FINAL YEAR MINOR PROJECT REPORT (PROJECT STAGE – I & II)

ML – Driven Criminal Activities Prediction

TRACK – 4 (INDUSTRY)

Submitted in Partial Fulfillment for the Award of Degree of Bachelor of Technology in Computer Science and Engineering from Rajasthan Technical University, Kota

COORDINATOR:

Dr. Tapas Badal
(Dept. of Computer Science & Engineering)

MENTOR:

Dr. Basant Agarwal
(Dept. of Computer Science & Engineering)

SUBMITTED BY:

Ishan Kalra (15ESKCS726)

Chandra Prakash (15ESKCS716)

Bhumika Saxena (15ESKCS715)

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

SWAMI KESHWANAND INSTITUTE OF TECHNOLOGY, MANAGEMENT & GRAMOTHAN
Ramnagaria (Jagatpura), Jaipur – 302017

SESSION 2018 - 19

SWAMI KESHWANAND INSTITUTE OF TECHNOLOGY, MANAGEMENT & GRAMOTHAN
Ramnagaria (Jagatpura), Jaipur – 302017

CERTIFICATE

This is to certify that Final Year Minor Project Report (Project Stage – II) entitled

"**AI – Based Automated System for Testing Driving Skills**" has been duly submitted by

- Ishan Kalra (15ESKCS726)
- Chandra Prakash (15ESKCS716)
- Bhumika Saxena (15ESKCS715)

For partial fulfillment of the Degree of Bachelor of Technology of Rajasthan Technical University. It has been found satisfactory and hence approved for submission as Minor Project during academic session 2018-2019.

Date: 8 October 2018

COORDINATOR:	MENTOR:	HEAD OF DEPARTMENT:
Mr. Tapas Badal (Dept. of Computer Science & Engineering)	**Dr. Basant Agarwal** (Dept. of Computer Science & Engineering)	**Prof. Dr. C.M.Choudhary** (Dept. of Computer Science & Engineering)

SWAMI KESHWANAND INSTITUTE OF TECHNOLOGY, MANAGEMENT & GRAMOTHAN
Ramnagaria (Jagatpura), Jaipur – 302017

ABSTRACT

As the population is increasing so are the crimes. It's extremely difficult for authorities to keep check on the crimes and deploy in all the areas of the city. This project aims to use data science and machine learning techniques to identify the potential areas which are prone to increased crime rates along with the types of crimes being committed. This project also aims to predict patterns of crimes being committed, to help the authorities deploy efficient resources to keep these crimes in check. The data from the project is taken from Boston Police Department criminal report datasets. Model generated in this project can also be generalized for cities in India if proper dataset is supplied.

Domain Descriptors:

7 CS 7 Web Development Lab

5 CS 4 Database Management Systems

4 CS 3 Software Engineering

Keywords:

Machine Learning, Data Science

SWAMI KESHWANAND INSTITUTE OF TECHNOLOGY, MANAGEMENT & GRAMOTHAN
Ramnagaria (Jagatpura), Jaipur – 302017

DECLARATION

We hereby declare that the report of the project entitled **ML – Driven Criminal Events Prediction** is a record of an original work done by us at **Swami Keshvanand Institute of Technology, Management & Gramothan, Jaipur** under the mentorship of **Dr. Basant Agarwal** (Dept. of Computer Science & Engineering) and coordination of **Mr. Tapas Badal** (Dept. of Computer Science & Technology). This project report has been submitted as the proof of original work for the partial fulfillment of the requirement for the award of the degree of **Bachelor of Technology** (B.Tech) in the **Department of Computer Science & Technology.** It has not been submitted anywhere else, under any other program to the best of our knowledge and belief.

Team Members: **Signatures:**

(15ESKCS726) Ishan Kalra

(15ESKCS716) Chandra Prakash

(15ESKCS715) Bhumika Saxena

ACKNOWLEDGMENT

A project of such a vast coverage cannot be realized without help from numerous sources and people in the organization. We take this opportunity to express our gratitude to all those who have been helping us in making this project successful.

