1. INTRODUCTION

1.1 PROBLEM STATEMENT

Numerous issues related to crime keep ravaging cities all over the world. Violence, theft, drug trafficking, and cybercrime are ongoing problems that compromise the security and wellbeing of metropolitan areas. Fear-based environments due to high crime rates affect the quality of life of residents and restrict investment and economic growth. Socio-economic inequalities, a lack of proper law enforcement resources, gang activity, and the growth of illegal firearms are all factors that contribute to these issues. A comprehensive strategy that combines community involvement, focused law enforcement efforts, investment in education and job prospects, and the use of cutting-edge technology to prevent and solve crimes is all needed to combat crime.

The development and education of individuals within a community are significantly impacted by crime. Crime generates a climate of fear and insecurity that has a negative impact on people's general wellbeing and personal growth. Here are some specific ways that crime may affect development and education:

- High crime rates can cause disruptions in the learning environment because schools may need to put in place safety measures that obstruct the teaching and learning process. It might be challenging for students and teachers to concentrate on learning when there are frequently violent or theft-related situations.
- People who live in underprivileged communities, where there may already be socio-economic difficulties, are disproportionately affected by crime. Crime hampers access to opportunities and resources, including good schools, extracurricular activities, and educational support programmes. This restricts opportunities for career advancement and personal improvement.
- Criminal activity may reinforce the cycle of poverty. Economic investment and potential are frequently reduced in high-crime areas because companies may be hesitant to operate there. The cycle of poverty is maintained by the absence of economic security and work opportunities, which prevents people from enrolling in high-quality schools and hampers their overall development.

- ❖ Direct or indirect exposure to crime can have long-lasting psychological impacts on people, particularly on youngsters. It may result in more anxiety, depression, and trauma, which may impede the growth of the brain and academic achievement.

It is crucial to address the core causes of crime, enhance community safety, and offer support structures that foster a safe and conducive learning environment in order to lessen the harmful effects of crime on development and education. In order to provide safe places that foster growth and guarantee equal access to high-quality education for everyone, this calls for comprehensive efforts by law enforcement, community organizations, lawmakers, and educational institutions.

Thus, there is a need for a website that predicts the safety of a place based on crime because it enables communities to take preventative action against criminal activity. Such a website can offer insights and forecasts on possible crime hotspots or patterns by analyzing past crime data, demographic data, and other pertinent aspects. With the help of this data, law enforcement organizations, decision-makers, and community members may better allocate resources, put into practice targeted prevention measures, and increase public safety. A crime prediction website can help reduce criminal events by foreseeing and preventing crimes, improving the safety and well-being of people and communities.

1.2 PROJECT SPECIFICATION

1.2.1 PROJECT SCOPE

In order to determine the probabilities of crimes occurring in a specific location, this project aims to create a website that makes use of a variety of features, including per capita income, male literacy rate, and the overall population. The website will offer insightful information to people, law enforcement organizations, and decision-makers, assisting in the creation of strategies to successfully prevent and handle criminal activity. It will also help people to get the insights of crime happened in a particular area and the type of crime happened in percentage value, represented through pie-chart along with the facility of helpline number and links which will redirect users directly to state government websites, where they can easily file e-FIR if they wish or get any other related information.

1.2.2 KEY FEATURES

- Data collection: To procure data on per capita income, male literacy rate, and overall population, the website will collect data from reputable sources such as government databases, surveys, and research papers. The main input for the models used to predict crime will be this data.

- Data cleaning and preparation: For the purpose of creating reliable crime prediction models, machine learning algorithms will be used. These models will examine the gathered data to find trends and connections between the chosen characteristics and crime rates. Linear regression, Support vector classifier(SVC), decision trees, and K-nearest neighbor(KNN) are just a few examples of the various algorithms that can be tested to produce precise predictions.

- Feature Selection: For the purpose of predicting crime rates, identify the appropriate features like per capita income, literacy rate, and total population, on which crime is highly dependent.

- Model Training: Apply machine learning algorithms to train the model that is built using the chosen features.

- Model Evaluation: Analyze the precision and reliability of the predictive model.

- Interface development: Users will be able to enter specific geographic areas for which they want to check safety measures using the website's user-friendly interface, which will allow for crime prediction. Additionally, users will have the option to enter the per-capita income, male literacy and total population for more accurate predictions. They can also check for some details like male and female population, number of literates, etc. of the city they choose from the listed cities. Alongwith this, users can get the insights of types of crime occurred in a particular year, in the city in the form of pie-charts and they get the helpline contact details and links to government websites where they can easily file an e-FIR and get other related information. The interface will be simple to use, visually appealing, and offer thorough and useful information.

- Information circulation: Users of the website will have access to interactive crime visualizations, and predictions that are simple to grasp. In order to enable individuals and communities to take preventative action, it will also provide safety advice, crime prevention techniques, and helpline contacts.

- Real-time updates: The website will contain real-time data updates to guarantee both the accuracy and relevance of the forecasts. This can be accomplished by integrating with trustworthy data sources and applying a suitable rate of change in per capita income, literacy rate and total population within the years.

- Feedback and user engagement: A feedback feature on the website will encourage user participation by enabling users to report any errors in the predictions. This feedback will make the website work more efficiently overall and enhance the precision of future crime predictions. Also there is a

helpline section in the website from where users can contact officials in case of any emergency.

- Privacy and security: By taking the necessary steps to protect user information and comply with applicable data protection laws, the project will give data privacy and security utmost priority. We will treat and retain any personally identifiable information with the highest care.

- Scalability and maintenance: It includes making sure that the website is capable of developing and being maintained. Scalable infrastructure must be used, software and algorithms must be updated frequently, and any flaws or technical problems must be fixed.

The crime prediction website will be a useful resource for citizens, law enforcement organizations, and decision-makers to help them decide how to best deploy resources. The website will offer insights into crime patterns by utilizing features like per capita income, male literacy rate, and total population, assisting in crime prevention and guaranteeing community safety. The goal of the website is to create a comprehensive and intelligent platform that makes use of geospatial technology, machine learning algorithms, and data analytics to anticipate and discourage criminal activity. It gives the public useful information, enabling people to take safeguards and decide with knowledge to improve their own safety. To learn about potential threats in their neighbourhoods and to take the necessary precautions to protect themselves and their property, residents can consult the website. The insights from crime forecasting can be used to improve urban planning, allocate resources for community development projects, and develop crime prevention techniques.

1.3 HARDWARE SPECIFICATIONS

Name	Minimum requirement
OS	Windows 10/ Intel Core I3/8th Gen
Processor	1Ghz (recommended 2 Ghz or more)
Installed Memory	4GB RAM
Internet connection	1.2 Mbps
IDE	Jupyter Notebook, Google Colab Notebook
Hard Disk	40 GB
Mouse	Logitech

Table 1: Hardware Specifications

1.5 SOFTWARE SPECIFICATIONS

Technology	Purpose
React Js	For creating interactive user interfaces
Django	Python-web framework for connecting machine learning section with frontend section.
Visual Studio	For complete overall development of the project in one place.
Jupyter Notebook	For developing predictive model using machine learning algorithms.
Bootstrap	For making our website responsive.
Javascript	For making web pages more dynamic.
HTML & CSS	Core technologies for creating websites.
Node js	For connecting the backend section with the database.
MongoDb	For creating databases.

Table 2: Software Specifications

1.6 MODEL SPECIFICATION:

Data Sources	Historical crime records (2011 census data)
Feature Selection	Per-capita income, total population, male and female literacy rate, types of crimes in numbers, area(state and districts)
Model Selection	K-Nearest Neighbour Algorithm(KNN)
Type of learning	Supervised Learning
Training and testing Model	Training data: 80% of dataset Testing data: 20% of dataset
Integrating model	Using Django
Accuracy	Almost 100%
F1 Score	

Table 3: Model Specification

2. LITERATURE SURVEY

2.1 EXISTING SYSTEM

A prevalent social issue that has an impact on a society's standard of living and economic development is crime. It is regarded as a crucial element in deciding whether or not individuals relocate to a new city and which locations should be avoided when travelling.

Websites that predict crime rates in a particular area or location often use the total number of crimes committed and the total population of the area or location. While this is the actual way of calculating crime rate i.e., total crimes divided by the total population; users cannot obtain the information needed by the websites which is total number of various crimes such as arson, theft, murders, rapes etc.

The existing crime prediction models use crime information derived from law enforcement agencies, such as police reports and records. These databases include details on the crime's kind, scene, timing, and other pertinent factors.

Traditional crime rate predictor websites do not use any machine learning algorithm and if they do, it's not necessary at all since all you have to do is divide the total crimes which are entered by the user and the total population which is again given by the user. So all the work that is to be done is to calculate the crime rate and determine if it's more than the average crime rate or not. There is not much of a machine learning process in its prediction.

Law enforcement must take preventive action in order to lower the crime rate. There is a need for cutting-edge systems and fresh ideas for enhancing crime analytics in order to protect communities from criminal activity. Since crime occurs as a result of numerous complicated causes, accurate real-time crime predictions serve to lower the crime rate but remain a difficult topic for the scientific community.

2.2 PROPOSED SYSTEM

As we have discussed earlier in the previous section, total crimes done are not easily procurable and sometimes impossible to do because of the lack of data or inaccurate data. Our system tries to overcome these issues by using different parameters such as per capita income, literacy rate and total population of the city or area which are very easily available on a search away on the Internet.

The system designed by us takes in per capita income of the city, male literacy rate and the total population of the city which is then fed to our prediction algorithm and thus the result is obtained i.e., one of the 3 values- Safe, Moderate safe, Not safe.

The parameters chosen by us for our system were chosen on a series of experimentation and discussions. Other parameters that could not be included in the proposed system were unemployment rate and poverty levels since these are not easily available, it couldn't be included in the final model of the system.

Other features in our system are-

1. Emergency helpline numbers: In case of any emergency situations, users can select the state whose emergency contact information they need. This is done to get a whole coverage of everything crime related and so that our system becomes complete. On the basis of this information, we can also determine which cities have called on the emergency contacts to further improve our system in the future.
2. e- FIR links : This feature will redirect users to government websites directly, which help them to register their complaints and check the status of their registered complaints and also seek any governmental aid.
3. Crime distribution chart : This feature enables user to check for distribution of different types of crimes that happened in a city in a particular year, in the form of pie-cahrt like robbery, murder, theft, rape, burglary, etc. This will help users to make decisions whether to relocate to that place or not.

4. Feedback form: This feature enables users to report any bugs in the system so that we can learn from them and modify our system likewise in future versions.
5. Data availability: This is to provide users with our database if they require so. Users can select the states,cities and the information they need about the cities.
6. Updation: This feature is for the admin to carefully go through users feedback and update the changes if needed in the dataset. Moreover, they add any new information about the cities or remove entries.

Since crimes are rising day by day, it is necessary to deploy steps ensuring that crimes are stopped and since we cannot take part in this task, we have designed our website to do our part in properly informing users about the safety of cities they wish to visit or relocate.

Advantages of Proposed System:

- Reduce paper work
- User friendliness and interactive
- All crime details at one place
- Helpline facility
- Direct links to government portal, where one can easily register their complaint
- Feedback
- Help police department and government to take necessary actions
- Data security
- Interactive charts to display distribution of crimes at different place
- Easy to use

2.3 FEASIBILITY STUDY

Predicting and preventing crime have grown to be major problems for communities all around the world in recent years. The concept of a Crime glance website has arisen as a viable tool for boosting public safety because of recent developments in technology and data analysis. The purpose of this feasibility study is to assess the viability and potential advantages of creating a website that predicts crime and provides crime distribution at a glance. To decide whether such a website is practical and desirable, the study evaluates the technical, economic, and social elements.

Technical Feasibility

There are a number of important factors to take into account while assessing the technical viability of a crime prediction website, like:

- Data Availability: For precise projections, relevant crime data must be available and of high quality. It is vital to have access to detailed crime records that include the place, kind, and time of the incident. To guarantee data availability and integrity, cooperation with government organizations and law enforcement agencies is crucial.
- Data Processing and Analysis: For the purpose of making accurate predictions, sophisticated data processing and analytical procedures must be used. Data visualization tools, statistical models, and machine learning algorithms will be used to find patterns and trends in crime data.
- Scalability and Performance: The website should be able to handle enormous amounts of data and offer forecasts in real time. To handle increased user traffic and retain maximum performance, it is crucial to deploy scalable technologies and build a reliable infrastructure.

Economic Feasibility

The financial viability of creating and sustaining a crime prediction website is examined.

- Development Cost: The website, database systems, data integration, and analytics infrastructure were all developed as part of the initial

expenditure. Additional expenses could result from working with data suppliers and technological partners.

Social Feasibility

The social viability looks at the influence and impact on people's lives of a crime prediction website.

