

ANALYSIS OF CRIME USING MACHINE LEARNING

at

**Sathyabama Institute of Science and Technology
(Deemed to be University)**

**Submitted in partial fulfillment of the requirements for the award of
Bachelor of Engineering Degree in Computer Science and Engineering**

By

**Tumu Renusri
(Reg.No.38110605)
Sai Tejaswini
(Reg.No. 38110596)**



DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

SCHOOL OF COMPUTING

SATHYABAMA INSTITUTE OF SCIENCE AND TECHNOLOGY JEPPIAAR NAGAR,

RAJIV GANDHI SALAI, CHENNAI – 600119, TAMIL NADU

MARCH 2022



**SATHYABAMA
INSTITUTE OF SCIENCE AND TECHNOLOGY**

(DEEMED TO BE UNIVERSITY)

Accredited with Grade “A” by NAAC
(Established under Section 3 of UGC Act, 1956)
JEPPIAAR NAGAR, RAJIV GANDHI SALAI, CHENNAI– 600119
www.sathyabamauniversity.ac.in

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

BONAFIDE CERTIFICATE

This to certify that this Project Report is the bonafide work of **Tumu Renu Sri** (Reg. No: 38110605) and **Sai Tejaswini**(Reg.No: 38110596) who carried out the project entitled “**Analysis of crime using machine learning**” under my supervision from December 2021 to March 2022.

Internal Guide
Mrs. M. D. Anto Praveena, M.C.A.,M.E.,(Ph.D.,).

Head of the Department
Dr. S.VIGNESHWARI, M.E., Ph.D.,

Submitted for Viva voce Examination held on_____

Internal Examiner

External Examiner

DECLARATION

I, **Tumu Renu Sri (38110605)** hereby declare that the project report entitled **Analysis of crime using machine learning** was done by me under the guidance of **Mrs. M. D. Anto Praveena** is submitted in partial fulfillment of the requirements for the award of Bachelor of Engineering Degree in Computer Science and Engineering.

DATE: 13/11/2021

PLACE: Chennai

SIGNATURE OF THE CANDIDATE

ACKNOWLEDGEMENT

I am pleased to acknowledge my sincere thanks to the **Board of Management of Sathyabama** for their kind encouragement in doing this project and for completing it successfully. I am grateful to them.

I convey my thanks to **Dr. T. Sasikala M.E., Ph.D, Dean**, School of Computing, **Dr. S. Vigneshwari, M.E., Ph.D.** and **Dr. L. Lakshmanan, M.E., Ph.D., Heads of the Department of Computer Science and Engineering** for providing me necessary support and details at the right time during the progressive reviews.

I would like to express my sincere and deep sense of gratitude to my Project Guide **Mrs. M. D. Anto Praveena, M.C.A.,M.E.,(Ph.D.,)** for his valuable guidance, suggestions and constant encouragement paved the way for the successful completion of my project work.

I wish to express my thanks to all Teaching and Non-teaching staff members of the **Department of Computer Science and Engineering** who were helpful in many ways for the completion of the project.

ABSTRACT

Crime analysis and prediction is a systematic approach for identifying the crime. This system can predict regions which have high probability for crime occurrences and visualize crime prone areas. Using the concept of data mining we can extract previously unknown, useful information from unstructured data. The extraction of new information is predicted using the existing datasets. Crimes are treacherous and common social problems faced worldwide. Crimes affect the quality of life, economic growth and reputation of a nation. With the aim of securing the society from crimes, there is a need for advanced systems and new approaches for improving the crime analytics for protecting their communities. Propose a system which can analyze, detect, and predict various crime probability in a given region. Explains various types of criminal analysis and crime prediction using several data mining techniques.

TABLE OF CONTENTS

CHAPTER NO	TITLE	PAGE NO
	Abstract	5
	List of Figures	8
	List of Tables	6
	List of Abbreviation	9
1	INTRODUCTION	
	1.1 Outline of theProject	10
	1.2 Objective	10
2	LITERATURE SURVEY	
3	AIM AND SCOPE OF THE PROJECT	
	3.1 Existing System	13
	3.2 Proposed System	13
4	METHODOLOGY	
	4.1 Introduction To ML	15
	4.2 Training The Data	15
	4.2.1 Supervised Learning	16
	4.2.2 Unsupervised Learning	16
	4.3 Methods in Supervised Learning	16
	4.3.1 Classification	17
	4.3.2 Regression	18
	4.4 System Architecture	18

	4.5 KNN(K-Nearest Neighbors)	19
	4.6 Decision Tree	20
	4.7 Random Forest	20
	4.8 Datasets	21
	4.9 Data Manipulation Packages	22
	4.9.1 Pandas	22
	4.9.1 Numpy	22
	4.10 Modal Building Package	23
	4.10.1 Scikit-Learn	23
	4.10.2 Scikit-Plot	23
	4.10.3 Matplotlib's Pyplot	23
	4.11 Module Descriptions	24
	4.12 Python	25
	4.13 Data Mining	30
5	RESULTS AND DISCUSSION	35
6	CONCLUSION AND FUTURE WORK	36
7	APPENDICES	
	A)Source Code	39
	B) Screenshots	44
	C) Publication With Plagiarism Report	49

LIST OF FIGURES

FIGURE NO	FIGURE NAME	PAGE NO
4.1	Classification vs Regression	17
4.2	System Architecture	19
4.3	K-Nearest neighbors	19
4.4	Decision Tree	20
4.5	Random Forest	21
4.6	Activity Diagram	34

LIST OF ABBREVIATIONS

ABBREVIATION	EXPANSION
CSS	CASCADING STYLE SHEET
HTML	HYPER TEXT MARKUP LANGUAGE
HTTP	HYPertext TRAnsfer PROTOCOL
KNN	K-NEAREST NEIGHBORS
URL	UNIFORM RESOURCE LOCATOR
WWW	WORLD WIDE WEB

CHAPTER 1

INTRODUCTION

1.1 OUTLINE OF THE PROJECT

Day by day crime data rate is increasing because the modern technologies and hi-tech methods are helps the criminals to achieving the illegal activities .according to Crime Record Bureau crimes like burglary, arson etc have been increased while crimes like murder, sex, abuse, gang rap etc have been increased.crime data will be collected from various blogs, news and websites. The huge data is used as a record for creating a crime report database. The knowledge which is acquired from the data mining techniques will help in reducing crimes as it helps in finding the culprits faster and also the areas that are most affected by crime .

1.2 OBJECTIVE

This system gives the most trending technology-based skills used at the present. To help police to detect the crime type based on location. Provides the user with the technology he is saving a life and saves a lot of time. Html (Hypertext Markup Language) and CSS (Cascading Style Sheet) are two of the core technologies for building Web pages.

HTML provides the structure of the page, CSS the (virtual and aural) layout for a variety of devices. Along with graphics and scripting HTML and CSS are the basis of building Web pages and Web Applications. HTML gives authors the means to : Publish online documents with headings, text, tables, lists, photos. Retrieve online information via hypertext links, at the click of a button. Design forms for conducting transactions with remote services, for use in searching for information etc. Include spreadsheets, video clips, Sound clips and other applications directly in their documents. Flask is the most

popular framework of python for web development. It is free, open source and server-side(the code is executed on the server).

Machine learning is a process that is widely used for prediction. N number of algorithms are available in various libraries which can be used for prediction. In this article, we are going to build a prediction model on historic data using different machine learning algorithms and classifiers, plot the results and calculate the accuracy of the model on the testing data. Building/Training a model using various algorithms on a large dataset is one part of the data. But using these models within different applications is the second part of deploying machine learning in the real world. To put it to use in order to predict the new data, we have to deploy it over the internet so that the outside world can use it. In this article, we will talk about how we have trained a machine learning model, created a web application on it using Flask.

