A Thesis/Project/Dissertation Report

on

RAKSITR: Crime Rate Prediction and Analysis

Submitted in partial fulfillment ofthe

requirement for the award ofthe

degree of

B.TECH
(COMPUTER SCIENCE ENGINEERING)

Under The
Supervision :
MS. KIRAN
SINGH

Submitted By
Subham Singh
18SCSE1010108
Sudhir Kumar Chaurasiya
18SCSE1010828

SCHOOL OF COMPUTING SCIENCE AND ENGINEERING
DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING /
DEPARTMENT OF COMPUTERAPPLICATION
GALGOTIAS UNIVERSITY, GREATER NOIDA
INDIA

SCHOOL OF COMPUTING SCIENCE AND ENGINEERING
GALGOTIAS UNIVERSITY, GREATER NOIDA

CANDIDATE'S DECLARATION

We hereby certify that the work which is being presented in the thesis/project/dissertation, entitled **"RAKSITR: CRIME PREDICTION AND ANALYSIS"** in partial fulfillment of the requirements for the award of the **B-TECH** submitted in the School of Computing Science and Engineering of Galgotias University, Greater Noida, is an original work carried out during the period of **08/2021 to 12/2021,** under the supervision of **MS.KIRAN SINGH**, Department of Computer Science and Engineering/Computer Application and Information and Science, of School of Computing Science and Engineering , Galgotias University, Greater Noida

The matter presented in the thesis/project/dissertation has not been submitted by me/us for the award of any other degree of this or any other places.

Subham Singh (18SCSE1010108)

Sudhir Kumar Chaurasiya (18SCSE1010828)

This is to certify that the above statement made by the candidates is correct to the best of my knowledge.

Supervisor

GALGOTIAS UNIVERSITY

CERTIFICATE

Certified that the project report entitled, "**RAKSITR: Crime Rate Prediction and Analysis** " is a bonafide work done under my guidance **MS. KIRAN SINGH** in partial fulfillment of the requirements for the award of degree of Bachelor of Engineering in Computer Science and Engineering.

Signature of Examiner(s) Signature of Supervisor(s)

Signature of Dean

Date :

Place :

CGALGOTIAS UNIVERSITY

ACKNOWLEDGEMENT

We have immense pleasure in expressing our sincerest and deepest sense of gratitude towards our guide Ms. Kiran Singh for the assistance, valuable guidance and co- operation in carrying out this Project successfully. We have developed this project with the help of Faculty members of our institute and we are extremely grateful to all of them. We also thank to our Dean of Galgotias university of CSE department, for providing the required facilities in completing this project. We are greatly thankful to our parents, friends and faculty members for their motivation, guidance and help whenever needed.

Name and signature of team Members:

1. Sudhir Kumar Chaurasiya
2. Subham Singh

Abstract

To be better prepared to respond to criminal activity, it is important to understand patterns in crime. In our project, we analyze crime data from the city of India, scraped from publicly available website of Indian Police.

At the outset, the task is to predict which category of crime is most likely to occur given a time and place in the city.
The use of AI and PHP to detect from our website which , is proven to work, and expected to continue to expand.

The use of AI in predicting crimes or an individual's likelihood for committing a crime has promise but is still more of an unknown. The biggest challenge will probably be "proving" to politicians that it works. When a system is designed to stop something from happening, it is difficult to prove the negative. Companies that are directly involved in providing governments with AI tools to monitor areas or predict crime will likely benefit from a positive feedback loop. Improvements in crime prevention technology will likely spur increased total spending on this technology.

We also attempt to make our classification task more meaningful by merging multiple classes into larger classes. Finally, we report and reflect on our results with different classifiers, and dwell on avenues for future work.

List of Tables

TABLE OF CONTENTS

Chapter-1

Introduction

Raksitr which is an Sanskrit word meaning protector, many important questions in public safety and protection relate to crime, and a better understanding of crime is beneficial in multiple ways: it can lead to targeted and sensitive practices by law enforcement authorities to mitigate crime, and more concerted efforts by citizens and authorities to create healthy neighborhood environments.

With the advent of the Big Data era and the availability of fast, efficient algorithms for data analysis, understanding patterns in crime from data is an active and growing field of research.

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

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

The output is the class of crime that is likely to have occurred.

We also perform multiple classification tasks – we first try to predict which of 6 classes of crimes are likely to have occurred, and later try to differentiate between violent and non-violent crimes.

1.1 Rationale

As of 2019, a total of 51.5 lakh cognizable crimes comprising 32.2 lakh Indian Penal Code (IPC) crimes and 19.4 lakh Special and Local Laws (SLL) crimes were registered nationwide. Showing a 1.6% annual increase in the registration of cases (50.7 lakh cases), the crime rate per 100,000 population has increased from 383.5 in 2018 to 385.5 in 2019. More than a fifth of all registered crime (10.5 lakh) were classified as offences affecting the human body, which included violent acts such as murder, kidnapping, assault and death by negligence.

With the rapid urbanization and development of big cities and towns, the graph of crimes is also on the increase. This phenomenal rise in offences and crime in cities is a matter of great concern and alarm to all of us.

There are robberies, murders, rapes and what not. The frequent and repeated thefts, burglaries, robberies, murders, killings, rapes, shoplifting, pick pocketing, drug- abuse, illegal trafficking, smuggling, theft of vehicles etc., have made the common citizens to have sleepless nights and restless days.

They feel very insecure and vulnerable in the presence of anti-social and evil elements. The criminals have been operating in an organized way and sometimes even have nationwide and international connections and links.

1.2 Goal

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

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

1.3 Objective

The objective of our work is to:

- Predicting crime before it takes place.
- Predicting hotspots of crime.
- Understanding crime pattern.

- Classify crime based on location.
- Analysis of crime in Indore.

1.4 Methodology

1.4.1 PHP

PHP is a general-purpose scripting language geared towards web development. It was originally created by Danish-Canadian programmer Rasmus Lerdorf in 1994. The PHP reference implementation is now produced by The PHP Group. PHP originally stood for Personal Home Page,[8] but it now stands for the recursive initialism PHP: Hypertext Preprocessor.

PHP code is usually processed on a web server by a PHP interpreter implemented as a module, a daemon or as a Common Gateway Interface (CGI) executable. On a web server, the result of the interpreted and executed PHP code – which may be any type of data, such as generated HTML or binary image data – would form the whole or part of an HTTP response. Various web template systems, web content management systems, and web frameworks exist which can be employed to orchestrate or facilitate the generation of that response. Additionally, PHP can be used for many programming tasks outside the web context, such as standalone graphical applications and robotic drone control. PHP code can also be directly executed from the command line.

The standard PHP interpreter, powered by the Zend Engine, is free software released under the PHP License. PHP has been widely ported and can be deployed on most web servers on a variety of operating systems and platforms.

The PHP language evolved without a written formal specification or standard until 2014, with the original implementation acting as the de facto standard which other implementations aimed to follow. Since 2014, work has gone on to create a formal PHP specification.

W3Techs reports that, as of April 2021, "PHP is used by 79.2% of all the websites whose server-side programming language we know. PHP version 7.4 is the most used versions, then 7.3 which lost support on 6 December 2021.

PHP Version 7.1.33-17+ubuntu20.04.1+deb.sury.org+1

System	Linux arg 5.4.0-48-generic #52-Ubuntu SMP Thu Sep 10 10:58:49 UTC 2020 x86_64
Build Date	Aug 7 2020 14:47:49
Server API	Built-in HTTP server
Virtual Directory Support	disabled
Configuration File (php.ini) Path	/etc/php/7.1/cli
Loaded Configuration File	/etc/php/7.1/cli/php.ini
Scan this dir for additional .ini files	/etc/php/7.1/cli/conf.d
Additional .ini files parsed	/etc/php/7.1/cli/conf.d/10-mysqlnd.ini, /etc/php/7.1/cli/conf.d/10-opcache.ini, /etc/php/7.1/cli/conf.d/10-pdo.ini, /etc/php/7.1/cli/conf.d/15-xml.ini, /etc/php/7.1/cli/conf.d/20-calendar.ini, /etc/php/7.1/cli/conf.d/20-ctype.ini, /etc/php/7.1/cli/conf.d/20-curl.ini, /etc/php/7.1/cli/conf.d/20-dom.ini, /etc/php/7.1/cli/conf.d/20-exif.ini, /etc/php/7.1/cli/conf.d/20-fileinfo.ini, /etc/php/7.1/cli/conf.d/20-ftp.ini, /etc/php/7.1/cli/conf.d/20-gd.ini, /etc/php/7.1/cli/conf.d/20-gettext.ini, /etc/php/7.1/cli/conf.d/20-iconv.ini, /etc/php/7.1/cli/conf.d/20-json.ini, /etc/php/7.1/cli/conf.d/20-mysqli.ini, /etc/php/7.1/cli/conf.d/20-pdo_mysql.ini, /etc/php/7.1/cli/conf.d/20-phar.ini, /etc/php/7.1/cli/conf.d/20-posix.ini, /etc/php/7.1/cli/conf.d/20-readline.ini, /etc/php/7.1/cli/conf.d/20-shmop.ini, /etc/php/7.1/cli/conf.d/20-simplexml.ini, /etc/php/7.1/cli/conf.d/20-sockets.ini, /etc/php/7.1/cli/conf.d/20-sysvmsg.ini, /etc/php/7.1/cli/conf.d/20-sysvsem.ini, /etc/php/7.1/cli/conf.d/20-sysvshm.ini, /etc/php/7.1/cli/conf.d/20-tokenizer.ini, /etc/php/7.1/cli/conf.d/20-wddx.ini, /etc/php/7.1/cli/conf.d/20-xmlreader.ini, /etc/php/7.1/cli/conf.d/20-xmlwriter.ini, /etc/php/7.1/cli/conf.d/20-xsl.ini, /etc/php/7.1/cli/conf.d/20-zip.ini
PHP API	20160303
PHP Extension	20160303
Zend Extension	320160303
Zend Extension Build	API320160303,NTS
PHP Extension Build	API20160303,NTS
Debug Build	no
Thread Safety	disabled
Zend Signal Handling	enabled
Zend Memory Manager	enabled
Zend Multibyte Support	disabled
IPv6 Support	enabled
DTrace Support	available, disabled
Registered PHP Streams	https, ftps, compress.zlib, php, file, glob, data, http, ftp, phar, zip
Registered Stream Socket Transports	tcp, udp, unix, udg, ssl, tls, tlsv1.0, tlsv1.1, tlsv1.2
Registered Stream Filters	zlib.*, string.rot13, string.toupper, string.tolower, string.strip_tags, convert.*, consumed, dechunk, convert.iconv.*

