Final Report

Technology Stack: Google Artificial Intelligence and Machine Learning

Project Title: A Crime Type and Occurrence Prediction Using Machine Learning Algorithm

Team ID: LTVIP2024TMID13105

Team Size: 04

Team Members:

1. Maddikunta Priyanka
2. Are Manoj Kumar
3. Pagidipalli Vamsi
4. Mandala Pavan Kumar

College: Kallam Haranadhareddy Institute of Technology

ABSTRACT:

Crime analysis and prediction is a systematic approach for identifying and analyzing patterns and trends in crime. The proposed system can predict regions that have high probability for crime occurrence and can visualize crime-prone areas. With the increasing advent of computerized systems, crime data analysts can help Law enforcement officers to speed up the process of solving crimes. Using the concept of machine learning we can extract previously unknown, useful information from unstructured data The proposed project presents an approach between computer science and criminal justice to develop a machine learning algorithm that can help to solve crimes faster. Instead of focusing on causes of crime occurrence like the criminal background of the offender, political enmity, etc. is focusing mainly on crime factors of each day. Even though it cannot predict who all may be the victims of crime but can predict the place that has probability for its occurrence. This system can also be used by any crime department for reducing and solving crimes within no time. So, to predict the area of crimes in different types of crime that is going to happen in nearby future and use decision tree algorithm

KEYWORDS: Crime detection, Machine Learning, Deep Learning, Decision tree, Random Forest.

CHAPTER 1: INTRODUCTION

1.1 Introduction

Crimes are the significant threat to the humankind. There are many crimes that happens regular interval of time. Perhaps it is increasing and spreading at a fast and vast rate. Crimes happen from small village, town to big cities. Crimes are of different type – robbery, murder, rape, assault, battery, false imprisonment, kidnapping, homicide. Since crimes are increasing there is a need to solve the cases in a much faster way. The crime activities have beenincreased at a faster rate and it is the responsibility of police department to control and reduce the crime activities. Crime prediction and criminal identification are the major problems to the police department as there are tremendous amount of crime data that exist. There is a need of technology through which the case solving could be faster.

The increase in crime data recording coupled with data analytics resulted in the growth of research approaches aimed at extracting knowledge from crime records to better understand criminal behavior and ultimately prevent future crimes.

Crime is a complex social phenomenon that has grown due to major changes in the society. Law enforcement agencies need to learn the factors that lead to an increase in crime tendency. To curb this, there is always a need of strategies and policies to prevent crime. As a result of technology development, science and information, machine learning and artificial intelligence tools are increasingly prevalent in the law enforcement community.

The above problem made me to go for research about how can solving a crime case made easier. Through many documentations and cases, it came out that machine learning and data science can make the work easier and faster. The aim of this project is to make crime prediction using the features present in the dataset. The dataset is extracted from the official sites. With the help of machine learning algorithm, using python as core we can predict thetype of crime which will occur in a particular area.

The objective would be to train a model for prediction. The training would be done using the training data set which will be validated using the test dataset. Building the model will be done using better algorithm depending upon the accuracy. The machine learning algorithms will be used for crime prediction. Visualization of dataset is done to analyze the crimes which may have occurred in the country. This work helps the law enforcement agencies to predict and detect crimes in a particular state with improved accuracy and thus reduces the crime rate.

The traditional way for analyzing and predicting the crime rate of a particular area is by referring all the First Information Report (FIR) that has been filed from that area. This process takes a lot of work and time. Even a minor mistake can lead to wrong result. To overcome this problem faced by the police department the proposed system includes the following objectives:

- Creating a website that helps the police department to analyses and predict the crime rate of aparticular area using Machine Learning technique.
- Help people travelling to different place to understand the crime trends of a particular area.

For the correct analysis and prediction of crime rate the data must be carefully observed because the false data entry will reflect erroneous result. For the correct prediction of crime, the data must be trained and then the machine learning model is used. The main intention of this project is to predict the crimes in particular location before they happen in order toincrease safety in that area by using Machine Learning techniques so that would ease the work,reduce time for the above and provide a more informative data to the public users.

1.2 Features

Machine learning techniques can be used to analyze and predict crime occurrences based on various features.

Location: The location of a crime is one of the most significant features. This could be a specific address, a neighbourhood, or even a larger area like a city or a county. Machine learning algorithms can analyze patterns in crime locations to predict where future crimes might occur.

Crime Type: The nature of the crime itself can also be a predictive feature. For example, machine learning algorithms can analyze patterns in different types of crimes, such as theft, assault, or vandalism, to predict the likelihood of each type occurring in a given location and time.

Demographic Information: Demographic information about the population in a given area can also be used as a feature in crime prediction. This could include factors like age, gender, income level, and education level.

Temporal Factors: Incorporating temporal aspects such as time of day, day of the week, month, seasonality, and trends over time can provide valuable insights into when crimes are more likely to occur.

User-friendly Interface Design: A user-friendly interface design that prioritizes ease of use and accessibility. The interface is intuitively designed, with clear navigation pathways and intuitive controls, ensuring a seamless user experience. Users can easily input queries, interpret results, and navigate through various features and functionalities without the need for extensive training or technical expertise.

1.3. Problem Statement

Crimes now days are increasing day by day and with different level of intensity and versatility. The result is great loss to society in terms of monitory loss, social loss and further it enhances the level of threat against the smooth livelihood in the society. To overcome this problem the computing era can help to reduce the crime or even may be helpful in predicting the crime so that sufficient measures can be taken to minimize the loss to property and life. The crime rate prediction strategies can be applied on historical data available in the police records by examining the data at various angles like reason of crime, frequency of similar kind of crimes at specific location with other parameters to prepare model the crime prediction. It is the major challenge to understand the versatile data available with us then model it to predict the future incidence with acceptable accuracy and further to reduce the crime rate.

1.4. Objective

The primary objective of this project is to develop a robust and accurate machine learning model capable of predicting the occurrence of specific types of crimes within a given geographical area and time frame. By leveraging historical crime data, demographic information, environmental factors, and other relevant features, the model aims to provide actionable insights and forecasts that enable law enforcement agencies and policymakers to:

Enhance Public Safety: Predictive models will identify high-risk areas and times for specific types of crimes, allowing law enforcement to deploy resources proactively and implement targeted crime prevention strategies.

Optimize Resource Allocation: By accurately predicting crime occurrence, law enforcement agencies can allocate personnel, patrol routes, and other resources more efficiently, maximizing their effectiveness in crime prevention and response.

Inform Policy Decisions: Insights derived from the predictive model will inform policymaking efforts aimed at addressing underlying socio-economic factors contributing to crime, such as poverty, unemployment, and lack of access to education or social services.

Empower Community Engagement: Transparent communication of crime predictions and prevention strategies fosters trust and collaboration between law enforcement agencies and the communities they serve, empowering citizens to take proactive measures to enhance their safety.

Reduce Crime Rates: By identifying patterns and trends in crime occurrence and implementing targeted interventions, the ultimate goal is to reduce overall crime rates and improve quality of life for residents in the affected areas.

1.5. Scope

Law Enforcement Agencies: Law enforcement agencies can utilize predictive models to allocate resources effectively. They can deploy patrols, surveillance, and other resources in areas and times identified as high-risk for specific types of crimes.

Policymakers: Policymakers can use predictive analytics to inform policy decisions aimed at reducing crime rates and improving public safety. Insights from predictive models can guide the development of targeted interventions and crime prevention strategies.

Community Organizations: Community organizations can leverage predictive models to engage with residents and implement localized crime prevention initiatives. By collaborating with law enforcement and policymakers, they can address community-specific concerns and improve overall safety.

Media and Public Awareness: Media outlets can raise public awareness about the benefits and limitations of crime prediction technologies. They can facilitate informed discussions on ethical considerations, privacy concerns, and the potential impact of predictive analytics on society.

Crime Analysts: Crime analysts can use predictive models to identify emerging crime trends and hotspots. They can provide actionable insights to law enforcement agencies and policymakers to facilitate proactive responses and resource allocation.

Research Institutions: Research institutions can conduct studies and experiments to improve the accuracy and effectiveness of predictive models. They can explore new machine learning techniques, feature engineering approaches, and data sources to enhance crime prediction capabilitie

CHAPTER 2. LITERATURE SURVEY

1.3 Alkesh Bharati, Dr SarvaNaguru RA, "Crime Analysis and Prediction Using Fuzzy C-Means Algorithm" [1], The data was/is presented as Crime research is a tool used to define criminal activities and study them. If the research conducted so far can be seen to be more specifically useful, it is mostly because it indicates which criminal types are useful in controlling crime, then, mostly they would it be places where violent crimes are reduced. It is an excellent method for measuring the crime rate because of each region can be broken down by procedure and the data is collected for any of each process to be examined. Through the rapid increase in information technology, crime analysts will be able to continue to enhance the investigations and help them interpret the evidence. on the sample clustering and preprocessing to get unstructured evidence, and then look for crimes inside it Thus, persons formerly investigated and then arrested or identified as having committed the same criminal behaviour may often be looked at for patterns such as criminal history, or incident reports, rather than only offences themselves. This is simply intended to direct law enforcement resources to where crimes can occur, without attention to identifying who is responsible. Bayesian classifiers were used as the current scheme was in use in place in the current methodology, the fuzzy C-Means algorithm will be used to group the crime data for all items that are apprehensible, apprehension of, physical assault, larceny-theft, and crime of women, as well as all criminal offences such as kidnapping, in the dataset.

Advantages:

Flexibility of Fuzzy C-Means (FCM) Algorithm: FCM is a soft clustering algorithm that allows data points to belong to multiple clusters with varying degrees of membership. This flexibility is beneficial for analyzing crime data, which may exhibit complex and overlapping patterns.

