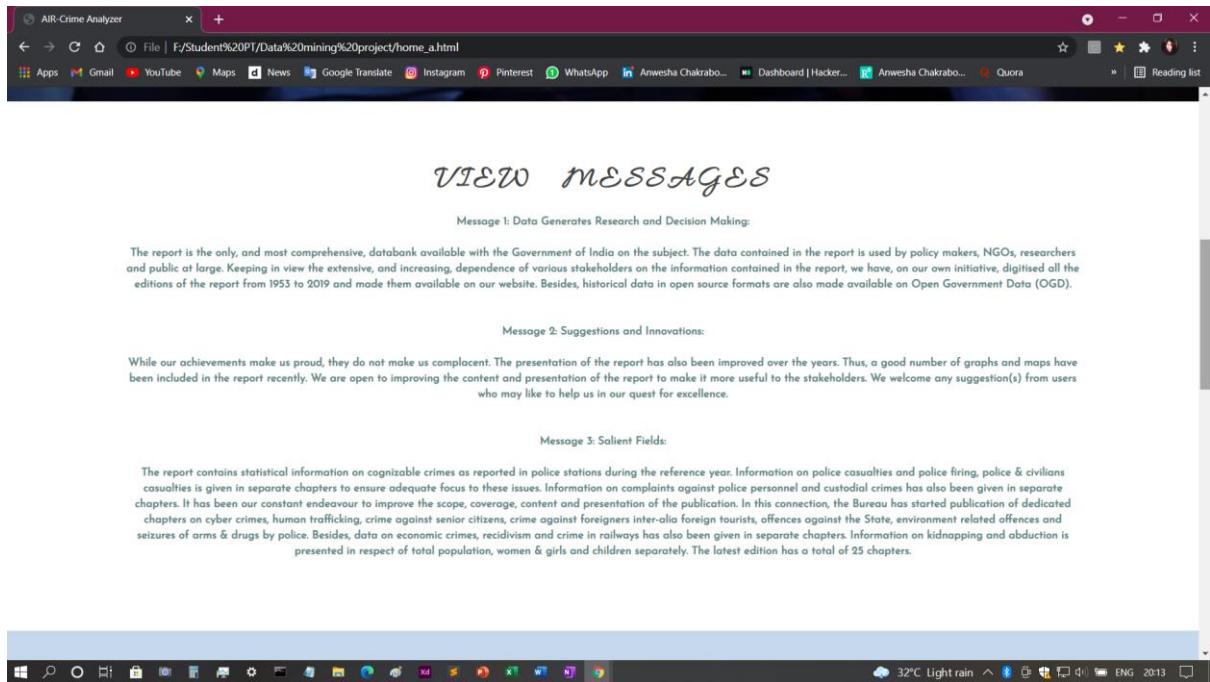


Figure 17: Admin-View Messages

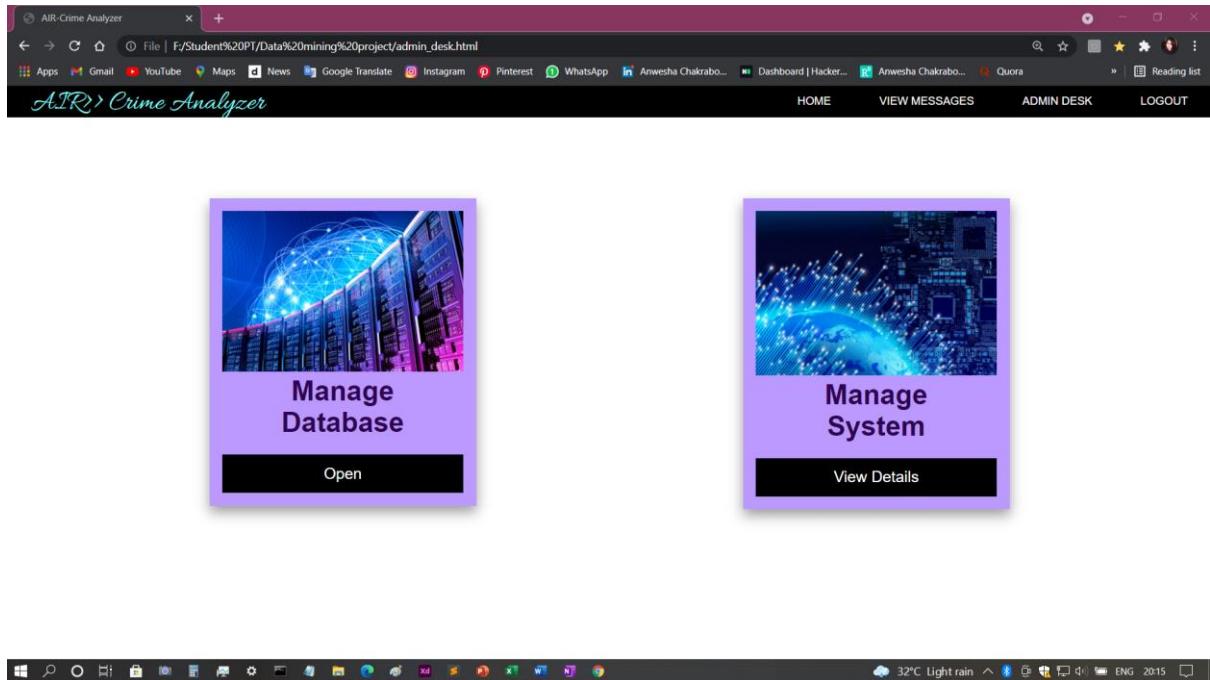


Figure 18: Admin Desk

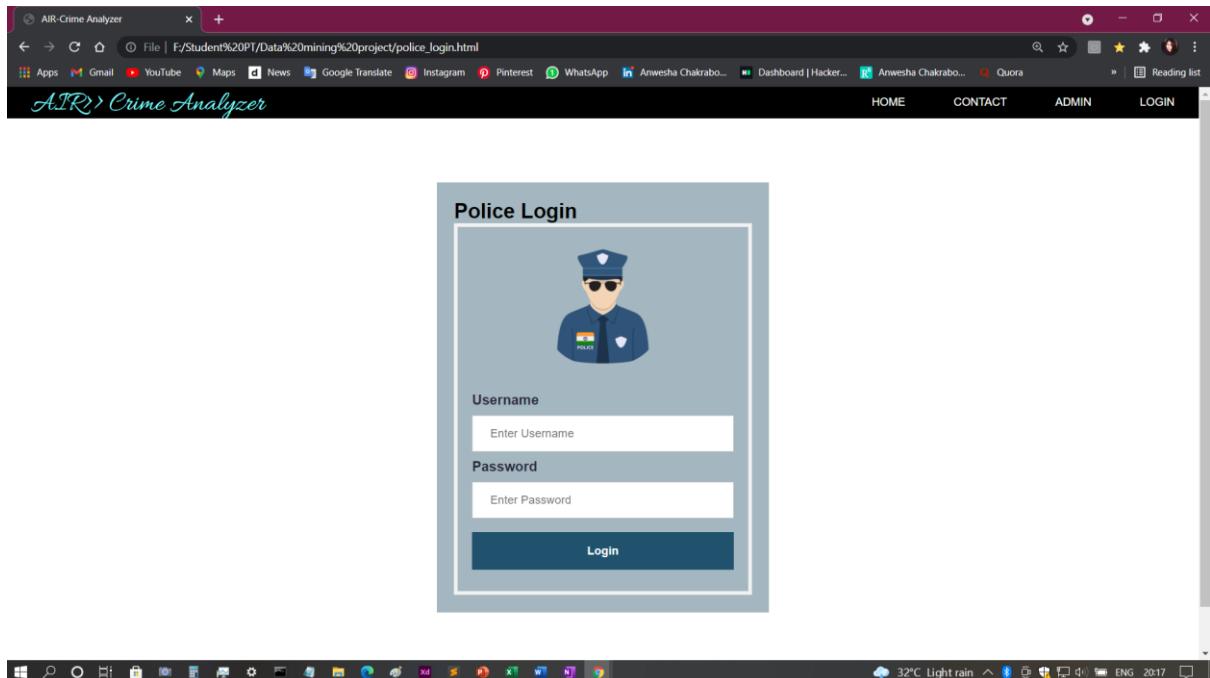


Figure 19: Police Login Page

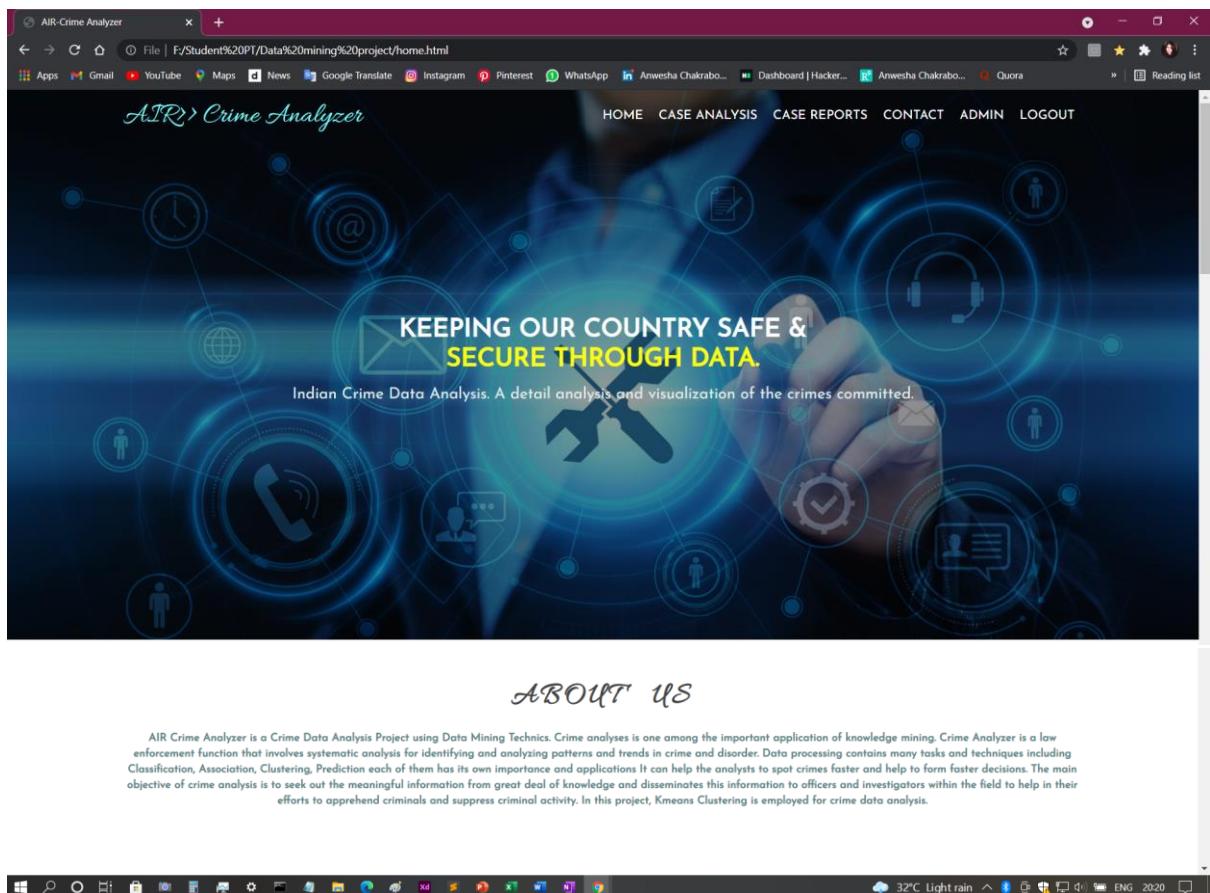


Figure 20: Police Home Page

The screenshot shows a web browser window titled "AIR-Crime Analyzer". The main content area is divided into three sections: "Case Updates", "Case Analysis", and "Solve Cases". Each section contains an image, a title, and a brief description. Below these sections is a "CONTACT US" form with fields for Email, Subject, and Message, and a Submit button.

**SERVICES**

**Case Updates**  
Officers can view all the latest case Updates and add any available current cases to keep updating the registry.

**Case Analysis**  
Officers can make uses of Machine learning algorithm for data analysis to make accurate predictions that can very much help them to solve cases.

**Solve Cases**  
Acquire the knowledge of the case through updates and combine it with analytics to quickly solve cases and reduce crime rate.

**CONTACT US**

Email  
Subject  
Message

Submit

© Anwesha-Ishika-Rohan 2021

Figure 20: Police-Contact Admin

The screenshot shows a web browser window titled "AIR-Crime Analyzer". The main content area displays three cards with titles: "Major Crime Analytical Results", "Crime Against Women", and "Crime Against Child". Each card has a "View" button at the bottom. The top navigation bar includes links for HOME, CASE ANALYSIS, CASE REPORTS, CONTACT, ADMIN, and LOGOUT.