This program makes use of the Zend Scripting Language Engine:
Zend Engine v3.1.0, Copyright (c) 1998-2018 Zend Technologies
with Zend OPcache v7.1.33-17+ubuntu20.04.1+deb.sury.org+1, Copyright (c) 1999-2018, by Zend Technologies

Fig 1.1-PHP learning process

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

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

The output is the class of crime that is likely to have occurred

1.4.2 Our Dataset

Dataset which we are using is scraped daily from website of police which is publicly available and the people which they add to our website.

But the dataset is Hindi and in order to perform PHP for some instance so this data cannot be used as it is.

Hence the data needs to be processed Features of this dataset

- थाना : Police Station
- थाना अपराध/मगग क्रमाांक : Police Station identification number
- धारा : I.P.C. act number
- फररयादी का नाम एवाां पता : Complainant name & address
- आरोपी का नाम एवाां पता : Accused name & address
- घटना स्थल : Incident place
- घटनाोा ददनाोाोाोा ोाोंक व समय : Incident date & time
- काोायमाी ददनाोाोाोा ोाोंक व समय : Reporting date & time
- ददलाोा ोाोंब साा काोायमाी काोा काोारण : Reason of Time delay in reporting to police
- घटनाोा काा काोारण सदहत सांक्षक्षप्त ददवरण : Incident information in brief

थाना	थाना अपराध/मग ग क्रमाांक	धारा	फररयादी का नाम एवा पता	आरोपी का नाम एवा पता	घटना स्थल	घटना दिन समय	कायमी दिन व समय	दिल कायमी का कारण	घटना क कारण सद त दिवर ण
थाना जूनी इाांदौर	89/18	379,	सुनील अनिज उम्र २८ वर्ग दिपत त सिर अनिज दिनविास ५६ टाईप २ पीएनटी कालोनी खातीवालाटैंक इन्दौर	अज्ञात -,	४६ टाईप २ बीएसएनएल क्राटगर खातीवालाटैंक इन्दौर	08-02-18 11:0 क 12:0 बीच	2/10/2018 12:45:00 PM	फररयादी क न आन पर	कोई अज्ञात व्यक्ति फररयादी की दुकान नम्बर की मोटर साकयकल को रख स्थान स चोरी कर ल गया
थाना राऊ	64/18	13 जुआ एक्ट	शासन तफ पिर दिल स सउदिन मिश्र श्रीवास्तव	दिनेश - का शरदस मिश्र , सुरेश रमेश यादव	बोडोडी मोटो मिल् राऊ	10-02-18 19:10क 20:0 बीच	2/10/2018 8:10:00 PM		घटना दिनेा ाक को आरोपीयो को तास

			¤पत ा ा/प त भगवान दास ¤नव ा ास ा ी प T ुलस थाना राऊ	मुक` श - सत्यनारायण प ं व ार,					पत्तो सें T ा ारज ीत काा दाव लगात`    हुव` पकडा

Table 1.1: Police Dataset

1.4.3 Preprocessing

Before implementing machine learning algorithms on our data, we went through a series of preprocessing steps with our classification task in mind. These included:

- Dropping features such police station, station number, Complainant name & address

  ,Accused name & address

- Dropping features such as Resolution, Description and Address: The resolution and description of a crime are only known once the crime has occurred, and have limited significance in a practical, real-world scenario where one is trying to predict what kind of crime has occurred, and so, these were omitted. The address was dropped because we had information about the latitude and longitude, and, in that context, the address did not add much marginal value.

- The timestamp contained the year, date and time of occurrence of each crime. This was decomposed into five features: Year (2018), Month (1-12), Date (1-31), Hour (0-23) and Minute (0-59).

Following these preprocessing steps, we ran some out-of-the box learning algorithms as a part of our initial exploratory steps. Our new feature set consisted of 9 features, all of which were now numeric in nature.

timestamp	act379	act13	act279	act323	act363	act302	latitude	longitude
28-02-2018 21:00	1	0	0	0	0	0	22.73726	75.87599
28-02-2018 21:15	1	0	0	0	0	0	22.72099	75.87608
28-02-2018 10:15	0	0	1	0	0	0	22.73668	75.88317
28-02-2018 10:15	0	0	1	0	0	0	22.74653	75.88714

Table 1.2: Dataset after Preprocessing

1.4.4 Methodology

After the preprocessing described in the previous sections, we had three different classifications problems to solve, which we proceeded to attack with an assortment of classification algorithms. The following are the algorithms which we are using:


php

1.5 Contribution of Project

1.6.1 Market potential

The use of AI and and PHP crime via soud or cameras currently exists, is proven to w ork, and expected to continue to expand.

The use of AI in predicting crimes or an individual's likelihood for committing a crime has promise but is still more of an unknown.The biggest challenge will probably be "proving" to politicians that it works. When a system is designed to stop something from happening, it is difficult to prove the negative.Companies that are directly involved in providing governments with AI tools to monitor areas or predict crime will likely benefit from a positive feedback loop. Improvements in crime prevention technology will likely spur increased total spending on this technology.

Possible avenues through which to extend this work include time-series modeling of the data to understand temporal correlations in it, which can then be used to predict surges in different categories of crime. It would also be interesting to explore relationships between surges in different categories of crimes.

For Example: it could be the case that two or more classes of crimes surge and sink together, which would be an interesting relationship to uncover. Other areas to work on include implementing a more accurate multi-class classifier, and exploring better ways to visualize our results.

1.6.2 Innovativeness

The idea behind this project is that crimes are relatively predictable; it just requires being able to sort through a massive volume of data to find patterns that are useful to law enforcement.This kind of data analysis was technologically impossible a few decades ago, but the hope is that recent developments in machine learning are up to the task.

1.6.3 Usefulness

Public safety and protection relate to crime, and a better understanding of crime is beneficial in multiple ways: it can lead to targeted and sensitive practices by law enforcement authorities to mitigate crime, and more concerted efforts by citizens and authorities to create healthy neighborhood environments. With the advent of the Big Data era and the availability

of fast, efficient algorithms for data analysis, understanding patterns in crime from data is an active and growing field of research.

1.6 Report Organization

The remaining section of the report is structured as follows:

- **Chapter 2** provides detailed business and technical requirements
- **Chapter 3** provides analysis and design of this project
- **Chapter 4** provides Construction, implementation details of this project
- **Chapter 5** provides Conclusion and future scope as well as future application of this project

Chapter-2

Requirement Engineering

2.1 Functional Requirement

The functional requirements describe the core functionality of the application

2.1.1 Interface Requirement:

- Screen 1 to accept user inputs.
- Field 2 accepts date & time.
- Button 1 overall analysis.
- Submit button to send data of Field 1 & 2 to Kernel.
- Screen 2 displays predicted values.
- Screen 3 displays analysis.
- Screen 4 police verification
- Screen 5 police action regarding the register crime.

2.2 Non Functional Requirement

Non function requirement are those requirement of the system which are not directly concerned with specific functional delivered by the system. They may be related to emergent properties such as reliability, extendibility, usability,etc.

- To provide prediction of crime.
- To provide maximum accuracy.
- Provide visualized analysis.
- Ease of use.
- Availability
- Reliability
- Maintainability

Chapter-3

Analysis and Design

3.1 Use case diagram

Use case diagram represent the overall scenario of the system. A scenario is nothing but a sequence of steps describing an interaction between a user and a system.

Thus use case is a set of scenario tied together by some goal. The use case diagram are drawn for exposing the functionalities of the system.