Identification of Crime Patterns: By applying the FCM algorithm to crime data, the study aims to identify patterns and clusters related to different types of crimes, such as apprehension, physical assault, larceny-theft, crimes against women, and kidnapping. This can provide valuable insights for law enforcement agencies to understand crime dynamics and allocate resources effectively.

Disadvantages:

Data Preprocessing Challenges: Preprocessing unstructured data, such as incident reports and criminal histories, can be challenging and time-consuming. It may require extensive data cleaning, normalization, and feature extraction to prepare the data for analysis with the FCM algorithm.

Interpretation of Fuzzy Clusters: Fuzzy clustering algorithms like FCM can produce clusters with overlapping boundaries, making it challenging to interpret the results. It may be difficult to clearly define and separate clusters corresponding to different types of crimes, leading to ambiguity in the analysis.

2.2 Shubham Agarwal, Lavish Yadav, "Crime Prediction based on Statistical Models" [2], The data was/is presented as in societies that have less criminal activity, the increase in the number of various offences is still a matter of concern. [Through] developments in technology, freely accessible records, and services,] these people manage to go undetected by society and continue their illegal deeds, even in the act of involving far more people. As a result, crime is rising in countries with a steep increase in incidence in [either/many] developed and [or/in] and [under] developing Because of the preceding year's criminal events in Indian states, we provide two models —Working Average Geometric Progression and WAGP and Seasonal Augmented EP WAG with known past criminal occurrences in order to anticipate the crime activity that is likely to occur in following years. These recent examples are used to help researchers understand how crime has changed in Indian states between the years 2001 and 2013. It was discovered that information from police reports covering the years 2001 to 2011 was a valuable in forecasting data that followed. The calculated expected crime values were compared to crime data for the same years, as well as for 2014 and 2015. To estimate our prediction to be within 85% of the correct value obtained from real data, the difference between the real and expected values of each needs to be doubled.

Advantages:

Prediction of Future Crime Activity: By analyzing past criminal occurrences in Indian states between 2001 and 2013, the study aims to predict future crime activity. This predictive analysis can provide valuable insights for law enforcement agencies and policymakers to anticipate and address potential crime trends.

Evaluation of Prediction Accuracy: The study compares the predicted crime values generated by the statistical models with actual crime data for validation. By assessing the accuracy of the predictions against real-world data for the years 2014 and 2015, the study aims to evaluate the effectiveness of the models in forecasting crime activity.

Disadvantages:

Limited Predictive Accuracy: The study mentions that the predicted crime values need to be within 85% of the correct values obtained from real data to be considered accurate. This suggests that the statistical models may have limitations in accurately predicting future crime activity, especially with a high degree of precision.

Dependency on Historical Data: The effectiveness of the statistical models relies heavily on historical crime data from previous years. Changes in societal factors, law enforcement strategies, or other external variables that are not accounted for in the models may affect their predictive accuracy.

CHAPTER 3: SYSTEM ANALYSIS

3.1 EXISTING SYSTEM:

The existing systems for crime prediction were not accurate or where too complicated with the algorithms used. Drawbacks like simplicity, flexibility, connectivity, processing time and inefficient parameters were keen points of concern. Previously developed system used algorithms like Fuzzy algorithm or Bayesian algorithm which were either complex or basic algorithms. Parameters needed to predict criminal was main point of concern as inaccurate or insufficient parameters led to inaccurate results. Following table gives the comparison of previously built system along with the disadvantages of algorithm being used.

3.2 PROPOSED SYSTEM:

In the proposed system the proposed system, we are introducing the application which will predict the crime that criminals can do in the future.

We will store all the previous record of crimes and by mining the previous record we can calculate the possibility and prediction of the crime he is likely to perform in that particular region.

This prediction is based on types of crimes like robbery, smuggling, etc. By using the previous data of a particular region. So, this project predicts the future types of crime of a specific region with the help of previous data.

Many efforts have used automated techniques to analyze different types of crimes, but without a systematic approach describing how to apply them. In general, understanding the relationship between analysis capability and crime type, clustering helps investigators more effectively use proposed techniques to identify trends and patterns, address vulnerable areas, and even predict crimes.

3.3 Advantages of Proposed System

Evidence-Based Decision Making: The use of data-driven predictive analytics provides law enforcement agencies with objective insights into crime patterns and trends. This allows decision-makers to make informed decisions and prioritize actions based on empirical evidence rather than subjective judgment.

Early Warning System: The predictive model serves as an early warning system for identifying emerging crime trends and patterns. By detecting patterns of criminal behavior before they escalate, law enforcement agencies can intervene early to prevent further criminal activity and mitigate potential risks.

Improved Investigation Techniques: The predictive model can assist law enforcement agencies in focusing their investigation efforts on specific types of crimes and potential suspects. This streamlines the investigation process and improves the likelihood of identifying and apprehending perpetrators.

3.4 Process Models

There are so many ways to predict crime locations. But we consider one of the way by using the past historical data and will design a model to predict the future where crime rate will increase in a particular region. Here we took the technology which is machine learning algorithms like decision tree and random forest. We design a model for prediction purpose which is trained by giving the hereditary dataset.

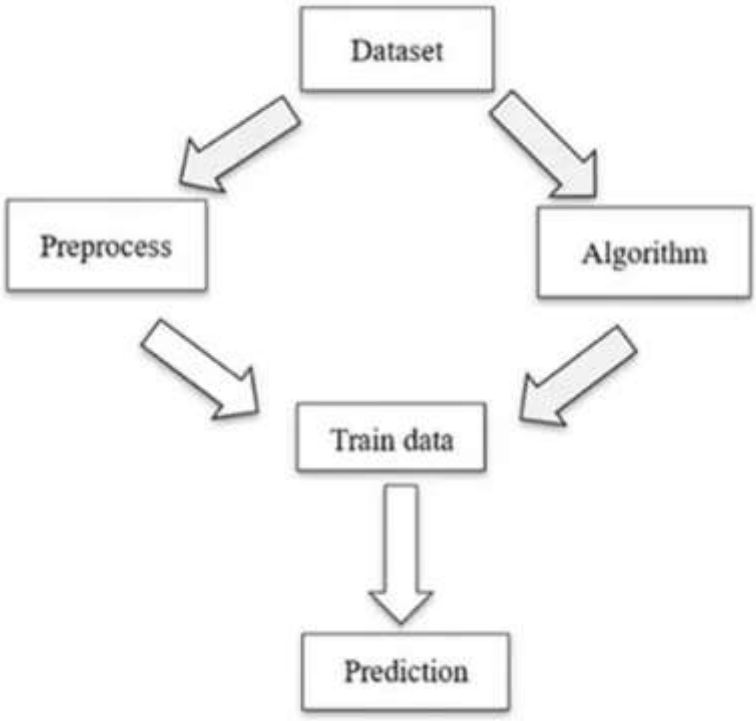


fig3.4.1: Process Diagram

In the above Process diagram, it shows a detailed explanation how our project will execute, and it predicts whether the person inherits the disease or not.

Here we explain it in the stepwise format,

To give input to the system/project we have to gather data from different repositories, and the dataset is named as crime pattern dataset which is a csv file.

In the next step data will be processed, data pre-processing is a method of preparing the raw data and making it suitable for a machine learning model. It is the first and crucial step while creating a machine learning model.

When creating a machine learning project, it is not always a case that we come across the clean and formatted data. And while doing any operation with data, it is mandatory to clean it and put in a formatted way. So, for this, we use data pre-processing task. The main part in processing the data is handling missing data removal and encoding categorical data.

Missing data removal: In this process, the null values consisting of missing values are eliminated the usage of imputer library.

Encoding Categorical data: That categorical data is described as variables with a finite set of label values. That maximum machine learning algorithms require numerical input and output variables. That an integer and one hot encoding is used to transform categorical data to integer data.

After processing the dataset, we divide our dataset into a training set and test set. This is one of the crucial steps of data preprocessing as by doing this, we can enhance the performance of our machine learning model.

After splitting the dataset, we apply classification techniques to the train data. Supervised machine learning algorithms like decision tree and random forest are applied to the dataset.

The prediction will be done by the test data provided and accuracy will be calculated, and we will provide the proof that will show which algorithm is best suitable based on accuracy.

3.5. Software Requirements Specifications

A Software Requirements specification (SRS) – a requirements specification for a software system is a complete description of behavior of a system to be developed. It includes a set of cases that describe all the interactions users will have with the software. In addition to use cases, the SRS also contains non-functional requirements. Non- functional requirements are requirements which impose constraints on the design or implementation (such as performance engineering requirements, quality standards, or design constraints). System Requirements Specification It is a collection of information that embodies the requirements of a system. A business analyst, sometimes titled system analyst, is responsible for analyzing the business needs of their clients and stakeholders to help identify business problems and propose solutions. Projects are subject to three sorts of require elements.

Business requirements describe in business terms what must be delivered or accomplished to provide value.

Product requirements describe properties of a system or product (which could be one of several ways to accomplish a set of business requirements.)

Process requirements describe activities performed by the developing organization. For instance, process requirements could specify methodologies that must be followed, and constraints that the organization must obey.

Product and process requirements are closely linked. Process requirements often specify the activities that will be performed to satisfy a product requirement. For example, a maximum development cost requirement (a process requirement) may be imposed to help achieve a maximum sales price requirement (a product requirement) a requirement that the product be maintainable (a product requirement) often is addressed by imposing requirements to follow development styles.

A system engineering, a requirement can be a description of what a system must do, referred to as Functional Requirement. This type of requirement specifies something that the delivered system must be able to do. Another type of requirement specifies something about the system itself, and how well it performs its functions. Such requirements are often called Non-functional requirements, or ‘Performance requirements’ or ‘Quality of service requirements. Examples of such requirements include usability, availability, reliability,

supportability, testability and maintainability. A collection of requirements defines the characteristics or features of the desired system. A 'good' list of requirements as far as possible avoids saying how the system should implement the requirements, leaving such decisions to the system designer. Specifying how the system should be implemented is called "implementation bias" or "solution engineering". However, implementation constraints on the solution may validly be expressed by the future owner, for example for required interfaces to external systems; for interoperability with other systems; and for commonality with other owned products.