- User Acceptance: Surveys, user feedback sessions, etc. can be used to get information on how the community feels about and is willing to utilize a certain website. To win the public's trust, it will be crucial to guarantee user privacy and data security.
- Ethical Considerations: The advantages of crime prediction must be weighed against any potential ethical issues. It's important to address and reduce data bias, profiling, and privacy violations. Maintaining community trust will be made easier by being transparent about the prediction process and regularly assessing its fairness and accuracy.

A crime prediction website shows strong viability based on the consideration of technical, economic, and social issues. It is a worthy endeavour because crime data is readily available, data processing methods have advanced, and there is a chance for beneficial societal impact. Data privacy, community involvement, and ethical considerations must, however, be carefully considered throughout the development and implementation phase. The successful implementation and administration of the crime prediction website depends on a cooperative strategy comprising law enforcement organizations, community groups, and technology professionals.

3. SYSTEM ANALYSIS & DESIGN

3.1 REQUIREMENT SPECIFICATIONS

This section contains a list of all the requirements for our Crime Glace website. It is categorized into functional and non-functional requirements. Every requirement is listed in a list format along with a small description for easier readability and understanding.

3.1.1 FUNCTIONAL REQUIREMENTS

REQ 1: User Details: Should be able to store the details of all users and also allow editing at all times.

REQ 2: User Storage: Should be able to store the details of more than 100,000 users.

REQ 3: Chart: Displays crime distribution pie charts for every state in every year from 2001-2022.

REQ 4: Verification Link: Sends a verification link to the mail address provided to complete the registration process for new users.

REQ 5: Helpline: Users can get helpline numbers of their respective states by selecting from a drop down.

REQ 6: FIR: Links attached to official government websites in case any individual wishes to register an FIR.

REQ 7: Safety Measure: Should be able to display the safety of a city or area from the given inputs by the user.

REQ 8: City Details: Should be able to display the details of the selected city from the drop down.

REQ 9: Feedback: Should be able to take feedbacks from users using the website regarding reports on bugs or any other important official information

3.1.2 NON-FUNCTIONAL REQUIREMENTS

REQ 1: Response Time: Response to user choices within 2 seconds.

REQ 2: Link Latency: The latency for the verification link should not be more than a minute.

REQ 3: Con-current users: System should be able to handle and process queries for 10,000 users at the same time.

REQ 4: New login: Notify the user when the account is logged in from another system and allow him to block access to his/her account.

REQ 5: Confirmations: Ask for confirmation from the user every time an input is taken from them.

REQ 6: Security: The website uses the necessary security measures to safeguard sensitive user information and shield it from intrusion. This includes regular security audits, secure user authentication, etc.

REQ 7: Compatibility: The website works on a variety of operating systems, web browsers, and hardware (such as computers, laptops, tablets, and mobile phones). It is designed responsively to ensure optimal user experience across different platforms.

REQ 8: Privacy: The website upholds user privacy and adheres to all relevant data protection laws.

REQ 9: Maintainability: To make it simpler to maintain and improve in the future, the website is developed using modular and well-documented code.

REQ 10: Accessibility: It can be easily accessed by any users from anywhere, to get the details of crime in their area or the area they want to relocate.

3.2. DIAGRAMS

A prediction website's development depends heavily on diagrams. They offer visual depictions of data, models, and procedures so that users can more easily comprehend complex information and make informed choices. To represent and visualize huge datasets, diagrams are frequently utilized. Data can be presented in an easy-to-understand way using charts, graphs, and infographics. Users are able to quickly understand the data, find correlations, and gain insights according to these visualizations.

To predict future results, prediction websites use models and algorithms. Diagrams are essential for demonstrating the organization and validity of these concepts. To illustrate the logical course taken by the algorithm, decision trees can be utilized to show a set of if-else circumstances. Similar to this, flowcharts can explain how input data is turned into predictions by showing the many steps used by a predictive model.

For user trust and comprehension, forecasts must be clear and understandable. Diagrams can be used to illustrate how predictions are made and the variables that affect them. A bar chart, for instance, can show how important various features are to the forecast and help consumers understand the underlying logic. The user experience is improved and forecast confidence is increased as a result of this transparency.

Diagrams help to visualize the overall design and individual elements of the prediction website. They give a general picture of the system's data flow and interactions between various parts. A block diagram, for instance, can show how the user interface, prediction model, database, and external APIs interact. This graphic depiction aids website developers in efficiently creating and managing the website.

Diagrams can be used to better understand the types and patterns of errors that happen. Developers and data scientists can use this visual analysis to pinpoint the causes of errors and adjust the prediction models accordingly.

In conclusion, graphics are crucial to the creation of websites that make predictions. Prediction websites can increase user knowledge, trust, and engagement by employing diagrams efficiently, ultimately offering insightful information and precise forecasts.

3.2.1. FLOWCHARTS

A flowchart is a diagram that employs different symbols and connectors to show how an algorithm works or how the various processes or activities fit together. The website's structure and navigational flow can be described using a flowchart. It aids in the planning and organization of the many pages, parts, and features of a website before the actual design and development process is initiated.

We have made a flowchart of our model as well as shows the working of login and authentication and sign up page.

Steps of flowchart of model is described as follows-

Step1: Start the website, and select the option whether you want to log in as user or admin.

Step2: If you are an existing user, sign in into your account with your login credentials, after proper authentication your home page will open. If you are new to our website, you have to create an account in order to continue with our service.

Step3: In the Crime Info page, you are given with 4 options: Check Safety Measures, Get City Details, Crime Distribution and Helpline/e-FIR. If you want to check safety measures, enter city per capita income, male literacy rate and total population to get the result. Else if you want to know the city details like total male and female population, per capita income, male and female literacy rate, etc. ,select the feature and city to get the result. Else if you want to check the distribution of crime, select the year and state, you want to know. Else if you want helpline contacts or want to file eFIR, select the state to get the information.

Step4: If you want to give feedback to our team, you can do so by filling the feedback form and then exit. If you don’t want to do so, you may exit from the page.

After this, we have 2 more diagrams explaining how we use OTP verification to authenticate users and steps to create a new account on our website, to protect the privacy of our users and prevent our website from hackers and unauthorized users. OTP will be received on the user's registered email id, which he/she uses at the time of sign up.