CHAPTER 2

LITERATURE SURVEY

Literature survey is the most important step in the software development process. Before developing the tool it is necessary to determine the time factor, economy and company strength. Once these things are satisfied, then the next step is to determine which operating system and language can be used for developing the tool. Once the programmers start building the tool the programmers need a lot of external support. This support can be obtained from senior programmers, from books or from websites. Before building the system the above considerations are taken into account for developing the proposed system. The major part of the project development sector considers and fully surveys all the required needs for developing the project. For every project Literature survey is the most important sector in the software development process. Before developing the tools and the associated designing it is necessary to determine and survey the time factor, resource requirement, man power, economy, and company strength. Once these things are satisfied and fully surveyed, then the next step is to determine about the software specifications in the respective system such as what type of operating system the project would require, and what all the necessary software is needed to proceed with the next step such as developing the tools, and the associated operations.

CHAPTER 3

AIM AND SCOPE OF THE PROJECT

3.1 EXISTING SYSTEM

Data mining in the study and analysis of criminology can be categorized into main areas, crime control and crime suppression. De Bruin et. al. introduced a framework for crime trends using a new distance measure for comparing all individuals based on their profiles and then clustering them accordingly. Manish Gupta et. al. highlights the existing systems used by Indian police as e-governance initiatives and also proposes an interactive query based interface as crime analysis tool to assist police in their activities. He proposed an interface which is used to extract useful information from the vast crime database maintained by National Crime Record Bureau (NCRB) and find crime hot spots using crime data mining techniques such as clustering etc. The effectiveness of the proposed interface has been illustrated on Indian crime records. Sutapat Thiprungsri examines the application of cluster analysis in the accounting domain, particularly discrepancy detection in audit. The purpose of his study is to examine the use of clustering technology to automate fraud filtering during an audit. He used cluster analysis to help auditors focus their efforts when evaluating group life insurance claims.

3.2 PROPOSED SYSTEM

In this project, we will be using the technique of machine learning and data science for crime prediction of crime data sets. The crime data is extracted from the official portal of police. It consists of crime information like location description, type of crime, date, time, latitude, longitude. Before training the model data preprocessing will be done following this feature selection and scaling will be done so that the accuracy obtained will be high. The K-Nearest Neighbor (KNN) classification and various other algorithms (Decision

Tree and Random Forest) will be tested for crime and propose one with better

query-based use for training. Visualization of the dataset will be done in terms of graphical representation of many cases, for example at which time the crime rates are high or at which month the criminal activities are high. The sole purpose of this project is to give a just idea of how machine learning can be used by law enforcement agencies to detect, predict and involve crimes at a much faster rate and thus reduce the crime rate. This can be used in other states or countries depending upon the availability of the dataset.

CHAPTER 4

METHODOLOGY

4.1 INTRODUCTION TO MACHINE LEARNING

A comparative study was carried out between violent crime patterns from the Communities and Crime Unnormalized Dataset versus actual crime statistical data using the open source data mining software Waikato Environment for Knowledge Analysis (WEKA). Three algorithms, namely, linear regression, additive regression, and decision stump, were implemented using the same finite set of features on communities and actual crime datasets. Test samples were randomly selected. The linear regression algorithm could handle randomness to a certain extent in the test samples and thus proved to be the best among all three selected algorithms. The scope of the project was to prove the efficiency and accuracy of ML algorithms in predicting violent crime patterns and other applications, such as determining criminal hotspots, creating criminal profiles, and learning criminal trends.

When considering WEKA , the integration of a new graphical interface called Knowledge Flow is possible, which can be used as a substitute for Internet Explorer. IT provides a more concentrated view of data mining in association with the process orientation, in which individual learning components (represented by java beans) are used graphically to show a certain flow of information. The authors then describe another graphical interface called an experimenter, which as the name suggests, is designed to compare the performance of multiple learning schemes on multiple data sets.

4.2 TRAINING THE DATA

There are basically two widely-used types of training that can be done to create a model:

- i. Supervised Learning
- ii. Un-supervised Learning

4.2.1 SUPERVISED LEARNING

Supervised learning is the machine learning task of learning a function that maps an input to an output based on example input-output pairs. It infers a function from labeled training data consisting of a set of training examples. In supervised learning, each example is a pair consisting of an input object (typically a vector) and a desired output value (also called the supervisory signal). A supervised learning algorithm analyzes the training data and produces an inferred function, which can be used for mapping new examples. An optimal scenario will allow for the algorithm to correctly determine the class labels for unseen instances. This requires the learning algorithm to generalize from the training data to unseen situations in a "reasonable" way.

4.2.2 UNSUPERVISED LEARNING

Unsupervised machine learning is the machine learning task of inferring a function that describes the structure of "unlabeled" data (i.e., data that has not been classified or categorized). Since the examples given to the learning algorithm are unlabeled, there is no straightforward way to evaluate the accuracy of the structure that is produced by the algorithm—one feature that distinguishes unsupervised learning from supervised learning and reinforcement learning. The type of training used in this model is SUPERVISED LEARNING.

4.3 METHODS IN SUPERVISED LEARNING

Supervised Learning mainly consists of two methods,

- Classification

- Regression

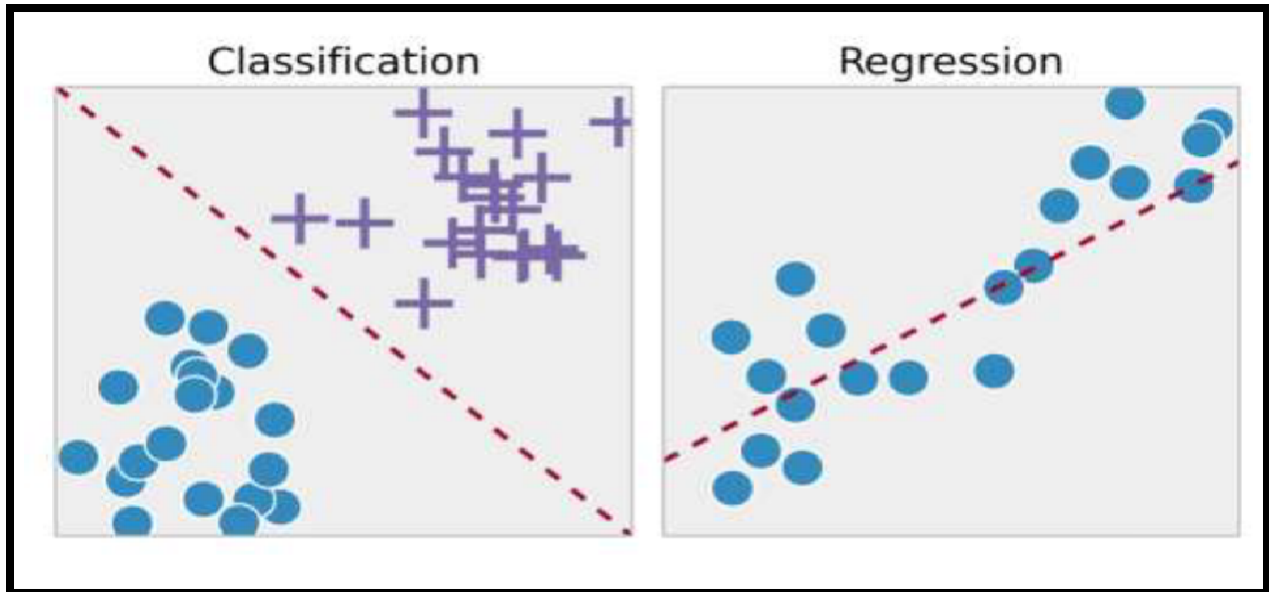


Fig 4.1 Classification vs Regression

4.3.1 CLASSIFICATION

In machine learning, classification(Fig 4.1) is the problem of identifying to which of a set of categories (sub-populations) a new observation belongs, on the basis of a training set of data containing observations (or instances) whose category membership is known. Examples are assigning a given email to the "spam" or "non-spam" class, and assigning a diagnosis to a given patient based on observed characteristics of the patient (gender, blood pressure, presence or absence of certain symptoms, etc.). Classification(Fig 4.1) is an example of pattern recognition. An algorithm that implements classification(Fig 4.1), especially in a concrete implementation, is known as a classifier. The corresponding unsupervised procedure is known as clustering, and involves grouping data into categories based on some measure of inherent similarity or

distance.