AIR> Crime Analyzer

HOME CASE ANALYSIS CASE REPORTS CONTACT ADMIN LOGOUT

**Major Crime Analytical Results**

**Crime Against Women**

**Crime Against Child**

View View View

Figure 21: Case Analysis Home

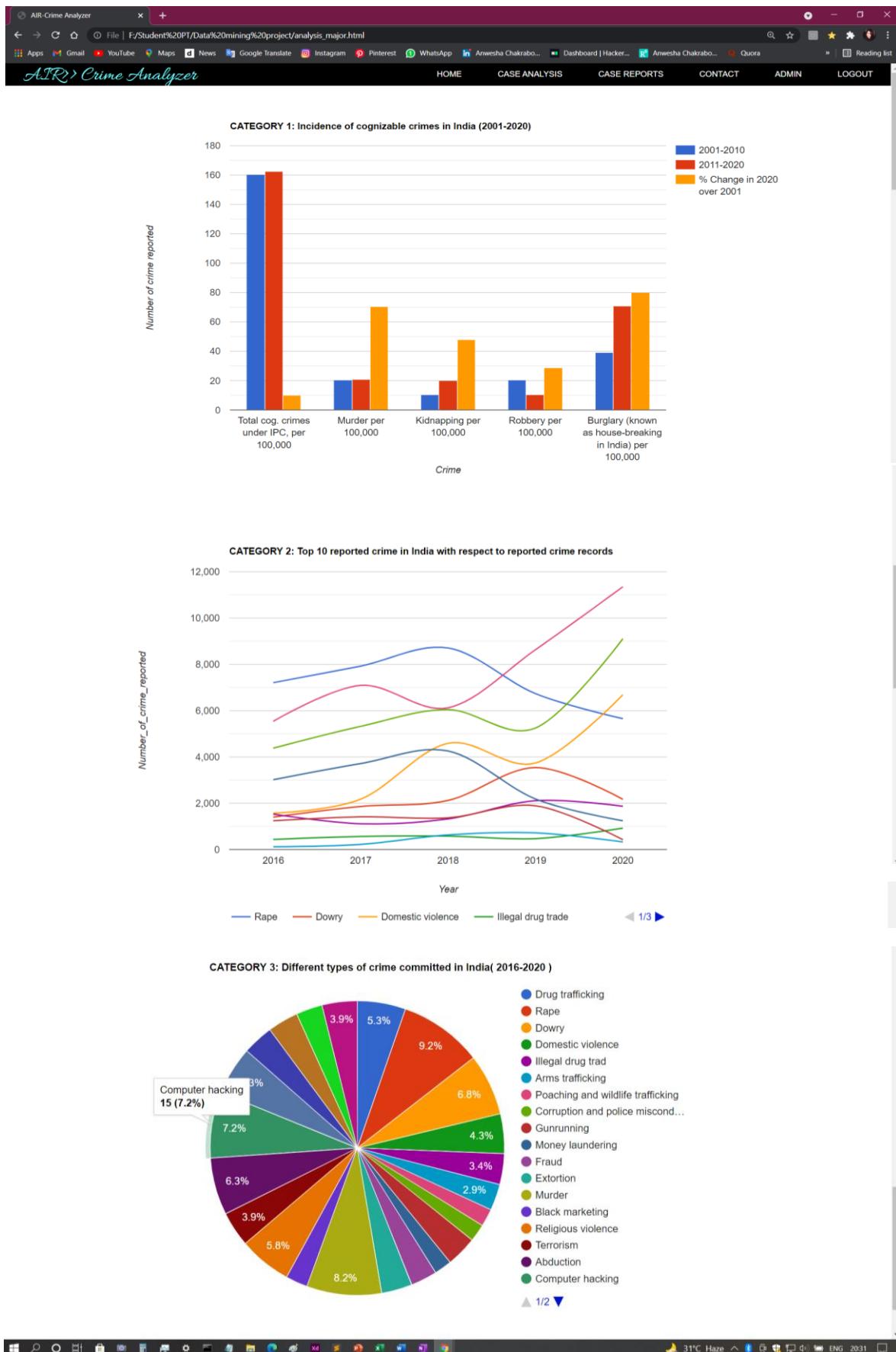


Figure 22: Major Crimes in India

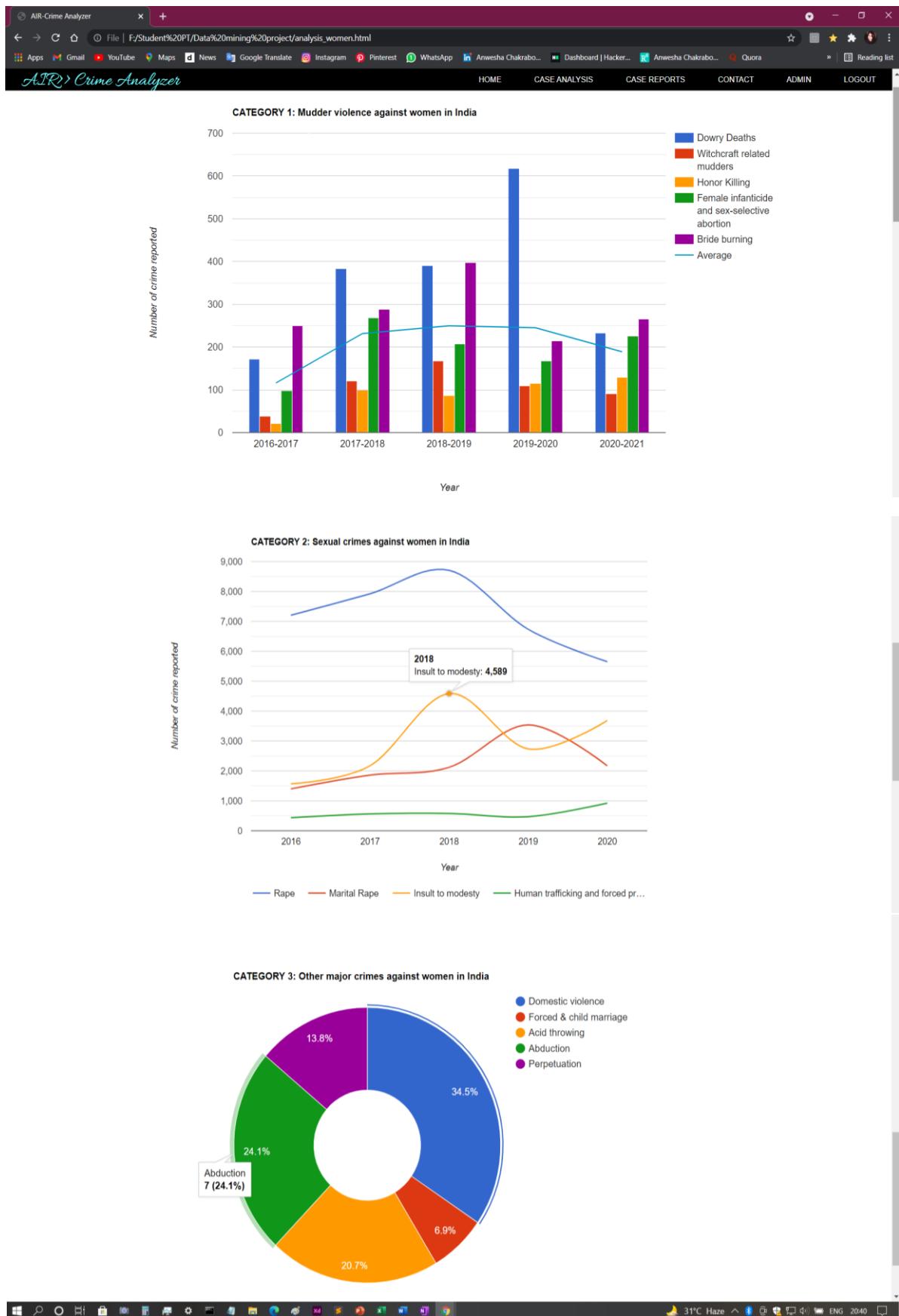


Figure 23: Crime against Women

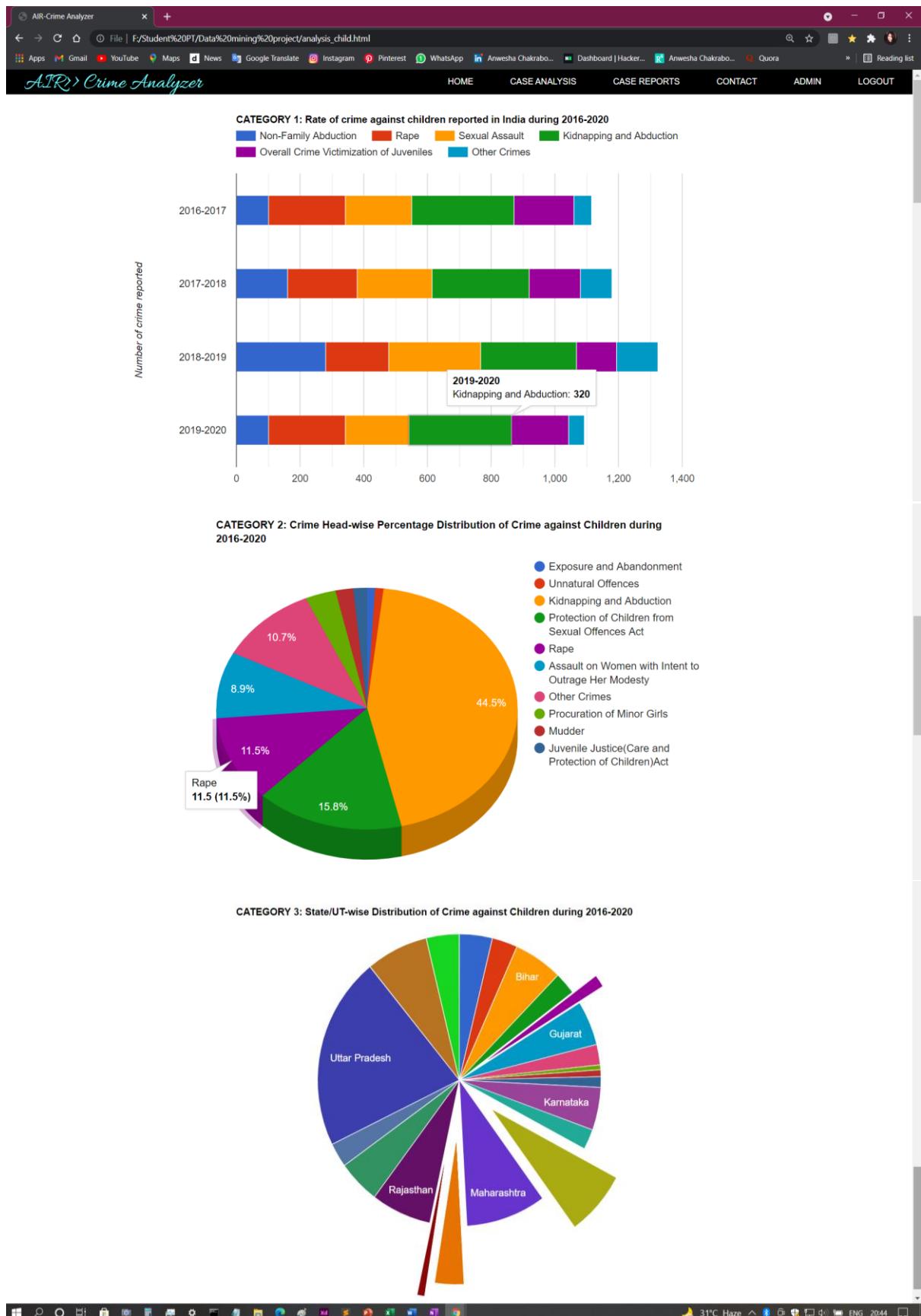


Figure 24: Crime against Child

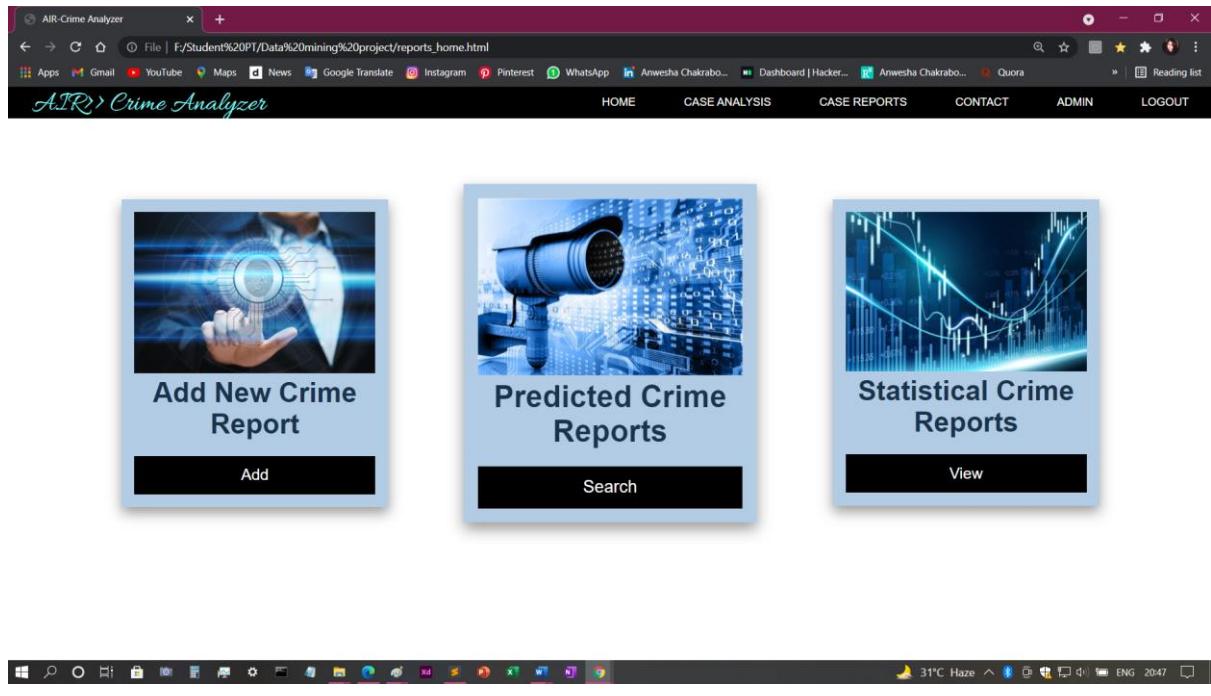


Figure 25: Crime Reports Home

A screenshot of a web browser showing the 'reports.add.html' page. The title bar says 'reports.add.html'. The main content area is a form titled 'Provide more information to be added to a Crime Report'. It includes fields for 'Victims First Name' (text input), 'Victims Last Name' (text input), 'Victims Gender' (radio buttons for Male, Female, Other), 'Contact Details' (text input for Phone No.), 'Current Address' (text input), 'Details about the Crime' (text area), 'Police Station' (text input for Crime reported police station), and a 'Submit' button at the bottom.