3.5.1. Functional Requirements

In software engineering, a functional requirement defines a function of a software system or its component. A function is described as a set of inputs, the behavior, and outputs (see also software). Functional requirements may be calculations, technical details, data manipulation and processing and other specific functionality that define what a system is supposed to accomplish. Behavioral requirements describing all the cases where the system uses the functional requirements are captured in use cases. Generally, functional requirements are expressed in the form "system shall do <requirement>". The plan for implementing functional requirements is detailed in the system design. In requirements engineering, functional requirements specify particular results of a system. Functional requirements drive the application architecture of a system. A requirements analyst generates use cases after gathering and validating a set of functional requirements. The hierarchy of functional requirements is: user/stakeholder request -> feature -> use case -> business rule.

Functional requirements drive the application architecture of a system. A requirements analyst generates use cases after gathering and validating a set of functional requirements. Functional requirements may be technical details, data manipulation and other specific functionality of the project is to provide the information to the user.

The following are the Functional requirements of our system:

- It should meet the functional requirements as mentioned in Objectives.
- It should be able to find all places those are registered in Weather database.

3.5.2.NON-FUNCTIONAL REQUIREMENTS

In systems engineering and requirements engineering, a non-functional requirement is a requirement that specifies criteria that can be used to judge the operation of a system, rather than specific behavior.

The project nonfunctional requirements include the following.

- Updating Work status.
- Problem resolution.
- Error occurrence in the system.
- Customer requests.

Availability: A system's "availability" or "uptime" is the amount of time that is operational and available for use. It's related to the server providing the service to the users in displaying images. As our system will be used by thousands of users at any time our system must be available always. If there are any cases of updates, they must be performed in a short interval of time without interrupting the normal services made available to the users.

Efficiency: Specifies how well the software utilizes scarce resources: CPU cycles, disk space, memory, bandwidth etc. All of the above-mentioned resources can be effectively used by performing most of the validations at client side and reducing the workload on server by using JSP instead of CGI which is being implemented now.

Flexibility: If the organization intends to increase or extend the functionality of the software after it is deployed, that should be planned from the beginning; it influences choices made during the design, development, testing and deployment of the system. New modules can be easily integrated to our system without disturbing the existing modules or modifying the logical database schema of the existing applications.

Portability: Portability specifies the ease with which the software can be installed on all necessary platforms, and the platforms on which it is expected to run. By using appropriate server versions released for different platforms our project can be easily operated on any operating system, hence can be said highly portable.

Scalability: Software that is scalable can handle a wide variety of system configuration sizes. The nonfunctional requirements should specify the ways in which the system may be expected to scale up (by increasing hardware capacity, adding machines etc.). Our system can be easily expandable. Any additional requirements such as hardware or software which increase the performance of the system can be easily added. An additional server would be useful to speed up the application.

Integrity: Integrity requirements define the security attributes of the system, restricting access to features or data to certain users and protecting the privacy of data entered into the software. Certain features access must be disabled to normal users such as adding the details of files, searching etc which is the sole responsibility of the server. Access can be disabled by providing appropriate logins to the users for only access.

3.5.2.1. SYSTEM REQUIREMENTS

HARDWARE REQUIREMENTS

- Processor - i3(min)
- RAM - 2GB (min)
- Hard Disk - 25GB

SOFTWARE REQUIREMENTS

- Operating System : Windows/Linux
- Modules : Django
  : Matplotlib
  : NumPy
  : Urllib
  : Pandas
  : Seaborn
- Language : Python

3.6. Feasibility Study

Feasibility Study is a high-level capsule version of the entire process intended to answer several questions like: What is the problem? Is there any feasible solution to the given problem? Is the problem even worth solving? Feasibility study is conducted once the problem clearly understood. Feasibility study is necessary to determine that the proposed system is Feasible by considering the technical, Operational, and Economical factors. By having a detailed feasibility study the management will have a clear-cut view of the proposed system.

The following feasibilities are considered for the project in order to ensure that the project is variable, and it does not have any major obstructions. Feasibility study encompasses the following things.

- Technical Feasibility
- Economic or financial feasibility
- Operational feasibility

In this phase, we study the feasibility of all proposed systems, and pick the best feasible solution for the problem. The feasibility is studied based on three main factors as follows.

3.6.1. Technical Feasibility

In this step, we verify whether the proposed systems are technically feasible or not. i.e., all the technologies required to develop the system are available readily or not.

Technical Feasibility determines whether the organization has the technology and skills necessary to carry out the project and how this should be obtained. The system can be feasible because of the following grounds.

- All necessary technology exists to develop the system.
- This system is flexible, and it can be expanded further.
- This system can give guarantee of accuracy, ease of use, and reliability.
- Our project is technically feasible because, all the technology needed for our project is readily available.

3.6.2. Economic or financial feasibility

In this step, we verify which proposal is more economical. We compare the financial benefits of the new system with the investment. The new system is economically feasible only when the financial benefits are more than the investments and expenditure.

Economic Feasibility determines whether the project goal can be within the resource limits allocated to it or not. It must determine whether it is worthwhile to process with the entire project or whether the benefits obtained from the new system are not worth the costs. Financial benefits must be equal or exceed the costs. In this issue, we should consider:

- The cost to conduct a full system investigation.
- The cost of h/w and s/w for the class of application being considered.
- The development tools.
- The cost of maintenance etc.

Our project is economically feasible because the cost of development is very minimal when compared to financial benefits of the application.

3.6.3. OPERATIONAL FEASIBILITY

In this step, we verify different operational factors of the proposed systems like manpower, time etc., whichever solution uses less operational resources, is the best operationally feasible solution. The solution should also be operationally possible to implement. Operational Feasibility determines if the proposed system satisfied user objectives could be fitted into the current system operation.

- The methods of processing and presentation are completely accepted by the clients since they can meet all user requirements.
- The clients have been involved in the planning and development of the system.
- The proposed system will not cause any problem under any circumstances.
- Our project is operationally feasible because the time requirements and personnel requirements are satisfied. We are a team of four members, and we worked on this project for three wor
-

CHAPTER 4 : SYSTEM DESIGN

4.1 System Architecture

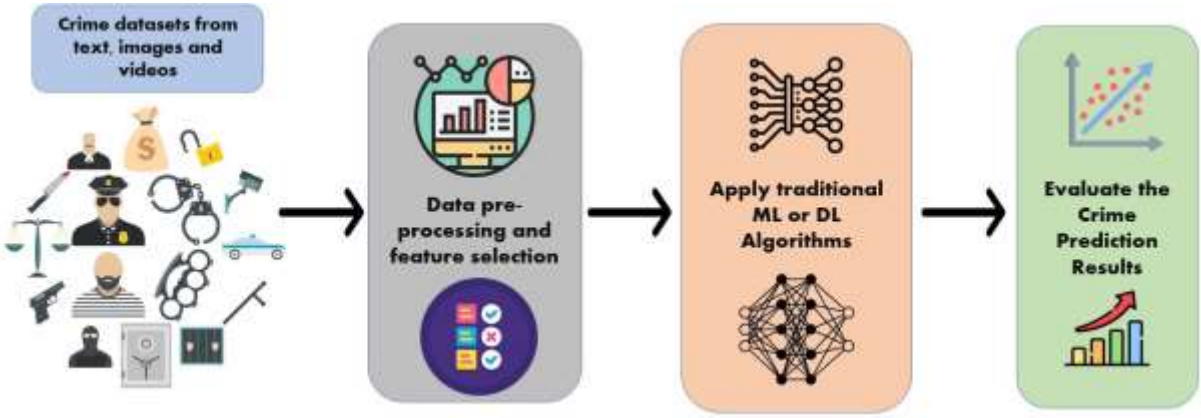


fig4.1.1: System Architecture

Crime prediction using machine and deep learning involves several major steps as shown in Figure. The first step is data collection, which involves gathering relevant data such as crime statistics, demographics, and weather patterns. The next step is data preprocessing, which includes cleaning and transforming the data into a usable format. After data preprocessing, the data is split into training and testing sets for model development and evaluation. The next step is featuring engineering, which involves selecting relevant features from the data that can be used to train the model. Once the features are selected, various machine and deep learning algorithms can be applied to the data for training and prediction purposes. Finally, the trained models are evaluated using various performance metrics to assess their accuracy and effectiveness in predicting crime. The results can be used to support decision-making in law enforcement and crime prevention efforts.

4.2 UML Diagrams

Uml is a method for describing the system architecture in detail using the blueprint. UML represents a collection of best engineering practice that has proven successful in the modeling of large and complex systems. The UML is very important parts of developing object-oriented software and the software development process. The UML uses mostly graphical notations to express the design of software projects. Using the helps UML helps project teams communicate explore potential designs and validate the architectural design of the software.

4.2.1. CLASS DIAGRAM

Class-based Modeling, or more commonly class-orientation, refers to the style of object-oriented programming in which inheritance is achieved by defining classes of objects as opposed to the objects themselves.

The most popular and developed model of OOP is a class-based model, as opposed to an object-based model. In this model, objects are entities that combine state (i.e., data), behavior and identity. The structure and behavior of an object are defined by a class, which is a definition, or blueprint, of all objects of a specific type. An object must be explicitly created based on a class and an object thus created is an instance of that class. An object is similar to a structure, with the addition of method pointers, member access control, and an implicit data member which locates instances of the class (i.e., actual objects of that class) in the class hierarchy.