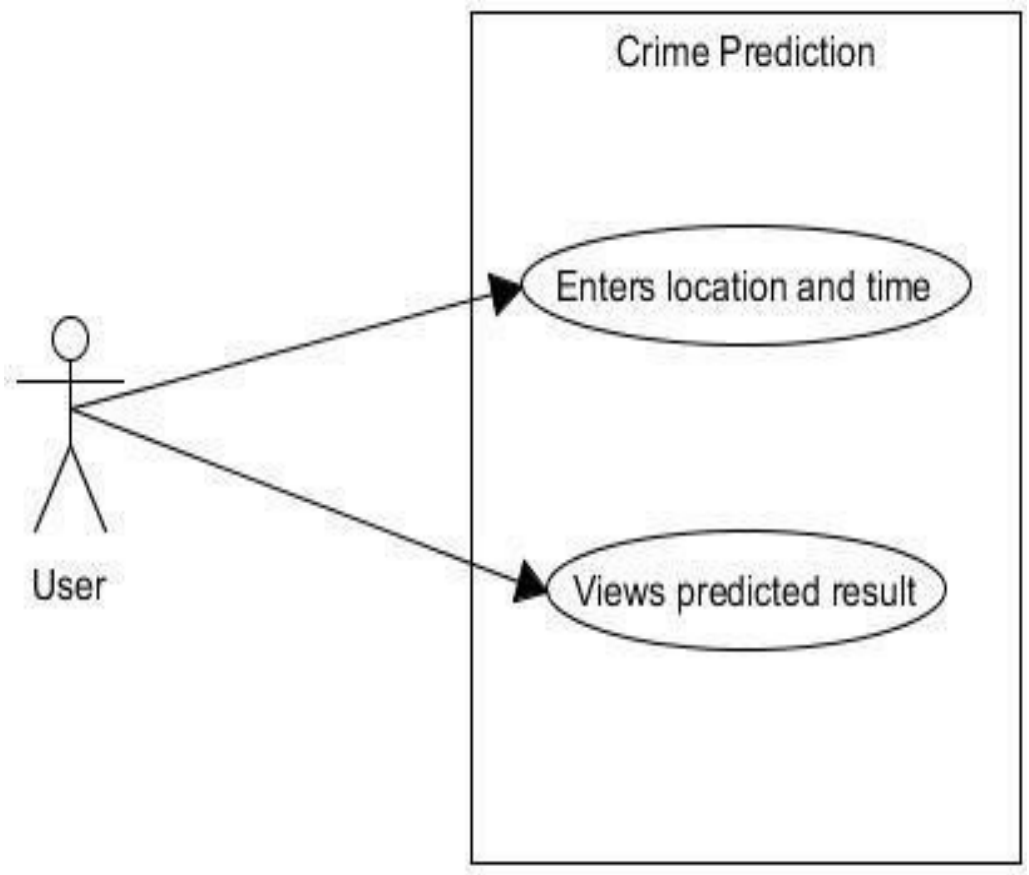


Fig 3.1-Use case diagram of Raksitr

3.2 Activity diagram

The activity diagram is a graphical representation for representing the flow of interaction within specific scenatios. It is similar to a flowchart in which various activities that can be performed in the system are represented.

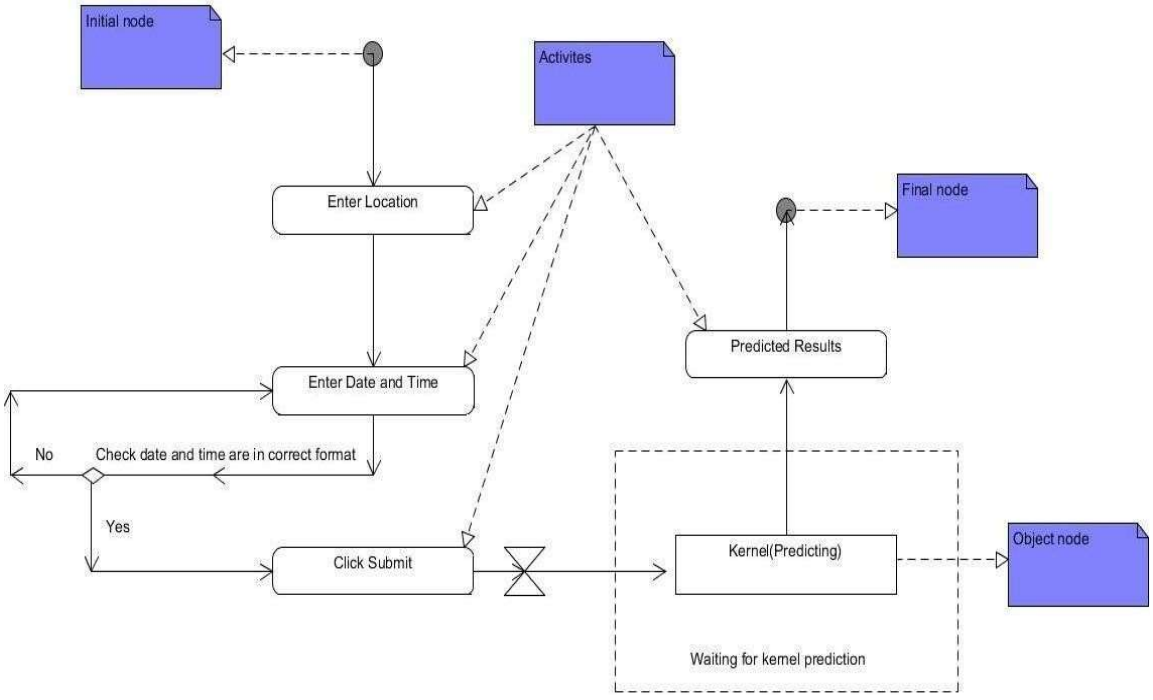


Fig 3.2-Activity diagram of Raksitr

3.3 Sequence diagram

In the sequence diagram how the object interacts with the other object is shown. There are sequence of events that are represented by a sequence diagram.

It is a time oriented view of the interation between objects to accomplish a behavioural goal of the system.

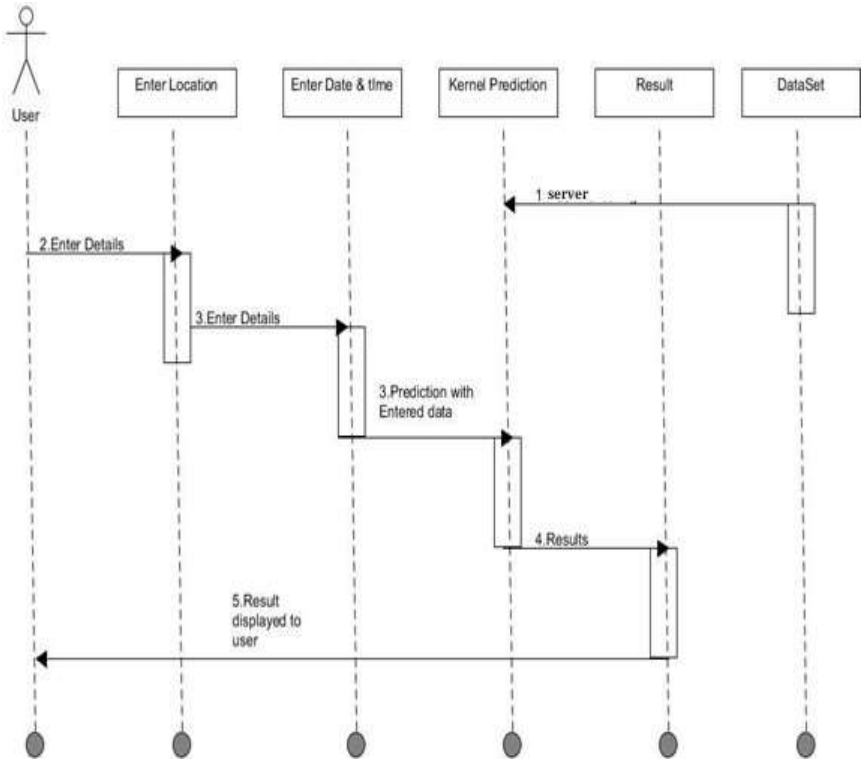


Fig 3.3-Sequence diagram of Raksitr

3.4 System architecture

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

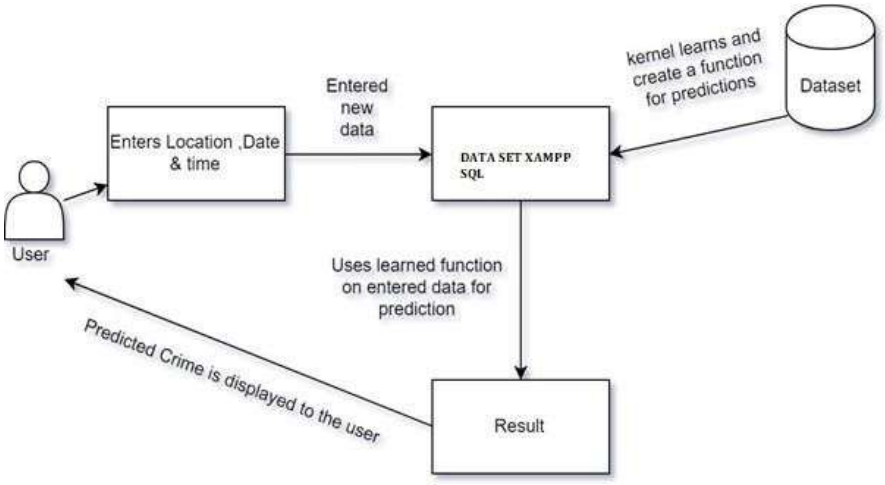


Fig 3.4-System architecture of Raksitr

Chapter-4

Construction

4.1 Implementation

The implementation of the project is done with the help of SQL. To be particular, for the purpose of PHP with XAMPP Software.

XAMPP is one of several Apache Server distributions. XAMPP is a new distribution of the Server.
XAMPP is an abbreviation for cross-platform, Apache, MySQL, PHP and Perl, and it allows you to build WordPress site offline, on a local web server on your computer. This simple and lightweight solution works on Windows, Linux, and Mac – hence the "cross-platform" part.