4.3.2 REGRESSION

Regression(Fig 4.1) analysis estimates the conditional expectation of the dependent variable given the independent variables – that is, the average value of the dependent variable when the independent variables are fixed. Less commonly, the focus is on a quantile, or other location parameter of the conditional distribution of the dependent variable given the independent variables. In all cases, a function of the independent variables called the regression(Fig 4.1) function is to be estimated. Regression(Fig 4.1) analysis is widely used for prediction and forecasting. It is also used to understand which among the independent variables are related to the dependent variable, and to explore the forms of these relationships. In restricted circumstances, regression(Fig 4.1) analysis can be used to infer causal relationships between the independent and dependent variables. However, this can lead to illusions or false relationships, so caution is advisable; for example, correlation does not prove causation. The type used in this model is **CLASSIFICATION**(Fig 4.1) and so, more focus will be given on it.

4.4 SYSTEM ARCHITECTURE

The system architectural design(Fig 4.2) is the design process for identifying the subsystems making up the system and framework for subsystem control and communication. The goal of the architectural design is to establish the overall structure of the software system.

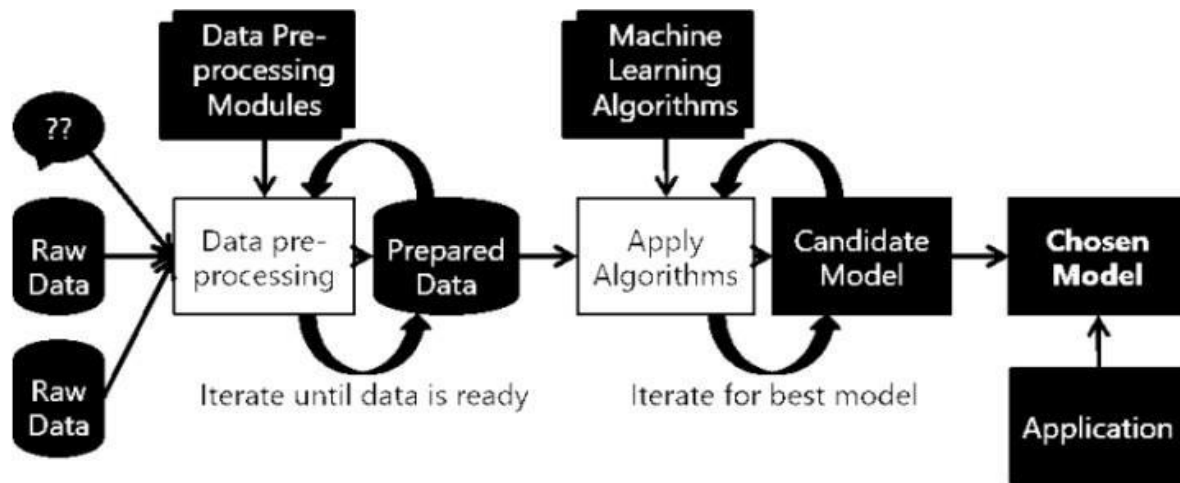


Fig 4.2 System Architecture

4.5 KNN (K-Nearest neighbors) A powerful classification algorithm used in pattern recognition K nearest neighbors(Fig 4.3) stores all available cases and classifies new cases based on a similarity measure (e.g distance function).One of the top data mining algorithms used today. A non-parametric lazy learning algorithm (An Instance based Learning method). KNN: Classification Approach • An object (a new instance) is classified by a majority vote for its neighbor classes. • The object is assigned to the most common class amongst its K nearest neighbors(Fig 4.3). measured by distance function).

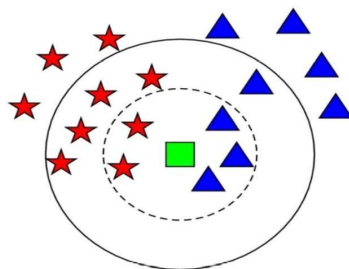


Fig 4.3 K-Nearest neighbors

4.6 DECISION TREE

As the name says all about it, it is a tree which helps us by assisting us in decision making. Used for both classification and regression, it is a very basic and important predictive learning algorithm.

- It is different from others because it works intuitively i.e., taking decisions one-by-one.
- Non-parametric: Fast and efficient.
- It consists of nodes which have parent-child relationship

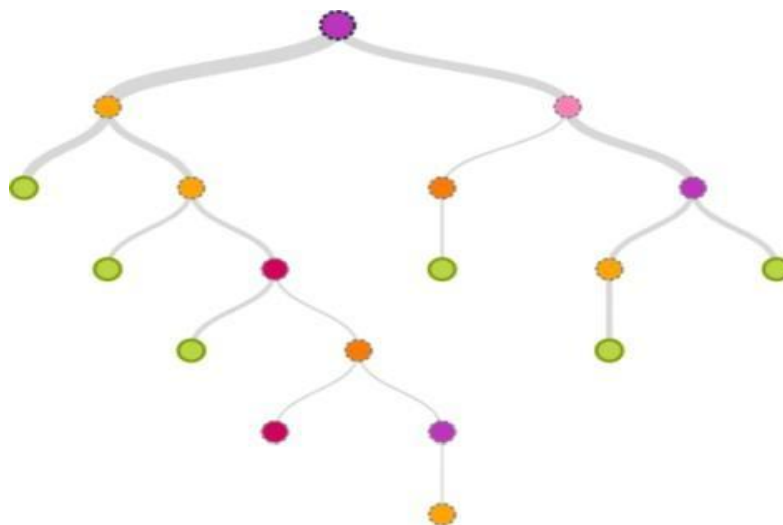


Fig 4.4 Decision Tree

Decision tree(Fig 4.4) considers the most important variable using some fancy criterion and splits the dataset based on it. It is done to reach a stage where we have homogenous subsets that are giving predictions with utmost surety.

4.7 RANDOM FOREST

Random Forests(Fig 4.5) is a very popular ensemble learning method which builds a

number of classifiers on the training data and combines all their outputs to make the best predictions on the test data. Thus, the Random Forests(Fig 4.5) algorithm is a variance minimizing algorithm that uses randomness when making split decisions to help avoid overfitting on the training data. A random forests(Fig 4.5) classifier is an ensemble classifier, which aggregates a family of classifiers $h(x|\theta_1), h(x|\theta_2), \dots, h(x|\theta_k)$. Each member of the family, $h(x|\theta)$, is a classification tree and k is the number of trees chosen from a model random vector. Also, each θ_k is a randomly chosen parameter vector. If $D(x,y)$ denotes the training dataset, each classification tree in the ensemble is built using a different subset $D_{\theta_k}(x,y) \subset D(x,y)$ of the training dataset.

$$y = \operatorname{argmax}_{p \in \{h(x_1) \dots h(x_k)\}} \left\{ \sum_{j=1}^K (I(h(x|\theta_j) = p)) \right\}$$

Fig 4.5 Random Forest

Thus, $h(x|\theta_k)$ is the k th classification tree which uses a subset of features $x_{\theta_k} \subset x$ to build a classification model. Each tree then works like regular decision trees: it partitions the data based on the value of a particular feature (which is selected randomly from the subset), until the data is fully partitioned, or the maximum allowed depth is reached. The final output y is obtained by aggregating the results thus: where I denote the indicator function.

