A

Mini Project

On

**CRIME RATE PREDICTION & ANALYSIS USING**

**K-MEANS CLUSTERING ALGORITHM**

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

**BACHELOR OF TECHNOLOGY**

in

**COMPUTER SCIENCE AND ENGINEERING**

by

T.NAGA SURENDAR REDDY(217R1A05J6)

T.RASHMITHA(227R5A0514)

S.MAHENDAR(227R5A0517)

Under the Guidance of

**S.V SUJI APARNA**

(Assistant Professor)

## DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

**CMR TECHNICAL CAMPUS**

**UGC AUTONOMOUS**

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

Kandlakoya (V), Medchal Road, Hyderabad-501401.

**2021-25**

## DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

## 

## 

**CERTIFICATE**

This is to certify that the project entitled “**CRIME RATE PREDICTION AND ANALYSIS USING K-MEANS CLUSTERING ALGORITHM**” being submitted by **T.NAGA SURENDAR REDDY(217R1A05J6), T.RASHMITHA(227R5A0514) & S.MAHENDAR(227R5A0517)** in partial fulfillment of the requirements for the award of the degree of B.Tech in **Computer Science and Engineering** to the **Jawaharlal Nehru Technological University,Hyderabad** is a record of bonafide work carried out by them under our guidance and supervision during the year 2024-25.

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

#### **S.V SUJI APARNA Dr. A. RAJI RE****DDY**

#### **Assistant Professor DIRECTOR**

**INTERNAL GUIDE**

**Dr. N.BHASKAR EXTERNAL EXAMINER**

**HEAD OF THE DEPARTMENT**

**Submitted for viva voice Examination held on**

**ACKNOWLEGDEMENT**

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

**S.V SUJI APARNA,** Assistant Professor for his exemplary guidance, monitoring and constant encouragement throughout the project work. The blessing, help and guidance given by him shall carry us a long way in the journey of life on which we are about to embark.

We also take this opportunity to express a deep sense of gratitude to Project Review Committee (PRC) Coordinators: **Dr. K.Maheswari, Dr. J.Narasimha rao, Ms. K.Shilpa, Mr. K.Ranjith Reddy,** for their cordial support, valuable information and guidance, which helped us in completing this task through various stages.

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

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

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

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

**T.NAGA SURENDAR REDDY (217R1A05J6)**

**T.RASHMITHA (227R5A0514)**

**S.MAHENDAR(227R5A0517)**

**ABSTRACT**

In India, the crime rate is increasing each day. In the current situation, recent technological influence, effects of social media and modern approaches help the offenders to achieve their crimes. Both analysis and prediction of crime is a systematized method that classifies and examines the crime patterns. There exist various clustering algorithms for crime analysis and pattern prediction but they do not reveal all the requirements. Among these, K means algorithm provides a better way for predicting the results. The proposed research work mainly focused on predicting the region with higher crime rates and age groups with more or less criminal tendencies. This proposes an optimized K means algorithm to lower the time complexity and improve efficiency in the result.

## LIST OF FIGURES

|  |  |  |
| --- | --- | --- |
| **FIGURE NO** | **FIGURE NAME** | **PAGE NO** |
| Figure 3.1 | Project Architecture | 8 |
| Figure 3.2 | Use case diagram | 10 |
| Figure 3.3 | Class diagram | 11 |
| Figure 3.4 | Sequence diagram | 12 |
| Figure 3.5 | Activity diagram | 13 |

**LIST OF SCREENSHOTS**

|  |  |  |
| --- | --- | --- |
| **SCREENSHOT NO** | **SCREENSHOT NAME** | **PAGE NO**. |
| Screenshot 5.1 | Training Dataset | 20 |
| Screenshot 5.2 | CMD.exe | 21 |
| Screenshot 5.3 | Home Page | 22 |
| Screenshot 5.4 | Login Page | 23 |
| Screenshot 5.5 | Login Successful Page | 24 |
| ScreenShot 5.6 | Upload Dataset | 25 |
| Screenshot 5.7 | Cluster Prediction | 26 |
| Screenshot 5.7.1 | Cluster Prediction Result | 27 |
| Screenshot 5.8 | Future Prediction | 28 |
| Screenshot 5.8.1 | Future Prediction Result | 29 |
| Screenshot 5.9  Screenshot 5.9.1 | Analysis  Analysis Result | 30  31 |

**LIST OF TABLES**

TABLE NO TABLE NAME PAGENO

Table 6.3.1 Uploading Dataset 34

Table 6.3.2 Classification 34

|  |  |  |
| --- | --- | --- |
|  |  |  |

**TABLE OF CONTENTS**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **ABSTRACT** | |  |  | i |
| **LIST OF FIGURES** | | |  | ii |
| **LIST OF SCREENSHOTS** | | | | iii |
| **1.** | **INTRODUCTION** | | | 1 |
|  | 1.1 | PROJECT SCOPE | | 1 |
|  | 1.2 | PROJECT PURPOSE | | 1 |
|  | 1.3 | PROJECT FEATURES | | 1 |
| **2.** | **SYSTEM ANALYSIS** | | | 2 |
|  | 2.1 | PROBLEM DEFINITION | | 2 |
|  | 2.2 | EXISTING SYSTEM | | 2 |
|  |  | 2.2.1 | LIMITATIONS OF THE EXISTING SYSTEM | 3 |
|  | 2.3 | PROPOSED SYSTEM | | 3 |
|  |  | 2.3.1 | ADVANTAGES OF PROPOSED SYSTEM | 4 |
|  | 2.4 | FEASIBILITY STUDY | | 4 |
|  |  | 2.4.1 | ECONOMIC FESIBILITY | 4 |
|  |  | 2.4.2 | TECHNICAL FEASIBILITY | 5 |
|  |  | 2.4.3 | SOCIAL FEASIBILITY | 5 |
|  | 2.5 | HARDWARE & SOFTWARE REQUIREMENTS | | 5 |
|  |  | 2.5.1 | HARDWARE REQUIREMENTS | 5 |
|  |  | 2.5.2 | SOFTWARE REQUIREMENTS | 6 |
| **3.** | **ARCHITECTURE** | | | 7 |
|  | 3.1 | PROJECT ARCHITECTURE | | 7 |
|  | 3.2 | DESCRIPTION | | 8 |
|  | 3.3 | USECASE DIAGRAM | | 9 |
|  | 3.4 | CLASS DIAGRAM | | 10 |
|  | 3.5 | SEQUENCE DIAGRAM | | 11 |
|  | 3.6 | ACTIVITY DIAGRAM | | 12 |
| **4.** | **IMPLEMENTATION** | | | 13 |
|  | 4.1 | SAMPLE CODE | | 13 |
| **5.** | **RESULTS** | | | 19 |
| **6.** | **TESTING** | |  | 31 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 6.1 | INTRODUCTION TO TESTING | | 31 |
|  | 6.2 | TYPES OF TESTING | | 31 |
|  |  | 6.2.1 | SYSTEM TESTING | 31 |
|  |  | 6.2.2 | UNIT TESTING | 32 |
|  |  | 6.2.3 | INTEGRATION TESTING | 32 |
|  |  | 6.2.4 | FUNCTIONAL TESTING | 32 |
|  |  | 6.2.5 | WHITE BOX TESTING | 33 |
|  |  | 6.2.6 | BLACK BOX TESTING | 33 |
|  |  | 6.2.7 | ACCEPTANCE TESTING | 33 |
|  | 6.3 | TEST CASES | | 34 |
|  |  | 6.3.1 | UPLOADING DATASET | 34 |
|  |  | 6.3.2 | CLASSIFICATION | 34 |
| **7.** | **CONCLUSION & FUTURE SCOPE** | | | 35 |
|  | 7.1 | PROJECT CONCLUSION | | 35 |
|  | 7.2 | FUTURE SCOPE | | 36 |
| **8.** | **REFERENCES** | |  | 37 |
|  | 8.1 | REFERENCES | | 37 |
|  | 8.2 | WEBSITES | | 38 |

**1.INTRODUCTION**

**1.INTRODUCTION**

### 

### **1.1 PROJECT SCOPE**

This project is titled “Crime Rate Prediction and Analysis Using K-Means Clustering Algorithm.” The software allows users to input crime-related data and predicts the likelihood of criminal activities in specific regions and age groups. It provides the ability to analyze multiple crime data points, enabling law enforcement agencies to make informed decisions on crime prevention. The project leverages an optimized K-means clustering algorithm to improve prediction accuracy and reduce time complexity. The focus is on identifying regions with high crime rates and age groups more likely to commit crimes.

### **1.2 PROJECT PURPOSE**

The This system has been developed to facilitate the identification, retrieval, and analysis of crime-related data. The system is built with manually exclusive features that ensure accurate crime prediction and analysis. In all cases, the system will specify objects such as regions, offender age groups, and crime types, based on physical or performance characteristics. The collected data plays a vital role in identifying, accessing, storing, and matching crime patterns and records to ensure comprehensive prediction and analysis.

### **1.3 PROJECT FEATURES**

The main features of this project are that the system functions as a problem-solving tool, helping law enforcement agencies and policymakers identify and predict crime patterns effectively. The system analyzes the challenges related to crime management and offers predictive insights as solutions. These solutions are evaluated, and the best outcomes are selected through analytical methods. The predictions and analysis results are presented to the end-users (e.g., law enforcement officials) for their approval.