On XAMPP technology, we found out Anaconda to be easier. Since it helps with the following problems:

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

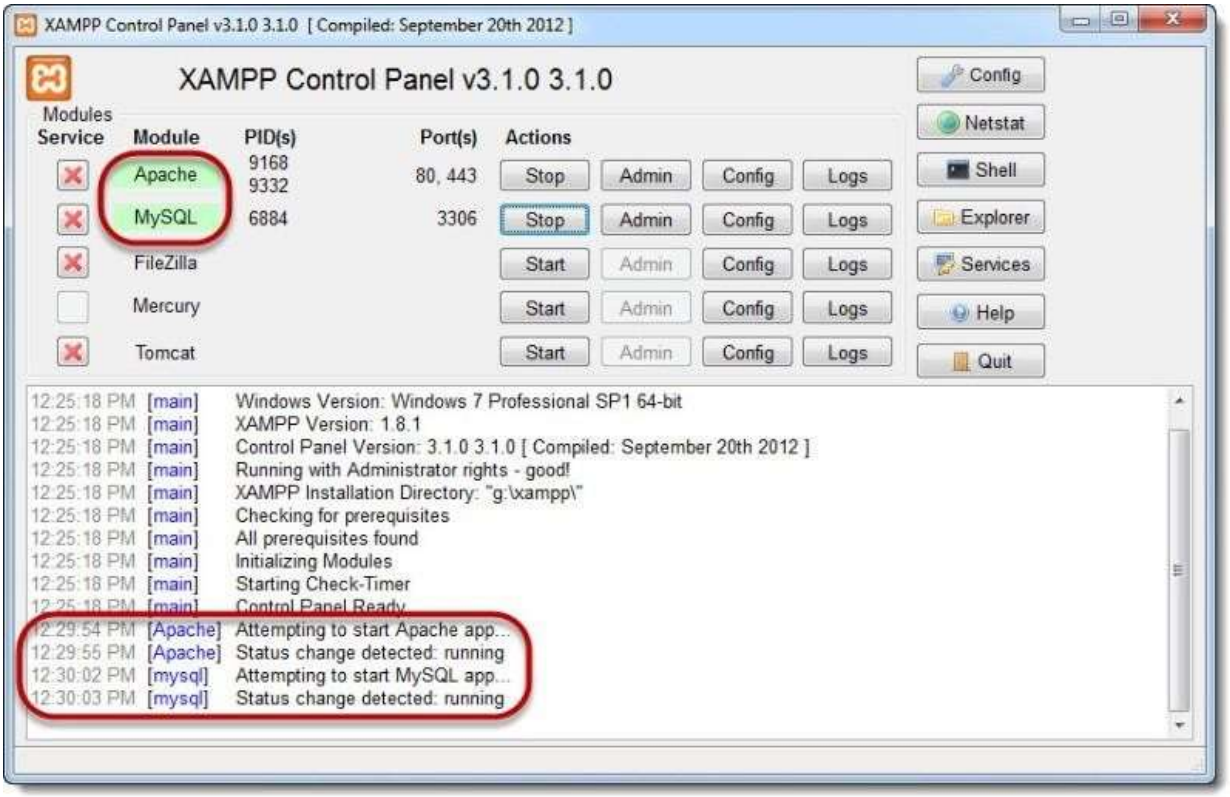
This data was scraped from the publically available data from police website which had been made by people in police station of different areas. Implementation of the idea started from the Indore city itself so as to limit an area for the prediction and making it less complex. The data was sorted and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

The entries was done just to make the machine learn what all it has to do with the data and what actually the output is being demanded. As soon as the

machine learnt the algorithms and the process, accuracy of different algorithms were measured & the algorithm with the most accuracy is used for the prediction kernel i.e. Random forest.

4.2 Implementation Details

For the purpose of proper implementation and functioning several Algorithms and techniques were used. Following are the algorithms used:



4.2.1 PHP INTEGER

An integer data type is a non-decimal number between -2,147,483,648 and 2,147,483,647.

Rules for integers:

- An integer must have at least one digit
- An integer must not have a decimal point
- An integer can be either positive or negative
- Integers can be specified in: decimal (base 10), hexadecimal (base 16), octal (base 8), or binary (base 2) notation

In the following example $x is an integer. The PHP var_dump() function returns the data type and value:

Example

```
<?php
$x = 5985;
var_dump($x);
?>
```

NULLABLE TYPE

As of PHP 7.1.0, type declarations can be marked nullable by prefixing the type name with a question mark (?). This signifies that the value can be of the specified type or null.

```
<?php
class C {}

function f(?C $c) {
    var_dump($c);
}

f(new C);
f(null);
?>
```

4.2.2 Union Types

A union type declaration accepts values of multiple different types, rather than a single one. Union types are specified using the syntax T1|T2| Union types are available as of PHP 8.0.0.

Nullable union types

The null type is supported as part of unions, such that T1|T2|null can be used to create a nullable union. The existing ?T notation is considered a shorthand for the common case of T|null.

False pseudo-type

The false literal type is supported as part of unions, and is included as for historical reasons many internal functions return false instead of null for failures. A classic example of such a function is strpos().

Duplicate and redundant types

To catch simple bugs in union type declarations, redundant types that can be detected without performing class loading will result in a compile-time error. This includes:
Each name-resolved type may only occur once. Types such as int|string|INT result in an error.
If bool is used, false cannot be used additionally.
If object is used, class types cannot be used additionally.
If iterable is used, array and Traversable cannot be used additionally.

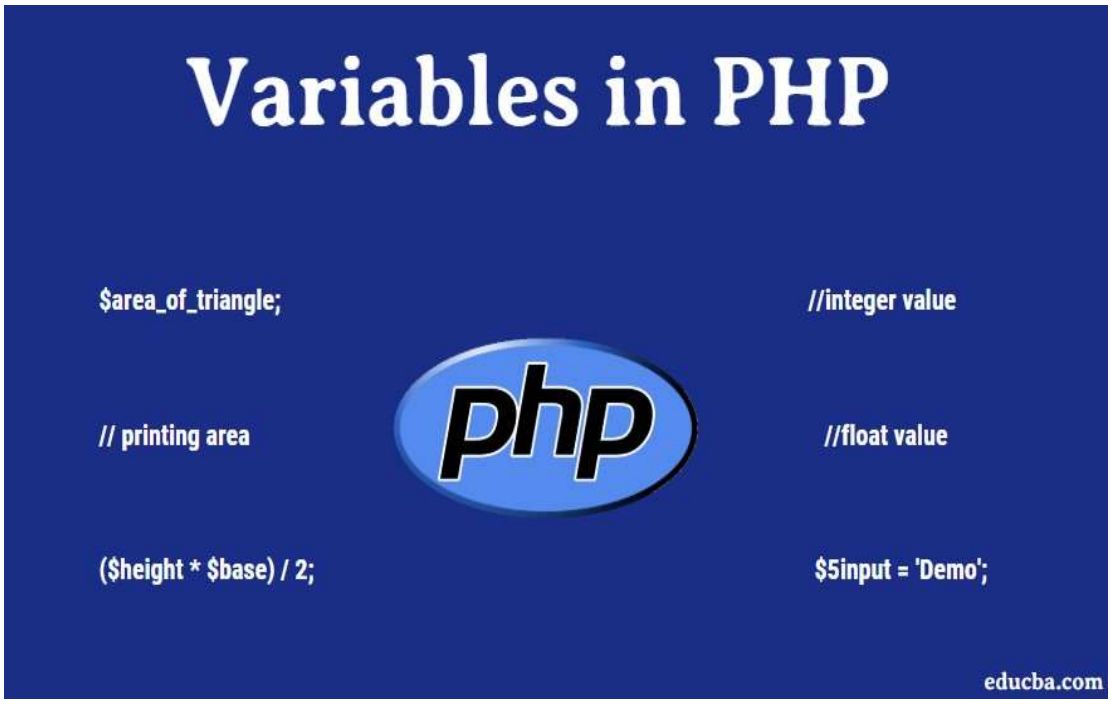


Fig 4.2.2 Variable In PHP

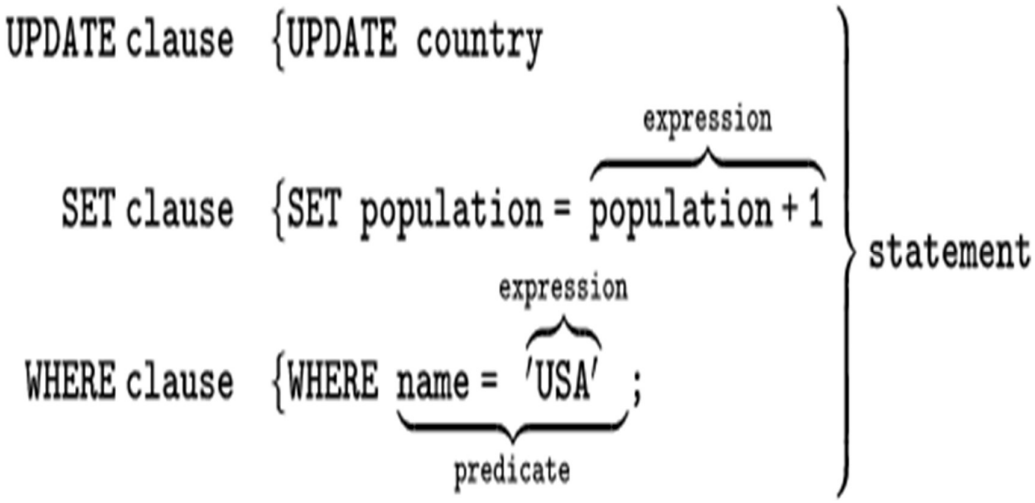
4.2.3 SQL

The SQL language is subdivided into several language elements, including:

- *Clauses*, which are constituent components of statements and queries. (In some cases, these are optional.)
- *Expressions*, which can produce either scalar values, or tables consisting of columns and rows of data
- *Predicates*, which specify conditions that can be evaluated to SQL three-valued logic (3VL) (true/false/unknown) or Boolean truth values and are used to limit the effects of statements and queries, or to change program flow.
- *Queries*, which retrieve the data based on specific criteria. This is an important element of *SQL*.
- *Statements*, which may have a persistent effect on schemata and data, or may control transactions, program flow, connections, sessions, or diagnostics.