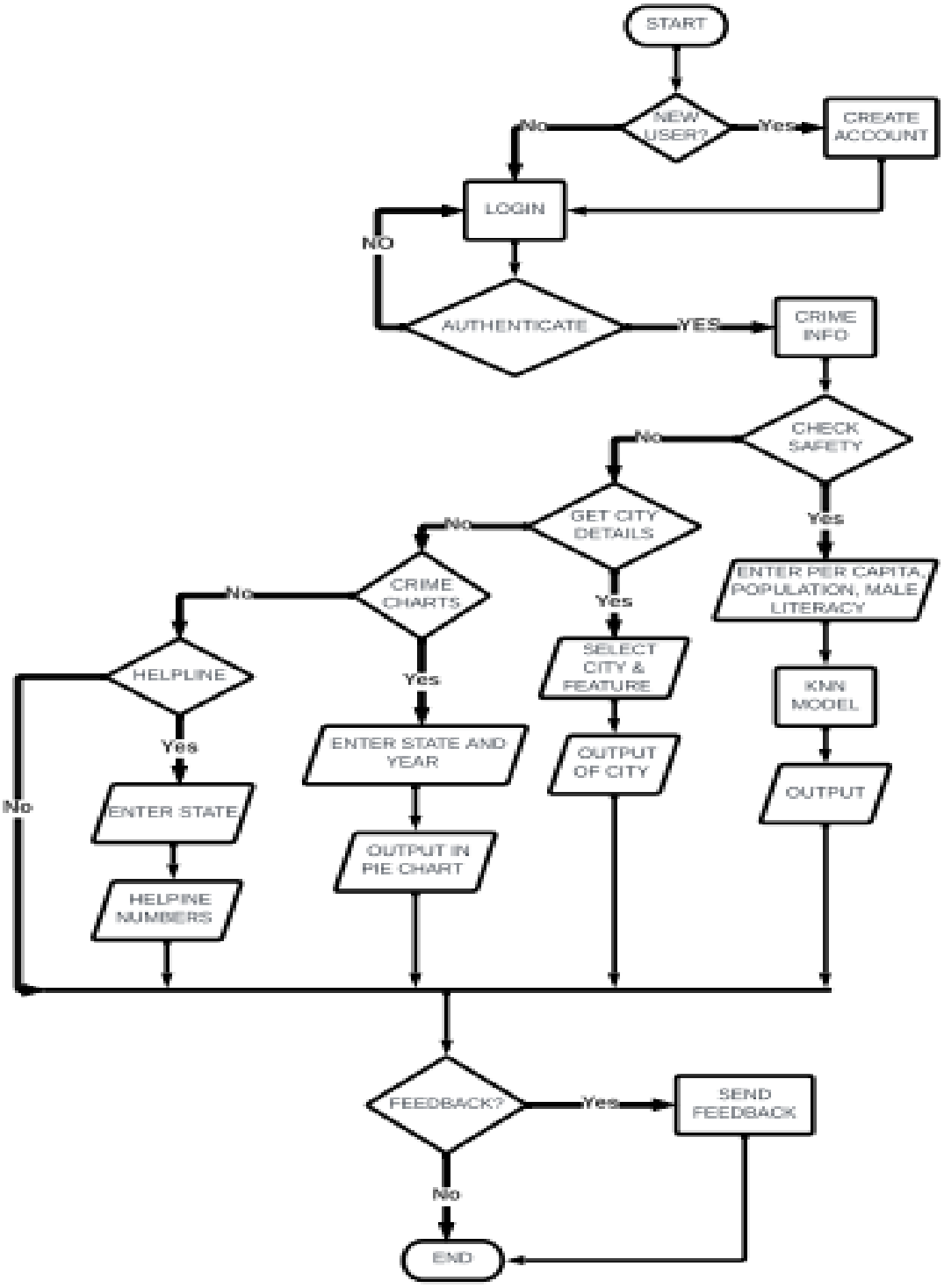

Fig 1: Flowchart of Crime Glance Model

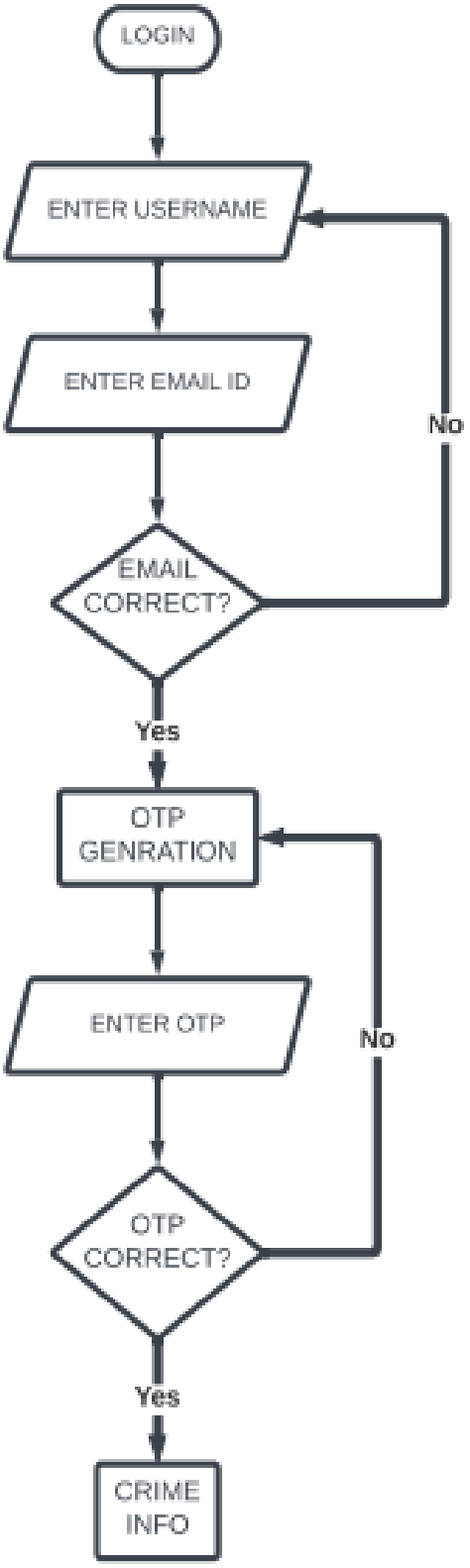

Fig 2: Flowchart of Login & Authentication

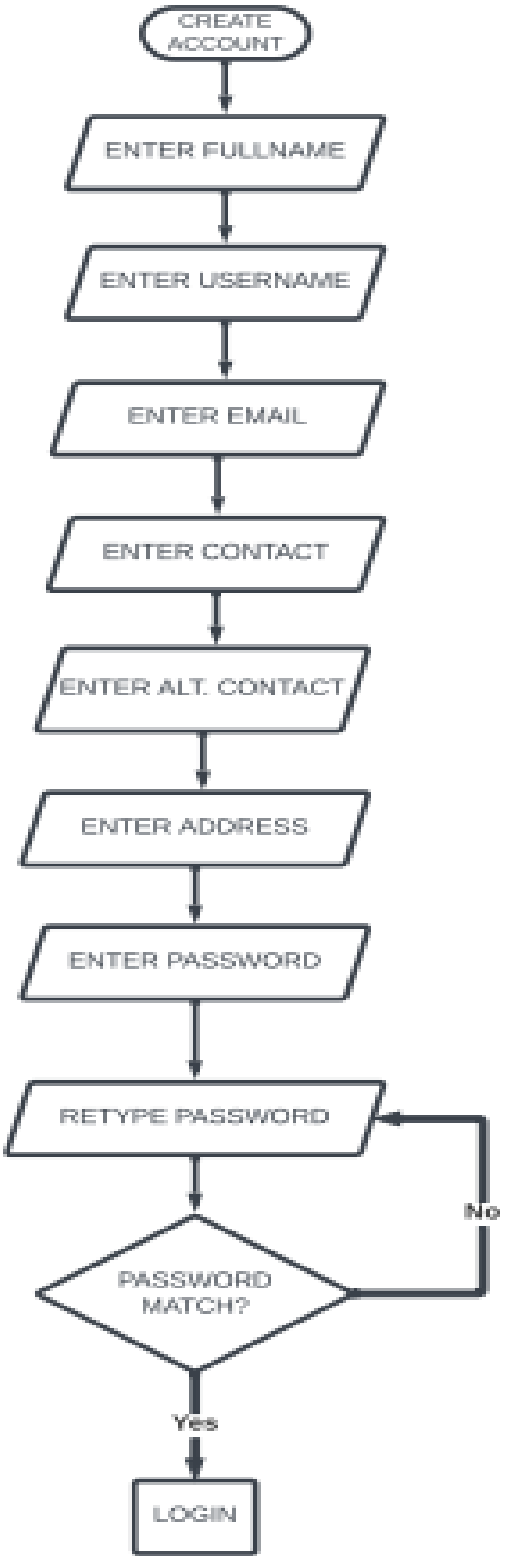

Fig 3: Flowchart of Create New Account

3.2.2. DATA FLOW DIAGRAMS

A data flow diagram, often known as a DFD, is a graphical depiction or modelling method that helps to see how data moves through a system or process. It demonstrates how data is transferred between different system elements, including processes, data stores, and external entities. DFDs provide a clear and comprehensive overview of how data is entered, processed, stored, and outputted inside the system by using symbols and notations to represent these parts and the data flows between them. Data flow diagrams (DFDs) are effective tools for displaying how data moves through a system. DFDs promote the comprehension and documenting of complicated processes with their intuitive graphical notations and methodical methodology, assisting in the detection of potential problems, stakeholder communication, and system improvement.

Data flow diagram elements:

The four primary parts of a typical data flow diagram are as follows:

1. Processes: The operations or transformations made to data within the system are represented by processes. In a DFD, they are shown as circles or rounded rectangles. Each process has a distinctive identification and a name that accurately describes what it does.
2. Data Flows: Data flows show how information is transferred between internal and external processes, data stores, and other entities. Arrows linking the various elements of the diagram serve as a representation of them. The information that data flows reflect and the data that they carry are labelled.
3. Data Stores: Data stores are the places in the system where data is kept in repositories. They might be virtual (like temporary storage) or actual (like databases and files). Rectangles with parallel lines on either side are used to represent data warehouses.
4. External Entities: Interacting with the system are external sources or destinations for data. They could be organisations, users, or different systems. Rectangles representing external entities are used to represent them, while incoming and outgoing data flows are used to indicate how those entities interact with the system.

Data Flow Diagrams (DFDs) come in a variety of forms or levels, each with a particular level of abstraction and level of detail. They are described as follows-

0-level DFD

0-level dfd is also known as Context Diagrams. By showing how the system interacts with outside elements, the context diagram gives a broad picture of the overall system. The system is portrayed as a single process or entity that is surrounded by other entities and engaged in data flows with them. The Context Diagram aids in defining the scope and bounds of the system and acts as a jumping off point for more in-depth DFDs.

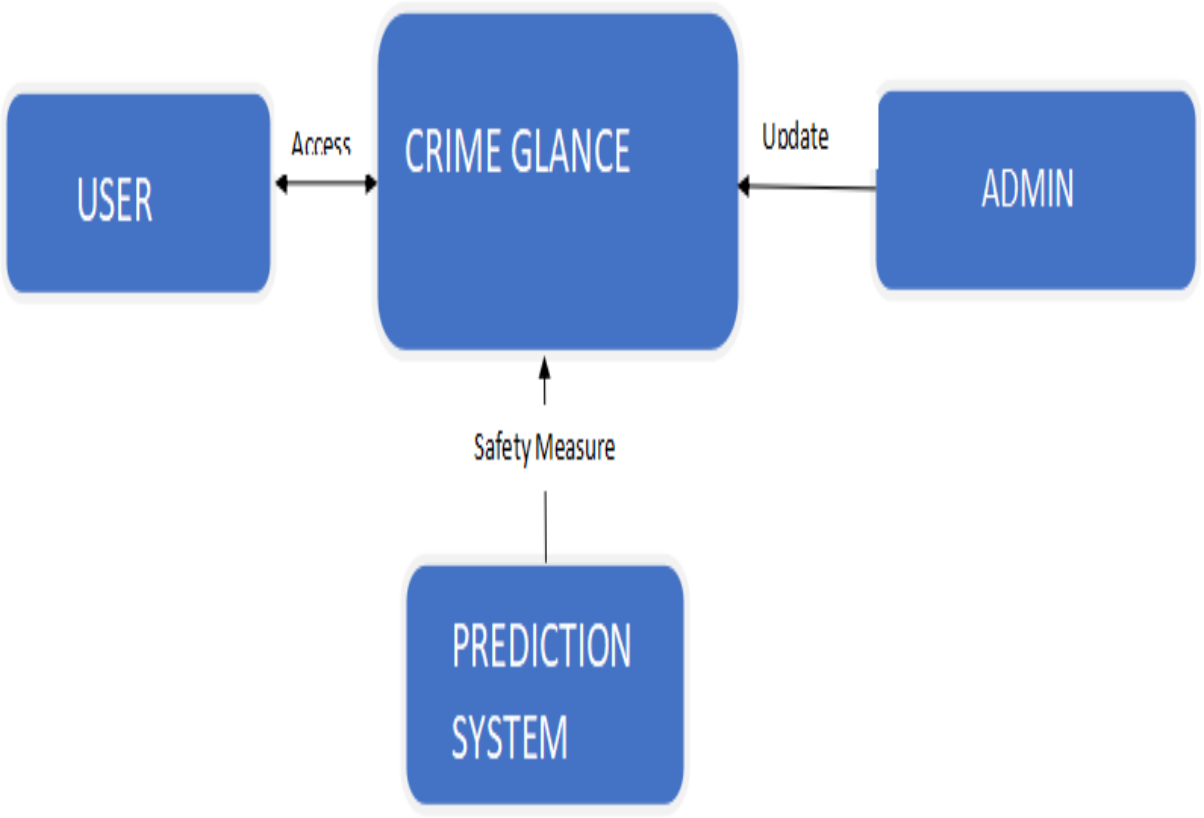

Fig 4: 0-level DFD

1-level DFD

Level 1 DFDs break down the major process in the Context Diagram into smaller processes to give a more in-depth understanding of the system. The main system processes are named, and the data flows between them are displayed. The core functions and interactions of the system are the focus of Level 1 DFDs.

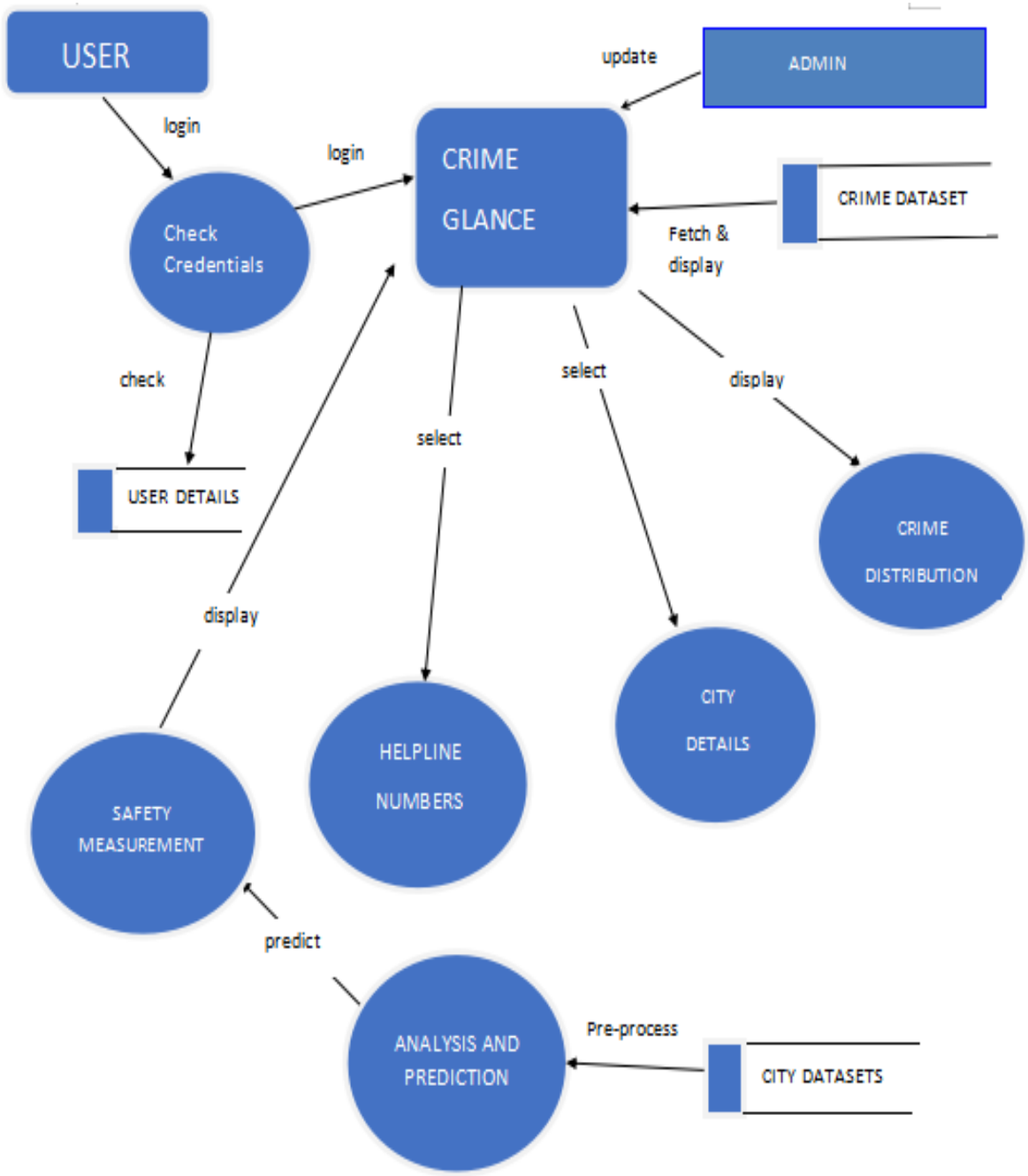

Fig 5: 1-level DFD

2-level DFD

The processes listed in Level 1 DFDs are subsequently broken down into more specific subprocesses in Level 2 DFDs. A more detailed perspective of the system is provided by this level of DFD, which also shows more processes, data flows, and data stores. Understanding the particular actions and modifications taking place within each process is much easier with the use of Level 2 DFDs.

3.2.3. ENTITY-RELATIONSHIP DIAGRAM

Entity-Relationship (ER) diagrams are a common modelling tool in the database design industry. They offer a visual depiction of the connections between the entities in a system and the properties connected to each entity. A database's structure and relationships can be understood, designed, and communicated using ER diagrams.

There are three primary components of ER diagrams:

1. Entities: Entities represent actual items, ideas, or things from the real world that are pertinent to the system being modelled. In an ER diagram, entities are commonly shown as rectangles or squares.
2. Attributes: The features or properties of an entity are described by its attributes. They offer more details about the entities that are being modelled. Typically, attributes are shown as ovals or ellipses related to the appropriate entities.
3. Relationships: Relationships show how different things are connected or associated with one another. They specify how things communicate with one another. One-to-one, one-to-many, or many-to-many relationships are all possible. They frequently have labels that describe the type of connection, such as "has," or "belongs to." Diamond shapes linking the linked entities are used to show relationships.

In ER diagrams, entities, characteristics, and relationships are represented by particular symbols and notations. Among the often employed symbols are:

1. Square or Rectangle: Depicts an object.
2. Ellipse or oval: Depicts an attribute.
3. Diamond: Symbolises a connection.
4. Lines: Entities and characteristics are connected to relationships via lines, which show their participation and cardinality.