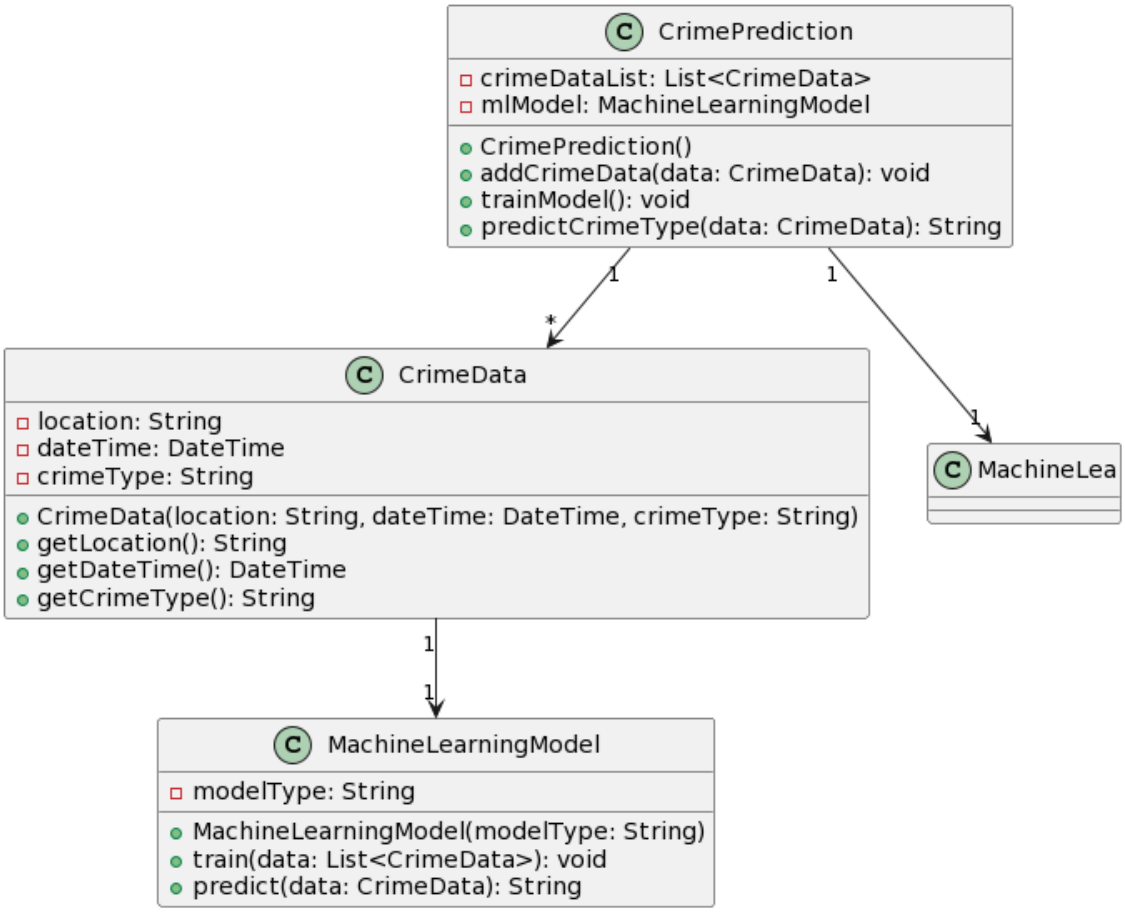


Fig4.2.1.1: Class Diagram for Crime Prediction

4.2.2. Usecase Diagram

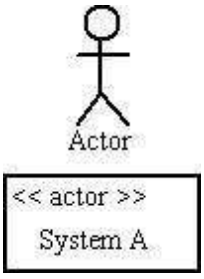
Use case diagram represents the functionality of the system. Use case focus on the behavior of the system from external point of view. Actors are external entities that interact with the system.

Use cases:

A use case describes a sequence of actions that provide something of measurable value to an actor and is drawn as a horizontal ellipse.

Actors:

An actor is a person, organization, or external system that plays a role in one or more interactions with the system.



The “user model view” encompasses a problem and solution from the preservative of those individuals whose problem the solution addresses. The view presents the goals and objectives of the problem owners and their requirements of the solution. This view is composed of “use case diagrams”. These diagrams describe the functionality provided by a system to external integrators. These diagrams contain actors, use cases, and their relationships.

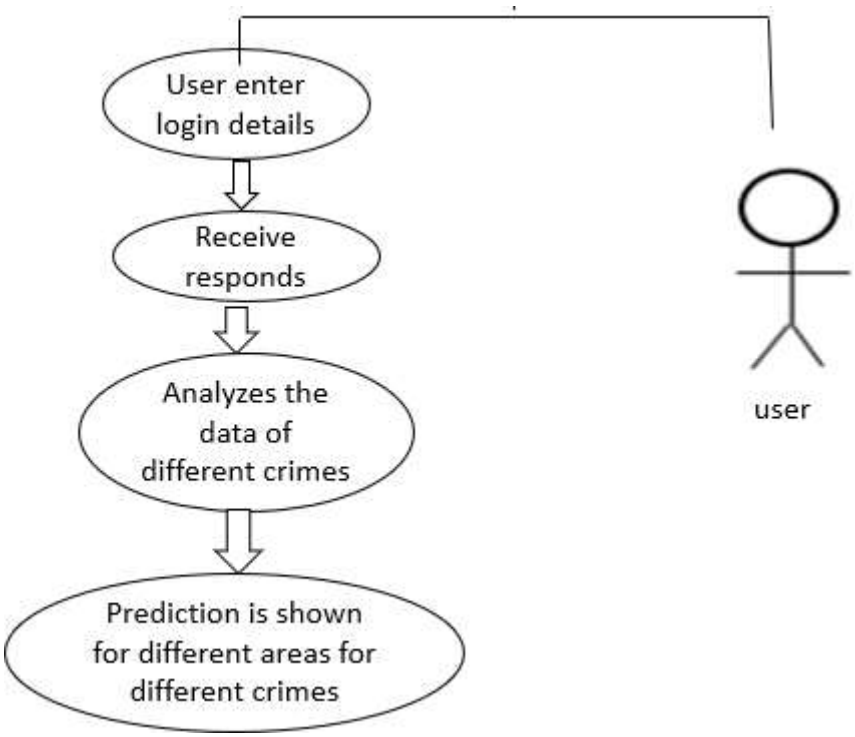


fig4.2.2.1: Class Diagram for Crime Prediction

4.2.3. Activity Diagram

Activity diagrams are graphical representations of workflows of stepwise activities and actions with support for choice, iteration, and concurrency. In the Unified Modeling Language, activity diagrams can be used to describe the business and operational step-by-step workflows of components in a system. An activity diagram shows the overall flow of control.

Activity diagrams are constructed with different types of shapes, connected with arrows. The most important shape types:

- Rounded rectangles represent activities.
- Diamonds represent decisions.
- Bars represent the start (split) or end (join) of concurrent activities.
- A black circle represents the start (initial state) of the workflow.
- An encircled black circle represents the end (final state).

Arrows run from the start towards the end and represent the order in which activities happen. However, the join and split symbols in activity diagrams only resolve this for simple cases, the meaning of the model is not clear when they are arbitrarily combined with the decisions or loops.

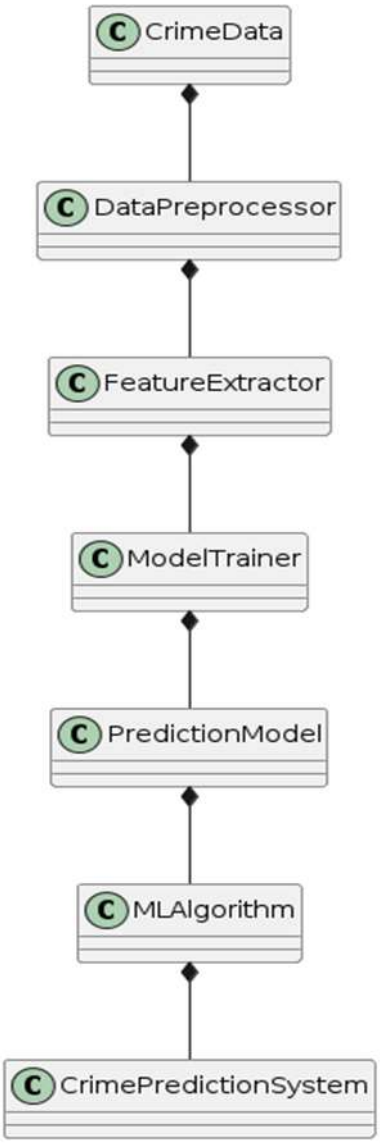


Fig4.2.3.1 Activity Diagram for Crime Prediction

4.2.4. Object Diagram

An object diagram is a type of UML (Unified Modeling Language) diagram that shows a snapshot of the instances of classes and their relationships at a specific moment in time. It is used to represent the static view of a system, displaying instances of classes and their relationships.

An object diagram consists of objects, which are instances of classes, and links, which represent the relationships between objects. Objects are represented as rectangles with the object's name, the object's class, and the object's attributes listed inside. Links are represented as lines connecting objects, with the number of participants on the link for each object listed at the end of the link.