Figure 26: Crime report form

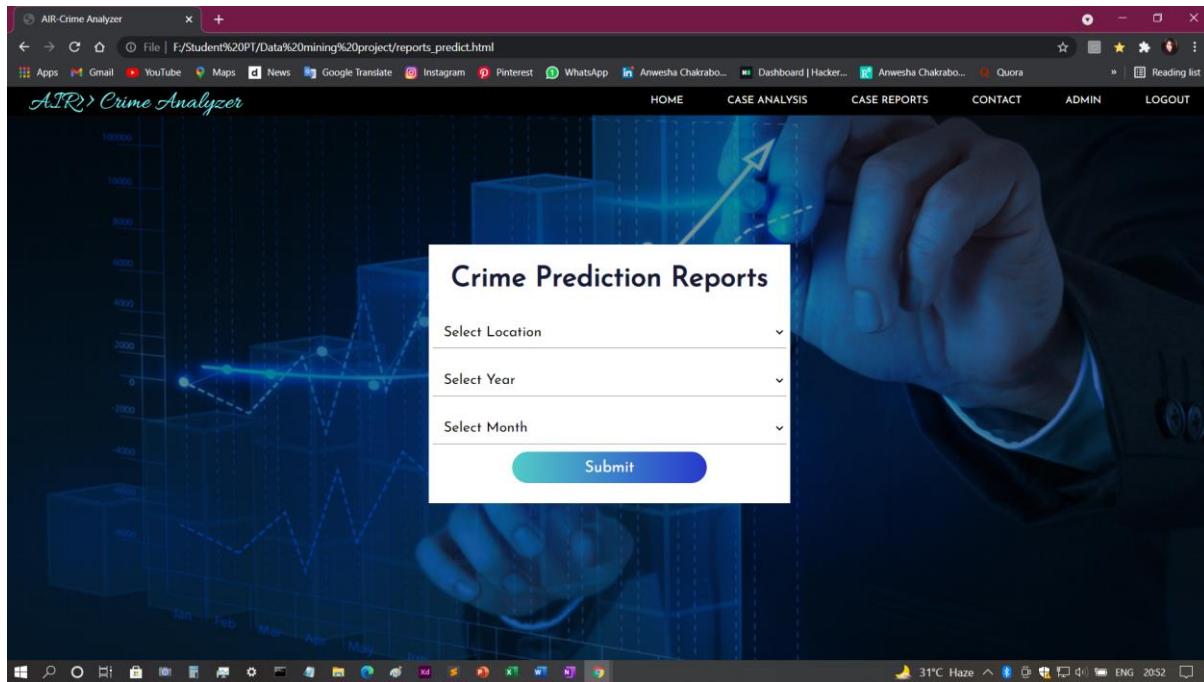


Figure 27: Crime Prediction Home

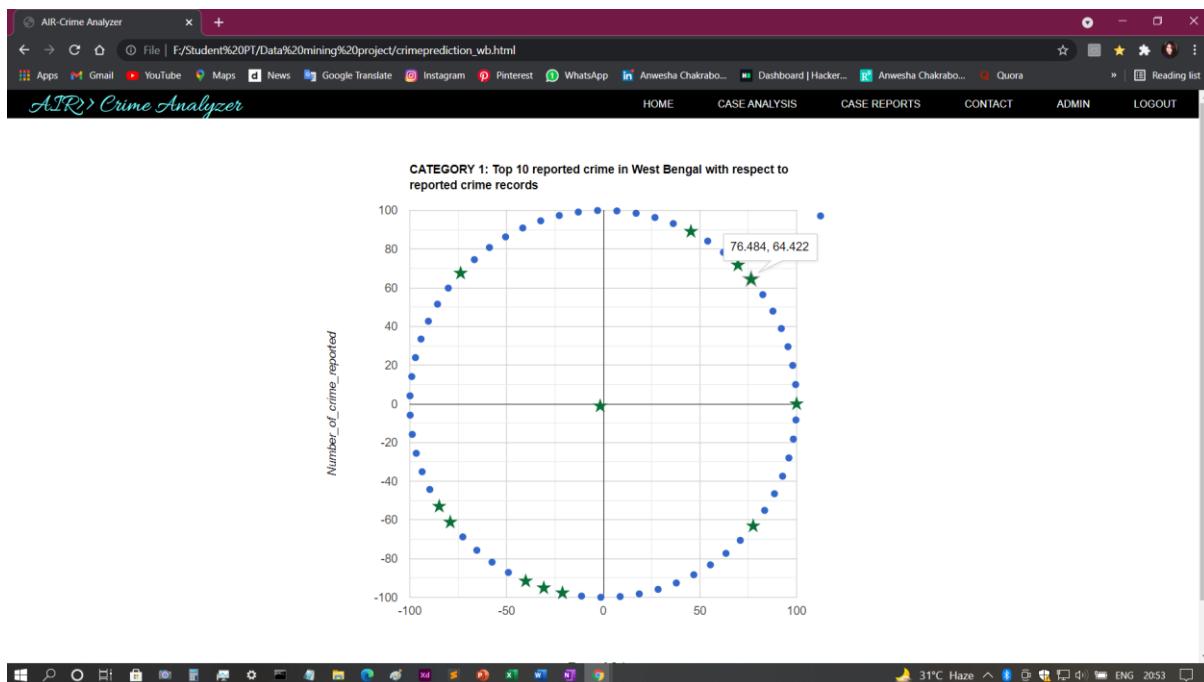


Figure 28: Crime prediction results

Absolute number of reported crimes & crime rates across India							
State/UT	2016	2017	2018	2019	Percentage share of state/UT (2019)	Crime rate in 2019 (IPC+SLL crimes)	
India	4831515	5007044	5074635	5156172	100.0	385.5	
Andhra Pradesh	129389	148002	144703	145751	2.8	278.6	
Arunachal Pradesh	2700	2746	2817	2877	0.1	190.9	
Assam	107014	109952	120572	123783	2.6	385.8	
Bihar	189696	236055	262815	269109	5.2	224.0	
Chhattisgarh	84192	90516	98233	96561	1.9	334.7	
Goa	3706	3943	3884	3727	0.1	241.5	
Gujarat	435422	334799	393194	431066	8.4	631.6	
Haryana	143111	224816	191229	166336	3.2	577.4	
Himachal Pradesh	17249	17796	19594	19924	0.4	272.4	
Jammu and Kashmir	26624	25608	27276	25408	0.5	187.8	
Jharkhand	47817	52664	55664	62206	1.2	165.5	
Karnataka	179479	184063	163416	163691	3.2	248.1	
Kerala	707870	653500	512167	453083	8.8	1287.7	
Madhya Pradesh	365154	379682	405129	395619	7.7	478.9	
Maharashtra	430866	467753	515674	509443	9.9	415.8	

Figure 29: Statistics of crime reports in India

## **5.2 Sample Codes:**

### 5.2.1 Frontend Codes

F:\Student PT\Data mining project\home.html (Data mining project) - Sublime Text (UNREGISTERED)

File Edit Selection Find View Goto Tools Project Preferences Help

**FOLDERS**

- Data mining project
- images
- admin\_desk.html
- admin\_image.jpg
- admin\_login.html
- analysis\_child.html
- analysis\_home.html
- analysis\_major.html
- analysis\_women.html
- case\_a.jpg
- case\_s.jpg
- case\_u.png
- chart\_s.css
- crimeprediction\_wb.htm
- data\_a1.jpg
- data\_m2.jpg
- data\_v.jpg