The number of instances in one entity that are connected to instances in another is known as the cardinality. It aids in defining the type of relationship that exists between entities. In ER diagrams, cardinality notations include:

1. One-to-One (1:1): Each instance in one entity has an identical counterpart in another.
2. One-to-Many (1:N): Each instance in one entity is connected to one or more instances in another entity through the one-to-many (1:N) relationship.

However, each instance in the second entity only has a single instance in the first entity as an association.

3. Many-to-One (N:1) refers to the relationship between several instances of one entity and a single instance of another entity.
4. Many-to-Many(N:N): Multiple instances of one entity are linked to multiple instances of another entity in a many-to-many (N:N) relationship. To connect the two entities in this relationship, an intermediary entity must be introduced.

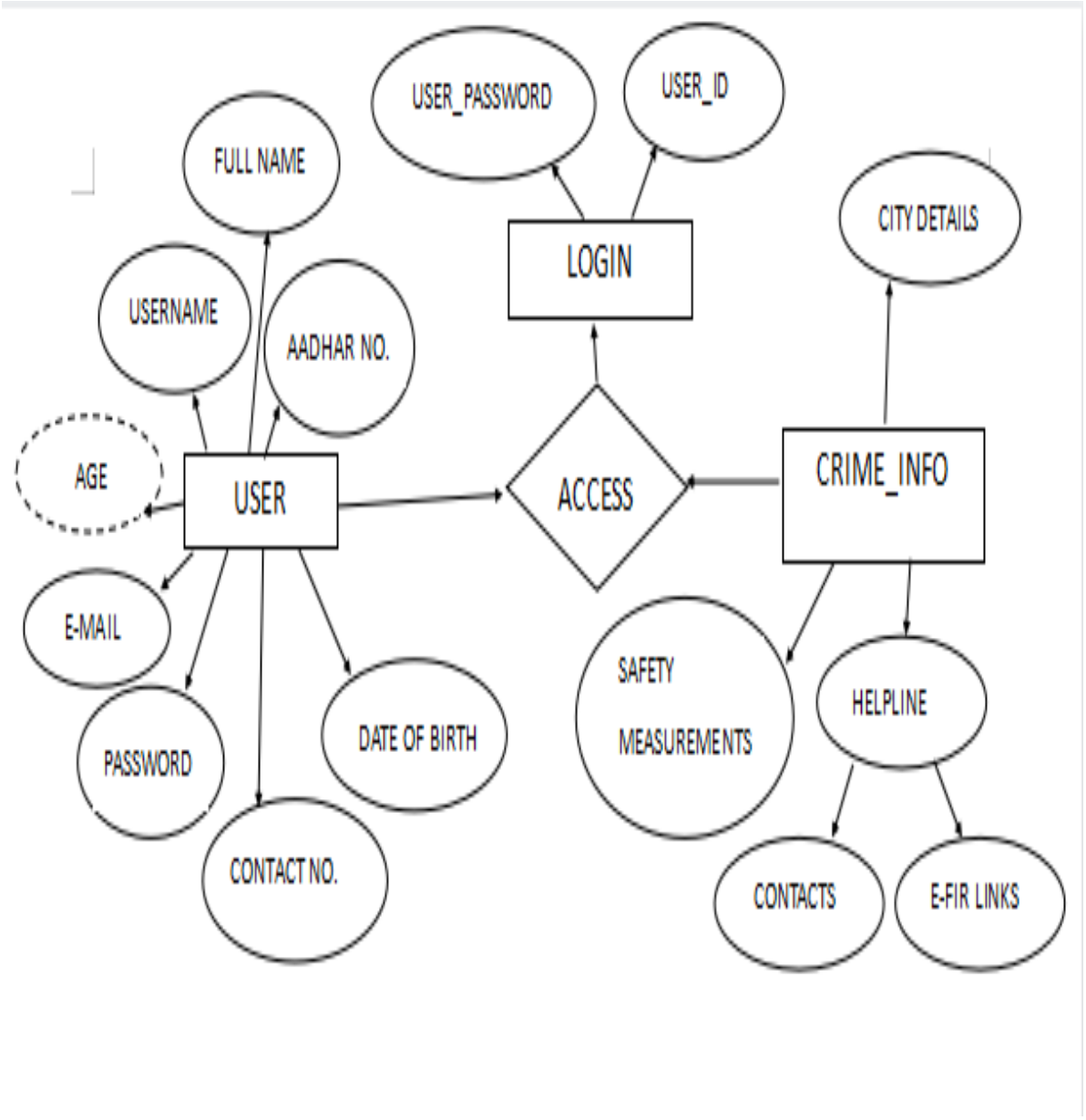

Fig 7: Entity-Relationship Diagram

3.2.4. Use Case Diagram

Use case diagrams are a common modelling method in the software development industry. They offer a graphic representation of a system's functional requirements as seen through the eyes of its users. Use case diagrams are useful for comprehending, deriving conclusions from, and explaining interactions between actors (users of the system) and the system itself. The functional requirements of a system are easier to elicit, organise, and comprehend when they are represented in diagrams. They record system users' objectives, deeds, and interactions and offer a framework for future system design and development. They help in pinpointing user objectives, system limitations, and the key features that must be implemented.

Several essential elements are present in use case diagrams:

1. Actors: Actors stand in for the outside parties that engage with the system. They could be users, different systems, or outside gadgets. Actors are shown as stick figures or icons that are outside the boundaries of the system.
2. Use Cases: Use Cases depict the particular actions or features that the system makes available to its users. They outline a series of interactions that take place between an actor and the system to accomplish a particular objective. Oval or elliptical shapes are used to depict use cases inside the system boundaries.
3. Relationships: Relationships show how actors and use cases are related and interact. The most typical kinds of relationships are:
    a. Association: Indicating that an actor interacts with or is dependent upon a use case, an association between an actor and a use case represents a general relationship between the two.
    b. Generalization/Inheritance: A specialised actor or use case that inherits the traits and behaviour of a more general actor or use case is represented by the generalization/inheritance concept. An "is-a" relationship is shown.
    c. Include: Denotes a connection between two use cases, showing that one use case incorporates the functionality of the other.
    d. Extend: Shows how two use cases are related to one another and how one use case may be expanded by another, resulting in the inclusion of new functionality.

3.3 Algorithms

We have used the K-nearest neighbour algorithm to build our predictive model. K-Nearest Neighbours (KNN) is a popular machine learning algorithm, utilized in many fields, including predicting crime. KNN can be a useful tool for analyzing trends in crime data and creating predictions based on analogous historical cases.

Instance-based learning, which includes the KNN algorithm, is a category of supervised machine learning techniques. Based on the similarity principle, it categorizes new occurrences according to how closely they resemble known examples in the training data. To estimate the risk of future crimes and safety measures in particular places, KNN can be used to find patterns and similarities among prior criminal episodes.

To build the predictive model, several steps are involved. They are explained below:

- Dataset: This is the very first step, in which we collect data from various authentic resources related to past recorded crimes. The dataset contains relevant attributes like state, districts, total population, Male and female literates, per capita income, types of crime occurring in particular year, etc.
- Preparing data: After collecting data, it is pre-processed to remove any noise or missing values in the dataset. This step is important to get accurate predictions.
- Train & test data: We divide our dataset into 80:20 ratio, in which 80% of our data is training dataset and 20% of data is used as testing data. The model is trained using KNN algorithm and then testing is performed to check for the accuracy of the model. In order to ensure that the model is trained on historical data and tested on future data in order to simulate real-world events, the dataset is frequently split in a way that retains the chronological order of the data.
- Safety measure: After the model is being trained, the user is supposed to enter the per capita income, male literacy rate and total population of the city they want to get the at a glance image of crime. Then, gdpp(i.e. Per capita income/total population),total population and male literacy used by our model

to predict safety measures. We have used gdpp as the population may increase the per capita of the city, which may result in inaccurate results. Safety measures are categorized into 3 clusters: ‘Safe’, ‘Moderately Safe’, and ‘Unsafe’ .

- Visuals: To make our platform interactive, we represent the distribution of crime in a particular state in a particular year in the form of pie-chart. The user has to select the state and year from the drop down and different types of crimes like murder, robbery, rape, burglarry, etc. will be shown in the form of pie-chart.
- Evaluation: The final step is to evaluate the performance of our model, using different performance metrics. The accuracy of our model is almost 100%.

We have used three main attributes- per capita income, total population and male literacy rate as the prime feature for our model as crime in any city has a very high dependency on all these three factors. If a city has very low per capita income, then there is a very obvious chance that people might be involved in activities like theft, robbery, murders, etc. Likewise, is with the total population and male literacy rate. The young youths are the future of our country and if they are illiterate, then crime will increase very rapidly in that place and it hinders the growth in personal as well as society level.

KNN Algorithm

For both classification and regression applications, the k-Nearest Neighbours (k-NN) technique is a straightforward but effective machine learning approach. It is a non-parametric algorithm, which means that it does not make any assumptions about the distribution of the underlying data.

KNN operates on the idea that similar data points frequently have related labels. By comparing new instances to the k closest neighbours in the training data and selecting the most prevalent class label among those neighbours, it categorizes new instances. The KNN algorithm simply saves the information during the training phase, and when it receives new data, it categorizes it into a category that is quite similar to the new data.

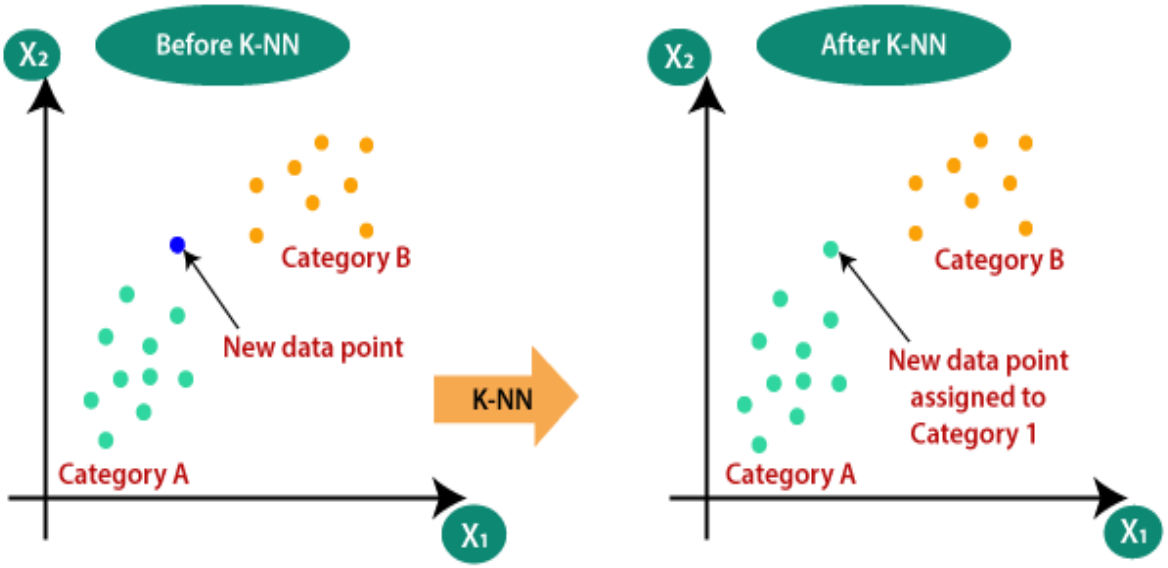

Fig 9: Illustration of KNN Algorithm

Steps of KNN algorithm are explained as follows:

Step1: Data Preparation: Gather the training data that has been labelled, which consists of examples and the classes that the instances belong to. To ensure that each characteristic contributes equally to the distance calculation, normalize or standardize the features in the data prior to processing.

Suppose we have 2 categories in which our data points are distributed and we have to classify a new data point in the required category.

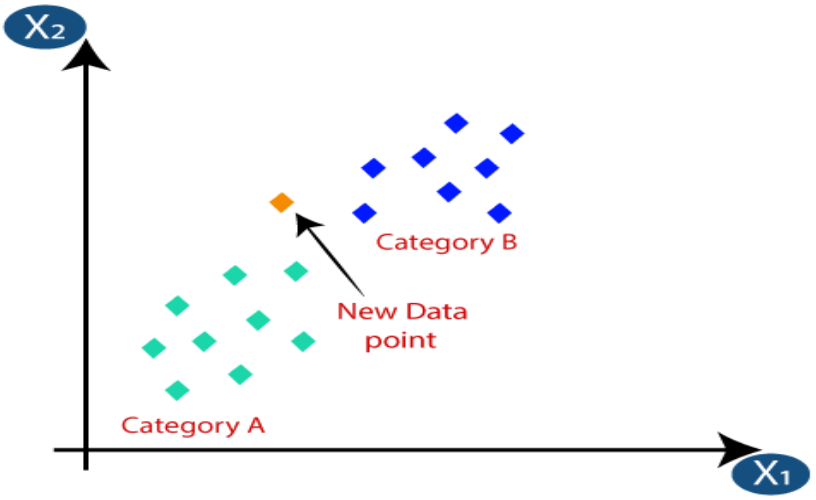

Fig 10: Figure illustrating new data point will fall into which category

Step2: Distance Calculation: Establish a distance metric to evaluate how similar two occurrences are. Although Euclidean distance is frequently employed, Manhattan or Minkowski distance metrics can also be applied. Measure the separation between the new instance and each training data instance.