- SQL statements also include the semicolon (";") statement terminator. Though not required on every platform, it is defined as a standard part of the SQL grammar.

- *Insignificant whitespace* is generally ignored in SQL statements and queries, making it easier to format SQL code for readability.



SQL implementations are incompatible between vendors and do not necessarily completely follow standards. In particular, date and time syntax, string concatenation, `NULL`s, and comparison case sensitivity vary from vendor to vendor. Particular exceptions are PostgreSQL and Mimer SQL which strive for standards compliance, though PostgreSQL does not adhere to the standard in all cases. For example, the folding of unquoted names to lower case in PostgreSQL is incompatible with the SQL standard,[24] which says that unquoted names should be folded to upper case. Thus, `Foo` should be equivalent to `FOO` not `foo` according to the standard.

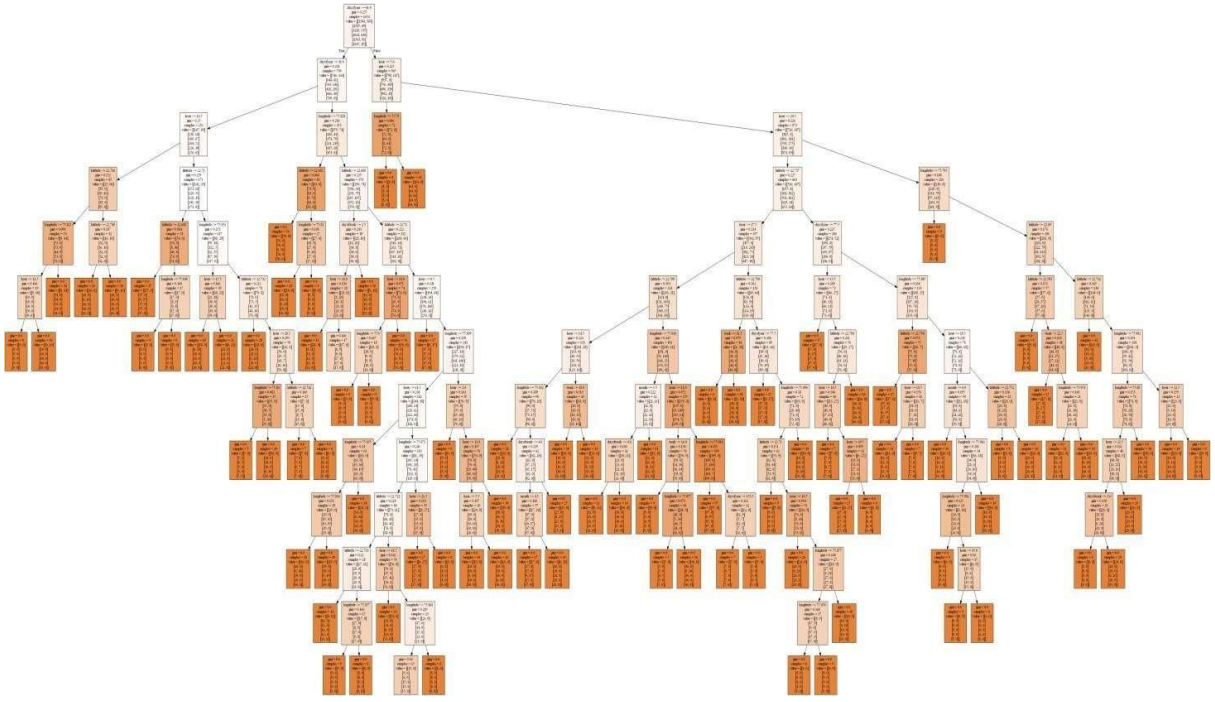


Fig 4.3.2 Decision Tree of Raksitr

4.2.4 Data Visualization

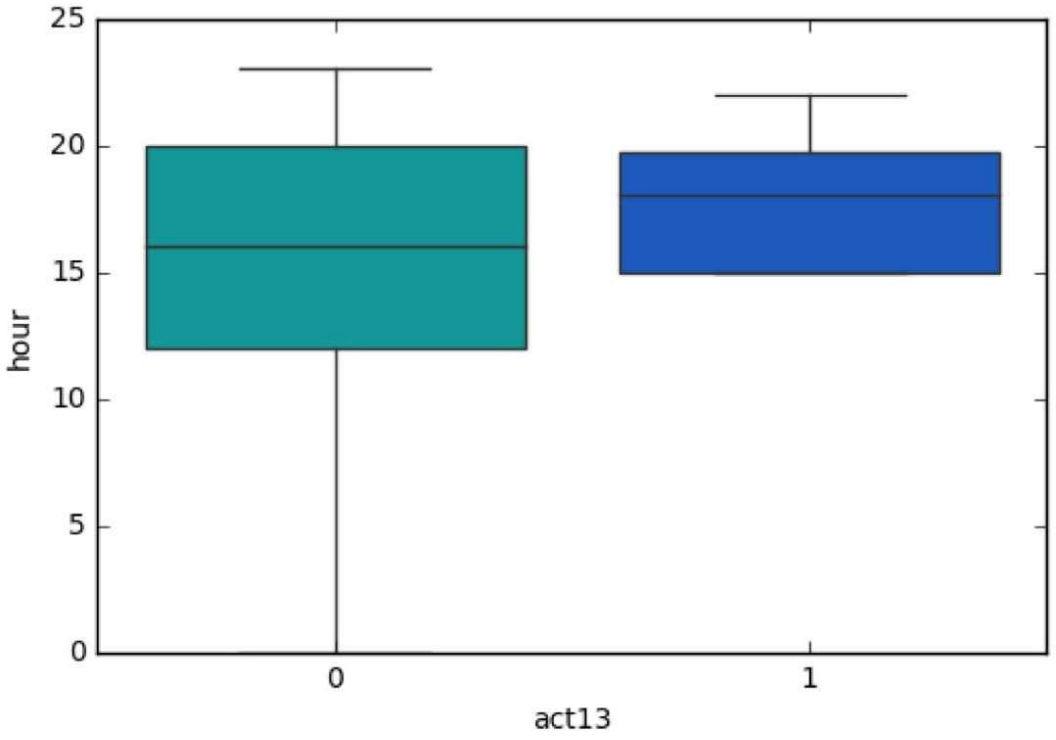


Fig 4.4.1 Act13(Gambling vs Hour)

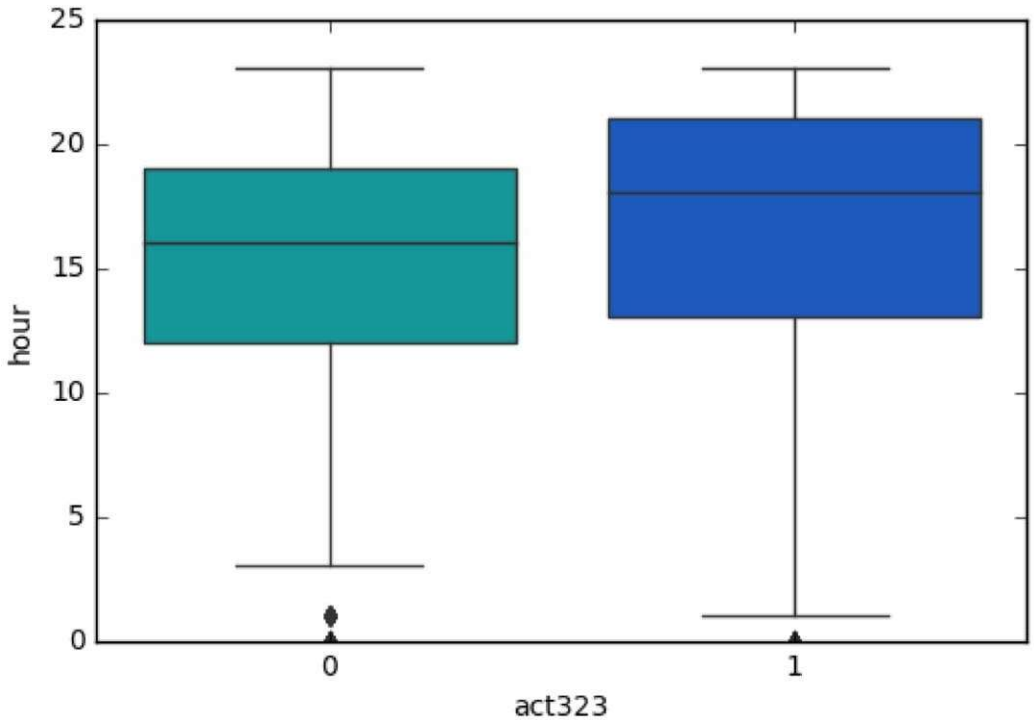


Fig 4.4.2 Act323(Violence vs Hour)