# **2.SYSTEM ANALYSIS**

## 2. SYSTEM ANALYSIS

### 

### **SYSTEM ANALYSIS**

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

**2.1 PROBLEM DEFINITION**

A detailed study of the process must be made by various techniques like Image processing, etc. The data collected by these sources must be scrutinized to arrive to a conclusion. The conclusion is an understanding of how the system functions. This system is called the existing system. Now the existing system is subjected to close study and problem areas are identified. The designer now functions as a problem solver and tries to sort out the difficulties that the enterprise faces. The solutions are given as proposals. The proposal is then weighed with the existing system analytically and the best one is selected. The proposal is presented to the user for an endorsement by the user. The proposal is reviewed on user request and suitable changes are made. This is loop that ends as soon as the user is satisfied with proposal.

### **2.2 EXISTING SYSTEM**

Crime for investigating crimes. Implementing a clustering algorithm on crime datasets enables analysis of crimes analysis tool is developed using various distinct data mining methods. It supports the police officers.It makes identification and analysis of

various criminality trends over the years through their conclusion. The random initial starting points produced by K-means which gives results in the form of cluster that helps in reaching the local optima . So to overcome this problem, the partitioned data along with the data axis with the highest variance for assigning the initial centroid for K-Means clustering was applied. So it is observed that the proposed technique uses a lesser .

**2.21 LIMITATIONS OF EXISTING SYSTEM**

* **Delayed Predictions:** Traditional systems often provide flood forecasts too late, leaving insufficient time for preventive actions
* **Less Accuracy:** Dependency on historical data, which might not capture new or emerging crime trends effectively.
* **Low Efficiency:** Iterative nature of K-means requires multiple passes through large datasets, increasing clustering time.
* **Overfitting to Local Optima:** The system may reach only local optima due to the random initialization of centroids, impacting the quality of cluster formation and leading to inconsistent results across different runs.
* **Limited Scalability:** As crime data grows in size and complexity, the existing system may struggle to scale efficiently, leading to performance degradation over time.

**2.3 PROPOSED SYSTEM**

We are working on Spyder for implementation. Here we use a Spyder 3.7 version. Spyder is an integrated development environment for systematic programming in Python. Here we implemented different packages like matplotlib ,numpy,sklearn, pandas,etc. Which helps to plot elbow graph and data frame table using a K-means clustering algorithm? Dataset is collected from Kaggle datasets and import datasets into Spyder in CSV format as shown in Fig 1. We perform normalization for finding the accurate number of clusters (k) using the elbow method. The elbow method performs k- means clustering on the obtained dataset for a range of values of k (2-15) and calculates the SSE. A line chart of the SSE is plotted for each value of k.

**2.3.1 ADVANTAGES OF THE PROPOSED SYSTEM**

* **High Accuracy**: The proposed System utilizes advanced techniques such as the elbow method for optimal cluster selection, ensuring that the number of clusters (K) is accurately determined.
* **High Efficiency:** By leveraging efficient libraries such as NumPy, Pandas, and scikit-learn, the system processes large datasets quickly. The integrated development environment, Spyder, facilitates streamlined coding and debugging, enhancing overall efficiency in implementation.
* **User-Friendly Interface:**  Using Spyder as the IDE offers a user-friendly experience with features such as an interactive console, variable explorer, and integrated documentation, making it accessible even for users with varying levels of programming expertise.

### **2.4 FEASIBILITY STUDY**

The feasibility of the project is analyzed in this phase and business proposal is put forth with a very general plan for the project and some cost estimates. During system analysis the feasibility study of the proposed system is to be carried out. Three key considerations involved in the feasibility analysis are

**• Economic Feasibility**

**• Technical Feasibility**

**• Social Feasibility**

### **2.4.1 ECONOMIC FEASIBILITY**

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

**2.4.2 TECHNICAL FEASIBILITY**

This study is completed to check the specialized possibility, or at least, the specialized necessities of the framework. Any framework created should not have a This study is carried out to check the technical feasibility, that is, the technical requirements of the system. Any system developed must not have a high demand on the available technical resources. The developed system must have a modest requirement, as only minimal or null changes are required for implementing this system

**2.4.3 SOCIAL FEASIBILITY**

The part of study is to actually take a look at the degree of acknowledgment of the framework by the client. This incorporates the method involved with preparing the client to productively utilize the framework. The client should not feel compromised by the framework, rather should acknowledge it as a need. The degree of acknowledgment by the clients exclusively relies upon the strategies that are utilized to teach the client about the system and to make him acquainted with it. His degree of certainty should be raised with the goal that he is additionally ready to make some valuable analysis, which is invited, as he is the last client of the framework.

### **2.5 HARDWARE & SOFTWARE REQUIREMENTS**

**2.5.1 HARDWARE REQUIREMENTS:**

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

* Processor : i3 or Above
* RAM : 4 GB (min)
* Hard Disk : 40 GB

### **2.5.2 SOFTWARE REQUIREMENTS:**

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

* Operating System : Windows 8 or above
* Coding language : Python
* Front End : HTML,CSS,JS
* Back End : Python
* Framework : Django

**3.ARCHITECTURE**

**3. ARCHITECTURE**

**3.1 PROJECT ARCITECTURE**

This project architecture shows the procedure followed for Crime Rate Prediction and Analysis using K-Means Clustering Algorithm

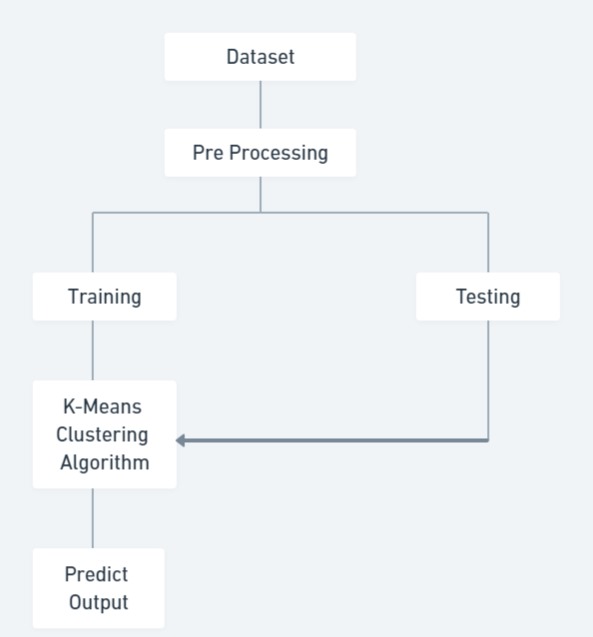


Figure 3.1:Project Arciteture

**3.2 DESCRIPTION**

* **Data Collection and Preparation:** Gather historical crime data from various sources, such as police reports, public databases, and community surveys.
* **K-Means Clustering Implementation:** Apply the K-Means clustering algorithm to categorize neighborhoods or regions based on crime rates and other relevant features
* **Crime Rate Prediction:** Develop a predictive model that utilizes clustering results to forecast future crime rates in specific areas.
* **Results Interpretation and Visualizatio**n**:** Analyze and interpret the clustering results to provide actionable insights for policymakers and law enforcement agencies.
* **Recommendations:** Formulate recommendations based on analysis results to improve community safety and resource allocation.

**3.3 USE CASE DIAGRAM**

In the use case diagram we have basically two actors who are the user and the Database System. A use case diagram is visual representation of how users interact with a system and is a key tool in the early stages of system design and development.

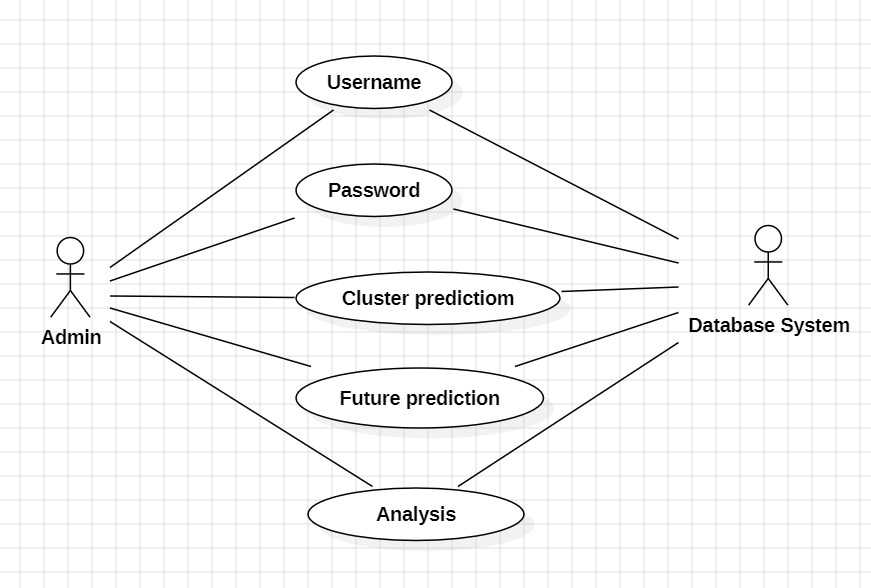


Figure 3.2: Use Case Diagram for Crime Rate Prediction and Analysis using K-Means Clustering Algorithm