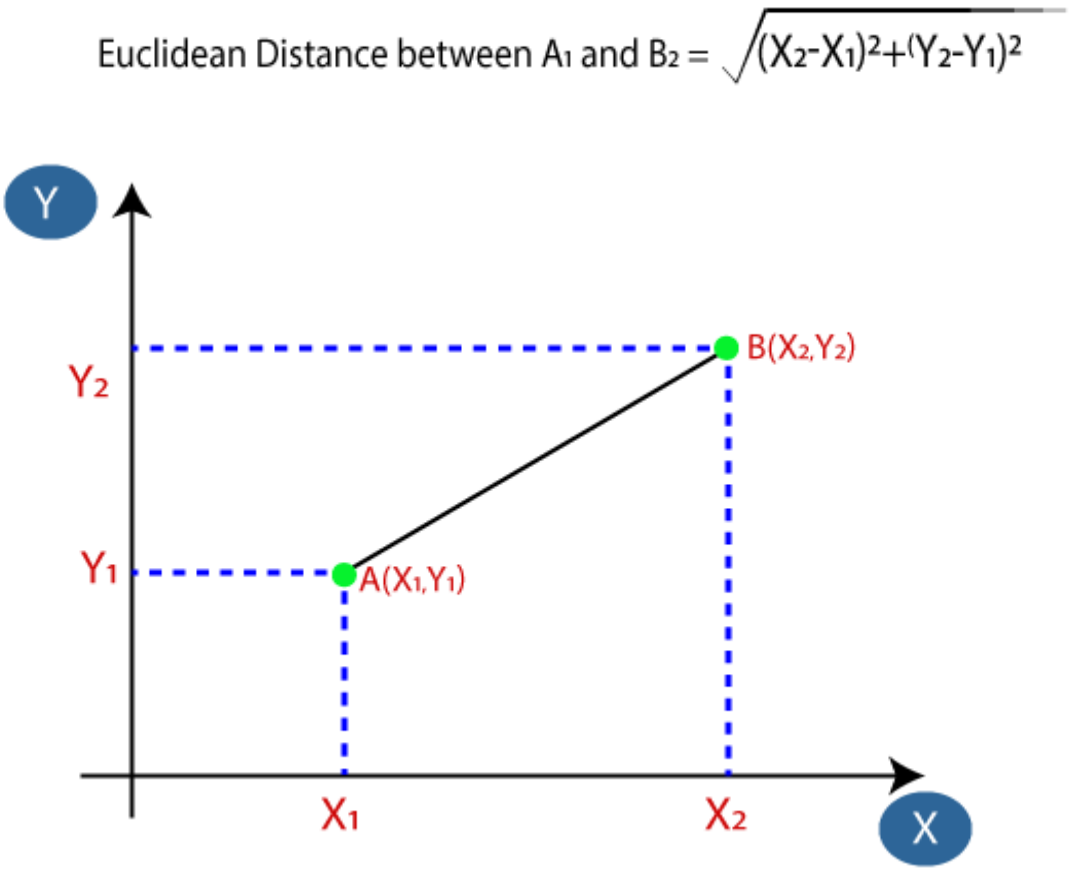

Fig 11: Distance calculation between data points and centroid of categories

Step3: Neighbour Selection: Based on the estimated distances, choose the k closest neighbours. The closest instances to the new instance are those in discussion.
The performance of the model is strongly influenced by the value of k. A greater k value takes into account a wider area of the feature space, while a smaller k value tends to catch local patterns.

Step4: Class Label Assignment: Determine the new instance's class label based on the class labels of its k closest neighbours while doing classification tasks. Utilizing majority voting, where the class label with the highest frequency among the new

instance's k neighbours is assigned, is one popular strategy. When performing regression tasks, the projected value can be determined by averaging the k nearest neighbours' goal values.

Step5: Model Evaluation: Utilize evaluation criteria such as accuracy, precision, recall, or F1 score to rate the effectiveness of the kNN model. To assess the model's capacity to generalize to new data, divide the labelled data into training and testing sets.

To discover the optimum value that offers the best performance on the test set, change the value of k and then repeat the process.

Step6: Predictions: The model can be used to generate predictions on fresh, unlabeled examples after it has been trained and assessed. The additional instances should be preprocessed in the same manner as the training set. Determine the distances between them and the training set's labelled instances.

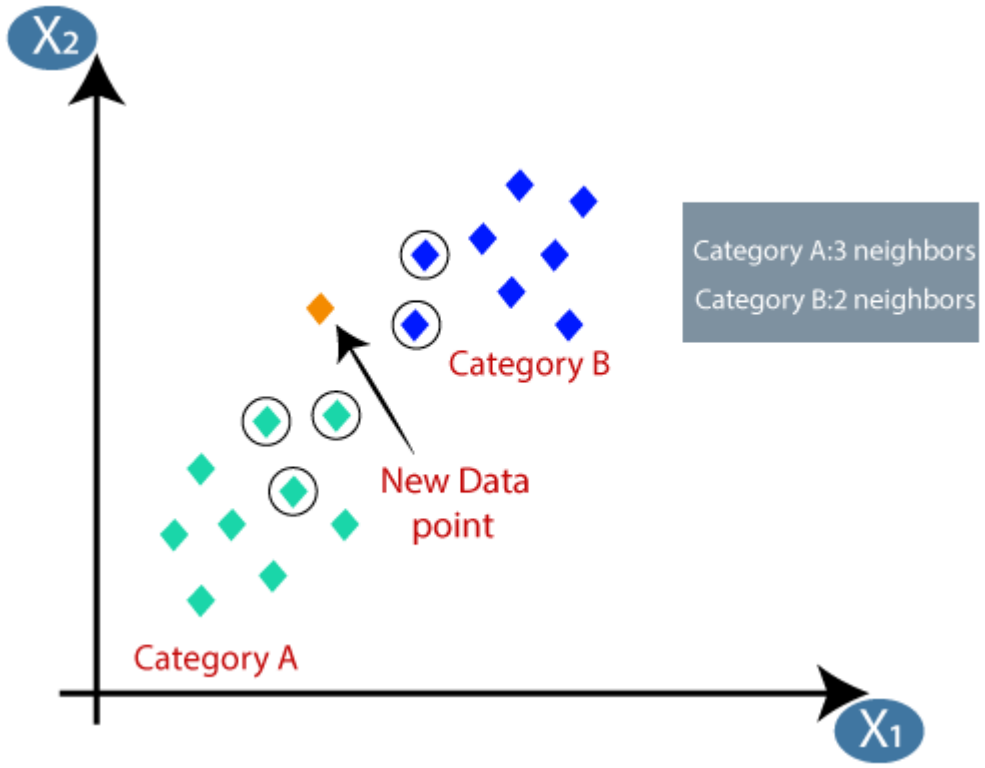

Fig 12: Predicting the output of KNN Algorithm

Thus, in this example the new data point will belong to category A.

Advantages of KNN

- Simplicity: Beginners in machine learning can use kNN since it is simple to comprehend and apply.
- No Training Phase: Since kNN is a lazy learner, it doesn't need a formal training phase. It simply stores the labelled instances as the training data.
- Flexibility: kNN is adaptable to a range of issues because it can handle both classification and regression jobs.
- Non-Parametric: kNN makes no assumptions about the distribution of the underlying data enabling it to effectively handle intricate and nonlinear patterns.
- Interpretable: Since kNN relies on its nearest neighbours, its decision-making process is transparent, making it simple to explain the predictions.
- Adaptability: kNN is suited for dynamic situations since it can adapt to new data by merely recalculating the distances and updating the nearest neighbours.

Overall, the robustness, simplicity and the flexibility of KNN makes it a very valuable and efficient algorithm to use.

Other Algorithm:

Any classification problem algorithm can be used in such types of problems that require classification into categories. Some of the popular classification algorithms are:

- SVC (Support Vector Classifier): SVC categories data points on the basis of their clusters. It falls under Support Vector Machines and its functionality is the same. Accuracy is 63.9%.
- Decision Tree: Decision Tree classifies data points based on criterions. It uses conditioning to follow a path which eventually comes to a leaf node which represents the final outputs and intermediate nodes represent the decisions which lead to the next path. Accuracy is 56.3%.
- Random Forest: Random forest uses the aggregate judgement of several trees to handle intricate interactions between features and enhance classification accuracy. Accuracy is 67.5%.

3.3 TESTING PROCESS

A crime prediction website's predictive models and user interface are tested to ensure its precision, dependability, and functionality. During the testing process, keep the following important factors in mind:

- Data Accuracy: Check the quality and correctness of the crime data that was used to develop and test the predictive models. To ensure that the website offers accurate predictions, confirm that the data sources are trustworthy and current.
- Model Evaluation: Examine how well the crime prediction website's predictive models are performing. To evaluate the model's success in foretelling crime episodes and safety measures, use evaluation metrics like accuracy and precision.
- Data Cleaning: Check if there are any data points which are out of normal bounds or are negatively affecting our system.

3.3.1 TEST CASES

To evaluate our model properly and efficiently, we cannot rely solely on accuracy metrics alone. We need to test the model on real life data to confirm the results for ourselves before we release the model. To test out our model accuracy, we have taken actual data of 12 cities namely Kanpur, Coimbatore, Kolkata, Delhi, Allahabad, Varanasi, Greater Noida, Jhansi, Ghaziabad, Gurugram, Kochi and Chennai.

Out of the 12 test cases selected, Coimbatore was the only one which did not yield desired results showing moderate safety when it is actually safe according to statistics provided by the Government of India.

Below are the test cases and the outputs received from the model.

- Kanpur-

```
#for kanpur(should be moderate)
gdp=87295*0.36
pop=3812000*0.87
lit=84.50*0.96
gdpp=gdp/pop
result=clf.predict([[gdpp,pop,lit]])
if result==0:
  print("not safe")
elif result==1:
  print("moderate safety")
else:
  print("safe")
```

```
moderate safety
```

Fig 13: Kanpur Safety prediction

- Coimbatore

```
#for coimbatore(should be safe)
gdp=189937*0.36
pop=3009000*0.87
lit=94.46*0.96
gdpp=gdp/pop
result=clf.predict([[gdpp,pop,lit]])
if result==0:
  print("not safe")
elif result==1:
  print("moderate safety")
else:
  print("safe")
```

```
moderate safety
```

Fig 14: Coimbatore Safety Prediction

❖ Kolkata

```
#for kolkata(should be safe)
gdp=339000*0.36
pop=14900000*0.87
lit=86.31*0.96
gdpp=gdp/pop
result=clf.predict([[gdpp,pop,lit]])
if result==0:
  print("not safe")
elif result==1:
  print("moderate safety")
else:
  print("safe")
```

```
safe
```

Fig 15: Kolkata Safety Prediction

❖ Delhi

```
#for delhi(should be not safe)
gdp=300000*0.36
pop=32066000*0.87
lit=90.94*0.96
gdpp=gdp/pop
result=clf.predict([[gdpp,pop,lit]])
if result==0:
  print("not safe")
elif result==1:
  print("moderate safety")
else:
  print("safe")
```

```
not safe
```

Fig 16: Delhi Safety Prediction

❖ Allahabad

```
#for allahabad(should be not safe)
gdp=74125*0.36
pop=1533000*0.87
lit=84.46*0.96
gdpp=gdp/pop
result=clf.predict([[gdpp,pop,lit]])
if result==0:
  print("not safe")
elif result==1:
  print("moderate safety")
else:
  print("safe")
```

```
not safe
```

Fig 17: Allahabad Safety Prediction

❖ Varanasi

```
#for varanasi(should be not safe)
gdp=59530*0.36
pop=1652000*0.87
lit=83.46*0.96
gdpp=gdp/pop
result=clf.predict([[gdpp,pop,lit]])
if result==0:
  print("not safe")
elif result==1:
  print("moderate safety")
else:
  print("safe")
```

```
not safe
```

Fig 18: Varanasi Safety Prediction

❖ Greater Noida

```
#for Greater Noida(should be not safe)
gdp=541709*0.36
pop=2290880*0.87
lit=88.06*0.96
gdpp=gdp/pop
result=clf.predict([[gdpp,pop,lit]])
if result==0:
  print("not safe")
elif result==1:
  print("moderate safety")
else:
  print("safe")
```

```
not safe
```

Fig 19: Greater Noida Safety Prediction

❖ Jhansi

```
#for Jhansi(should be safe)
gdp=79852*0.36
pop=2315581*0.87
lit=85.38*0.96
gdpp=gdp/pop
result=clf.predict([[gdpp,pop,lit]])
if result==0:
  print("not safe")
elif result==1:
  print("moderate safety")
else:
  print("safe")
```

```
moderate safety
```