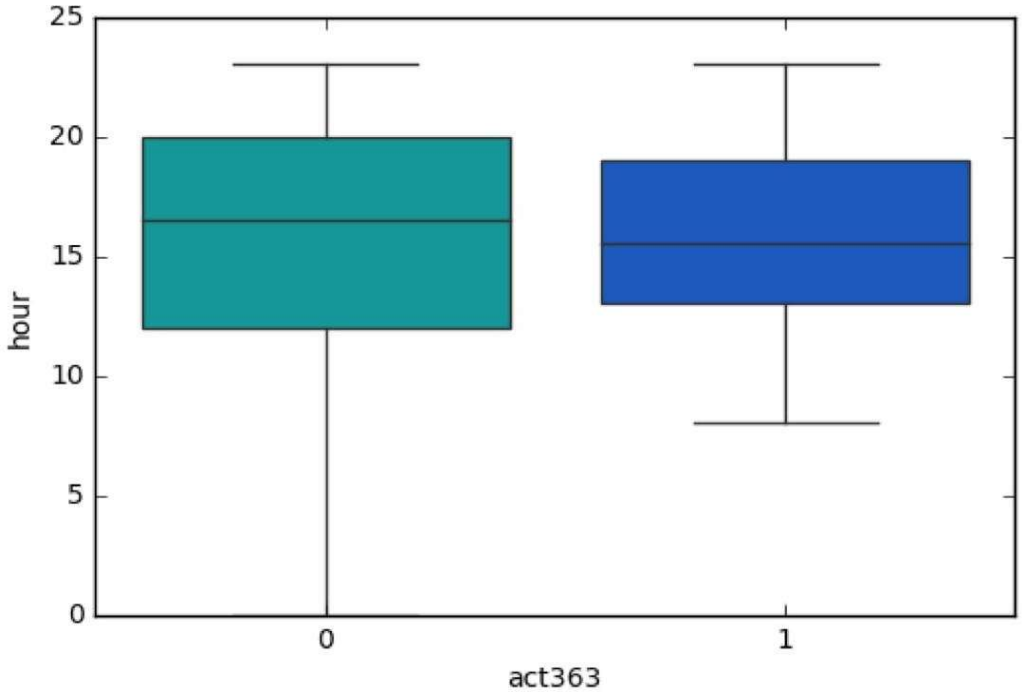


Fig 4.4.3 Act363(Kidnapping vs Hour)

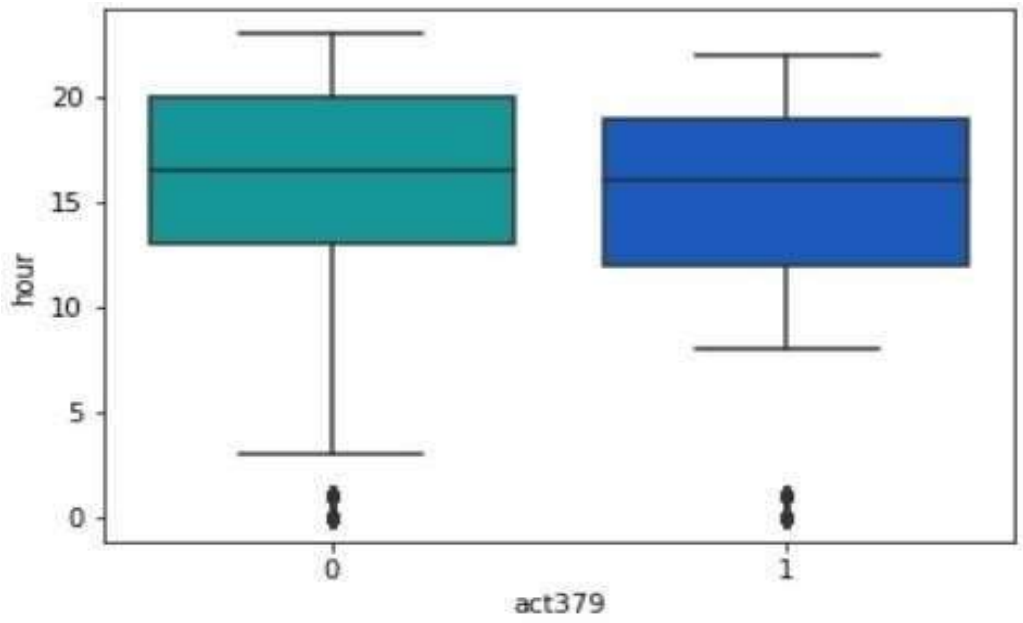


Fig 4.4.4 Act379(Robbery vs Hour)

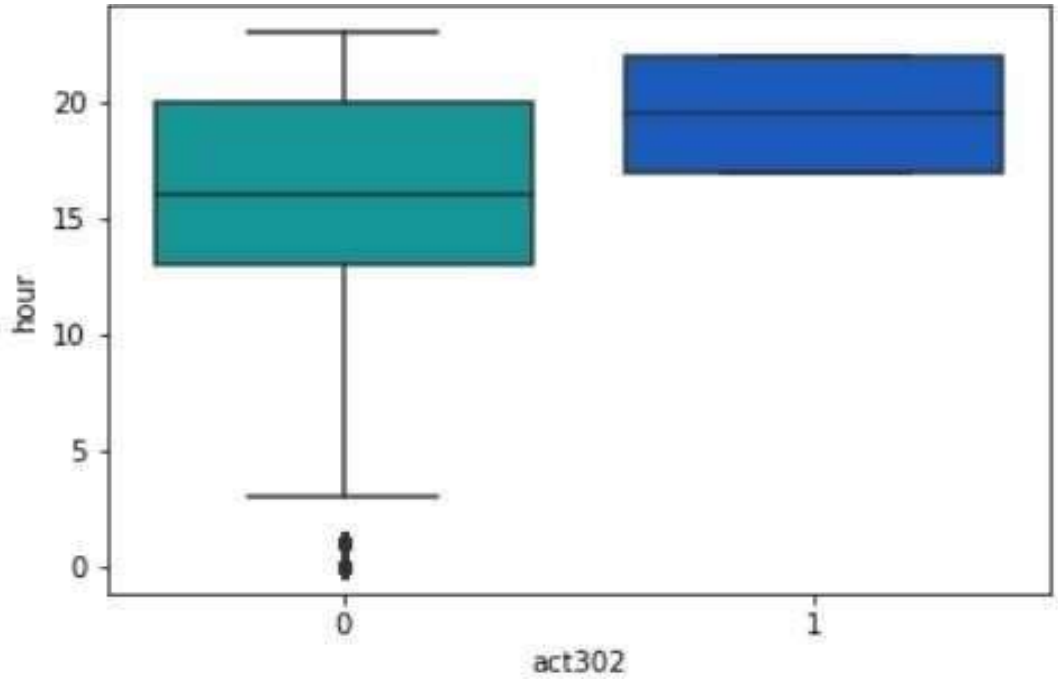


Fig 4.4.5 Act302(Murder vs Hour)

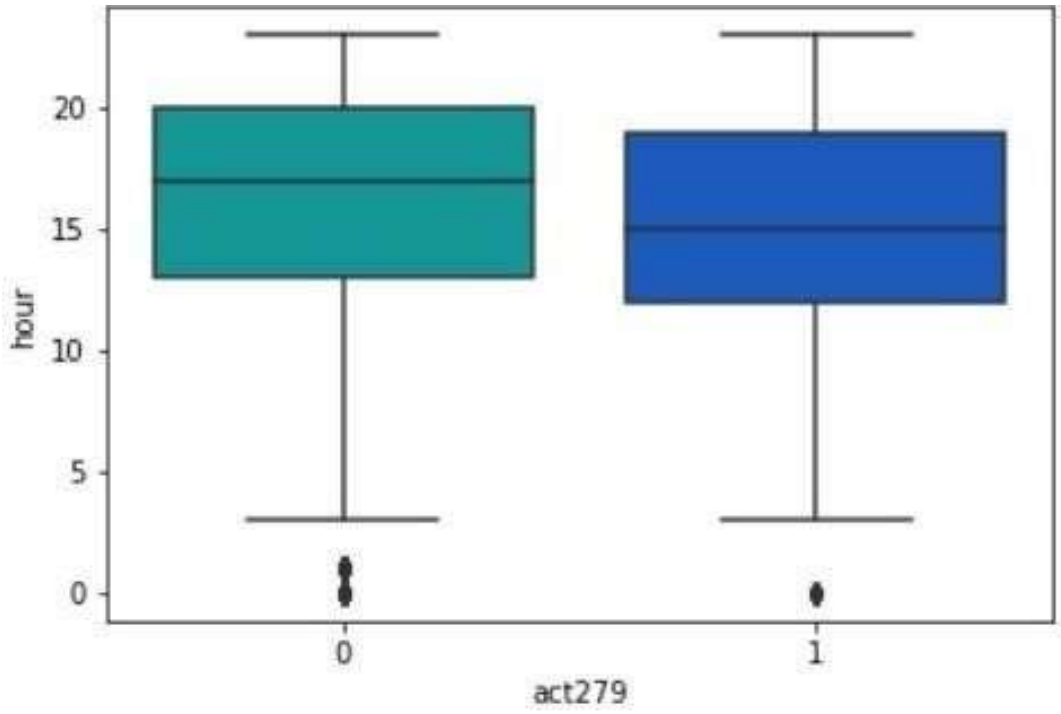


Fig4.4.6 Act279(Accident vs Hour)

4.3 Software Details

- XAMPP Control panel 3.2.4
- SQL
- HTML 5
- CSS 3
- APACHE SERVER
- PHP 1.8.0
- Phpmyadmin5.1.1

4.4 Hardware Details

- Operating system: Windows 7 or newer, 64-bit macOS 10.9+, or Linux.
- System architecture: 64-bit x86, 32-bit x86 with Windows or Linux.
- CPU: Intel Core 2 Quad CPU Q6600 @ 2.40GHz or greater.
- RAM: 4 GB or greater.