4.8 DATASETS

The dataset used for this experiment is real and authentic. The dataset is acquired from UCI machine learning repository website. The title of the dataset is 'Crime and Communities'. It is prepared using real data from socio-economic data from the 1990 US Census, law

enforcement data from the 1990 US LEMAS survey, and crime data from the 1995 FBI UCR. This dataset contains a total number of 9 attributes and 1994 instances. All data provided in this dataset is numeric and normalized. Crime dataset Dimensions: 9 x 2091 Attributes: 9 Names of attributes: timestamp, act379, act13, act279, act323, act363, act302, latitude, longitude.

4.9 DATA MANIPULATION PACKAGES

4.9.1 *Pandas*

Pandas is a Python package providing fast, flexible, and expressive data structures designed to make working with structured (tabular, multidimensional, potentially heterogeneous) and time series data both easy and intuitive. It aims to be the fundamental high-level building block for doing practical, real world data analysis in Python. Additionally, it has the broader goal of becoming the most powerful and flexible open-source data analysis / manipulation tool available in any language. It is already well on its way toward this goal.

4.9.2 NumPy

NumPy is a library for Python, adding support for large, multidimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays. NumPy is an open-source software and has many contributors. In comparison, MATLAB boasts a large number of additional toolboxes, notably Simulink, whereas NumPy is intrinsically integrated with Python, a more modern and complete programming language. Moreover, complementary Python packages are available; SciPy is a library that adds more MATLAB-like functionality and Matplotlib is a plotting package that provides MATLAB-like plotting functionality.

4.10 MODAL BUILDING PACKAGE

4.10.1 Scikit-learn

Scikit-learn (formerly scikits. learn) is a free software machine learning library for the Python programming language. It features various classification, regression and clustering algorithms including support vector machines(svm), random forest, gradient boosting, k means and DBSCAN, and is designed to interoperate with the Python numerical and scientific libraries NumPy and SciPy. It has been built on top of NumPy, SciPy and matplotlib.

4.10.2 Scikit-plot

Scikit-plot is the result of an unartistic data scientist's dreadful realization that visualization is one of the most crucial components in the data science process, not just a mere afterthought. Gaining insights is simply a lot easier when you're looking at a colored heatmap of a confusion matrix complete with class labels rather than a single-line dump of numbers enclosed in brackets. Besides, if you ever need to present your results to someone 15 (virtually any time anybody hires you to do data science), you show them visualizations, not a bunch of numbers in Excel. All in all, it is an intuitive library used to add plotting functionality to a scikit-learn object.

4.10.3 Matplotlib's pyplot

Matplotlib is a Python 2D plotting library which produces publication quality figures in a variety of hardcopy formats and interactive environments across platforms. Matplotlib can be used in Python scripts, the Python and IPython shells, the Jupyter notebook, web application servers, and four graphical user interface toolkits.

It provides an object-oriented API for embedding plots into applications using general-purpose GUI toolkits like Tkinter, wxPython, Qt or GTK.matplotlib. Pyplot provides a MATLAB-like plotting framework pylab combines pyplot with NumPy into a single namespace. This is convenient for interactive work, but for programming it is recommended that the namespaces be kept separate.

4.11 MODULE DESCRIPTIONS

4.11.1 Data collection Module

Crime dataset from kaggle is used in CSV format.

4.11.2 Data Preprocessing Module

10k entries are present in the dataset. The null values are removed using `df = df.dropna()` where `df` is the data frame. The categorical attributes (Location, Block, Crime Type, Community Area) are converted into numeric using Label Encoder. The date attribute is splitted into new attributes like month and hour which can be used as features for the model.

4.11.3 Feature selection Module

Features selection is done which can be used to build the model. The attributes used for feature selection are Block, Location, District, Community area, X coordinate , Y coordinate, Latitude , Longitude, Hour and month.

4.11.4 Building and Training Model

After feature selection location and month attribute are used for training. The dataset is divided into pairs of `xtrain` ,`train` and `xtest`, `y test`. The algorithm model is imported from `sklearn`. Building models is done using `models`. `Fit (xtrain, ytrain)`.

4.11.5 Prediction Module

After the model is built using the above process, prediction is done using `model.predict(xtest)`. The accuracy is calculated using `accuracy_score` imported from `metrics` - `metrics.accuracy_score (ytest, predicted)`.

4.11.6 Visualization Module

Using `matplotlib` library from `sklearn`. Analysis of the crime dataset is done by plotting various graphs.

4.12 Python

Python is an interpreted high-level general-purpose programming language. Its design philosophy emphasizes code readability with its use of significant indentation. Its language constructs as well as its object-oriented approach aim to help programmers write clear, logical code for small and large-scale projects. Python is dynamically-typed and garbage-collected. It supports multiple programming paradigms, including structured (particularly, procedural), object-oriented and functional programming. It is often described as a "batteries included" language due to its comprehensive standard library. Guido van Rossum began working on Python in the late 1980s, as a successor to the ABC programming language, and first released it in 1991 as Python 0.9.0. Python 2.0 was released in 2000 and introduced new features, such as list comprehensions and a garbage collection system using reference counting. Python 3.0 was released in 2008 and was a major revision of the language that is not completely backward-compatible. Python 2 was discontinued with version 2.7.18 in 2020. Python consistently ranks as one of the most popular programming languages.

4.12.1. GETTING PYTHON

The most up-to-date and current source code, binaries, documentation, news, etc., is available on the official website of Python <https://www.python.org>.

Windows Installation

Here are the steps to install Python on a Windows machine.

- Open a Web browser and go to <https://www.python.org/downloads/>.
- Follow the link for the Windows installer python-XYZ.msi file where XYZ is the version you need to install.
- To use this installer python-XYZ.msi, the Windows system must support Microsoft Installer 2.0. Save the installer file to your local machine and then run it to find out if your machine supports MSI.
- Run the downloaded file. This brings up the Python install wizard, which is really easy to use. Just accept the default settings, wait until the install is finished, and you are done.

The Python language has many similarities to Perl, C, and Java. However, there are some definite differences between the languages.

4.12.2. Flask Framework

Flask is a web application framework written in Python. Armin Ronacher, who leads an international group of Python enthusiasts named Pocco, develops it. Flask is based on the Werkzeug WSGI toolkit and Jinja2 template engine. Both are Pocco project Http protocol is the foundation of data communication on the world wide web. Different methods of data retrieval from specified URLs are defined in this protocol.

The following table summarizes different http methods –

Sr. No	Methods & Description
-----------	-----------------------

1	GET Sends data in unencrypted form to the server. Most common method.
2	HEAD Same as GET, but without response body
3	POST Used to send HTML form data to server. Data received by POST method is not cached by server.
4	PUT Replaces all current representations of the target resource with the uploaded content.
5	DELETE Removes all current representations of the target resource given by a URL

By default, the Flask route responds to the GET requests. However, this preference can be altered by providing methods argument to route() decorator.