Fig 20: Jhansi Safety Prediction

❖ Ghaziabad

```
#for Ghaziabad(should be not safe)
gdp=72996*0.36
pop=3088000*0.87
lit=89.54*0.96
gdpp=gdp/pop
result=clf.predict([[gdpp,pop,lit]])
if result==0:
  print("not safe")
elif result==1:
  print("moderate safety")
else:
  print("safe")
```

```
not safe
```

Fig 21: Ghaziabad Safety Prediction

❖ Gurugram

```
#for Gurugram(should be not safe)
gdp=446305*0.36
pop=1209000*0.87
lit=90.93*0.96
gdpp=gdp/pop
result=clf.predict([[gdpp,pop,lit]])
if result==0:
  print("not safe")
elif result==1:
  print("moderate safety")
else:
  print("safe")
```

```
not safe
```

Fig 22: Gurugram Safety Prediction

❖ Kochi

```
#for kochi(should be not safe)
gdp=251104*0.36
pop=29222000*0.87
lit=98.25*0.96
gdpp=gdp/pop
result=clf.predict([[gdpp,pop,lit]])
if result==0:
  print("not safe")
elif result==1:
  print("moderate safety")
else:
  print("safe")
```

```
not safe
```

Fig 23: Kochi Safety Prediction

❖ Chennai

```
#for chennai(should be safe)
gdp=209975*0.36
pop=5390674*0.87
lit=93.75*0.96
gdpp=gdp/pop
result=clf.predict([[gdpp,pop,lit]])
if result==0:
  print("not safe")
elif result==1:
  print("moderate safety")
else:
  print("safe")
```

```
safe
```

Fig 24: Chennai Safety Prediction

4. RESULTS

On the basis of the information and algorithms used by our website, we give the findings of our analysis of crime prediction in this area. Our crime prediction model seeks to pinpoint prospective crime hotspots and patterns, assisting citizens and law enforcement organisations in taking preemptive steps to improve public safety. We seek to promote initiatives aimed at boosting public safety and building more secure communities by utilising the power of data analysis and machine learning. Keep in mind that cooperation between law enforcement organisations, neighbourhood associations, and private citizens is essential for crime prevention programmes to be effective. We can work together to create a more secure future.

Here is the complete result of the functionality of how our system works.

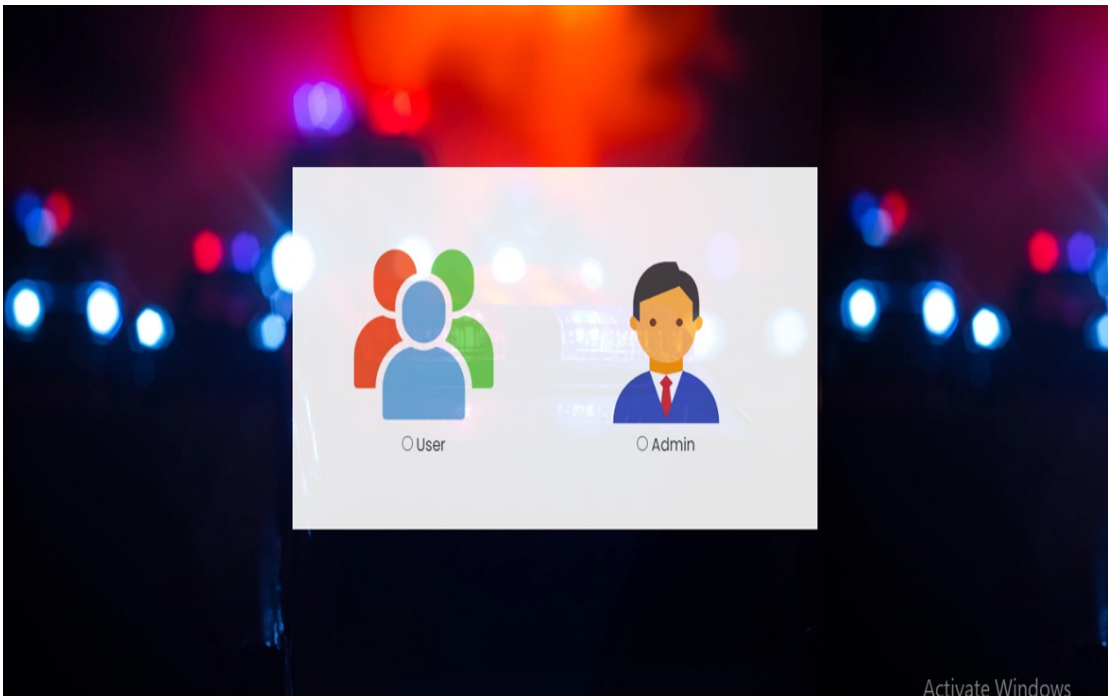

Fig 25: Home Page of Crime Glance

This is the homepage of our website. Whenever a user visits our website, he/she will use this interface. Here is two option provided-

User: people who want to use the facilities of our websites.

Admin: People responsible for making any changes and updation in this website.

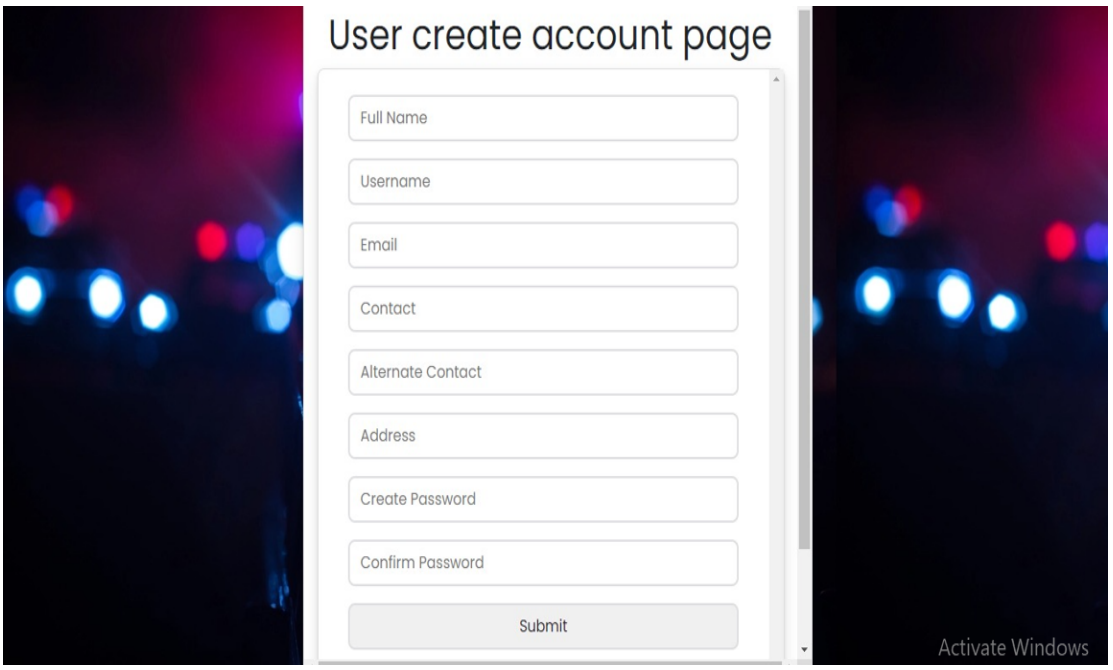

Fig 26: Sign-Up Page

When a new user enters our website, he/she first has to create an account to make use of our services. In order to create a new account, user have to provide some personal details like- Full Name, Username, Email, Contact, Address and Password.

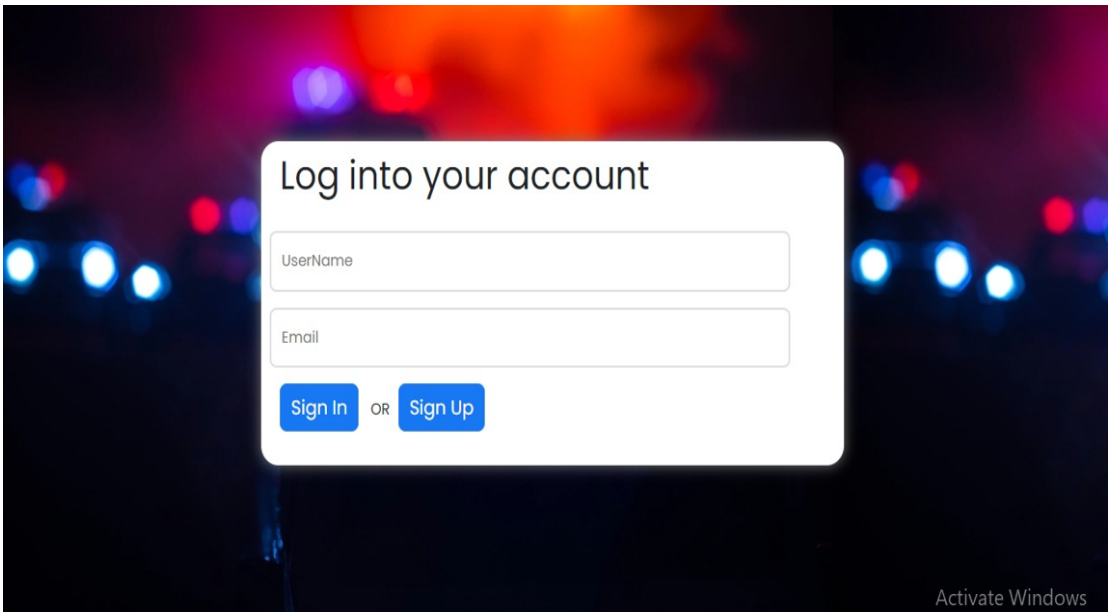

Fig 27: Login page of existing user

There are 2 fields to be filled by the existing users when they want to log in back to their account. They need to fill the same username and email id that was provided at the time of the sign up process.

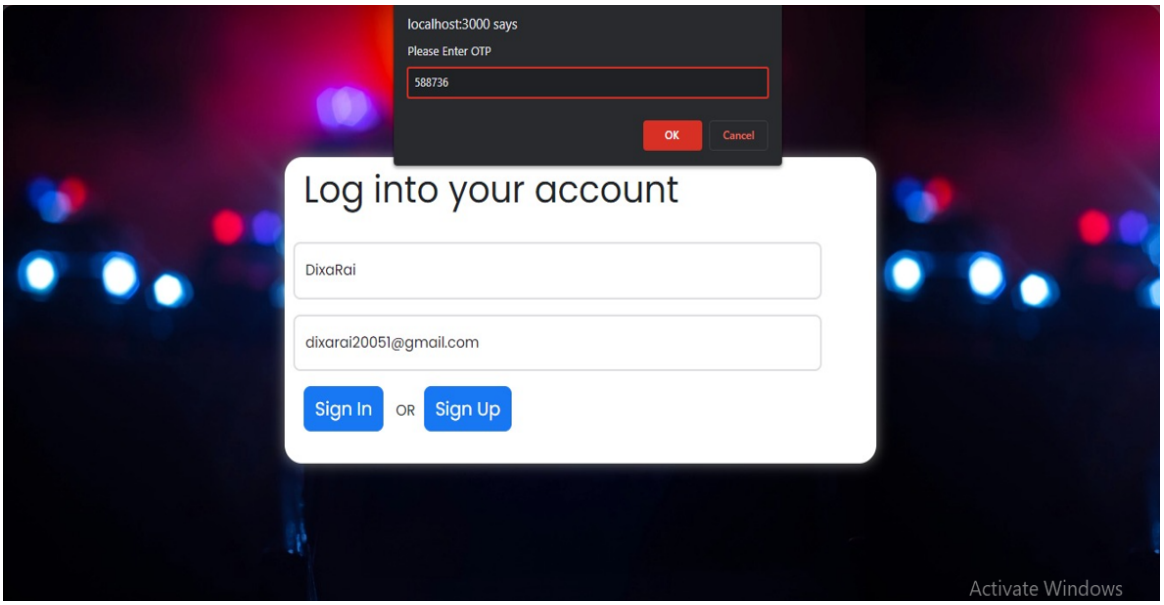

Fig 28: OTP Authentication

Once you enter the registered email id, you will receive an otp on your email id. You must enter the received otp to log in to your account.