**3.4 CLASS DIAGRAM**

Class Diagram is a collection of classes and objects

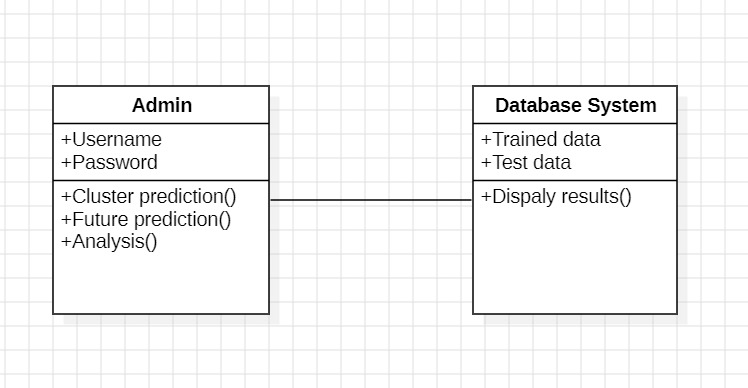


Figure 3.3: Class Diagram for Crime Rate Prediction and Analysis using K-

Means Clustering Algorithm

**3.5 SEQUENCE DIAGRAM**

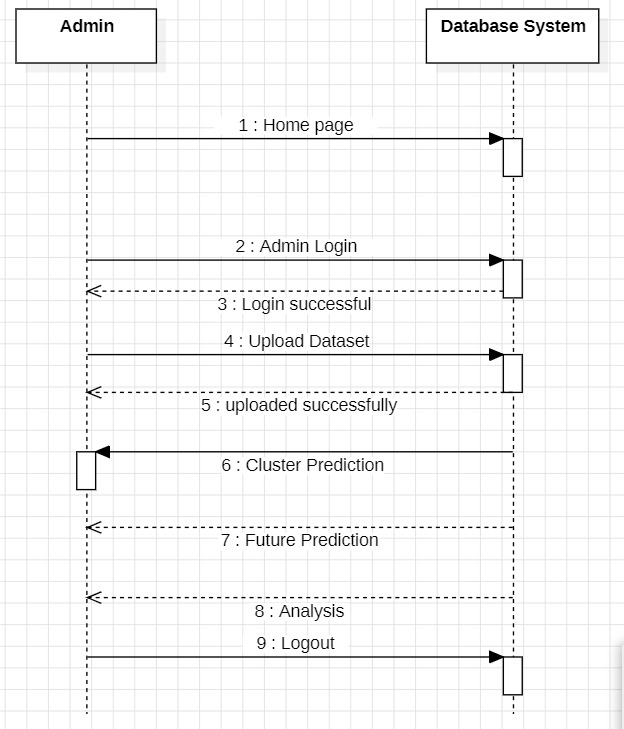
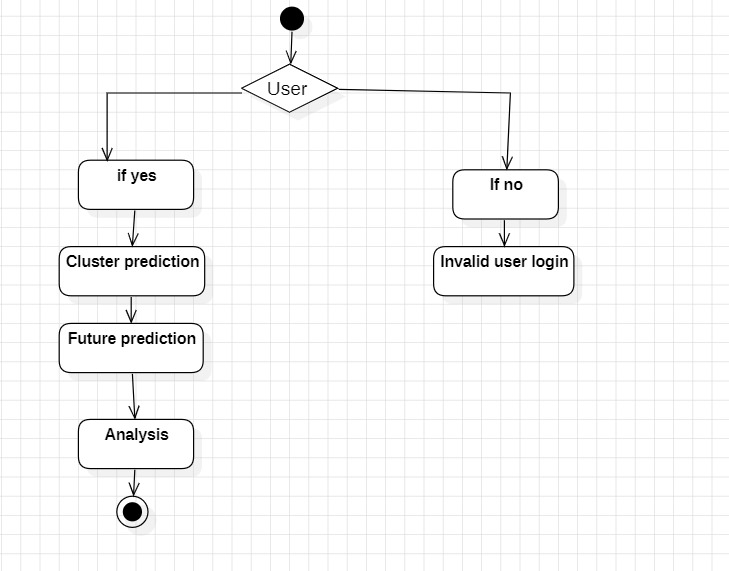


Figure 3.4: Sequence Diagram for Crime Rate Prediction and Analysis using

K-Means Clustering Algorithm

**3.6 ACTIVITY DIAGRAM**



**4.IMPLEMENTATION**

**4.1 SAMPLE CODE**

from django.shortcuts import render

from django.template import RequestContext

from django.contrib import messages

from django.http import HttpResponse

import os

import matplotlib.pyplot as plt

import pandas as pd

import numpy as np

from sklearn.preprocessing import LabelEncoder

from sklearn.preprocessing import MinMaxScaler

from sklearn.ensemble import RandomForestRegressor

from sklearn.model\_selection import train\_test\_split

from sklearn.cluster import KMeans

global dataset, kmeans\_cluster, theft\_cls, rape\_cls, murder\_cls

sc = MinMaxScaler(feature\_range = (0, 1))

le1 = LabelEncoder()

le2 = LabelEncoder()

le3 = LabelEncoder()

le4 = LabelEncoder()

def UploadDatasetAction(request):

    if request.method == 'POST':

        global dataset, kmeans\_cluster, theft\_cls, rape\_cls, murder\_cls

        myfile = request.FILES['t1']

        dataset = pd.read\_csv("Dataset/Dataset.csv", usecols=['States/UTs','District', 'Murder', 'Rape', 'Theft', 'Dowry\_Deaths', 'Year'])

        dataset.fillna(0, inplace = True)

        cols = ['States/UTs', 'District']

        dataset[cols[0]] = pd.Series(le1.fit\_transform(dataset[cols[0]].astype(str)))

        dataset[cols[1]] = pd.Series(le2.fit\_transform(dataset[cols[1]].astype(str)))

        X = dataset.values

        X = sc.fit\_transform(X)

        kmeans\_cluster = KMeans(n\_clusters=2, n\_init=1200)

        kmeans\_cluster.fit(X)

        dataset = pd.read\_csv("Dataset/Dataset.csv", usecols=['States/UTs','District', 'Year', 'Theft', 'Murder', 'Rape'])

        dataset.fillna(0, inplace = True)

        print(dataset)

        cols = ['States/UTs', 'District']

        dataset[cols[0]] = pd.Series(le3.fit\_transform(dataset[cols[0]].astype(str)))

        dataset[cols[1]] = pd.Series(le4.fit\_transform(dataset[cols[1]].astype(str)))

        theft\_Y = dataset['Theft'].values

        murder\_Y = dataset['Murder'].values

        rape\_Y = dataset['Rape'].values

        dataset.drop(['Theft'], axis = 1,inplace=True)

        dataset.drop(['Murder'], axis = 1,inplace=True)

        dataset.drop(['Rape'], axis = 1,inplace=True)

        X = dataset.values

        theft\_cls = RandomForestRegressor()

        theft\_cls.fit(X, theft\_Y)

        rape\_cls = RandomForestRegressor()

        rape\_cls.fit(X, rape\_Y)

        murder\_cls = RandomForestRegressor()

        murder\_cls.fit(X, murder\_Y)

        dataset = pd.read\_csv("Dataset/Dataset.csv")

        dataset.fillna(0, inplace = True)

        columns = list(dataset.columns)

        strdata = '<table border=1 align=center width=100%><tr><th><font size="" color="black">'+columns[0]+'</th>'

        for i in range(1,len(columns)):

            strdata+='<th><font size="" color="black">'+columns[i]+'</th>'

        strdata += "</tr>"

        dataset = dataset.values

        for i in range(len(dataset)):

            strdata += "<tr>"

            for j in range(len(dataset[i])):

                strdata+='<td><font size="" color="black">'+str(dataset[i,j])+'</td>'

            strdata += "</tr>"

        context= {'data':strdata}

        return render(request, 'ViewDataset.html', context)

def AdminLogin(request):

    if request.method == 'POST':

        user = request.POST.get('t1', False)

        password = request.POST.get('t2', False)

        if user == 'admin' and password == 'admin':

            context= {'data':user}

            return render(request, 'AdminScreen.html', context)

        else:

            context= {'data':"Invalid login details"}

            return render(request, 'Admin.html', context)

def index(request):

    if request.method == 'GET':

       return render(request, 'index.html', {})

def Admin(request):

    if request.method == 'GET':

       return render(request, 'Admin.html', {})

def UploadDataset(request):

    if request.method == 'GET':

       return render(request, 'UploadDataset.html', {})

def ClusterPrediction(request):

    if request.method == 'GET':

        dataset = pd.read\_csv("Dataset/Dataset.csv", usecols=['States/UTs','District', 'Year'])

        dataset.fillna(0, inplace = True)

        states = np.unique(dataset['States/UTs'].values)

        output = '<tr><td><font size="" color="black">States</b></td><td><select name="t1">'

        for i in range(len(states)):

            output += '<option value="'+states[i]+'">'+states[i]+'</option>'

        output += "</select></td></tr>"

        output += '<tr><td><font size="" color="black">District</b></td><td><select name="t2">'

        for i in range(len(states)):

            district = dataset[dataset['States/UTs'] == states[i]]['District']

            district = district.values

            output += '<option value="'+states[i]+'"><b>--'+states[i]+'--</b></option>'

            for j in range(len(district)):

                output += '<option value="'+district[j]+'">'+district[j]+'</option>'

        output += "</select></td></tr>"