- data\_v1.jpg
- home.html
- home\_a.html
- index.html
- nav.css
- police\_login.html
- reports\_add.html
- reports\_home.html
- reports\_predict.html
- reports\_statistics.html
- statisticsc.jpg
- stats.jpg
- styles.css
- system.jpg

**home\_a.html**

```
1 <!DOCTYPE html>
2 <html lang="en">
3 <head>
4   <meta charset="UTF-8">
5   <meta name="viewport" content="width=device-width, initial-scale=1.0"/>
6   <title>AIR-Crime Analyzer</title>
7   <style type="text/css">
8     @import url('
      https://fonts.googleapis.com/
      css?family=Allura|Josefin+Sans');
9
10  *{
11    margin: 0;
12    padding: 0;
13    /*box-sizing: border-box;*/
14    outline: none;
15    list-style: none;
16    text-decoration: none;
17    font-family: 'Josefin Sans', sans-serif;
18  }
19
20 h1.title{
21   margin-bottom: 20px;
22   color: #444444;
23   font-family: 'Allura', cursive;
24   text-transform: uppercase;
25   font-size: 38px;
26   word-spacing: 25px;
27 }
```

**home.html**

```
1 <!DOCTYPE html>
2 <html lang="en">
3 <head>
4   <meta charset="UTF-8">
5   <meta name="viewport" content="width=device-width, initial-scale=1.0"/>
6   <title>AIR-Crime Analyzer</title>
7   <style>
8     @import url(
      https://fonts.googleapis.com/
      css?family=Allura|Josefin+Sans);
9
10  *{
11    margin: 0;
12    padding: 0;
13    /*box-sizing: border-box;*/
14    outline: none;
15    list-style: none;
16    text-decoration: none;
17    font-family: 'Josefin Sans', sans-serif;
18  }
19
20 h1.title{
21   margin-bottom: 20px;
22   color: #444444;
23   font-family: 'Allura', cursive;
24   text-transform: uppercase;
25   font-size: 38px;
26   word-spacing: 25px;
27 }
```

Line 323. Column 40 Tab Size: No new notifications (Off)