4.5 Testing

The development of software involves a series of production activities were opportunities for injection of human fallibilities are enormous.

Error may begin to occur at very inspection of the process where the objective may be enormously or imperfectly specified as well as in lateral design and development stage. Because of human inability to perform and communicate with perfection, software development quality assurance activities.

Software testing is a crucial element of software quality assurances and represents ultimate review of specification, design and coding.

4.5.1 XAMPP

It focuses on the program control structure. Here all statement in the project have been executed at least once during testing and all logical condition have been exercised.

4.5.2 PhpMyAdmin

This is designed to uncover the error in functional requirements without regard to the internal working of the project. This testing focuses on the information domain of the project , deriving test case by partitioning the input and output domain of programming – A manner that provides through test coverage.

This testing contains are main local host admin page which shows the all data gatherd in the server ay that period of time

Test Case ID	Test Name	Test Description	Steps	Executed result	Actual result	Test case statement
01	Check for correct entered numeric values and date and time.	The entered values are in correct format.	1. Enter details in fields. 2. Click submit.	If format is correct details are sent to kernel successfully.	As expected.	Pass
02	Check for correct entered time.	The entered values are in correct format.	1. Enter details in fields. 2. Click submit.	If format is correct details are sent to kernel successfully	As expected.	Pass
03	Check for correct entered location	The entered values are correct.	1. Enter details in fields. 2. Click submit.	If format is correct details are sent to kernel successfully	As expected	Pass
04	Predicted Result	Output is displayed		If kernel predicts successfully output is then showed to the screen	As expected	Pass
05	Analysis Button	Data visualization is displayed.	1.Click Analysis	Shows the overall analysis on screen 3	As expected	Pass

Table 4.5.3 Tests

Chapter-5

Conclusion and future scope

5.1 Conclusion

The initial problem of classifying 6 different crime categories was a challenging multi-class classification problem, and there was not enough predictability in our initial data-set to obtain very high accuracy on it. We found that a more meaningful approach was to collapse the crime categories into fewer, larger groups, in order to find structure in the data. We got high accuracy and precision on Prediction. However, the Violent/Non-violent crime classification did not yield remarkable results with the same classifiers – this was a significantly harder classification problem. Thus, collapsing crime categories is not an obvious task and requires careful choice and consideration.

Possible avenues through which to extend this work include time-series modeling of the data to understand temporal correlations in it, which can then be used to predict surges in different categories of crime. It would also be interesting to explore relationships between surges in different categories of crimes – for example, it could be the case that two or more classes of crimes surge and sink together, which would be an interesting relationship to uncover. Other areas to work on include implementing a more accurate multi-class classifier, and exploring better ways to visualize our results.

5.2 Future Scope

The goal of any society shouldn't be to just catch criminals but to prevent crimes from happening in the first place

- **Predicting Future Crime Spots:** By using historical data and observing where recent crimes took place we can predict where future crimes will likely happen. For example a rash of burglaries in one area could correlate with more burglaries in surrounding areas in the near future. System highlights possible hotspots on a map the police should consider patrolling more heavily

Fig 5.1 Predicting Surges

- **Predicting Who Will Commit a Crime**: Using Face Recognition to predict if a individual will commit a crime before it happens. The system will detect if there are any suspicious changes in their behavior or unusual movements. For example if an individual seems to be walking back and forth in a certain area over and over indicating they might be a pickpocket or casing the area for a future crime. It will also track individual over time.

- **Pretrial Release and Parole:** After being charged with a crime, most individuals are released until they actually stand trial. In the past deciding who should be released pretrial or what an individual's bail should be set at is mainly now done by judges using their best judgment. In just a few minutes, judges had to attempt to determine if someone is a flight risk, a serious danger to society, or at risk to harm a witness if released. It is an imperfect system open to bias. The media organization's analysis indicated the system might indirectly contain a strong racial bias. They found, "That black defendants who did not recidivate over a two-year period were nearly twice as likely to be misclassified as higher risk compared to their white counterparts (45 percent vs. 23 percent)." The report raises the question of whether better AI/ML can eventually produce more accurate predictions or if it would reinforce existing problems. Any system will be based off of real world data, but if the real world data is generated by biased police officers, it can make the AI/ML biased.

5.3 Expected Outcome

The idea behind this project is that crimes are relatively predictable; it just requires being able to sort through a massive volume of data to find patterns that are useful to law enforcement. This kind of data analysis was technologically impossible a few decades ago, but the hope is that recent developments in machine learning are up to the task.

The use of AI and machine learning to detect crime via sound or cameras currently exists, is proven to work, and expected to continue to expand. The use of AI/ML in predicting crimes or an individual's likelihood for committing a crime has promise but is still more of an unknown. The biggest challenge will probably be "proving" to politicians that it works. When a system is designed to stop something from happening, it is difficult to prove the negative.

Companies that are directly involved in providing governments with AI tools to monitor areas or predict crime will likely benefit from a positive feedback loop. Improvements in crime prevention technology will likely spur increased total spending on this technology.

Possible avenues through which to extend this work include time-series modeling of the data to understand temporal correlations in it, which can then be used to predict surges in different categories of crime. It would also be interesting to explore relationships between surges in different categories of crimes – for example, it could be the case that two or more classes of crimes surge and sink together, which would be an interesting relationship to uncover. Other areas to work on include implementing a more accurate multi-class classifier, and exploring better ways to visualize our results.

References

[1] Bogomolov, Andrey and Lepri, Bruno and Staiano, Jacopo and Oliver, Nuria and Pianesi, Fabio and Pentland, Alex.2014. Once upon a crime: Towards crime prediction from demographics and mobile data, Proceedings of the 16th International Conference on Multimodal Interaction.

[2] Yu, Chung-Hsien and Ward, Max W and Morabito, Melissa and Ding, Wei.2011. Crime forecasting using data mining techniques, pages 779-786, IEEE 11th International Conference on Data Mining Workshops (ICDMW)

[3] Kianmehr, Keivan and Alhajj, Reda. 2008. Effectiveness of support vector machine for crime hot-spots prediction, pages 433-458, Applied Artificial Intelligence, volume 22, number 5.

[4] Toole, Jameson L and Eagle, Nathan and Plotkin, Joshua B. 2011 (TIST), volume 2, number 4, pages 38, ACM Transactions on Intelligent Systems and Technology

[5] Wang, Tong and Rudin, Cynthia and Wagner, Daniel and Sevieri, Rich. 2013. pages 515-530, Machine Learning and Knowledge Discovery in Databases

[6] Friedman, Jerome H. ”Stochastic gradient boosting.” Computational Statistics and Data Analysis 38.4 (2002): 367-378.sts

[7] Leo Breiman, Random Forests, Machine Learning, 2001,Volume 45, Number 1, Page 5