In order to demonstrate the use of POST method in URL routing, first let us create an HTML form and use the POST method to send form data to a URL.

Save the following script as login.html

```
<html>
```

```
<body>
```

```
<form action="http://localhost:5000/login" method="post">
```

<p>Enter Name:</p>

<p><input type="text" name="nm"/></p>

<p><input type="submit" value="submit"/></p>

</form>

</body>

</html>

Now enter the following script in Python shell.

```
from flask import Flask, redirect, url_for,
```

```
request app=Flask(__name__)
```

```
@app.route('/success/<name>')
```

```
def success(name):
```

```
    return 'welcome %s'% name
```

```
@app.route('/login', methods=['POST', 'GET'])
```

```
def login():
```

```
    if request.method == 'POST':
```

```
        user=request.form['nm']
```

```
        return redirect(url_for('success', name= user))
```

```
    else:
```

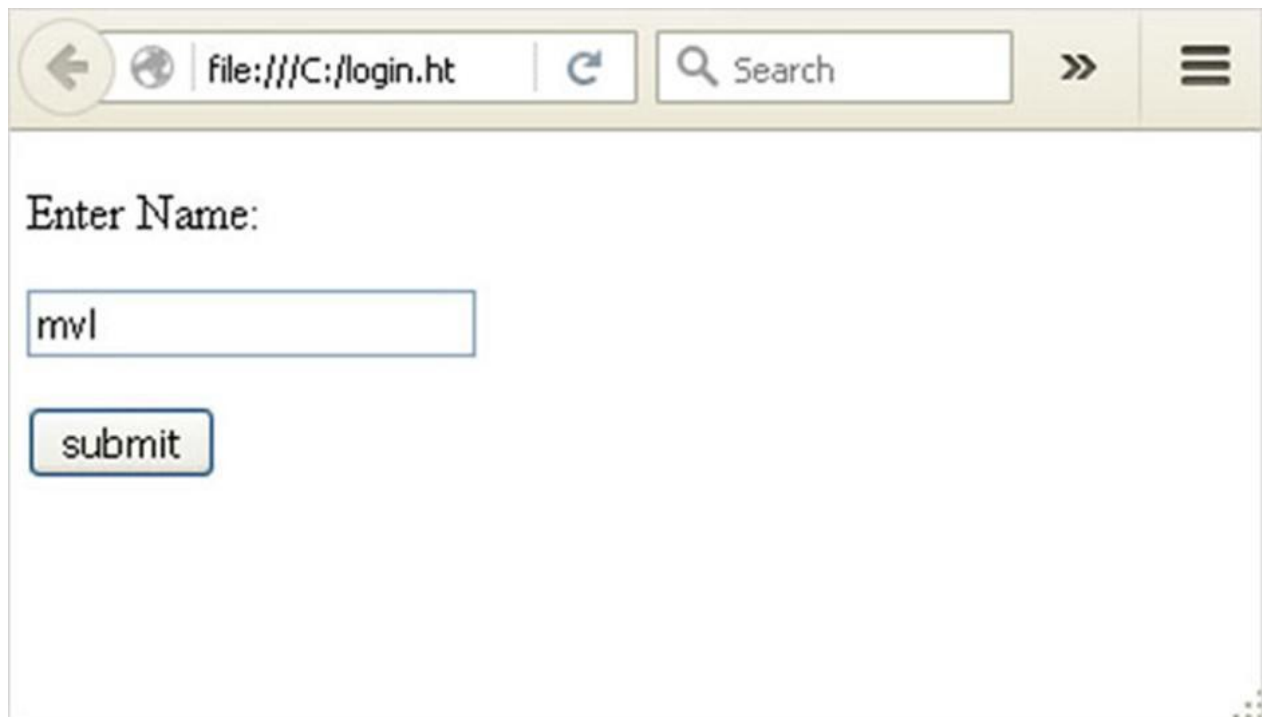
```
        user=request.args.get('nm')
```

```
        return redirect(url_for('success', name=
```

```
user)) if __name__ == '__main__':
```

```
    app.run(debug = True)
```

After the development server starts running, open login.html in the browser, enter name in the text field and click Submit.

A screenshot of a web browser window. The address bar shows the file path 'file:///C:/login.ht'. The page content includes the text 'Enter Name:', a text input field containing 'mvl', and a 'submit' button.

Form data is POSTed to the URL in action clause of form tag.

`http://localhost/login` is mapped to the `login()` function. Since the server has received data by POST method, value of 'nm' parameter obtained from the form data is obtained by –

```
user = request.form['nm']
```

It is passed to '/success' URL as variable part. The browser displays a welcome message in the window.

Change the method parameter to 'GET' in login.html and open it again in the browser. The data received on the server is by the GET method. The value of 'nm' parameter is now obtained by –

```
User = request.args.get('nm')
```

Here, args is a dictionary object containing a list of pairs of form parameters and its corresponding value. The value corresponding to 'nm' parameter is passed on to '/success' URL as before.

4.13 DATA MINING

4.13.1 An Exploration of Crime Prediction Using Data Mining on Open Data

The increase in crime data recording coupled with data analytics resulted in the growth of research approaches aimed at extracting knowledge from crime records to better understand criminal behavior and ultimately prevent future crimes. While many of these approaches make use of clustering and association rule mining techniques, there are fewer approaches focusing on predictive models of crime. In this paper, we explore models for predicting the frequency of several types of crimes by LSOA code (Lower Layer Super Output Areas — an administrative system of areas used by the UK police) and the frequency of antisocial behavior crimes. Three algorithms are used from different categories of approaches: instance-based learning, regression and decision trees. The data are from the UK police and contain over 600,000 records before preprocessing. The results, looking at predictive performance as well as processing time, indicate that decision trees (M5P algorithm) can be used to reliably predict crime frequency in general as well as anti-social behavior frequency.

4.13.2 Crime Analysis and Prediction Using Data Mining

Crime analysis and prevention is a systematic approach for identifying and analyzing patterns and trends in crime. Our system can predict regions which have high probability for crime occurrence and can visualize crime prone areas. With the increasing advent of computerized systems, crime data analysts can help the Law enforcement officers to speed up the process of solving crimes. Using the concept of data mining we can extract previously unknown, useful information from unstructured data. Here we have an approach between computer science and criminal justice to develop a data mining procedure that can help solve crimes faster. Instead of focusing

on causes of crime occurrence like criminal background of offenders, political enmity etc we are focusing mainly on crime factors of each day.

4.13.3 Crime Detection Techniques Using data Mining and K-Means

Crimes will somehow influence organizations and institutions when they occur frequently in a society. Thus, it seems necessary to study reasons, factors and relations between occurrences of different crimes and finding the most appropriate ways to control and avoid more crimes. The main objective of this paper is to classify clustered crimes based on occurrence frequency during different years. Data mining is used extensively in terms of analysis, investigation and discovery of patterns for occurrence of different crimes. We applied a theoretical model based on data mining techniques such as clustering and classification to real crime dataset recorded by police in England and Wales within 1990 to 2011. We assigned weights to the features in order to improve the quality of the model and remove low value of them. The Genetic Algorithm (GA) is used for optimizing Outlier Detection operator parameters using the RapidMiner tool.