Figure 30: Sample Come for Home pages

# St. Thomas' College of Engineering and Technology

```

admin_login.html
1 <!DOCTYPE html>
2 <html>
3 <head>
4   <title>AIR-Crime Analyzer</title>
5   <meta name="viewport" content="width=device-width, initial-scale=1">
6   <link rel="stylesheet" href="nav.css">
7   <style>
8     {
9       padding: 0px;
10      margin: 0px;
11    }
12   body {font-family: Arial, Helvetica, sans-serif;}
13   form {
14     border: 5px solid #f1f1f1;
15     font-size: 75px 519px;
16     color: #204060;
17   }
18   .form-box{
19     width:25%;
20     height: 70%;
21     position: relative;
22     margin: 8% auto;
23     background: #c2d6d6;
24     padding: 20px;
25     overflow: hidden;
26   }
27   input[type=text], input[type=password] {
28     width:100%
29     height: 30px;
30     border: 1px solid #ccc;
31     padding: 5px;
32     margin-bottom: 10px;
33   }
34   .button {
35     width: 100px;
36     height: 30px;
37     border: 1px solid #ccc;
38     background-color: #e0e0e0;
39     border-radius: 5px;
40     cursor: pointer;
41     font-size: 14px;
42     font-weight: bold;
43     color: black;
44   }
45   .button:hover {
46     background-color: #d0d0d0;
47   }

police_login.html
1 <!DOCTYPE html>
2 <html>
3 <head>
4   <title>AIR-Crime Analyzer</title>
5   <meta name="viewport" content="width=device-width, initial-scale=1">
6   <link rel="stylesheet" href="nav.css">
7   <style>
8     {
9       padding: 0px;
10      margin: 0px;
11    }
12   body {font-family: Arial, Helvetica, sans-serif;}
13   form {
14     border: 5px solid #f1f1f1;
15     font-size: 75px 519px;
16     color: #204060;
17   }
18   .form-box{
19     width:25%;
20     height: 70%;
21     position: relative;
22     margin: 8% auto;
23     background: #a4b6bf;
24     padding: 20px;
25     overflow: hidden;
26   }
27   input[type=text], input[type=password] {
28     width:100%
29     height: 30px;
30     border: 1px solid #ccc;
31     padding: 5px;
32     margin-bottom: 10px;
33   }
34   .button {
35     width: 100px;
36     height: 30px;
37     border: 1px solid #ccc;
38     background-color: #e0e0e0;
39     border-radius: 5px;
40     cursor: pointer;
41     font-size: 14px;
42     font-weight: bold;
43     color: black;
44   }
45   .button:hover {
46     background-color: #d0d0d0;
47   }

```

Figure 31: Sample Code for Login pages

```

analysis_major.html
1 <!DOCTYPE html>
2 <html lang="en">
3 <head>
4   <title>AIR-Crime Analyzer</title>
5   <meta name="viewport" content="width=device-width, initial-scale=1">
6   <link rel="stylesheet" href="nav.css">
7   <link rel="stylesheet" href="chart_css.css" >
8   <!-- chart I -->
9   <script type="text/javascript" src="https://www.gstatic.com/charts/loader.js"></script>
10  <script type="text/javascript">
11    google.charts.load('current', {'packages':['corechart']});
12    google.charts.setOnLoadCallback(drawBasic);
13  </script>
14  function

analysis_women.html
1 <!DOCTYPE html>
2 <html lang="en">
3 <head>
4   <title>AIR-Crime Analyzer</title>
5   <meta name="viewport" content="width=device-width, initial-scale=1">
6   <link rel="stylesheet" href="nav.css">
7   <link rel="stylesheet" href="chart_css.css" >
8   <!-- chart I -->
9   <script type="text/javascript" src="https://www.gstatic.com/charts/loader.js"></script>
10  <script type="text/javascript">
11    google.charts.load('current', {'packages':['corechart']});
12    google.charts.setOnLoadCallback(drawBasic);
13  </script>
14  function

analysis_child.html
1 <!DOCTYPE html>
2 <html lang="en">
3 <head>
4   <title>AIR-Crime Analyzer</title>
5   <meta name="viewport" content="width=device-width, initial-scale=1">
6   <link rel="stylesheet" href="nav.css">
7   <link rel="stylesheet" href="chart_css.css" >
8   <!-- chart I -->
9   <script type="text/javascript" src="https://www.gstatic.com/charts/loader.js"></script>
10  <script type="text/javascript">
11    google.charts.load('current', {'packages':['corechart', 'bar']});
12    google.charts.setOnLoadCallback(drawBasic);
13  </script>
14  function

```

Figure 32: Sample Code for Crime Analysis

```

reports_add.html
1 <!DOCTYPE html>
2 <html>
3 <head>
4   <meta name="viewport" content="width=device-width, initial-scale=1">
5   <link rel="stylesheet" href="nav.css">
6   <style>
7     body{
8       font-family: Calibri, Helvetica, sans-serif;
9       background-color: #ffffff;
10      }
11      .container {
12        padding:50px;
13        margin:0px;
14        margin-right:300px;
15        margin-left:300px;
16        background-color: #d1e0e0;
17      }
18      input[type=text], input[type=password], textarea {
19        width: 100%;
20        padding: 15px;
21      }

reports_predict.html
1 <html>
2 <title>AIR-Crime Analyzer</title>
3 <head>
4   <meta charset="UTF-8">
5   <meta http-equiv="X-UA-Compatible" content="IE=edge">
6   <meta name="viewport" content="width=device-width, initial-scale=1.0">
7   <title>Student Portal</title>
8   <link rel="stylesheet" href="nav.css">
9   <style>
10    *{
11      margin: 0;
12      padding: 0;
13      box-sizing: border-box;
14      outline: none;
15      text-decoration: none;
16      font-family: 'Josefin Sans';
17    }
18    .img {
19      margin: 70px;
20      padding-left:90px;
21    }
22    .img import url('https://fonts.googleapis.com/css?family=Allura|Josefin+Sans');

reports_statistics.html
1 <!DOCTYPE html>
2 <html>
3 <head>
4   <style type="text/css">
5     img {
6       margin: 70px;
7       padding-left:90px;
8     }
9   </style>
10  <title>AIR-Crime Analyzer</title>
11  <meta name="viewport" content="width=device-width, initial-scale=1">
12  <link rel="stylesheet" href="nav.css">
13  </head>
14  <body>
15    
16    <center><p><b>Source:</b> National Crime Records Bureau</p></center>
17    <div class="navbar">
18      <div class="logo">
19        <a href="#">AIR>> Crime Analyzer</a>
20    </div>

```

Figure 33: Sample Code for Crime Prediction

## 5.2.2 Backend Codes

```

# For some basic operations
import numpy as np
import pandas as pd
# For visualizations
import matplotlib.pyplot as plt
import seaborn as sns
import folium
import squarify
# To see the first 5 rows in the data set
print(data.head())
# To Describe the data set
print(data.describe())
# To check if there are any null values
print(data.isnull().sum())
# Filling the missing value in PdDistrict using
data['PdDistrict'].fillna(data['PdDistrict'].mode()[0], inplace=True)
print(data.isnull().any().any())
##### Data Visualization #####

```

Figure 35: Sample Code for Crime Analysis using Python

```

##### Data Visualization #####
#Different categories of crime—Figure 1
plt.rcParams['figure.figsize'] = (10, 10)
plt.style.use('seaborn')
plt.style.use('dark_background')
sns.countplot(data['Category'], palette = 'gnuplot')
plt.title('Major Crimes in West Bengal', fontweight='bold')
plt.xticks(rotation = 90)
plt.show()
#Plotting a tree map—Figure 2
plt.rcParams['figure.figsize'] = (10, 10)
plt.style.use('seaborn')
plt.style.use('dark_background')
y = data['Category'].value_counts().head(15)
plt.rcParams['figure.figsize'] = (10, 10)
color = plt.cm.magma(np.linspace(0, 1, 15))
squarify.plot(sizes = y.values, label = y.index,
              color = color)
plt.title('Crime Categories', fontweight='bold')
plt.show()

```

Figure 36: Sample Code for data visualization

```

from django.contrib import admin
# Register your models here.
from app.models import crimes_against_women
from app.models import murder
admin.site.register(crimes_against_women)
admin.site.register(murder)

```

Figure 37: Extracting Data from date sets using Django.