        output += '<tr><td><font size="" color="black">Year</b></td><td><select name="t3">'

        year = np.unique(dataset['Year'].values)

        for i in range(len(year)):

            output += '<option value="'+str(year[i])+'">'+str(year[i])+'</option>'

        output += "</select></td></tr>"

        context= {'states':output}

        return render(request, 'ClusterPrediction.html', context)

def ClusterPredictionAction(request):

    if request.method == 'POST':

        state = request.POST.get('t1', False)

        district = request.POST.get('t2', False)

        year = request.POST.get('t3', False)

        murder = request.POST.get('t4', False)

        rape = request.POST.get('t5', False)

        theft = request.POST.get('t6', False)

        dowry = request.POST.get('t7', False)

        temp = []

        temp.append([state, district, murder, rape, theft, dowry, year])

        test = pd.DataFrame(temp, columns=['States/UTs','District', 'Murder', 'Rape', 'Theft', 'Dowry\_Deaths', 'Year'])

        test.fillna(0, inplace = True)

        test['States/UTs'] = pd.Series(le1.transform(test['States/UTs'].astype(str)))

        test['District'] = pd.Series(le2.transform(test['District'].astype(str)))

        test = test.values

        test = sc.transform(test)

        predict = kmeans\_cluster.predict(test)

        print(predict)

        output = district+" Low Crime Rate Area"

        if predict == 1:

            output = district+" High Crime Rate Area"

        context= {'data':output}

        return render(request, 'index.html', context)

def FuturePrediction(request):

    if request.method == 'GET':

        dataset = pd.read\_csv("Dataset/Dataset.csv", usecols=['States/UTs','District', 'Year'])

        dataset.fillna(0, inplace = True)

        states = np.unique(dataset['States/UTs'].values)

        output = '<tr><td><font size="" color="black">States</b></td><td><select name="t1">'

        for i in range(len(states)):

            output += '<option value="'+states[i]+'">'+states[i]+'</option>'

        output += "</select></td></tr>"

        output += '<tr><td><font size="" color="black">District</b></td><td><select name="t2">'

        for i in range(len(states)):

            district = dataset[dataset['States/UTs'] == states[i]]['District']

            district = district.values

            output += '<option value="'+states[i]+'"><b>--'+states[i]+'--</b></option>'

            for j in range(len(district)):

                output += '<option value="'+district[j]+'">'+district[j]+'</option>'

        output += "</select></td></tr>"

        output += '<tr><td><font size="" color="black">Year</b></td><td><select name="t3">'

        year = np.unique(dataset['Year'].values)

        for i in range(len(year)):

            output += '<option value="'+str(year[i])+'">'+str(year[i])+'</option>'

        output += '<option value="2022">2022</option>'

        output += '<option value=2023>2023</option>'

        output += "</select></td></tr>"

        context= {'states':output}

        return render(request, 'FuturePrediction.html', context)

def FuturePredictionAction(request):

    if request.method == 'POST':

        state = request.POST.get('t1', False)

        district = request.POST.get('t2', False)

        year = request.POST.get('t3', False)

        classify\_type = request.POST.get('t4', False)

        temp = []

        temp.append([state, district, year])

        test = pd.DataFrame(temp, columns=['States/UTs','District', 'Year'])

        test.fillna(0, inplace = True)

        test['States/UTs'] = pd.Series(le1.transform(test['States/UTs'].astype(str)))

        test['District'] = pd.Series(le2.transform(test['District'].astype(str)))

        test = test.values

        predict = 0

        if classify\_type == "Theft":

            predict = "Future Predicted Thefts = "+str(int(theft\_cls.predict(test)[0]))

        if classify\_type == "Murder":

            predict = "Future Predicted Murders = "+str(int(murder\_cls.predict(test)[0]))

        if classify\_type == "Rape":

            predict = "Future Predicted Rapes = "+str(int(rape\_cls.predict(test)[0]))

        context= {'data':predict}

        return render(request, 'index.html', context)

def Analysis(request):

    if request.method == 'GET':

       return render(request, 'Analysis.html', {})

def AnalysisAction(request):

    if request.method == 'POST':

        classify\_type = request.POST.get('t1', False)

        strdata = '<table border=1 align=center width=100%>'

        if classify\_type == "Theft":

            strdata += '<tr><td><img src="static/analysis/theft\_bar.png" height="300" width="500"/></td></tr>'

            strdata += '<tr><td><img src="static/analysis/theft\_pie.png" height="300" width="500"/></td></tr>'

        if classify\_type == "Murder":

strdata += '<tr><td><img src="static/analysis/murder\_bar.png" height="300" width="500"/></td></tr>'

            strdata += '<tr><td><img src="static/analysis/murder\_pie.png" height="300" width="500"/></td></tr>'

        if classify\_type == "Rape":

            strdata += '<tr><td><img src="static/analysis/rape\_bar.png" height="300" width="500"/></td></tr>'

            strdata += '<tr><td><img src="static/analysis/rape\_pie.png" height="300" width="500"/></td></tr>'

        context= {'data':strdata}

        return render(request, 'ViewGraphs.html', context)

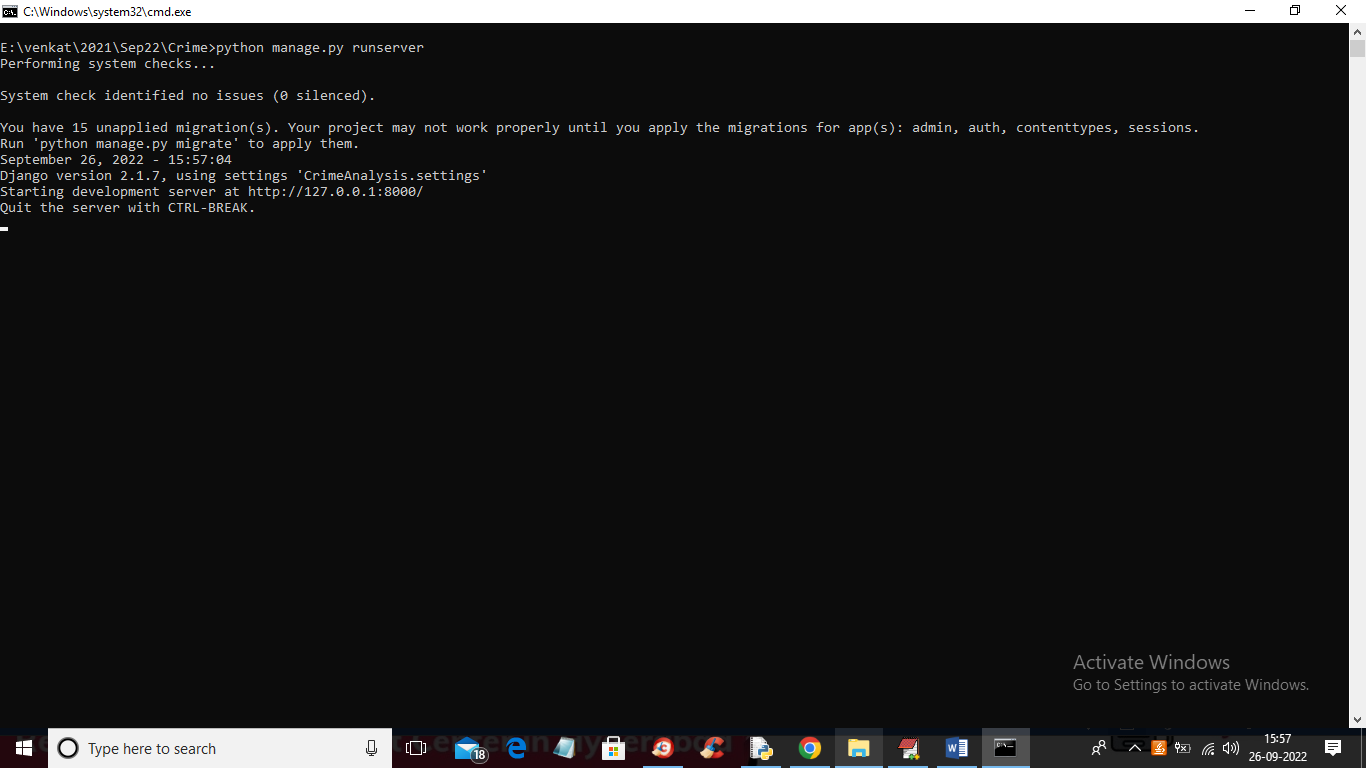
**5.RESULT**



Screenshot 5.1: Traning DataSet

Splitting DataSet Results for Crime Rate Predction and Analysis using K-

Means Clustering Algorithm.