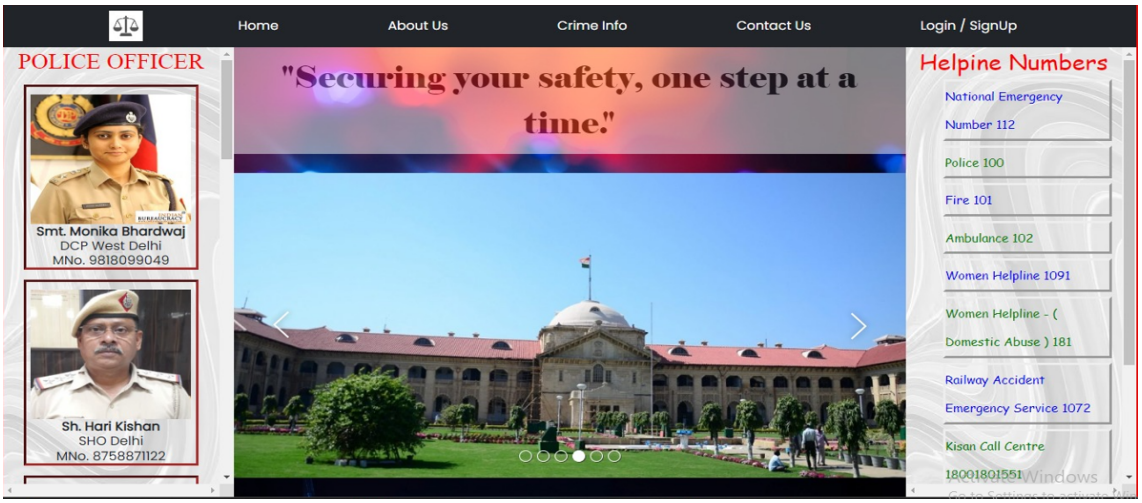

Fig 29: Home Page of User

This is how the home page of a user looks like. Homepage has links to several other related pages along with the details of some prominent police department officials and different helpline numbers to assist people in their need.

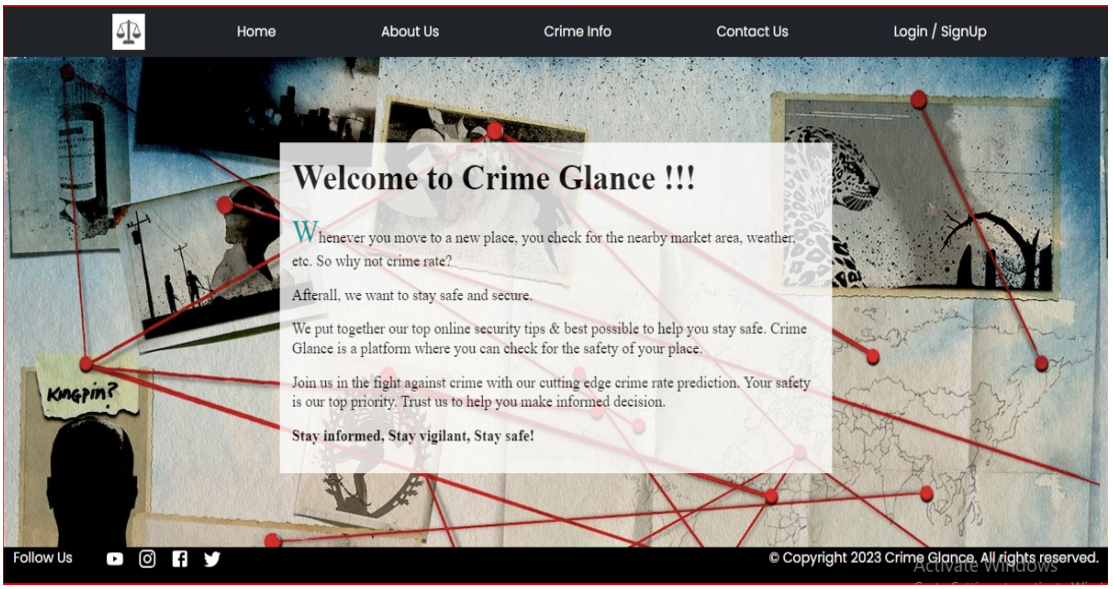

Fig 30: About Us Page

The about us page gives a brief description about the purpose of our website, our vision and mission.

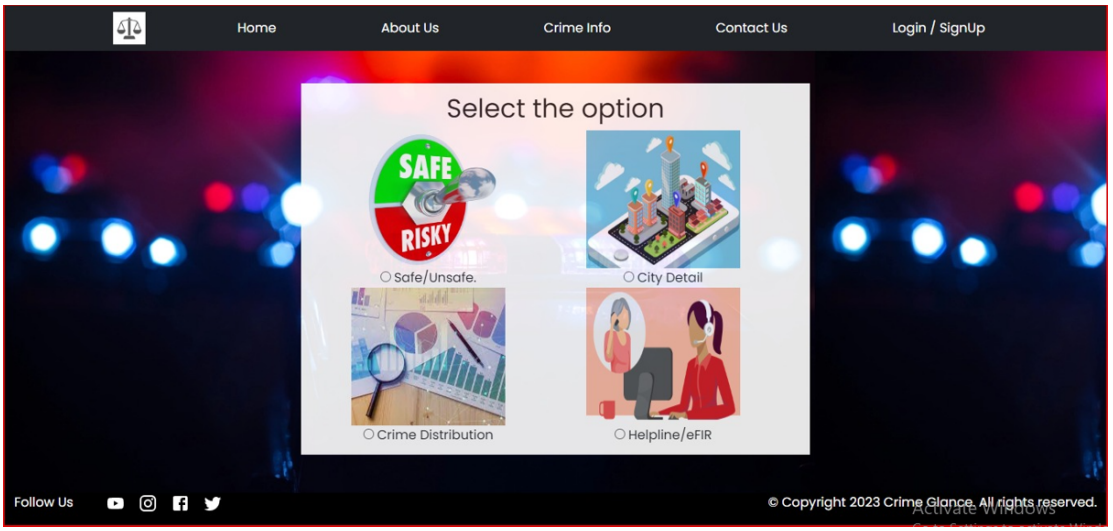

Fig 31: Crime Info Page

In the Crime Info page, user is provided with 4 facilities:

a. If a user wants to check the safety measure of a place.
b. If a user wants to know the details about their place, such as male and female literacy rate, per capita income, etc.
c. If a user wants to get the crime distribution of their place over the years.
d. If a user wants to get the helpline contact of their place or direct links to the government's website, where they can easily file an e-FIR.

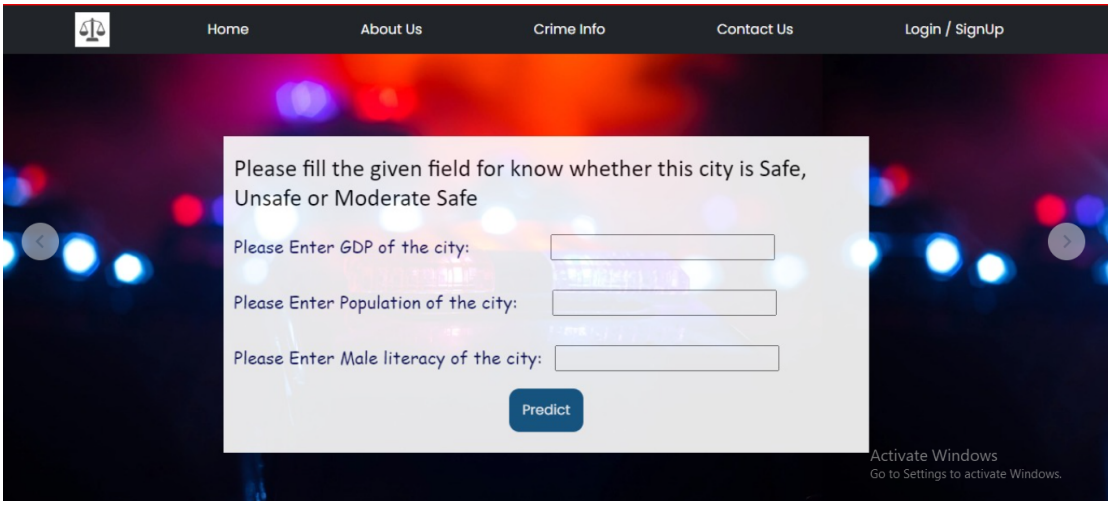

Fig 32: Safety Measure

If a user wants to check for the safety measure, he/she has to enter the gdp, total population and male literacy of their city, to get whether their city falls is safe, moderately safe or unsafe zone.

Fig 33: City Details

If the user wants to get the details of their place, he/she has to select the district name from the dropdown menu, to get the details like Total population, male and female literacy rate, per capita income, total crime and crime rate of the city.

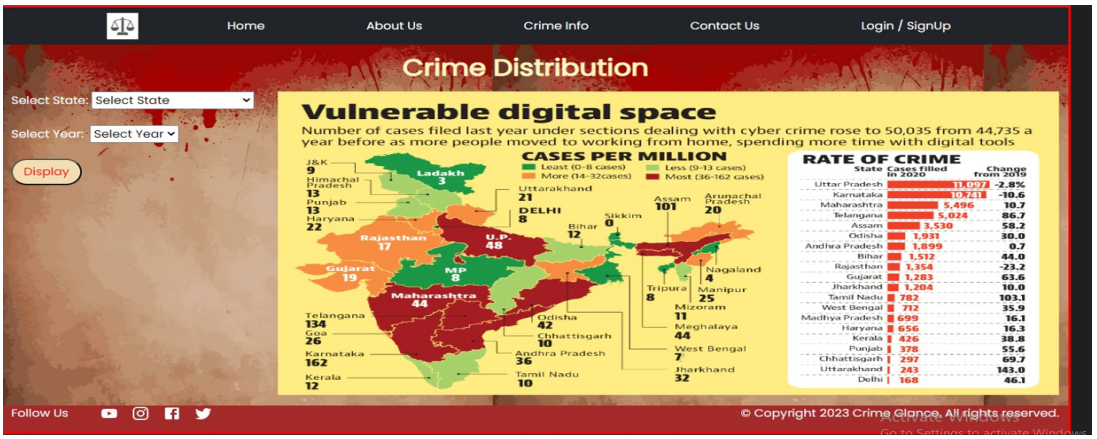

Fig 34: Crime Distribution

This will let the user know the types of crime that happened in a particular year in the city.

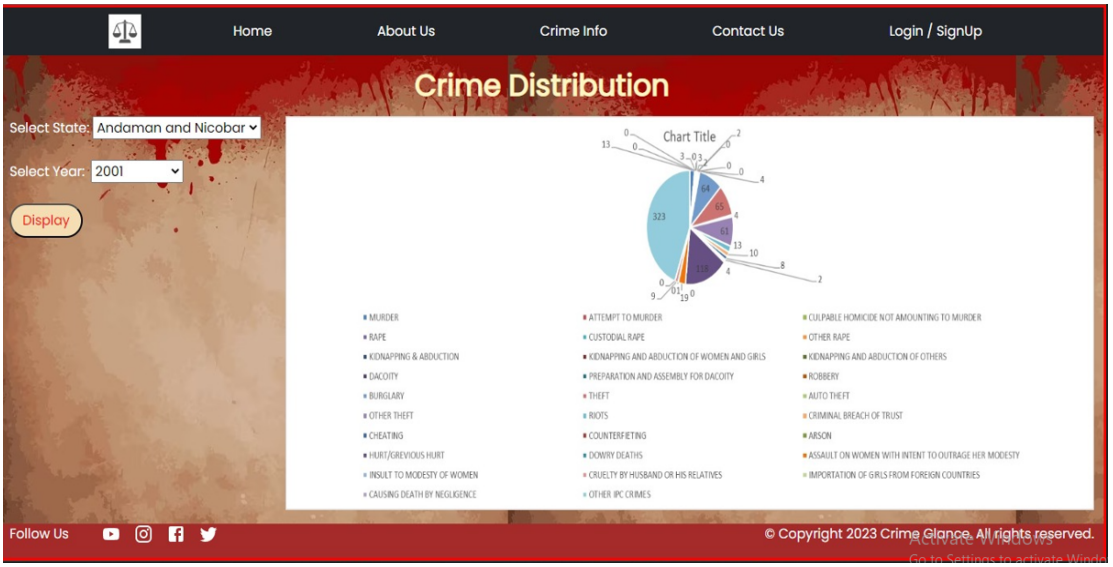

Fig 35: Crime Distribution of Andaman & Nicobar in 2001

This shows the types of crimes in numbers that happened in the year 2001 in Andaman & Nicobar.

Fig 36: Helpline Contacts of State along with e-FIR links

If the user needs to contact the emergency number of state, he/she finds the same from our website alongwith the links that redirects users directly to state government's website where they can easily file an e-FIR if they wish to do so aur get other relevant information.