4.13.4 Survey on crime analysis and prediction using data mining techniques

Data Mining is the procedure which includes evaluating and examining large pre-existing databases in order to generate new information which may be essential to the organization. The extraction of new information is predicted using the existing datasets. Many approaches for analysis and prediction in data mining had been performed. But, few efforts have been made in the criminology field. Many few have taken efforts for compare the information all these approaches produce. The police stations and other similar criminal justice agencies hold many large databases of information which can be used to predict or analyze the criminal movements and criminal activity involvement in the society. The criminals can also be predicted based on the crime data. The main aim of this work is to perform a survey on the supervised learning and unsupervised learning techniques that has been applied towards criminal

identification. This paper presents the survey on Crime analysis and crime prediction using several Data Mining techniques.

4.13.5 Crime Pattern Analysis, Visualizations And Prediction Using Data Mining

Crime against women these days has become problem of every nation around the globe. Many countries are trying to curb this problem. Prevents are taken to reduce the increasing number of cases of crime against women. A huge amount of data is generated every year on the basis of reporting of crime. This data can prove very useful in analyzing and predicting crime and help us prevent the crime to some extent. Crime analysis is an area of vital importance in the police department. Study of crime data can help us analyze crime patterns, inter-related clues and important hidden relations between the crimes. That is why data mining can be a great aid to analyze, visualize and predict crime using a crime data set. Classification and correlation of data set makes it easy to understand similarities & dissimilarities amongst the data objects. We group data objects using clustering techniques. Dataset is classified on the basis of some predefined condition. Here grouping is done according to various types of crimes against women taking place in different states and cities of India. Crime mapping will help the administration to plan strategies for prevention of crime, further using data mining techniques data can be predicted and visualized in various form in order to provide better understanding of crime patterns.

4.13.5 Crime Analysis And prediction using Data Mining Techniques

Crime analysis and prevention is a systematic approach for identifying and analyzing patterns and trends in crime. Our system can predict regions which have high probability for crime occurrence and can visualize crime prone areas. With the increasing advent of computerized systems, crime data analysts can help Law enforcement officers to speed up the process of solving crimes. Using the concept of data mining we can extract previously unknown, useful information from unstructured data. Here we have an approach between computer science and criminal justice to

develop a data mining procedure that can help solve crimes faster. Instead of focusing on causes of crime occurrence like the criminal background of the offender, political enmity, etc we are focusing mainly on crime factors of each day.

4.13.6 Systematic Review of Crime Data Mining

Crime analysis is a methodical approach for identifying and analyzing patterns and trends in crime. With the increasing origin of computerized systems, crime data analysts can help the Law enforcement officers to speed up the process of solving crimes. Using the concept of data mining, we can analyze previously unknown, useful information from unstructured data. Predictive policing means, using analytical and predictive techniques, to identify criminals and it has been found to be pretty much effective in doing the same. Because of the increased crime rate over the years, we will have to handle a huge amount of crime data stored in warehouses which would be very difficult to be analyzed manually, and also now a day's, criminals are becoming technologically advance, so there is need to use advanced technologies in order to keep police ahead of them. In this paper, the main focus is on the review of algorithms and techniques used for identifying the criminals.

4.13.7 Survey paper on Crime Prediction using Ensemble Approach

Crime is a foremost problem where the top priority has been concerned by individual, the community and government. This paper investigates a number of data mining algorithms and ensemble learning which are applied on crime data mining. This survey paper describes a summary of the methods and techniques which are implemented in crime data analysis and prediction. Crime forecasting is a way of trying to mining out and decreasing the upcoming crimes by forecasting the future crime that will occur. Crime prediction practices historical data and after examining data, predict the upcoming crime with respect to location, time, day, season and year. In present crime cases rapidly increase so it is an inspiring task to foresee upcoming crimes closely with better accuracy. Data mining methods are too important to resolving crime problems

with investigating hidden crime patterns. so the objective of this study could be analyzing and discussing various methods which are applied on crime prediction and analysis. This paper delivers reasonable investigation of Data mining Techniques and ensemble classification techniques for discovery and prediction of upcoming crime.

4.14 ACTIVITY DIAGRAM

Activity diagrams(Fig.4.6) are graphical representations of workflows of stepwise activities and actions with support for choice, iteration and concurrency. In the Unified Modeling Language, activity diagrams can be used to describe the business and operational step-by-step workflows of components in a system. An activity diagram shows the overall flow of control.

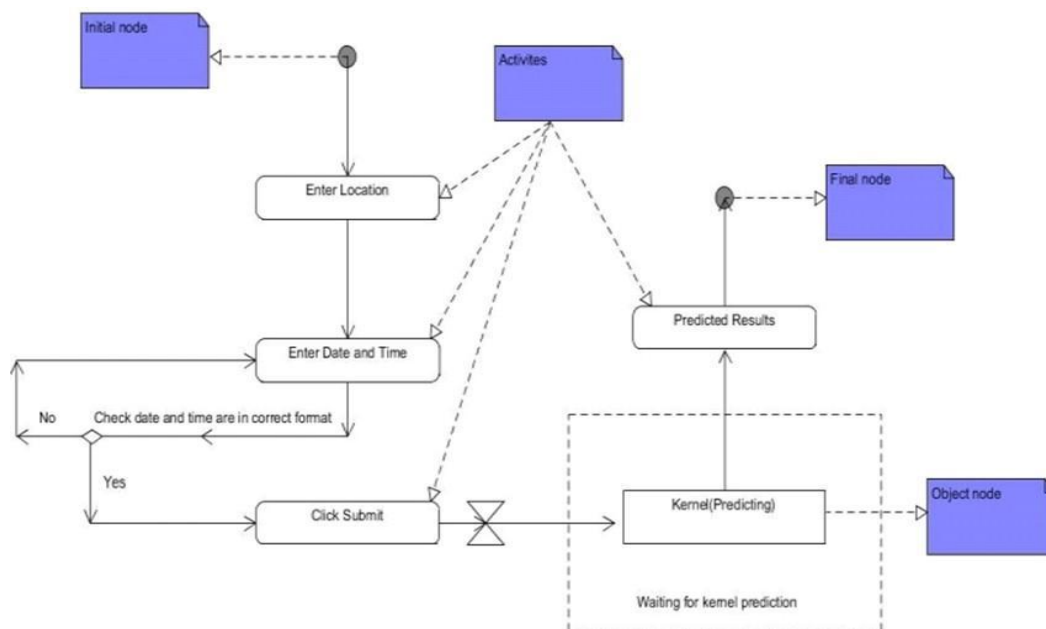


Fig. 4.6. Activity Diagram

CHAPTER 5

RESULTS AND DISCUSSION

This section presents all of the results from the implementations of the KNN ,DECISION TREE and RANDOM FOREST algorithms. The algorithms were run to predict each of the following features in the datasets: murders, murdPerPop, rapes, rapesPerPop, robberies, robbbPerPop, assaults, assaultPerPop, and ViolentCrimesPerPop. Note that perPop refers to for every 100K of people. The algorithm that gives the lowest error values for each feature and the highest correlation coefficient is highlighted in the results present. The application displays the results of the crime type prediction based on the input crime summary. The middle part of the application displays the results of the crime type prediction based on the input crime summary. The prediction results show the probability values for 21 crime types as a bar graph. the types of crimes with the highest probability values are displayed. Based on the predicted results, field personnel can quickly identify the type of crime. The predicted CRS is displayed at the bottom of the application. This prediction result is also displayed in real-time. The platform can predict crime type and CRS and display the prediction results in real-time; therefore, field staff, such as police officers, can easily check predictive information about crimes received through the platform.