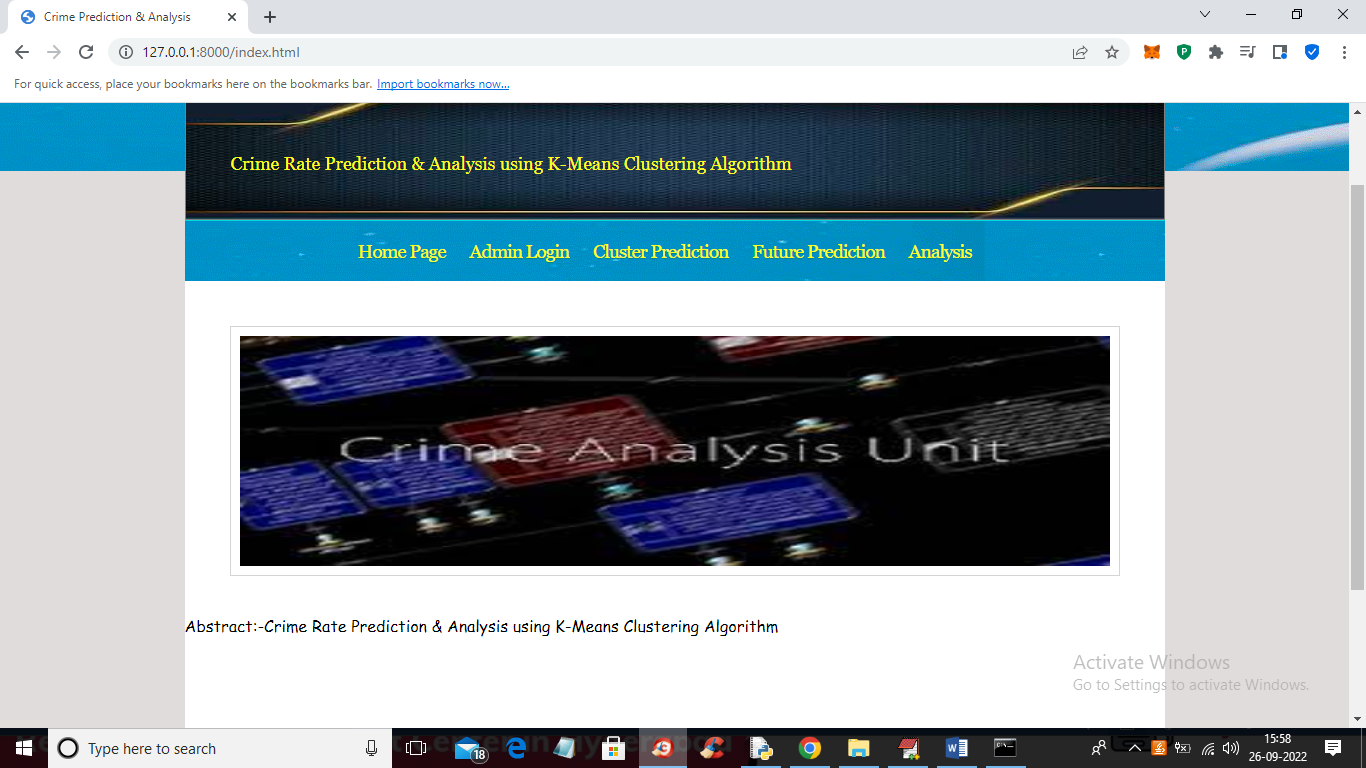


Screenshot 5.2: CMD.exe

In above screen python server started and now open browser and enter

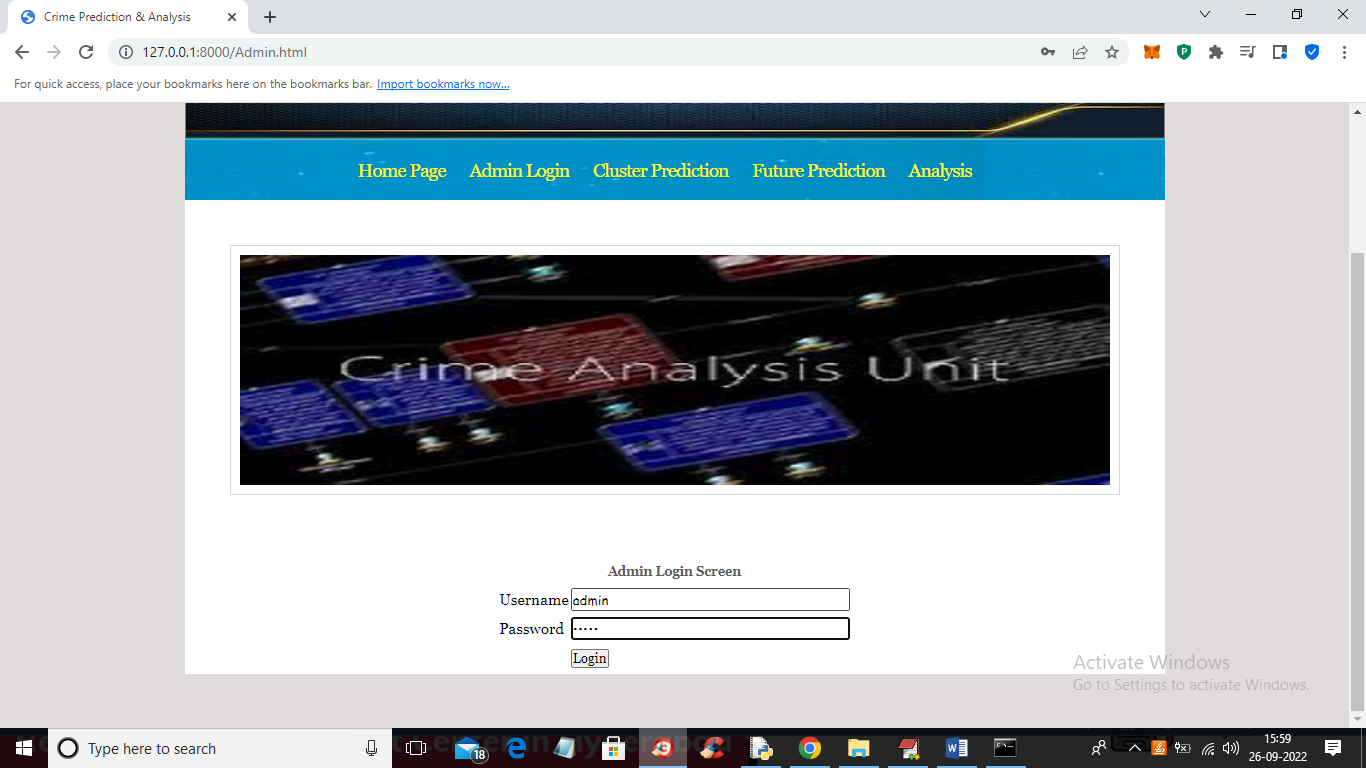
URL as ‘http://127.0.0.1:8000/index.html’ and press enter key to get

below page.



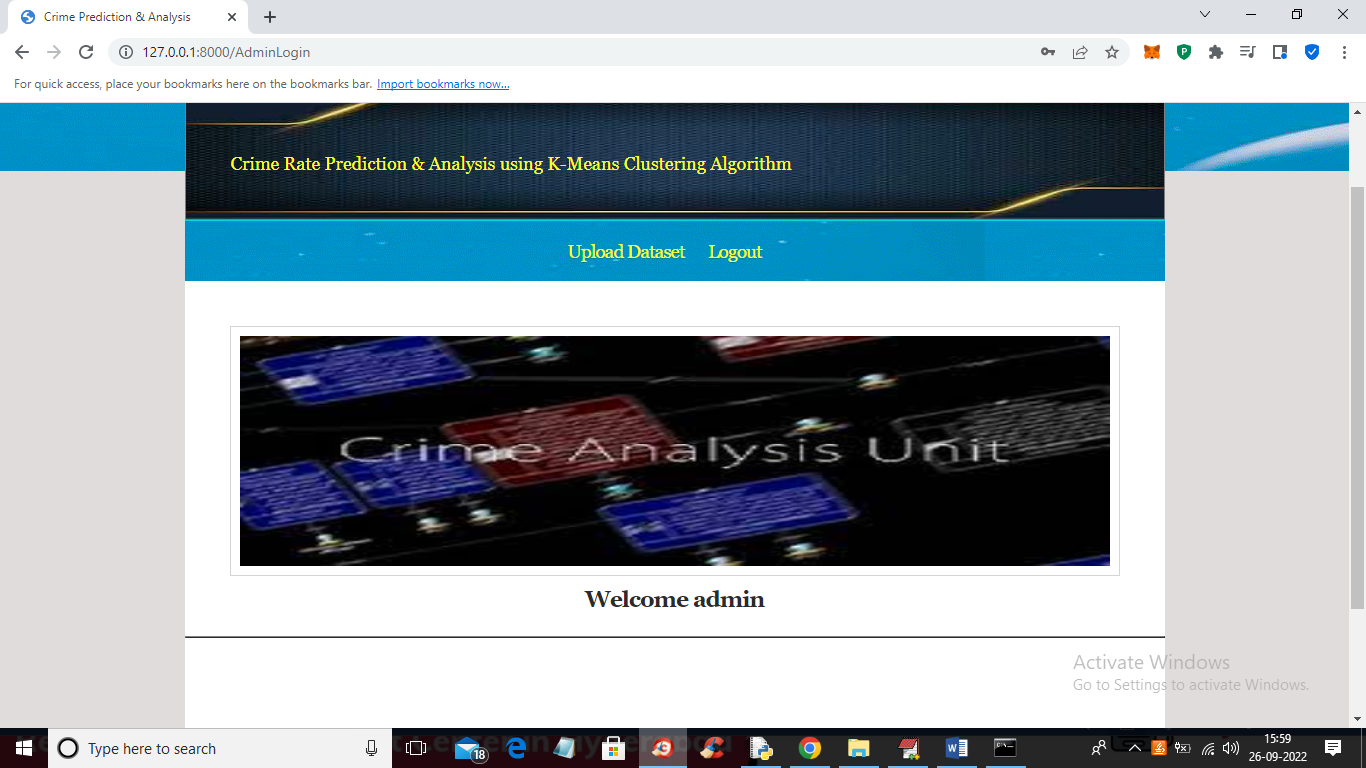
Screenshot 5.3: Home Page

In above screen click on ‘Admin Login’ link to get below login – page.