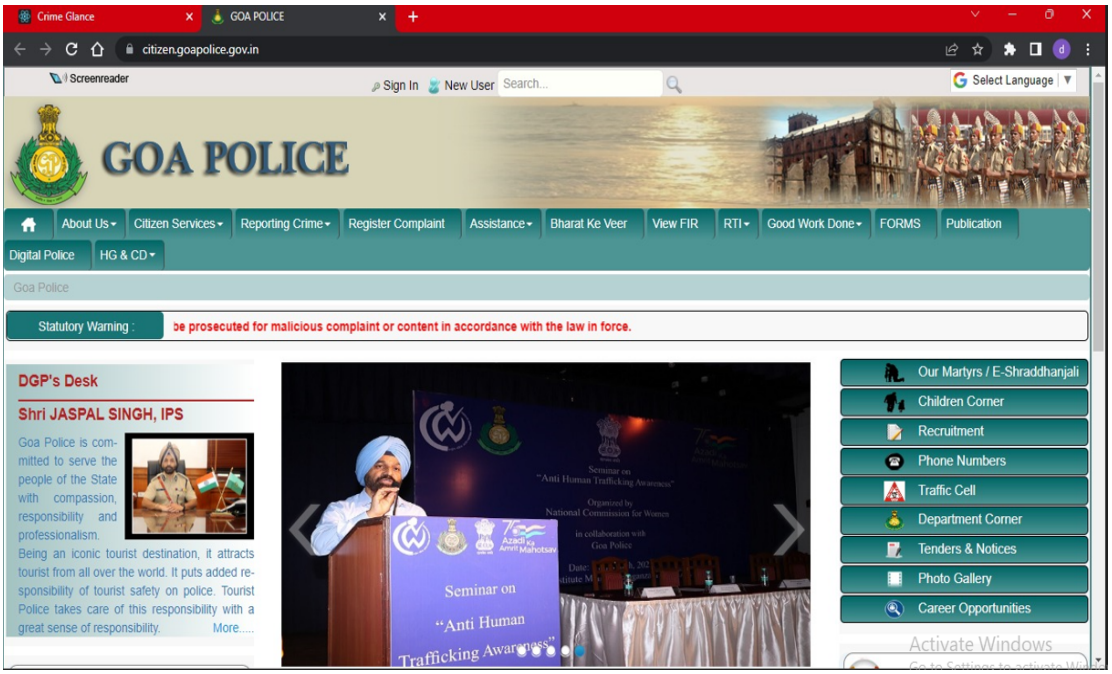

Fig 37: Goa Police Website

If a user belongs to a particular state, wants to visit the government website he/she just has to click on the button of their respective state, he automatically redirects to it, as shown in example of Goa state.

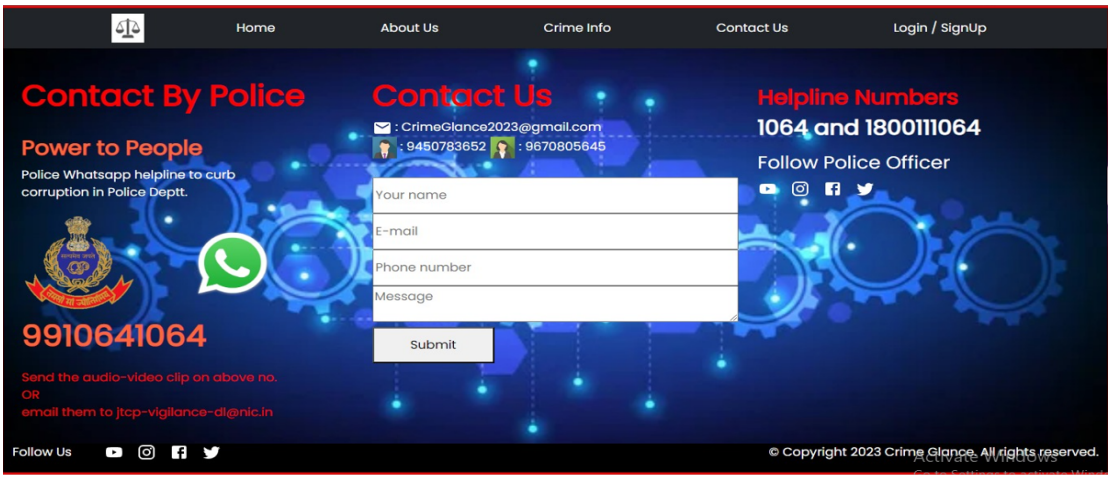

Fig 38: Contact Us Page

If a user wants to contact us for any further assistance, he/she can do so by filling a small contact us form. In this, they have to enter their full name, email, contact, and brief description of the assistance they need from us. All this information will get stored in our databases for future help.

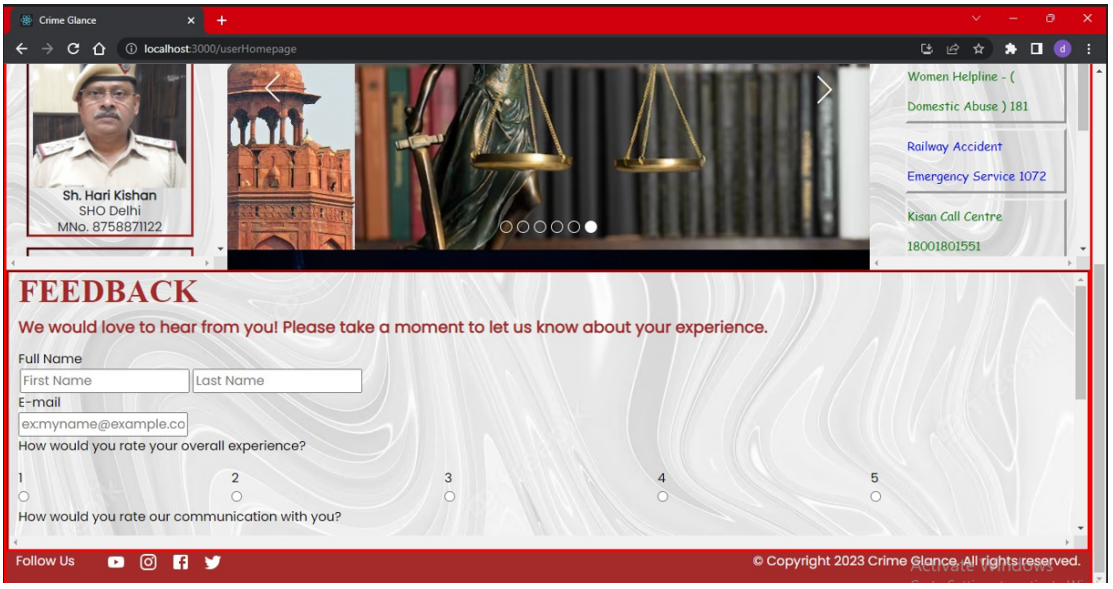

Fig 39: Feedback Form

In the feedback form, users can share their experiences with us through ratings and provide any suggestions regarding the betterment of our services. Suggestions are always welcomed and we would love to hear from our users. Their ratings motivate us to do more such tasks. They can also follow us on our different social media handles like facebook, instagram page and twitter.

In the Admin section, all the pages are almost the same. Only the difference lies in the log in page of admin and one more functionality is added in the homepage of admin, i.e., Update Records.

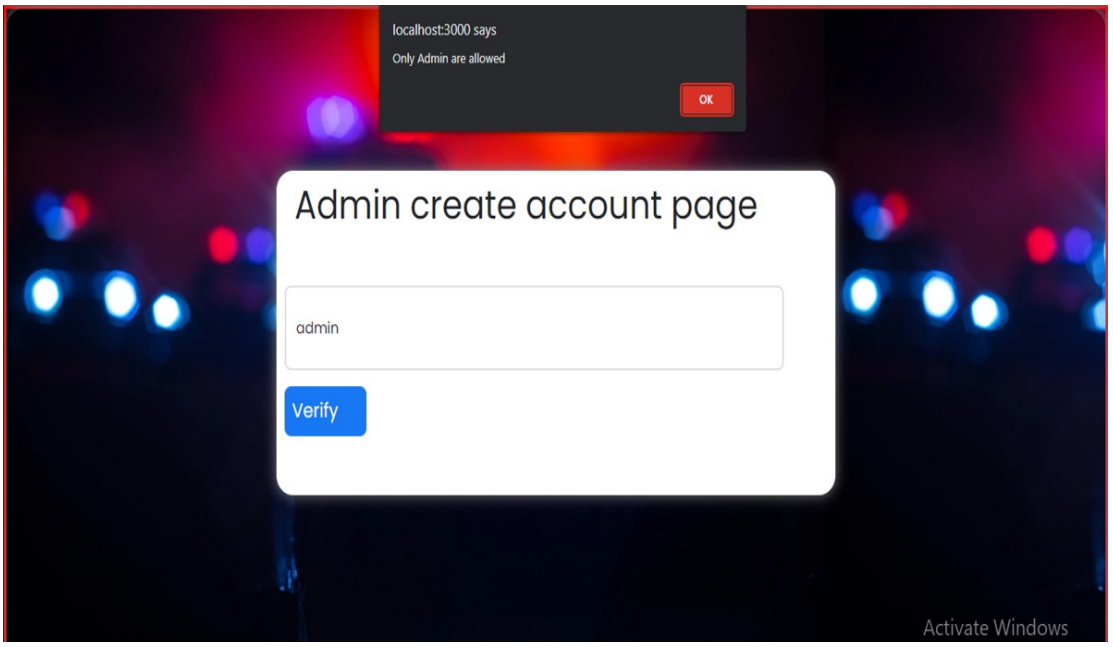

Fig 40: Admin Login Page

No one can log in to the admin's page. Admins are provided with a unique verification id which they have to enter to log in to their account else a pop up message will be displayed as : " Only admins are allowed". After entering the verification code, admin will also receive an otp to their registered email id to get access to their accounts.

Fig 41: Home Page of Admin

The homepage of admin has one additional feature to upadate the records in the databases.

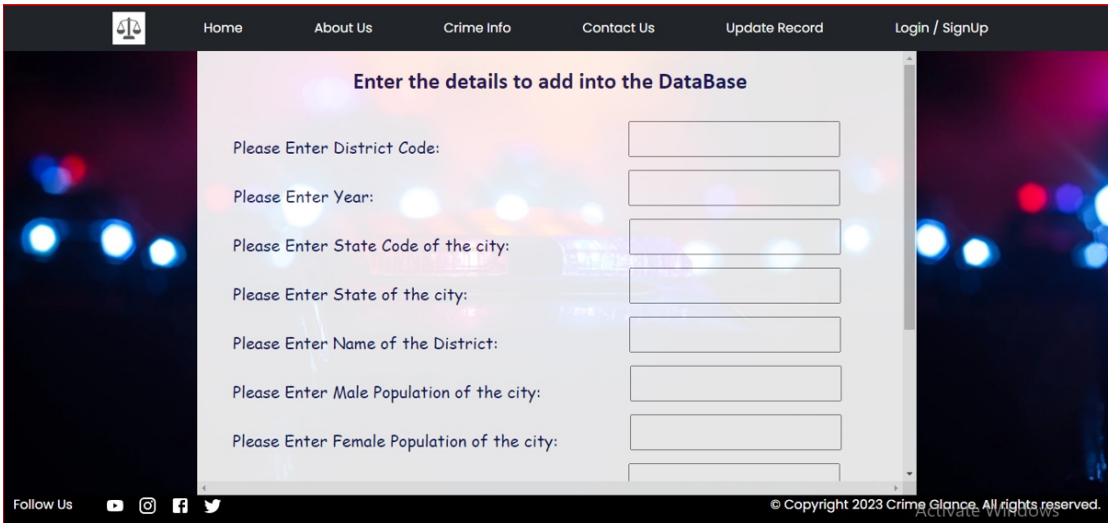

Fig 42: Update Record Page

In the update record page, admin can make any required changes in the existing databases and also update the records such as district code, year, state code, state, district, total population, male and female population, literacy rate, per capita income, crime rate, etc.

5. CONCLUSIONS

In conclusion, our website for crime prediction is a useful resource for comprehending and mitigating prospective criminal activity in a variety of contexts. We work to deliver precise insights and forecasts to improve public safety by utilising a large dataset, sophisticated algorithms, and machine learning techniques.

Although our algorithm makes an effort to deliver trustworthy results, it is crucial to recognise that crime prediction is a difficult undertaking. Although predictions are based on previous data and a number of other factors, they cannot absolutely ensure that future criminal activity will not occur. However, law enforcement organisations, neighbourhood associations, and individuals seeking to avoid crime proactively can benefit much from using our website. We want to make it easier to implement targeted interventions, resource allocation, and crime prevention methods by identifying crime hotspots, trends, and influencing factors.

It is imperative to stress that crime prediction should not take the position of law enforcement or individual discretion. As an additional tool to improve situational awareness and guide decision-making, it should be used instead.

In order to fully take advantage of the information offered by our website for crime prediction, we seek cooperation and active participation from law enforcement organisations, community stakeholders, and people. Together, we can create safer settings, use resources wisely, and put preventative measures in place to lower crime rates.

We understand how crucial it is to consistently enhance our crime prediction models by incorporating the most recent data, enhancing algorithms, and taking user input into account. By doing this, we want to improve the precision and dependability of our forecasts and hence have a positive influence on efforts to prevent crime.Finally, in the joint effort to create safer neighbourhoods, our crime prediction website is a valuable tool. We work to create a future where crime rates are lower, people can live more independently, and technology can help.

6. REFERENCES

7. APPENDICES