Object diagrams are used for various purposes, such as:

- Forward and reverse engineering
- Understanding object behavior and their relationships from a practical perspective
- Capturing a particular system's state at a specific moment
- Modeling complex data structures
- Making the prototype of a system

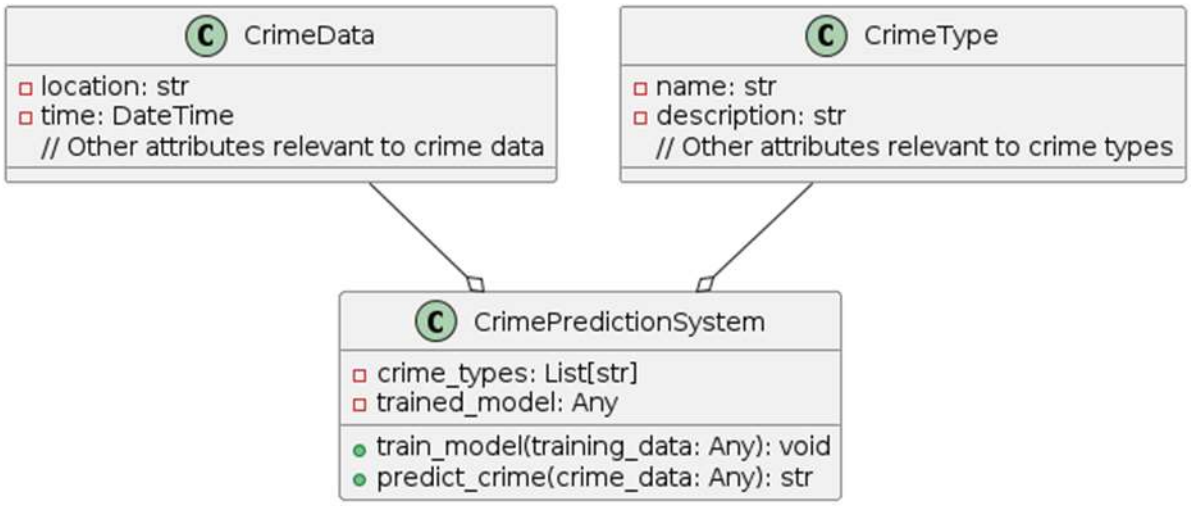


fig4.2.4.1: Object Diagram for crime prediction

4.2.5. Sequence Diagram

A sequence diagram in Unified Modeling Language (UML) is a kind of interaction diagram that shows how processes operate with one another and in what order. It is a construct of a Message Sequence Chart.

Sequence diagrams are sometimes called event diagrams, event scenarios, and timing diagrams. A sequence diagram shows, as parallel vertical lines (lifelines), different processes or objects that live simultaneously, and as horizontal arrows, the messages exchanged between them, in the order in which they occur. This allows the specification of simple run time scenarios in a graphical manner. If the lifeline is that of an object, it demonstrates a role. Note that leaving the instance name blank can represent anonymous and unnamed instances. To display interaction, messages are used. These are horizontal arrows with the message name written above them.

Activation boxes, or method-call boxes, are opaque rectangles drawn on top of lifelines to represent those processes are being performed in response to the message.

Objects calling methods on themselves use messages and add new activation boxes on top of any others to indicate a further level of processing. When an object is destroyed (removed from memory), an X is drawn on top of the lifeline, and the dashed line ceases to be drawn below it. It should be the result of a message, either from the object itself, or another.

A message sent from outside the diagram can be represented by a message originating from a filled-in circle or from a border of sequence diagram.

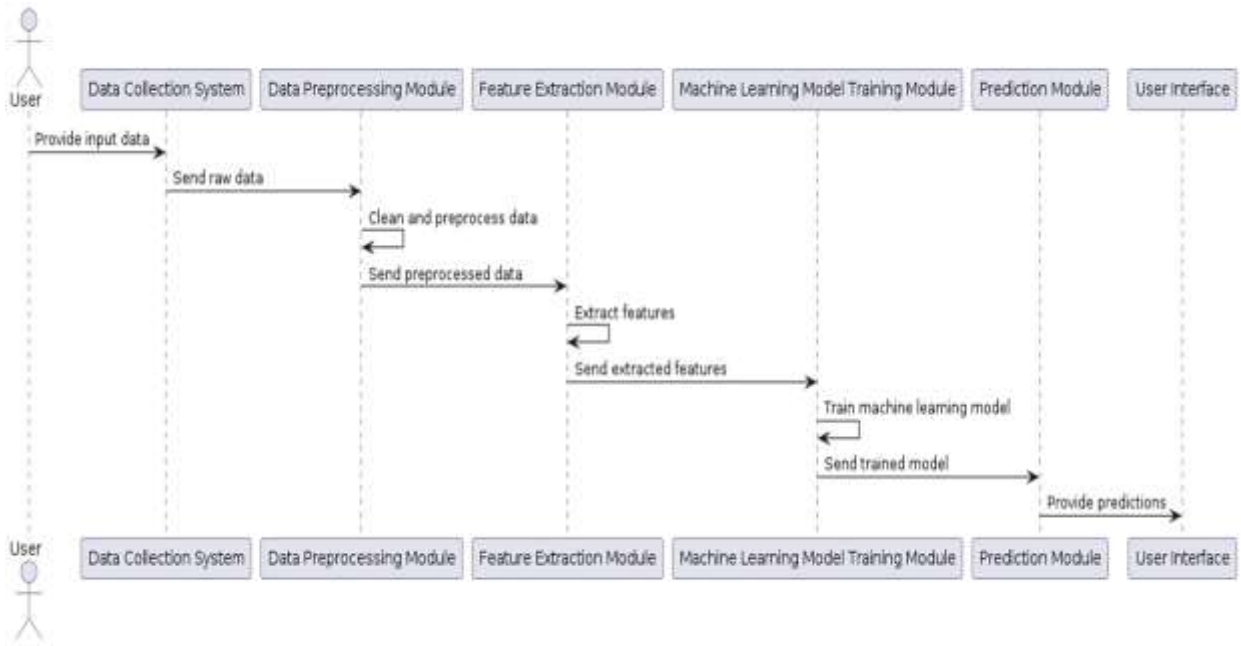


Fig4.2.5.1: SEQUENCE DAIGRAM FOR CRIME PREDICTION

4.2.6. Collaboration Diagram

The collaboration diagram is used to show the relationship between the objects in a system. Both the sequence and the collaboration diagrams represent the same information but differently. Instead of showing the flow of messages, it depicts the architecture of the object residing in the system as it is based on object-oriented programming. An object consists of several features. Multiple objects present in the system are connected to each other. The collaboration diagram, which is also known as a communication diagram, is used to portray the object's architecture in the system.

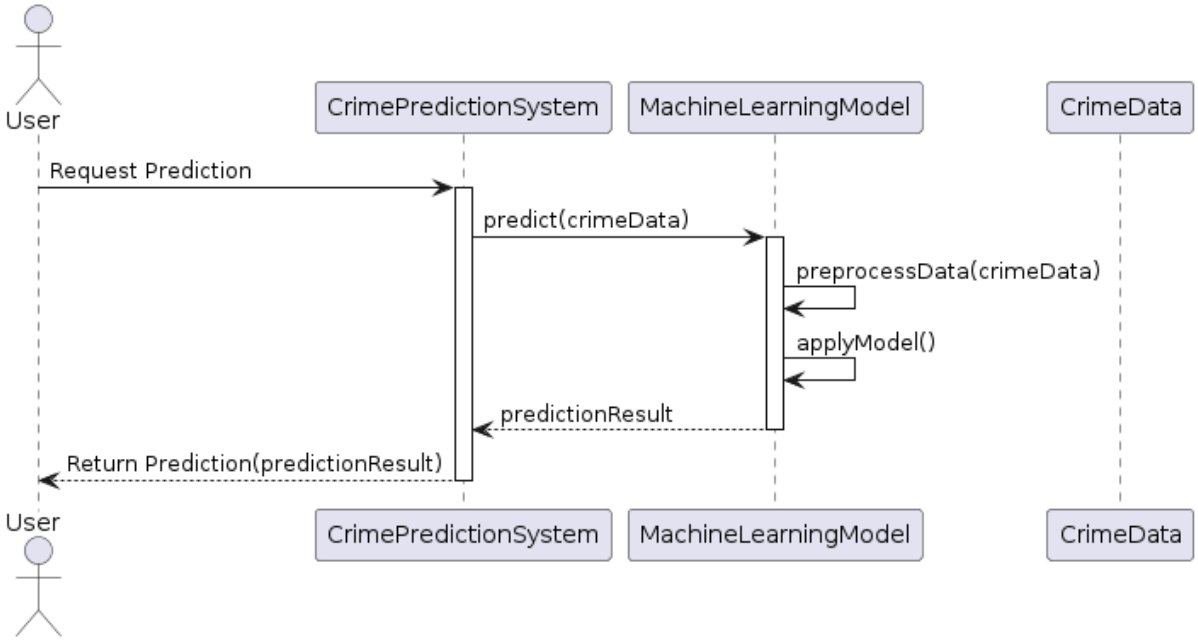


Fig4.2.6.1: Collaboration Diagram for crime prediction

4.4.7. Component Diagram

This section defines the term component and discusses the differences between object oriented, traditional, and process related views of component level design. Object Management Group OMG UML defines a component as “a modular, deploy able, and replaceable part of a system that encapsulates implementation and exposes a set of interfaces.”

A component contains a set of collaborating classes. Each class within a component has been fully elaborated to include all attributes and operations that are relevant to its implementation.

As part of the design elaboration, all interfaces that enable the classes to communicate and collaborate with other design classes must also be defined. To accomplish this, the designer begins with the analysis model and elaborates analysis classes and infrastructure classes.