Screenshot 5.4: Login Page

In above screen admin is login and after login will get below screen.

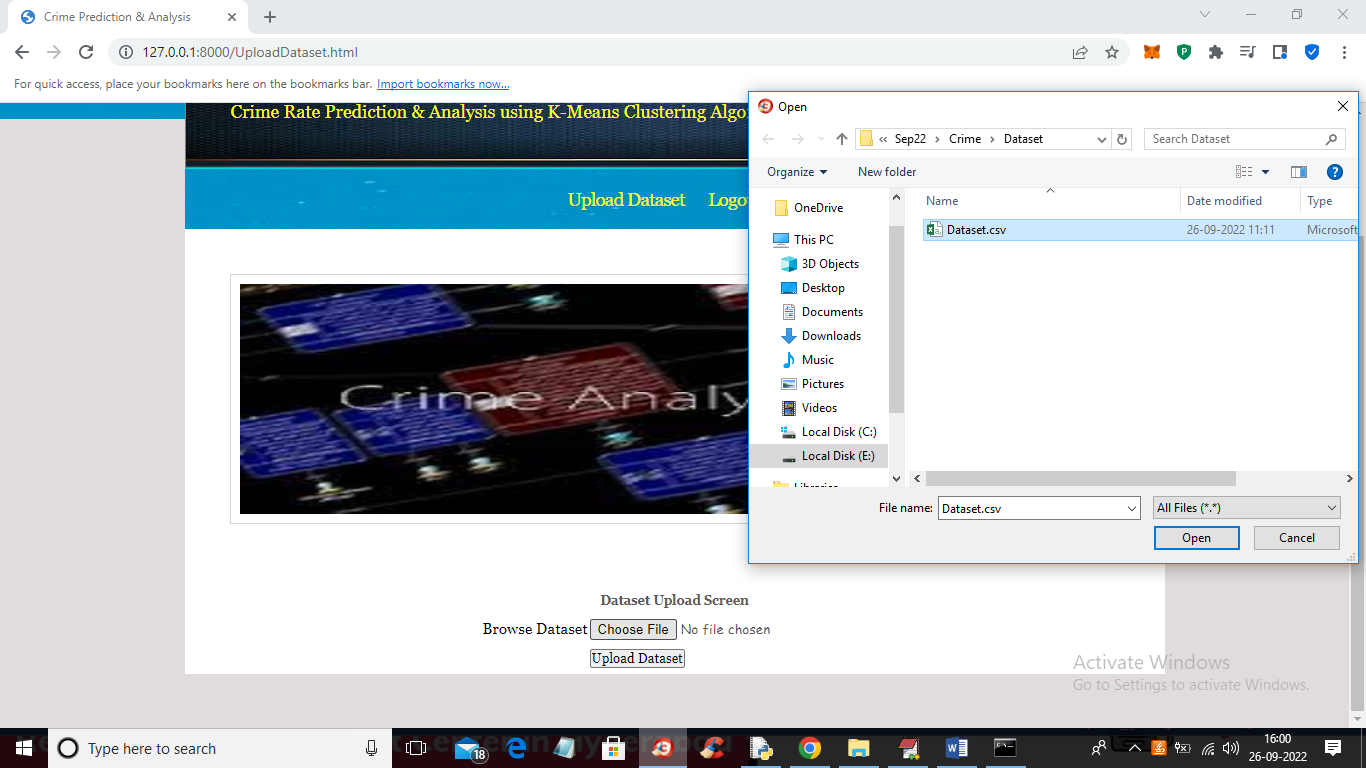


Screenshot 5.5: Login Successful Page

In above screen admin can click on ‘Upload Dataset’ link to upload dataset

and then click submit button to load dataset and then train it with machine

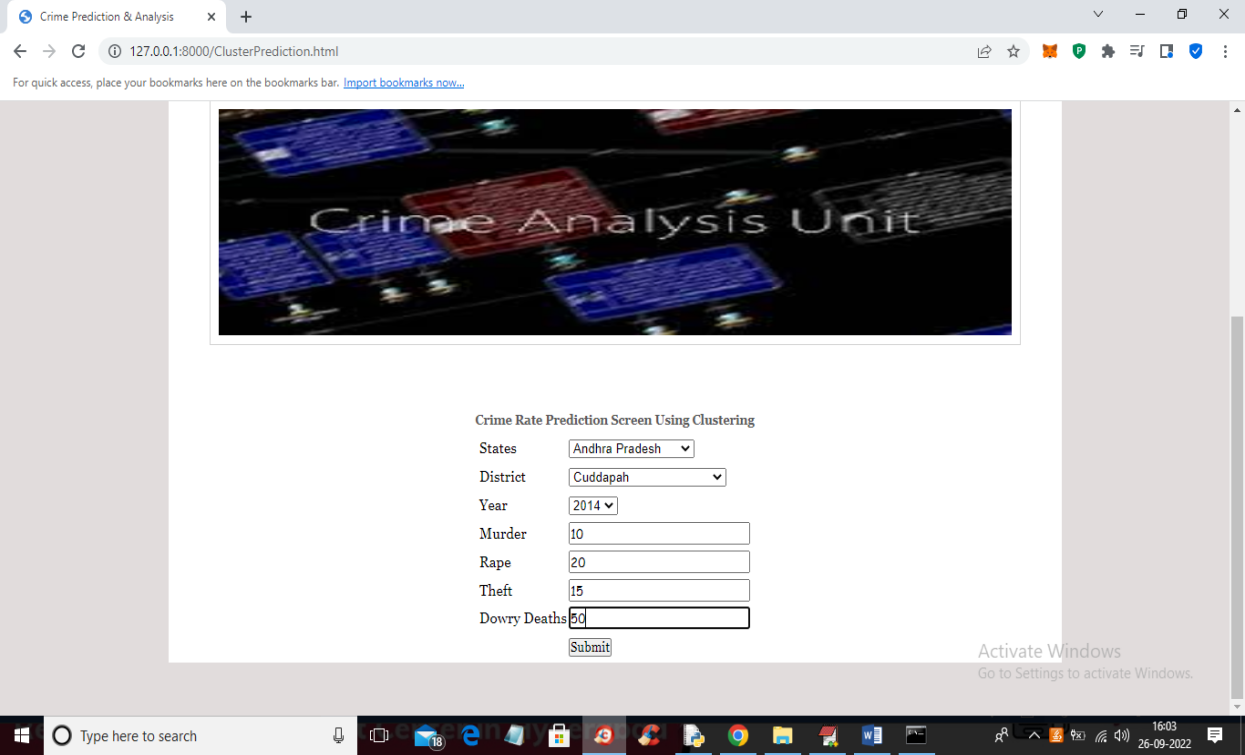
learning algorithms.



Screenshot 5.6: Upload Dataset

In above screen selecting and upload dataset and then click on ‘Open’

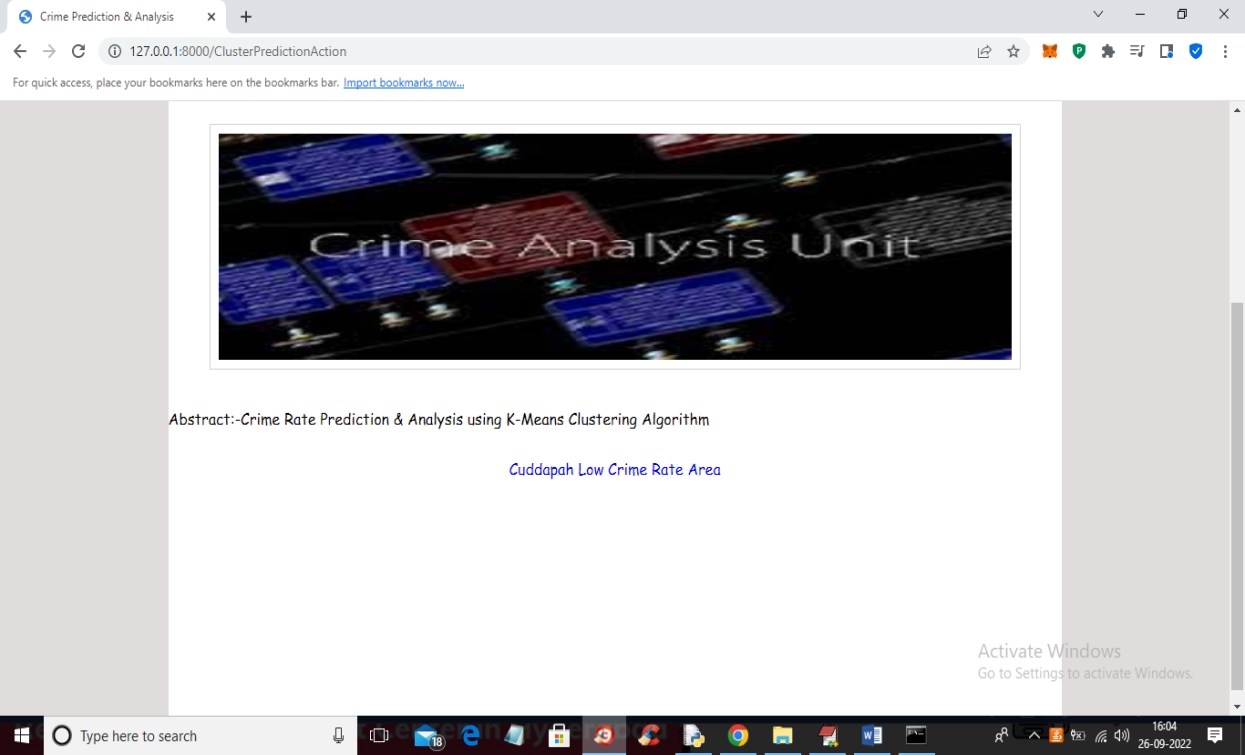
and ‘Upload Dataset’ button to load and complete training process.