CHAPTER 6

CONCLUSION AND FUTURE WORK

6.1 CONCLUSION

This is focused on building predictive models for crime frequencies per crime type per month. The crime rates in India are increasing day by day due to many factors such as increase in poverty, implementation, corruption, etc. The proposed model is very useful for both the investigating agencies and the police officials in taking necessary steps to reduce crime. The project helps the crime analysis to analyze these crime networks by means of various interactive visualizations. Future enhancement of this research work on training bots to predict the crime prone areas by using machine learning techniques. Since, machine learning is similar to data mining, advanced concepts of machine learning can be used for better prediction. The data privacy, reliability, accuracy can be improved for enhanced prediction.

6.2 FUTURE WORK

Even though the scope of this project was to prove how effective and accurate machine learning algorithms can be at predicting violent crimes, there are other applications of data mining in the realm of law enforcement such as determining criminal "hot spots", creating criminal profiles, and learning crime trends. Utilizing these applications of data mining can be a long and tedious process for law enforcement officials who have to sift through large volumes of data. However, the precision in which one could infer and create new knowledge on how to slow down crime is well worth the safety and security of people.

REFERENCES

- [1] Ginger Saltos and Mihaela Coacea, An Exploration of Crime Prediction Using Data Mining on Open Data, International journal of Information technology & Decision Making, 2017.
- [2] Shiju Sathyadevan, Devan M.S, Surya Gangadharan.S, Crime Analysis and Prediction Using Data Mining, First International Conference on networks & soft computing (IEEE) 2014.
- [3] Khushabu A.Bokde, Tisksha P.Kakade, Dnyaneshwari S. Tumasare, Chetan G.Wadhai B.E Student, Crime Detection Techniques Using Data Mining and K-Means, International Journal of Engineering Research & technology (IJERT) ,2018.
- [4] H.Benjamin Fredrick David and A.Suruliandi, Survey on crime analysis and prediction using data mining techniques, ICTACT Journal on Soft computing, 2017.
- [5] Tushar Sonawanev, Shirin Shaikh, Rahul Shinde, Asif Sayyad, Crime Pattern Analysis, Visualization And Prediction Using Data Mining, Indian Journal of Computer Science and Engineering (IJCSE), 2015.
- [6] RajKumar.S, Sakkarai Pandi.M, Crime Analysis and prediction using data mining techniques, International Journal of recent trends in engineering & research, 2019.
- [7] Sarpreet kaur, Dr. Williamjeet Singh, Systematic review of crime data mining, International Journal of Advanced Research in computer science , 2015.
- [8] Ayisheshim Almaw, Kalyani Kadam, Survey Paper on Crime Prediction using Ensemble Approach, International journal of Pure and Applied Mathematics, 2018.

- [9] Dr .M.Sreedevi, A.Harsha Vardhan Reddy, ch.Venkata Sai Krishna Reddy, Review on crime Analysis and prediction Using Data Mining Techniques, International Journal of Innovative Research in Science Engineering and technology ,2018.
- [10] K.S.N .Murthy, A.V.S.Pavan kumar, Gangu Dharmaraju, international journal of engineering, Science and mathematics, 2017.
- [11] Deepika k.K, Smitha Vinod, Crime analysis in India using data mining techniques , International journal of Engineering and technology, 2018.
- [12] Hitesh Kumar Reddy ToppyiReddy, Bhavana Saini, Ginika Mahajan, Crime Prediction And Monitoring Framework Based on Spatial Analysis, International Conference on Computational Intelligence Data Science (ICCIDS 2018).

APPENDICES

A. SOURCE

CODE Login.html

```
<!DOCTYPE html>
```

```
<html lang="en">
```

```
<head>
```

```
<meta charset="utf-8">
```

```
<meta content="width=device-width, initial-scale=1.0" name="viewport">
```

```
<title>login</title>
```

```
<meta content="" name="description">
```

```
<meta content="" name="keywords">
```

```
<!-- Favicons -->
```

```
<link href="../static/img/favicon.png" rel="icon">
```

```
<link href="../static/img/apple-touch-icon.png" rel="apple-touch-icon">
```

```
<!-- Google Fonts -->
```

<!-- Vendor CSS Files -->

<link href="../../static/vendor/bootstraps/css/bootstrap.min.css" rel="stylesheet">

<link href="../../static/vendor/icofont/icofont.min.css" rel="stylesheet">

<link href="../../static/vendor/animate.css/animate.min.css" rel="stylesheet">

<link href="../../static/vendor/font-awesome/css/font-awesome.min.css" rel="stylesheet">

<link href="../../static/vendor/nivo-slider/css/nivo-slider.css" rel="stylesheet">

<link href="../../static/vendor/owl.carousel/assets/owl.carousel.min.css" rel="stylesheet">

<link href="../../static/vendor/venobox/venobox.css" rel="stylesheet">

<!-- Template Main CSS File -->

<link href="../../static/css/styles.css" rel="stylesheet">

<!-- =====

* Template Name: eBusiness - v2.2.1

* Template URL: <https://bootstrapmade.com/ebusiness-bootstrap-corporate-template/>

* Author: BootstrapMade.com

* License: <https://bootstrapmade.com/license/>

===== -->

</head>

```
<body data-spy="scroll" data-target="#navbar-example">

<!-- ===== Header ===== -->

<header id="header" class="fixed-top">

  <div class="container d-flex">

    <div class="logo mr-auto">

      <h1 class="text-light"><a href="{{url_for('first')}}"><span></span>CRIME
TYPE</a></h1>

      <!-- Uncomment below if you prefer to use an image logo -->

      <!-- <a href="index.html"></a>-->

    </div>

    <nav class="nav-menu d-none d-lg-block">

      <ul>

        <li class="active"><a href="{{url_for('first')}}">Home</a></li>

        <li><a href="{{url_for('login')}}">Login</a></li>

      </ul>

    </nav>

  </div>

</header>

</body>
```


</nav><!-- .nav-menu -->

</div>

</header><!-- End Header -->

<!-- ===== Slider Section ===== -->

<div class="header-bg page-area">

<div class="home-overly"></div>

<div class="container">

<div class="row">

<div class="col-md-12 col-sm-12 col-xs-12">

<div class="slider-content text-center">

<div class="header-bottom">

<div class="layer2 wow zoomIn" data-wow-duration="1s"
data-wow-delay=".4s">

<h1 class="title2"></h1>

</div>

<title>login</title>

<meta content="" name="description">

<meta content="" name="keywords">

<!-- Favicons -->

<link href="../../static/img/favicon.png" rel="icon">

<link href="../../static/img/apple-touch-icon.png" rel="apple-touch-icon">

<!-- Google Fonts -->

<!-- Vendor CSS Files -->

<link href="../../static/vendor/bootstrap/css/bootstrap.min.css" rel="stylesheet">

<link href="../../static/vendor/icomfont/icomfont.min.css" rel="stylesheet">

<link href="../../static/vendor/animate.css/animate.min.css" rel="stylesheet">

<link href="../../static/vendor/font-awesome/css/font-awesome.min.css" rel="stylesheet">

<link href="../../static/vendor/nivo-slider/css/nivo-slider.css" rel="stylesheet">

<link href="../../static/vendor/owl.carousel/assets/owl.carousel.min.css" rel="stylesheet">