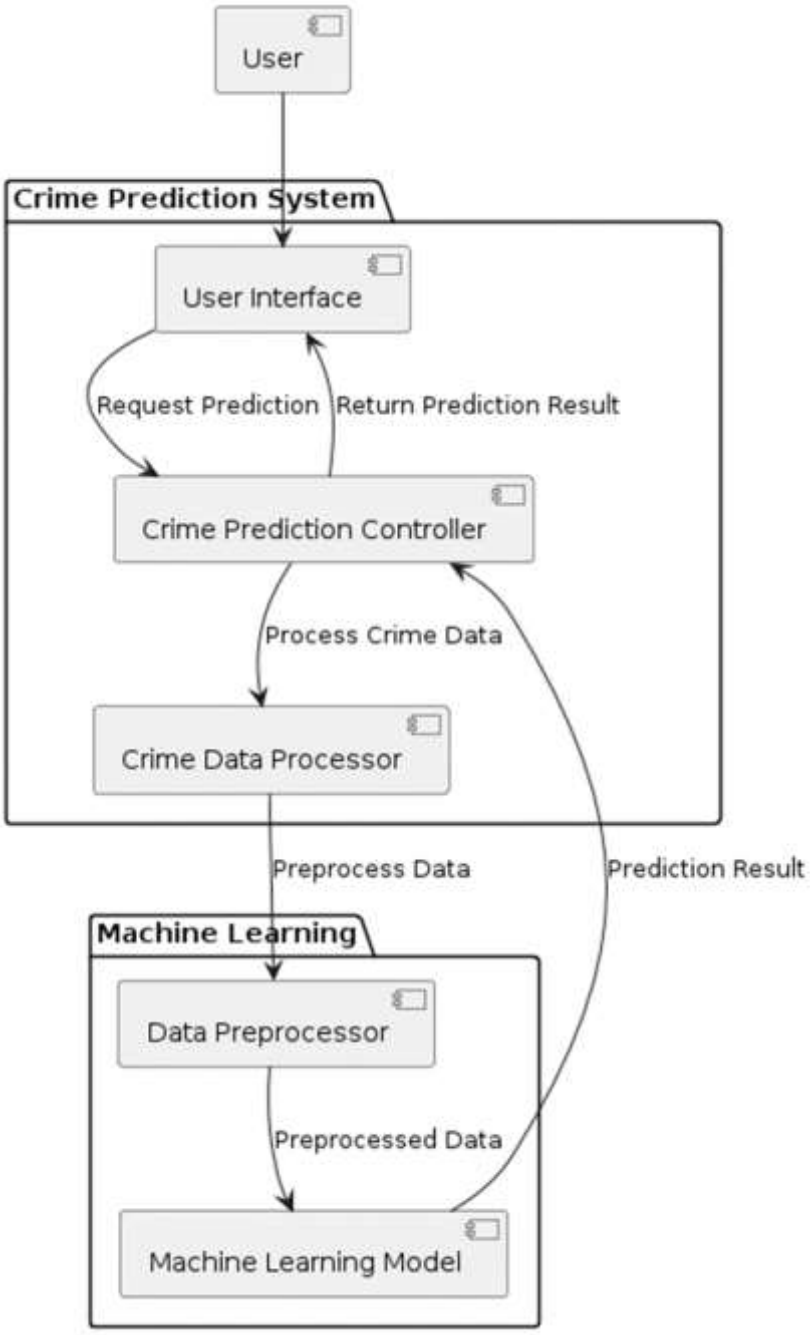


Fig4.2.7.1: Component Diagram for Crime Prediction

4.2.8 DEPLOYMENT DIAGRAM

Deployment diagrams are used to visualize the topology of the physical components of a system where the software components are deployed.

So, deployment diagrams are used to describe the static deployment view of a system. Deployment diagrams consist of nodes and their relationships.

Purpose:

The name Deployment itself describes the purpose of the diagram. Deployment diagrams are used for describing the hardware components where software components are deployed. Component diagrams and deployment diagrams are closely related.

Component diagrams are used to describe the components and deployment diagrams shows how they are deployed in hardware.

UML is mainly designed to focus on software artifacts of a system. But these two diagrams are special diagrams used to focus on software components and hardware components.

So, most of the UML diagrams are used to handle logical components but deployment diagrams are made to focus on hardware topology of a system. Deployment diagrams are used by the system engineers.

The purpose of deployment diagrams can be described as:

- Visualize hardware topology of a system.
- Describe the hardware components used to deploy software components.
- Describe run time processing nodes.

Deployment diagrams are useful for system engineers. An efficient deployment diagram is very important because it controls the following parameters

- Performance
- Scalability
- Maintainability
- Portability

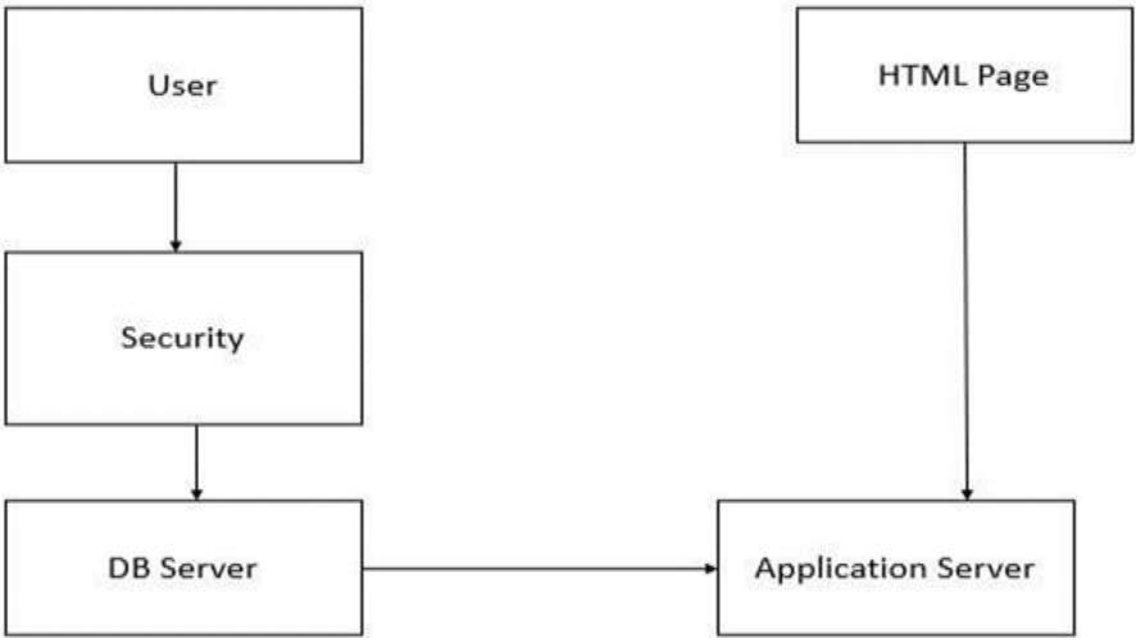


Fig 4.2.8.1: Deployment diagram for crime prediction

4.3 Modules

4.3.1. Data pre-processing

Data gathered from open sources must first be pre-processed to prevent unnecessary violations. Because it contains an enormous amount of crime data over six years, the dataset for Denver city was selected. The machine learning technique filter and wrapper is used to find the missing integral in the supplied attribute values. Both the training of a prediction model and the efficiency of the process require data purification. Datasets are filtered for instances, and unnecessary context is eliminated. The importance of the attributes is determined in part by the filtering techniques.

When choosing characteristics, the correlation with the dependent variables is taken into consideration. By building a prediction model on the feature subset, the imposed wrapper approach is utilised to ascertain how useful the feature subset is. The data is separated into test and training characteristics after pre-processing.

4.3.2. Mapping

The criminal characteristics such as crime type, date of occurrence, and order of occurrence are initially separated. Then, in order to make labelling easier, it is converted to an integer.

Following further analysis, the tagged features are utilised for graph charting. Python was selected as the programming language to do the required job since it works well with machine learning. Using the matplotlib package, a graph showing the prevalence of illegal behaviour is produced. The graph can be used to plot the crime that occurs the most frequently, which aids in making predictions.

4.3.3 Model Selection

In this work, the Random Forest classifier is utilised. When compared to previously composed works, the accuracy obtained was relatively high. In crime prediction, random forest algorithm can be used to analyze various factors that contribute to criminal activity, such as location, time of day, demographics, and historical crime data. By analyzing these factors, the algorithm can predict the likelihood of criminal activity in a particular area or at a particular time. The random forest classifier was implemented on the pre-processed dataset using the Scikit-learn machine learning library. It can handle large and complex of data set is resistant to overfitting, and can handle missing values and outliers. This method makes a prediction by combining multiple decision trees. To create decision trees, the algorithm randomly selects subsets of features and data points from the training data set. The individual decision trees are then integrated to create a final prediction.

4.3.4 Crime Prediction

Predicting the type of crime that might occur in a certain location at a specific time is the major objective of our research. With the random forest classifier technique, we were able to accomplish this objective with a respectable level of accuracy. Four related features must be provided to predict the type of crime. The following characteristics are required: the occurrence month, the time, place, and day of the week that the crime occurred. You can enter all characteristics together with their nominal values.

4.4.5. Evaluation

The performance of the proposed prediction is then evaluated in order to get a high degree of accuracy when compared to the previously used model. Cross validation is used in training to aid in training the data on a variety of training data sets. It will validate the overall divides in the assumed cross validation. To calculate the accuracy number in Python, we must provide data parameters such as the model's name, target set, and cv, which aid in showing the split occurrence. Finally, the mean and standard deviation of the average precision are computed.

CHAPTER 5 : ALGORITHMS

5.1. Algorithm

Step-1: Begin the tree with the root node, says S, which contains the complete dataset.

Step-2: Find the best attribute in the dataset using Attribute Selection Measure (ASM).

Step-3: Divide the S into subsets that contains possible values for the best attributes.

Step-4: Generate the decision tree node, which contains the best attribute.

Step-5: Recursively make new decision trees using the subsets of the dataset created in step -3. Continue this process until a stage is reached where you cannot further classify the nodes and called the final node as a leaf node Classification and Regression Tree algorithm.

5.2. Sample Code

```
#!/usr/bin/env python
"""Django's command-line utility for administrative tasks."""
import os
import sys
def main():
    os.environ.setdefault('DJANGO_SETTINGS_MODULE', 'crime_pattern.settings')
    try:
        from django.core.management import execute_from_command_line
    except ImportError as exc:
        raise ImportError(
            "Couldn't import Django. Are you sure it's installed and "
            "available on your PYTHONPATH environment variable? Did you "
            "forget to activate a virtual environment?"
        ) from exc
    execute_from_command_line(sys.argv)
if __name__ == '__main__':
    main()
```

CHAPTER 6 : TESTING

6.1. Testing Strategies

Testing is the process where the test data is prepared and is used for testing the modules individually and later the validation given for the fields. Then the system testing takes place which makes sure that all components of the system property function as a unit. The test data should be chosen such that it passed through all possible condition. Actually, testing is the state of implementation which aimed at ensuring that the system works accurately and efficiently before the actual operation commence. The following is the description of the testing strategies, which were carried out during the testing period.