Screenshot 5.7: Cluster Predction

In above screen select state and district name and then enter details of

crime and then press ‘Submit’ button to get below output.



Screenshot 5.7.1: Cluster Predction Result

In above screen in blue colour we got output as ‘Cuddapah is the Low

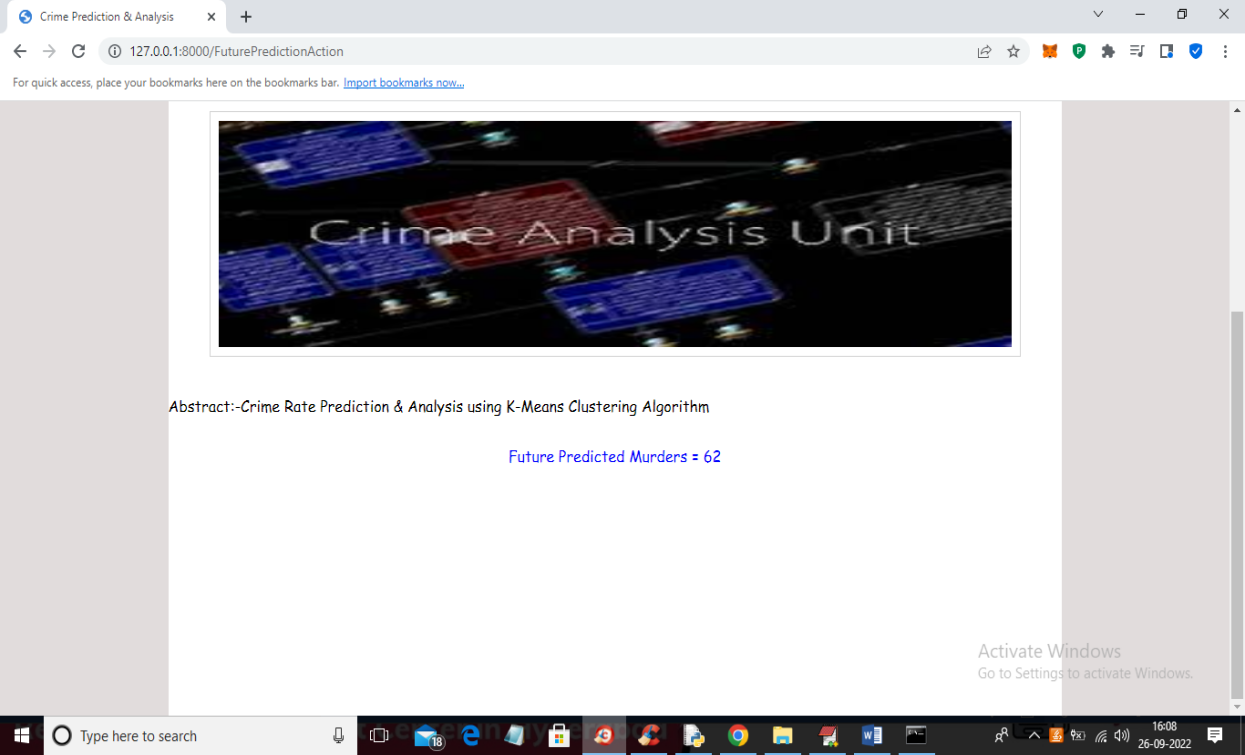
Crime Area’.



Screenshot 5.8:Future Prediction

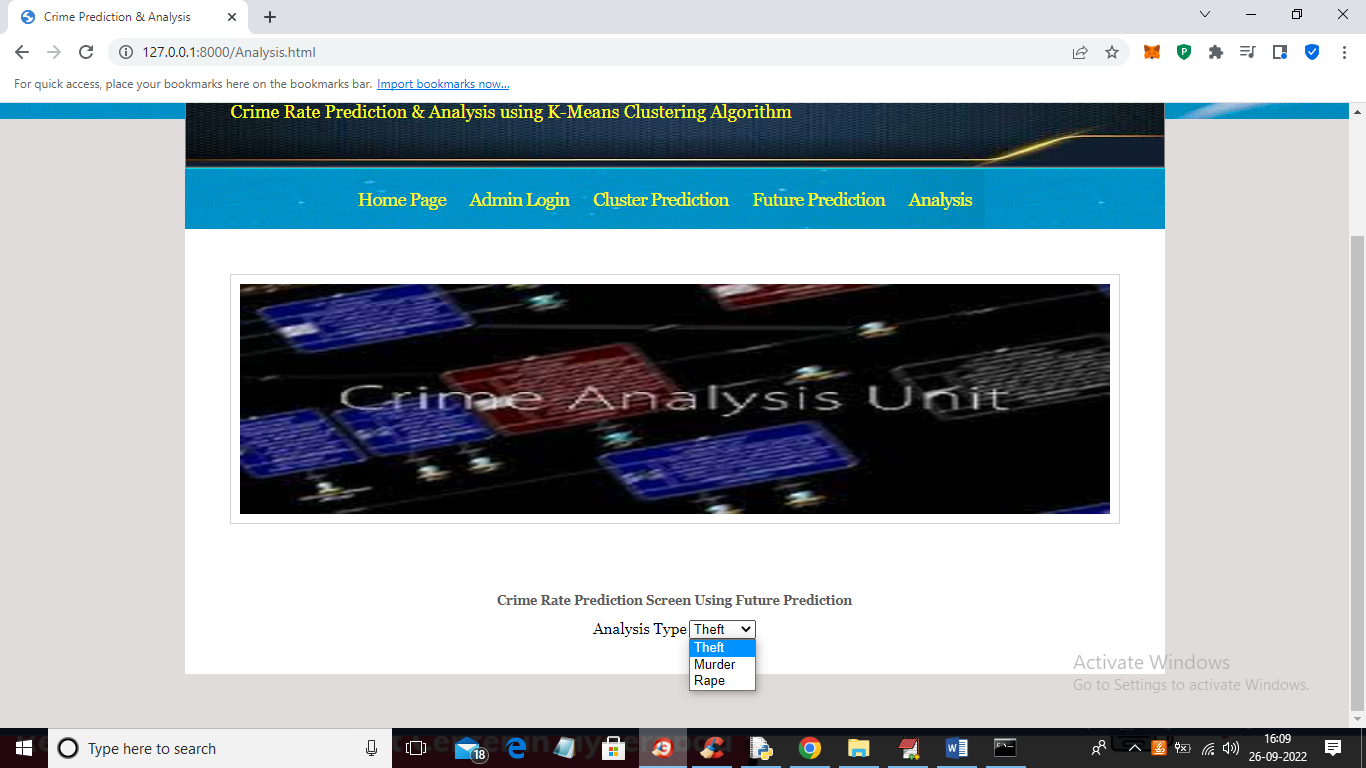
In above screen I selected state, district and then select crime as

‘Murder’ and then press ‘Submit’ button to get below output.