<link href="../../static/vendor/venobox/venobox.css" rel="stylesheet">

B. SCREENSHOTS

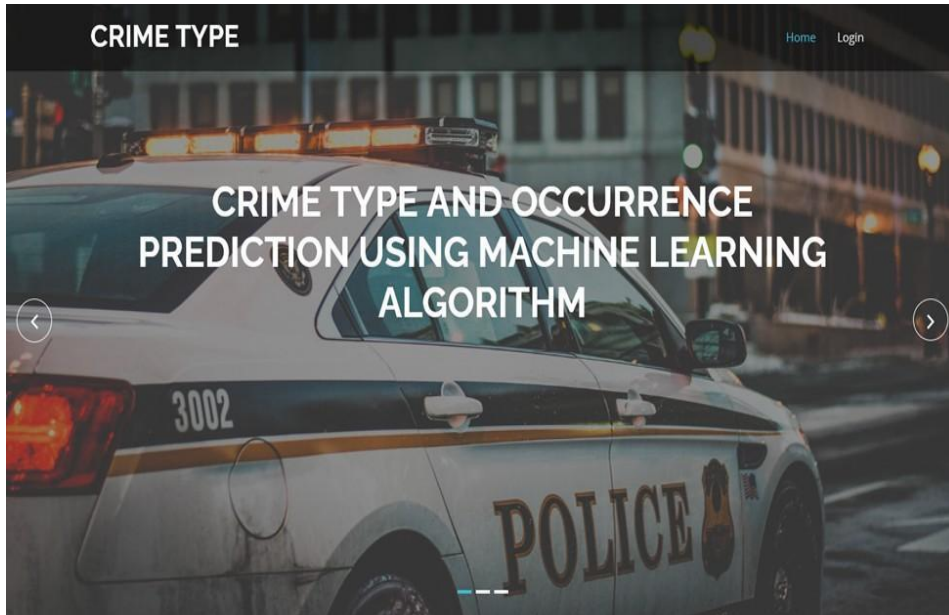


Fig 1 Home page

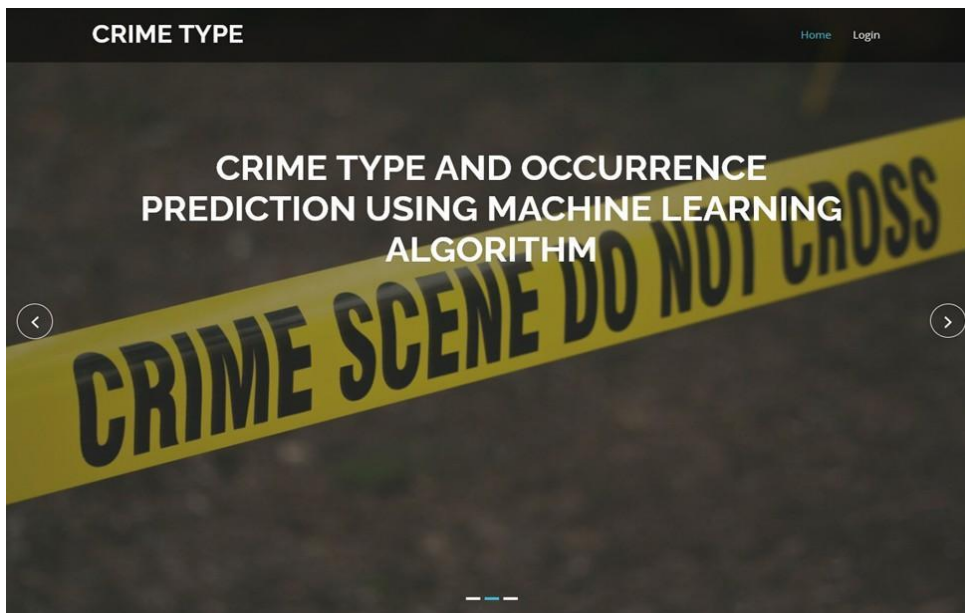


Fig 2 Upload page

CRIME TYPE

Home Login

CRIME TYPE AND OCCURRENCE PREDICTION USING MACHINE LEARNING ALGORITHM

Login

Username

Password

Login

Fig 3 Login Page

CRIME TYPE

Home Login Uploads

Preview

	Date	Primary Type	Location Description	Arrest	District	Year	Updated On	Latitude	Longitude	Location
Id										
1	05-03-2016 22:00	THEFT	RESIDENCE	False	15	2016	05-10-2016 15:56	41.886297	-87.761751	(41.886297242, -87.761750709)
2	05-03-2016 17:30	THEFT	OTHER	False	12	2016	05-10-2016 15:56	41.877812	-87.655758	(41.877811861, -87.655758012)
3	05-03-2016 09:00	THEFT	STREET	False	1	2016	05-10-2016 15:56	41.843017	-87.617227	(41.843016959, -87.617227059)

Fig 4 Preview Page

CRIME TYPE

[Home](#)[Login](#)[Uploads](#)[prediction](#)[performance_analysis](#)



CRIME TYPE AND OCCURRENCE PREDICTION USING MACHINE LEARNING ALGORITHM

CRIME TYPE PREDICTION

Year:

Month:

Day:

dayOfWeek:

Monday

Minute:

Second:

Latitude:

Longitude:

Predict

Prediction is :

Fig 5 Form Filling

CRIME TYPE

[Home](#)[Login](#)[Uploads](#)[prediction](#)[performance_analysis](#)



CRIME TYPE AND OCCURRENCE PREDICTION USING
MACHINE LEARNING ALGORITHM

CRIME TYPE PREDICTION

Year:

Month:

Day:

dayOfWeek:

Minute:

Second:

Latitude:

Longitude:

Predict

Prediction is : THEFT

Fig 6 Prediction Page



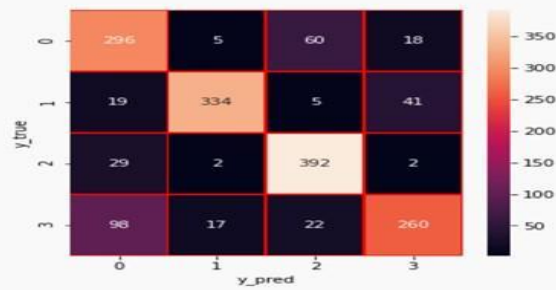
PERFORMANCE ANALYSIS

Precision and recall

Recall Precision

ROBBERY	0.67	0.78
SEX OFFENSE	0.93	0.84
THEFT	0.82	0.92
WEAPONS VIOLATION	0.81	0.65

Confusion Matrix



Percentage of crime occurrence with particular DayOfWeek

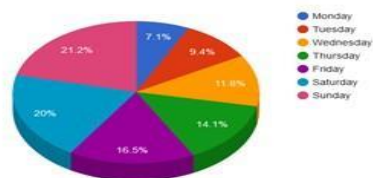


Fig 7 Analysis page

C. PUBLICATION WITH PLAGIARISM REPORT

Report <i>by</i> Tumu Renusri
Submission date: 07-Mar-2022 10:05AM (UTC-0600) Submission ID: 1778618827 File name: Untitled_document_1.docx (1.79M)

Submission date: 07-Mar-2022 10:05AM (UTC-0600) Submission ID: 1778618827 File name: Untitled_document_1.docx (1.79M) Word count: 2256 Character count: 13067	
<table border="1"><tr><td> SPOTTING AND SCORING ALL THE CRIME TYPES </td></tr></table>	SPOTTING AND SCORING ALL THE CRIME TYPES
SPOTTING AND SCORING ALL THE CRIME TYPES	

Report

ORIGINALITY REPORT

8%

SIMILARITY INDEX

6%

INTERNET SOURCES

6%

PUBLICATIONS

1%

STUDENT PAPERS

PRIMARY SOURCES

1

innovate.mygov.in

Internet Source

3%

2

Satti Rami Reddy, Satti Mouli Satti Reddy, Suja Cherukullapurath Mana, B.Keerthi Samhitha, Jithina Jose. "Implementation of a Privacy Preserving and Loss Data Retrieval System for Cloud Environment", 2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184), 2020

Publication

1%