SYSTEM TESTING

Testing has become an integral part of any system or project especially in the field of information technology. The importance of testing is a method of justifying, if one is ready to move further, be it to be check if one is capable to with stand the rigors of a particular situation cannot be underplayed and that is why testing before development is so critical.

When the software is developed before it is given to user to use the software must be tested whether it is solving the purpose for which it is developed. This testing involves various types through which one can ensure the software is reliable.

The program was tested logically and pattern of execution of the program for a set of data are repeated. Thus, the code was exhaustively checked for all possible correct data and the outcomes were also checked.

MODULE TESTING

To locate errors, each module is tested individually. This enables us to detect error and correct it without affecting any other modules. Whenever the program is not satisfying the required function, it must be corrected to get the required result.

Thus, all the modules are individually tested from bottom up starting with the smallest and lowest modules and proceeding to the next level. Each module in the system is tested separately. For example, the job classification module is tested separately. This module is tested with different job and its approximate execution time and the result of the test is compared with

the results that are prepared manually. The comparison shows that the results proposed system works efficiently than the existing system.

Each module in the system is tested separately. In this system the resource classification and job scheduling modules are tested separately and their corresponding results are obtained which reduces the process waiting time.

UNIT TESTING

Unit testing is usually conducted as part of a combined code and unit test phase of the software lifecycle, although it is not uncommon for coding and unit testing to be conducted as two distinct phases.

6.2. Test Cases

We now know, test cases are integral part of testing. So we need to know more about test cases and how these test cases are designed. The most desired or obvious expectation from a test case is that it should be able to find most errors with the least amount of time and effort

A software product can be tested in two ways. In the first approach only the overall functioning of the product is tested. Inputs are given and outputs are checked. This approach is called black box testing. It does not care about the internal functioning of the product.

The other approach is called white box testing. Here the internal functioning of the product is tested. Each procedure is tested for its accuracy. It is more intensive than black box testing. But for the overall product both these techniques are crucial. There should be enough tests in both categories to test the overall product.