Screenshot 5.8.1: Future prediction result

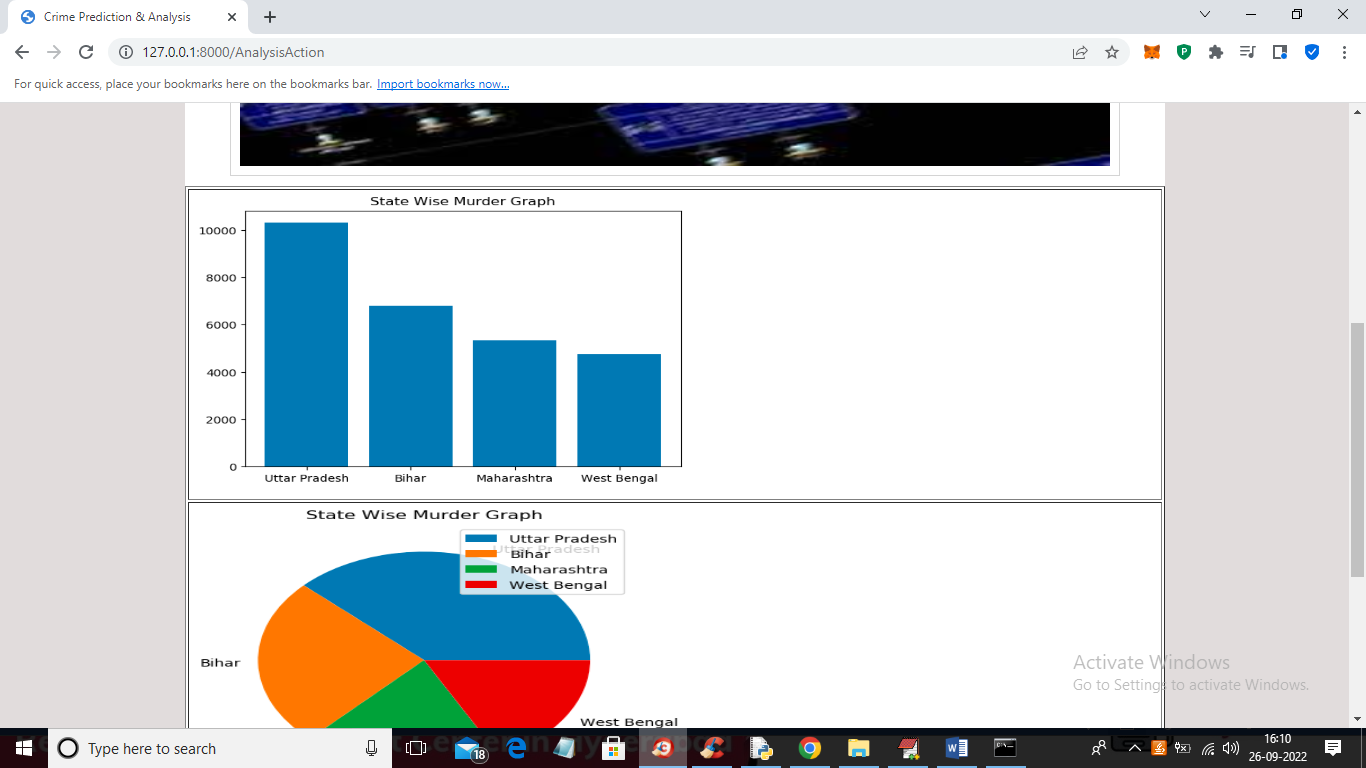
In above screen future predicted Murders for Gujarat state and Ahmadabad district is 62.



Screenshot 5.9: Analysis

In above screen select the type of analysis and press button to get

below graphs.



Screenshot 5.9.1: Analysis Result

The figure shows state-wise murder statistics in India, comparing Uttar

Pradesh, Bihar, Maharashtra, and West Bengal through a bar graph and a

pie chart.

**6.TESTING**

**6. TESTING**

**6.1 INTRODUCTION TO TESTING**

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

**6.2 TYPES OF TESTING**

1.System Testing

2. Unit testing

3. Integration Testing

4. Functional Testing

5. White box testing

6. black box testing

7. Acceptance testing

**6.2.1 SYSTEM TESTING**

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

**6.2.2 UNIT TESTING**

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

**6.2.3 INTEGRATION TESTING**

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

**6.2.4 FUNCTIONAL TESTING**

Functional tests provide systematic demonstrations that functions tested are available as specified by the business and technical requirements, system documentation, and user manuals.

Functional testing is centered on the following items:

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

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

Functions : identified functions must be exercised.

Output : identified classes of application outputs must be exercised.

Organization and preparation of functional tests is focused on requirements, key functions, or special test cases. In addition, systematic coverage pertaining to identify Business process flows; data fields, predefined processes, and successive processes must be considered for testing. Before functional testing is complete, additional tests are identified and the effective value of current tests is determined.

**6.2.5 WHITE BOX TESTING**

White Box Testing is a testing in which in which the software tester has knowledge of the inner workings, structure and language of the software, or at least its purpose. It is purpose. It is used to test areas that cannot be reached from a black box level.

**6.2.6 BLACK BOX TESTING**

Black Box Testing is testing the software without any knowledge of the inner workings, structure or language of the module being tested. Black box tests, as most other kinds of tests, must be written from a definitive source document, such as specification or requirements document, such as specification or requirements document. It is a testing in which the software under test is treated, as a black box .you cannot “see” into it. The test provides inputs and responds to outputs without considering how the software works.

**6.2.7 ACCEPTANCE TESTING**

User Acceptance Testing is a critical phase of any project and requires significant participation by the end user. It also ensures that the system meets the functional requirements.

**6.3 TESTCASES**

**TABLE NO-6.3.1 UPLOADING DATASET**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Test case ID** | **Test case name** | **Purpose** | **Input** | **Output** |
| 1 | Upload Crimes Data | User provides a Dataset of Crime rate | User provides a Dataset of Crime rate | Data uploaded successfully |

**TABLE NO-6.3.2 CLASSIFICATION**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Test case ID** | **Test case name** | **Purpose** | **Input** | **Output** |
| 1 | Cluster Prediction | To check the Low or High Crime Area. | Dataset with known crimes | High crime or low crime |
| 2 | Future Prediction | To check the crimes of specific district | Dataset with known crimes in various States and districts | Displays the future crime rate for the given input |
| 3 | Analysis | To check the crimes analysis through states | Select the specific crime | Displays the graph and pie chart state wise |

**7.CONCLUSION**

**7.CONCLUSION & FUTURE SCOPE**

**7.1 PROJECT CONCLUSION**

The crime rate analysis and prediction project utilizing the K-means clustering algorithm aims to uncover patterns in crime data to facilitate better resource allocation and policy-making. By employing K-means clustering, we can group regions with similar crime characteristics, identifying hotspots and trends over time. The project begins with data collection from reliable sources, such as law enforcement agencies and public records, covering various factors like crime type, location, and socioeconomic indicators. After preprocessing the data, including normalization and handling missing values, we apply the K-means algorithm to segment the data into clusters that reflect distinct crime patterns. Visualization techniques, such as scatter plots and heat maps, are used to present the clusters, enabling stakeholders to easily interpret the findings. Finally, we explore predictive modeling to forecast future crime rates based on historical data, enhancing the capacity of law enforcement to proactively address crime. This project not only demonstrates the utility of machine learning in social sciences but also contributes to safer communities through data-driven decision-making.

This information can significantly aid in optimizing resource allocation for police departments, allowing for targeted interventions in high-crime neighborhoods. While K-means clustering provided a strong foundation for understanding crime trends, its limitations, such as sensitivity to initial centroid placement and the assumption of spherical clusters, suggest the need for exploring more advanced techniques in future research. Additionally, the accuracy of our findings heavily depended on the quality of the data, emphasizing the importance of thorough data collection and preprocessing. Ultimately, the insights gained from this analysis can contribute to community safety by informing data-driven strategies that address crime effectively. Future directions could involve integrating socio-economic and demographic data, as well as employing more sophisticated predictive

models to enhance our understanding of crime dynamics. Overall, this project highlights the potential of data analytics in fostering safer communities through informed decision-making.

**7.2 FUTURE SCOPE**

The future scope of the crime rate analysis and prediction project using K-means clustering is broad and promising. One key direction is the integration of real-time data feeds, which would enhance prediction accuracy and enable swift responses to emerging crime trends. Additionally, exploring advanced machine learning techniques and other clustering algorithms could further refine the analysis. Expanding the dataset to include variables such as weather patterns, economic indicators, and community events would provide deeper insights into the factors influencing crime rates.

# 

# **8.BIBILOGRAPHY**

## 

## 8. BIBILOGRAPHY

### 

### **8.1 REFERENCES**

Here are some references for your project on crime rate prediction and analysis

1. \*\*De BruinJ.S.,Cocx\*\*,T.K,Kosters,W.A.,Laros,J.and Kok,J.N(2006) “Data mining approaches to criminal carrer analysis ,”in Proceedings of the Sixth International Conference on Data Mining (ICDM”06) ,Pp. 171-177.

2. \*\*Manish Gupta1\*\*, B.Chandra1 and M. P. Gupta1,”2007 Crime Data Mining for Indian Police Information System”.

3. \*\*Nazlena Mohamad Ali1\*\*, Masnizah Mohd2, Hyowon Lee3, Alan F. Smeaton3, Fabio Crestani4 and Shahrul Azman Mohd Noah2 ,”2010 Visual Interactive Malaysia Crime News Retrieval System”.

4. \*\*Sutapat Thirprungsri\*\* Rutgers University .USA ,2011 Cluster Analysis of Anomaly Detection in Accounting Data : An Audit Approach 1.

5. \*\*A.Malathi ,Dr.S.Santhosh Baboo\*\*. D.G. Vaishnav College,Chennai ,2011 “Algorithmic Crime Prediction Model Based on the Analysis of Crime Clusters”.

6. \*\*Malathi.A\*\* 1 ,Dr.S.Santhosh Baboo 2 and Anbarasi . A 31 Assistant professor ,Department of Computer Science ,Govt Arts College ,Coimbatore , India . 2 Readers , Department of Computer science , D.G. Vaishnav Collge ,Chennai , India, 2011 An intelligent Analysis of a city Crime Data Using Data Mining.

7. \*\*Malathi \*\*, A; Santhosh Baboo , S, 2011 An Enhanced Algorithm to Predict a Future Crime using Data Mining.

8. \*\*Kadhim B.Swadi al-Janabi \*\* . Department of Computer Science . Faculty of Mathematics and Computer Science .University of Kufa/Iraq , 2011 A Proposed Framework for Analyzing Crime DataSet using Decision Tree and Simple K-means Mining Algorithms.

**8.2 GITHUB LINK**

<https://github.com/MAHENDAR09/CRIME-RATE-PREDICTION-AND-ANALYSIS-USING-K-MEANSCLUSTERING-ALGORITHM>