## Chapter 7: FUTURE WORK AND CONCLUSION

### 7.1 Future Work

#### A. Criminal Profiling

In addition to this a new concept called Criminal profiling, which helps the crime investigators to record the characteristics of criminals. It is a very accurate tool for profiling the characteristics or details of offenders is a behavioural and investigative tool that is intended to help investigators to accurately predict and profile the characteristics of unknown criminal subjects or offenders [2],[13]. The main goal of doing criminal profiling is that:

- to provide crime investigators with a social and psychological assessment of the offender;
- to evaluate belongings found in the possession of the offender.

#### B. Snatching

We are concentrating more on crimes like snatching to get more details related to it like crime location, time, date, crime type (which type of snatching), victim and offender names etc. Currently we are getting crime details like: (1) Name of person (victims, offenders), (2) Location, (3) Organization, (4) Type of crime (whether murder, robbery), (5) Subcategories of crime type (for snatching there are other categories like chain snatching, purse snatching etc), (6) Type of vehicle offender used, (7) Whether any weapons used, (8) Time of incident, (9) Date, (10) Incident summary, (11) Criminal groups involved [10].

As a future extension of our work, we plan to apply more classification models to increase crime prediction accuracy and to enhance the overall performance. It is also a helpful extension for our study to consider the income information for neighbourhoods in order to see if there are relationships between neighbourhood's income level and their crime findings. Furthermore, we want to study other crimes datasets from new cities along with their demographic's datasets.

### 7.2 Conclusion

An acceptable model for data mining which comes up with excellent results of analysing crime data set; it requires huge historical data that can be used for creating and testing the model. More than 150500 crime records that were used in this work can give estimation and lead to an acceptable model. VS Code and Excel software were used to pre-process and analyse the collected crime and criminal data. In this paper we have tested the accuracy of classification and prediction based on different test sets. Classification is done based on the Bayes theorem which showed more than 90% accuracy [2]. Using this algorithm, we trained numerous news articles and build a model dataset. For testing we are inputting some test data into the model which shows better results. Our system takes factors/attributes of a place and Apriori algorithm gives the frequent patterns of that place. The pattern is used for building a model for decision tree. Till now we trained our system using certain attributes but we are planning to include more factors to improve accuracy.

## REFERENCES

- [1] Malathi. A and Dr. S. Santhosh Baboo. Article:an enhanced algorithm to predict a future crime using data mining. International Journal of Computer Applications, 21(1):1–6, May 2011. Published by Foundation of Computer Science.
- [2] Eibe Frank and Remco R. Bouckaert. Naive bayes for text classification with unbalanced classes. In Proceedings of the 10th European Conference on Principle and Practice of Knowledge Discovery in Databases, PKDD'06, pages 503–510, Berlin, Heidelberg, 2006. Springer-Verlag.
- [3] Ralf Hartmut Güting. Graphdb: Modeling and querying graphs in databases. In Proceedings of the 20th International Conference on Very Large Data Bases, VLDB '94, pages 297–308, San Francisco, CA, USA, 1994. Morgan Kaufmann Publishers Inc
- [4] Lior Rokach and Oded Maimon. Decision trees. In Oded Maimon and Lior Rokach, editors, The Data Mining and Knowledge Discovery Handbook, pages 165–192. Springer, 2005.
- [5] Tong Wang, Cynthia Rudin, Daniel Wagner, and Rich Sevieri. Detecting patterns of crime with series finder. In Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECMLPKDD 2013), 2013.
- [6] R. Sanjana, H. S. S, and S. Sruthi, ‘Big Data Approach for Crime Classification and Visualization Using Crime Dataset’, Int. J. Innov. Res. Sci. Eng. Technol., vol. 7, no. 2, pp. 25–29, 2018.
- [7] A. Jain and V. Bhatnagar, ‘Crime Data Analysis Using Pig with Hadoop’, Procedia Comput. Sci., vol. 78, no. December 2015, pp. 571–578, 2016.
- [8] N. G. Karthigaa and M. K. B. Janani, ‘Big Data Prediction on Crime Detection’, in National Conference on Computational Intelligence Systems, 2017, pp. 43–48.
- [9] V. N. Inukollu, S. Arsi, and S. Rao Ravuri, ‘Security Issues Associated with Big Data in Cloud Computing’, Int. J. Netw. Secur. Its Appl., vol. 6, no. 3, pp. 45–56, 2014.
- [10] H. V. Jagadish et al., ‘Big data and its technical challenges’, Commun. ACM, vol. 57, no. 7, pp. 86–94, 2014.
- [11] D. C. Ukpabi, ‘Statistical Analysis on Crime Rate in Nigeria’, 2018.
- [12] O. Faweya, A. T. Adeniran, and K. O. Balogun, ‘Principal Component Analysis of Crime Rate in Nigeria : A Case Study of Ekiti and Osun State’, Am. J. Math. Stat., vol. 8, no. 4, pp. 79–88, 2018.
- [13] A. Hills, ‘Partnership policing: Is it relevant in Kano, Nigeria?’, Criminol. Crim. Justice, vol. 14, no. 1, pp. 8–24, Feb. 2014.
- [14] Nurhayati, Busman, and V. Amrizal, ‘Big Data Analysis Using Hadoop Framework and Machine Learning as Decision Support System (DSS) (Case Study: Knowledge of Islam Mindset)’, in 2018 6th International Conference on Cyber and IT Service Management, CITSM 2018, 2018, no. Citsm, pp. 1–6.
- [15] M. P.K, ‘Technology’s Ways: Information Technology, crime analysis and the rationalizing of policing’, SAGE, vol. 1, no. 1, pp. 83–103, 2015.
- [16] K. Overland Park, ‘Definition and Types of Crime Analysis (White Paper 2014-02)’, 2014.