CHAPTER 7 : IMPLEMENTATIONS

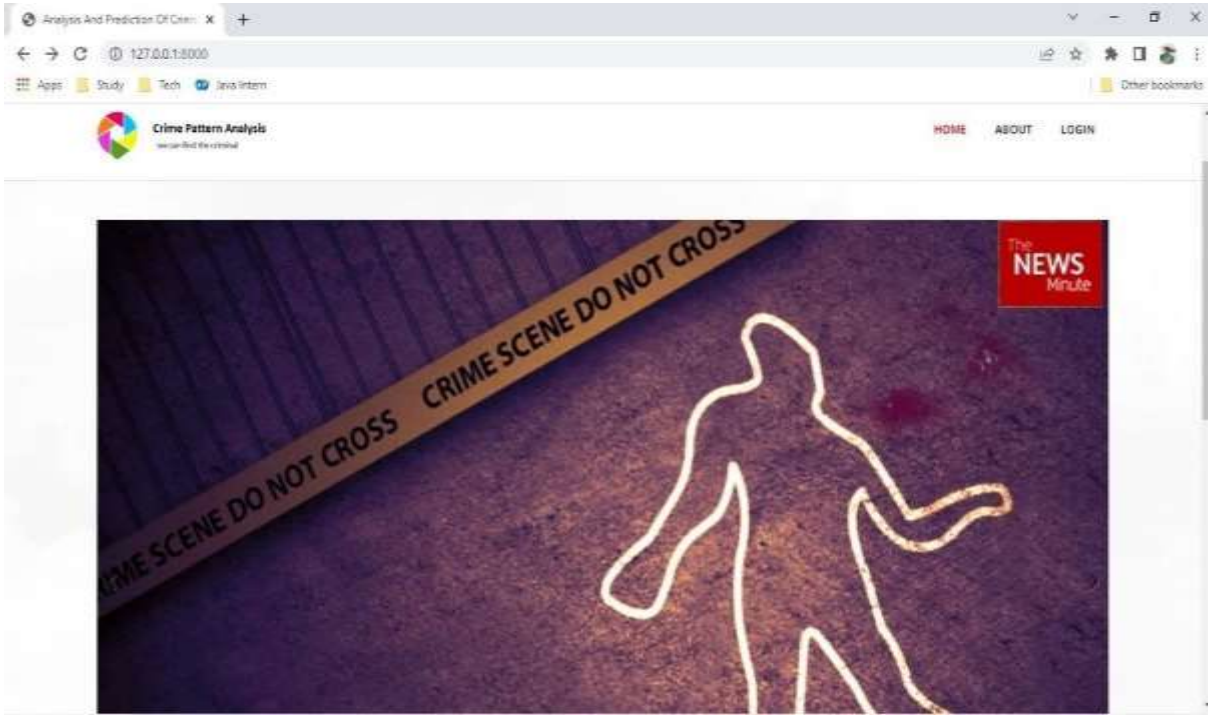


fig7.1: Homepage

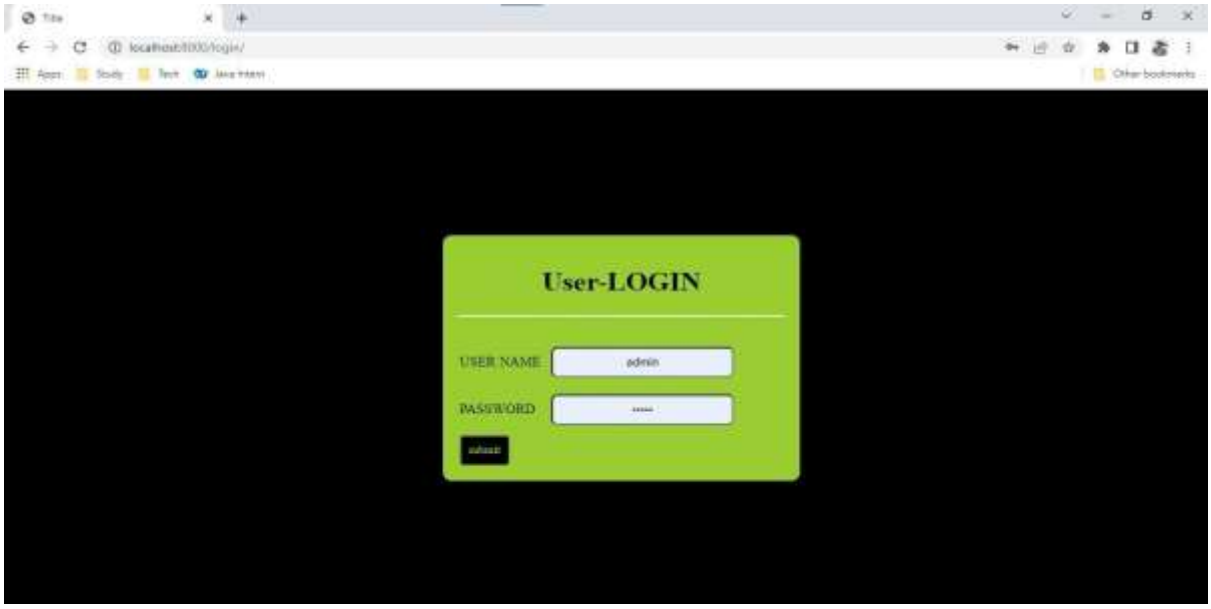


fig7.2: Login Page

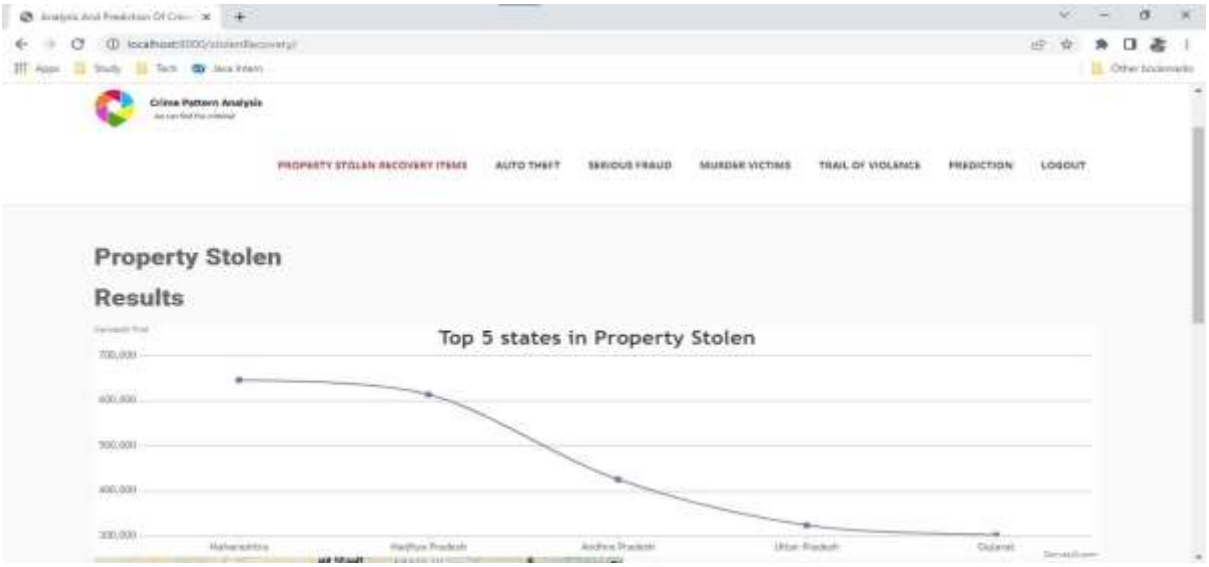


fig7.3: Property Stolen Output

Fig7.4: Serious Fraud

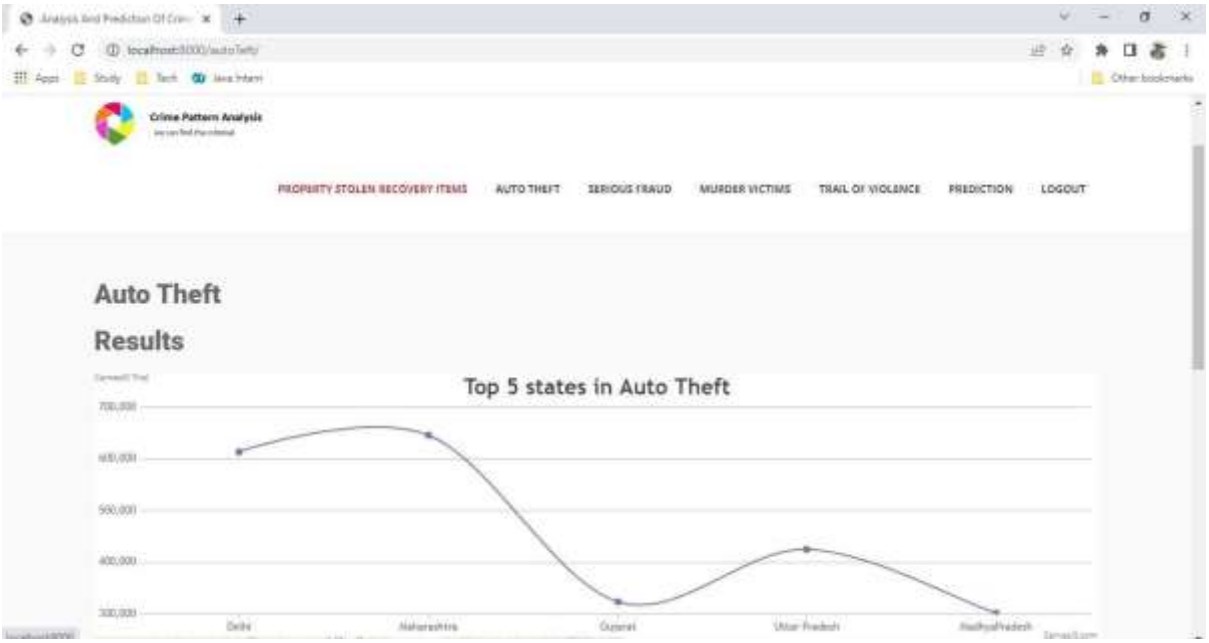


fig7.5: Auto Theft

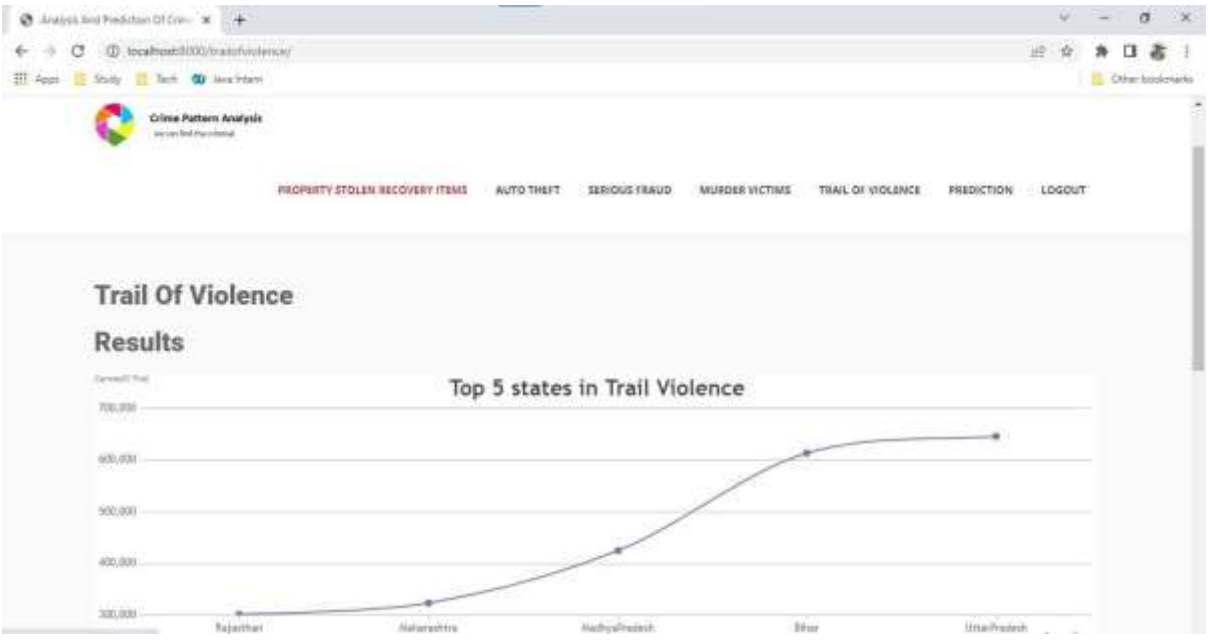


fig7.6: Trail of Violence

fig7.7: Murder Victim

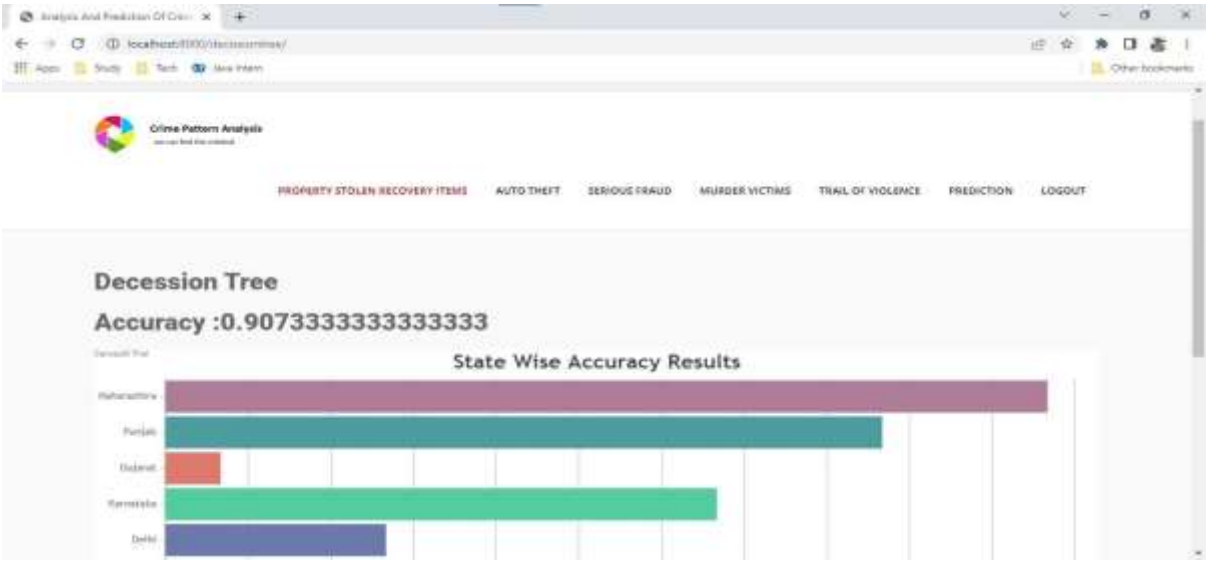


fig7.8: Prediction

CHAPTER 8 : CONCLUSION AND FUTURE ENHANCEMENTS

CONCLUSION

The main aim of the project is to make change by reducing the crime rate with the help of machine learning. The crime rates are increased day-by-day due to many factors such as theft, auto theft, trail of violence, murder etc. The proposed model is very useful for both investigating agencies and police officials in taking necessary steps to reduce crimes. The majority deals with the prediction of regions that have a high probability for crime occurrencesand visualize crime prone areas. This usually helps the crime force to draw conclusions about the crime that occurred at a time. This project helps the crime analysis to analyze these crime networks by means of various interactive visualizations.

FUTURE ENHANCEMENT

In the future, it is possible to provide extensions or modifications to the proposed analysis and prediction algorithms to achieve further increased performance. It is also possible to use other machine learning algorithms and can implement a real-world application for the user experience in a simple manner in sucha way that crimes can be reduced subsequently. The future enhancement for this project will be based on the crime rates that increase in a particular area. Machine learning and various other optimization techniques can also be used so that the evaluation results can again be increased. More different ways of normalizing the data can be used and the results can be compared.

REFERENCES

[1] http://www.cs.sunysb.edu/~cse634/lecture_notes/07apriori.pdf.

[2] http://www-users.cs.umn.edu/~kumar/dmbook/ch4.pdf.

[3] Lior Rokach and Oded Maimon. Decision trees. In Oded Maimon and Lior Rokach, editors, The DataMining and Knowledge Discovery Handbook, pages 165–192. Springer, 2005

[4] http://www.iiisci.org/journal/pdv/sci/pdfs/QN851PS.pdf

[5] http://cse.anits.edu.in/projects/projects1920C13.pdf

[6] Malathi. 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

[7] Tong Wang, Cynthia Rudin, Daniel Wagner, and Rich Sevier. 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.

[8] Wikipedia contributors. (9 July 2013), Stanford NLP. [Online]. Available: http://www- nlp.stanford.edu/software/dcoref.shtml. Last accessed: 24-Feb-2014, 10:00 AM.

[9] Eibe Frank and Remco R. Bruckert. 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, 2006Springer-Verlag.