Appendix-A

Fig A.1-Snapshot 1

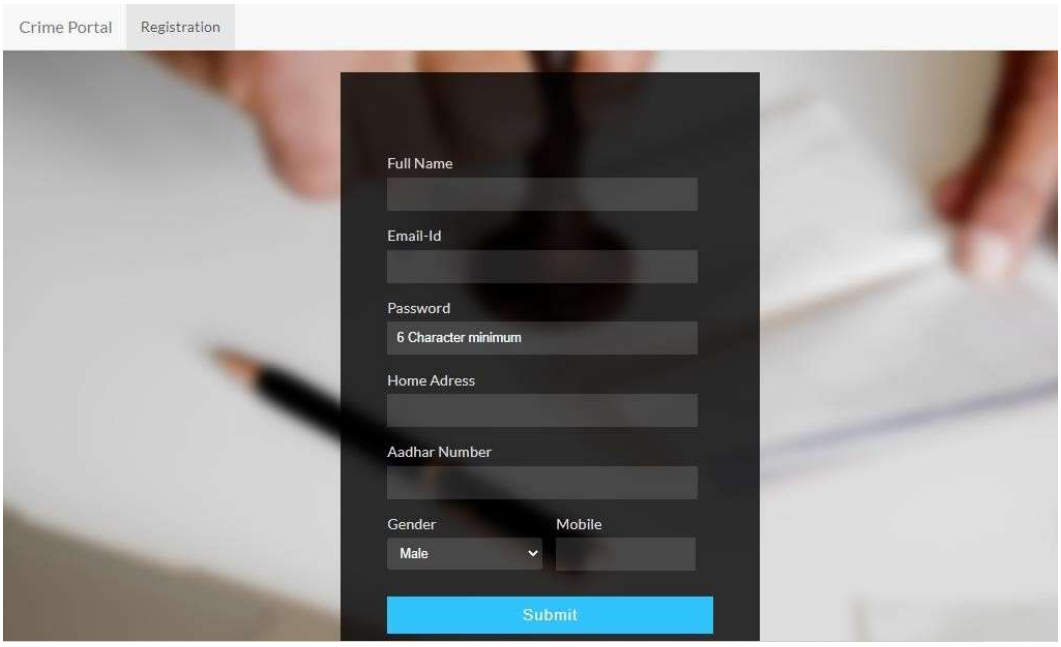
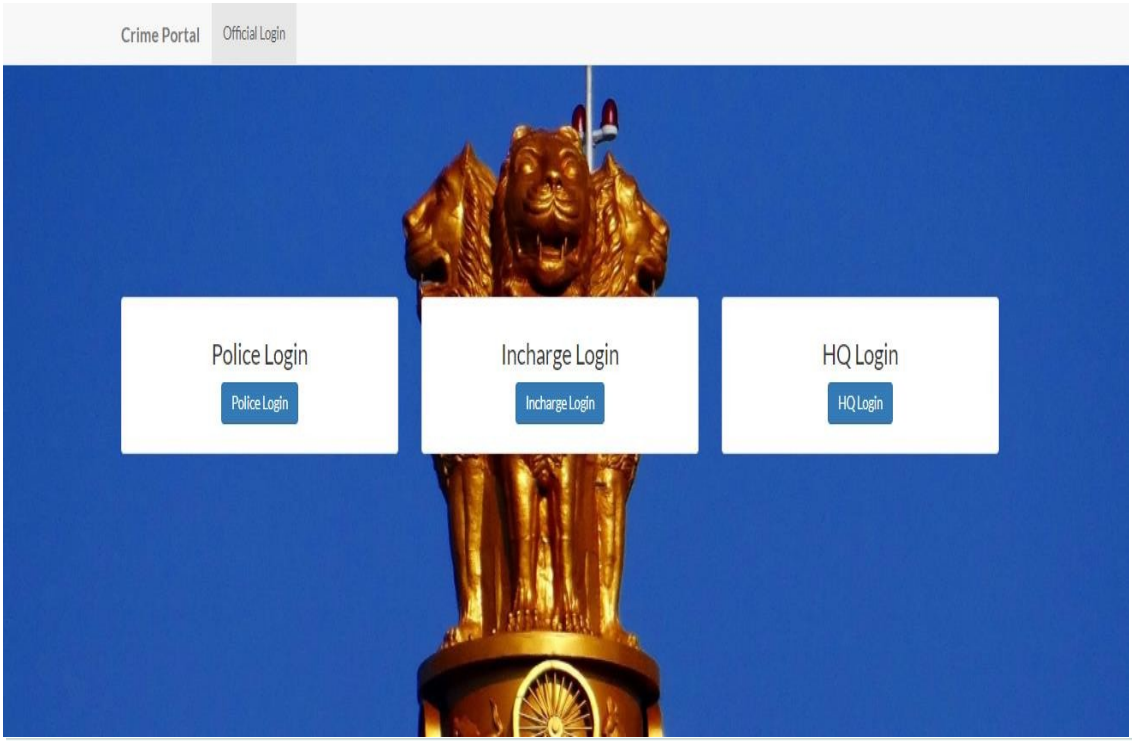


Fig A.2-Snapshot 2

Appendix-B

Fig B.1-Snapshot 3



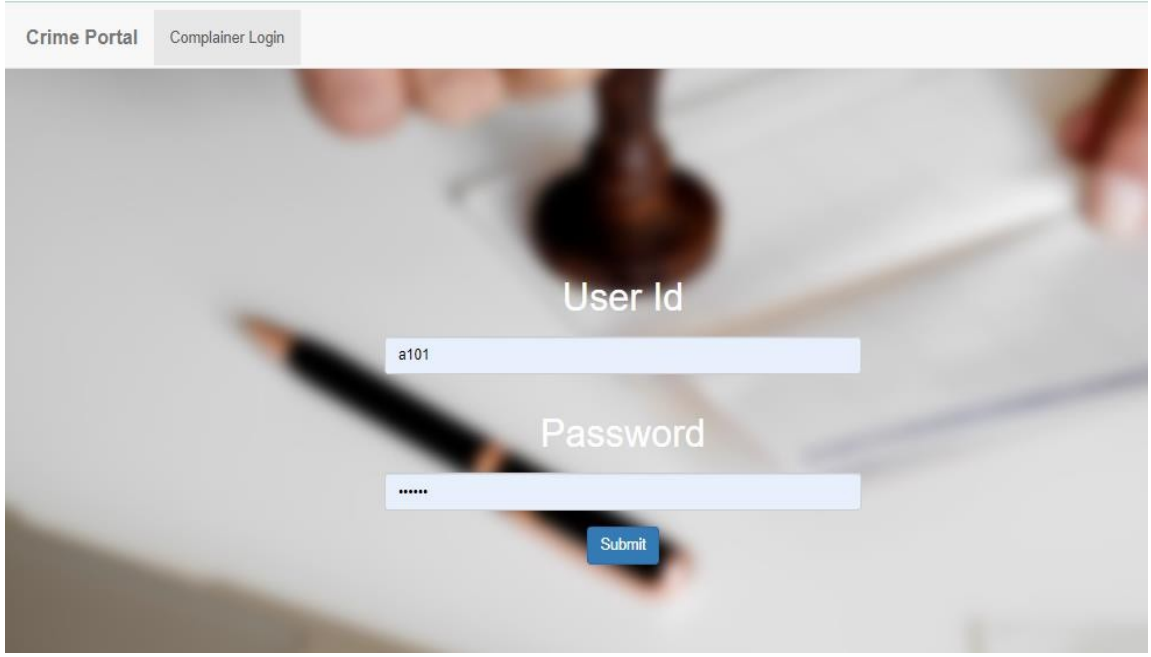


Fig B.2-Snapshot 4

Appendix-C

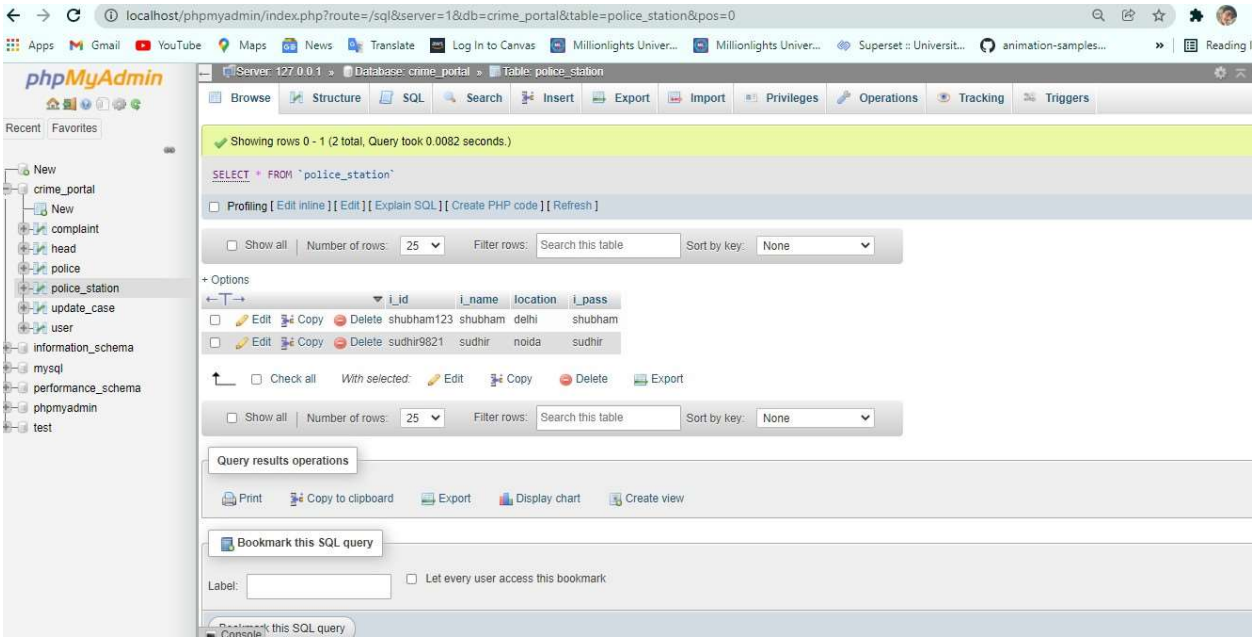


Fig C.1-Snapshot 5

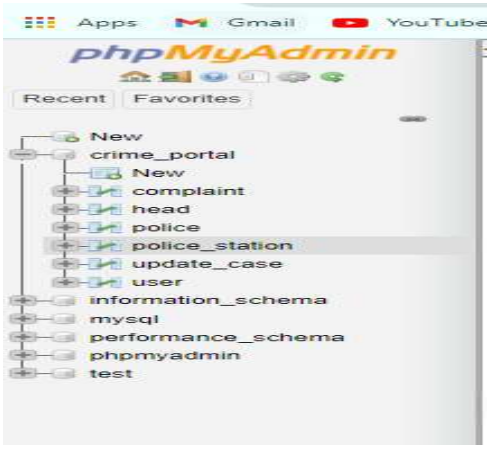


Fig C.2-Snapshot 6