We are highly indebted to our faculty mentor **Dr. Basant Agarwal.** He has been a guide, motivator & source of inspiration for us to carry out the necessary proceedings for the project to be completed successfully. We also thank our project coordinator **Mr. Tapas Badal** for his co-operation, encouragement, valuable suggestions and critical remarks that galvanized our efforts in the right direction.

We would also like to convey our sincere thanks to **Prof. C.M. Choudhary,** HOD, Department of Computer Science & Engineering, for facilitating, motivating and **s**upporting us during each phase of development of the project. Also, we pay our sincere gratitude to all the **Faculty Members** of Swami Keshvanand Institute of Technology, Management & Gramothan, Jaipur and all our Colleagues for their co-operation and support.

Last but not least we would like to thank all those who have directly or indirectly helped and cooperated in accomplishing this project.

Team Members:

(15ESKCS726) Ishan Kalra

(15ESKCS716) Chandra Prakash

(15ESKCS715) Bhumika Saxena

INDEX

INDEX OF FIGURES

INDEX OF TABLES

ABBREVIATIONS USED

ML	Machine Learning
DBMS	Database Management System
OS	Operating System
DMW	Data Mining & Warehousing

Chapter -1 PROJECT CHARTER

1.1 Problem Statement & Objective

Problem: Efficient use administrative resources in containing crime in the city.

Reasons: There are several reasons for the problem

- Lack of force to deal with large Indian population.
- Changing patterns and types of criminal events.
- Offline process slows the time for obtaining valuable insights.

Solution: A solution to this problem requires automation of the process of generating insights from vast amount of criminal databases. The first step is to digitize the process of filing criminal incident reports. The next and more challenging part is to automate the process of generating insights from this data.

To automate the generation of insights the data should be streamed from central database of police. This data then will cleaned, transformed and analysed to generate valuable insights from the data and also for prediction of future criminal events. Based on these predictions and insights the authorities then will be able to plan their course of action and deploy their resources effectively

Objective: To create a website to provide a simple interface for authorities to see insights and predictions along with an interface for general public to stay aware of their surroundings.

1.2 Literature Survey

The main purpose of this report is to review literature on crime measures which may be suitable for use in analyzing the criminal data and also generating forecasts for future criminal events. First, the main objectives of analyzing criminal data are defined. These are:

- To establish a framework which can analyze criminal events, and “flag” the events based on their severity.
- To generate forecasts for future criminal events that might happen with high accuracy.

The major details from any crime are – the nature of crime, the type of crime, who committed it, timestamps, places, day of week, month, year etc - are considered in detail.

The primary objectives of this project is to help authorities in planning their strategies and deploy their administrative resources effectively.

The degree to which a testing system attains these objectives is difficult to assess, for several reasons. Primarily the system uses various variables to forecast the possibility of future events. These prediction may not always be true and the efficiency of 70 percent is generally considered to be good in these types of systems.

It is apparent that testing procedures are to be properly evaluated and improved.

1.3 Introduction to Project

In a country where population is increasing rapidly and the ratio of civilian population vs police authorities is 28000:1, containing crime is a tough task. Like any other task of this scale comes the natural difficulties with human resources, proper infrastructure and efficient administration. All this is due to dependence on manual analysis of criminal reports and offline maintenance of records.

We aim to automate this process. We will incorporate the data from central crime database. This data will then be analyzed with the help of Data Science and Machine Learning models that will make the process of obtaining insights from the events fast and efficient.

The parameters we shall include are the nature of crime, the type of crime, who committed it, timestamps, places, day of week, month, year

1.4 Proposed Logic or Algorithm

The project is aimed at automating the generation of insights and forecasts for criminal events.

- The data will be imported from central crime database.
- Data Science and Machine Learning methods will be used to generate insights and forecasts..

- Based on the nature of crime, the type of crime, who committed it, timestamps, places, day of week, month, year the models will be applied.

Figure 1.1 – Architecture Diagram of the Proposed System

1.5 Scope of the Project

The project will automate the entire process of obtaining insights from criminal data.

- This project will automate the generation of insights from criminal data.
- The project can be extended for other locations as long as data is consistent with required format.
- The project will significantly reduce the time and help authorities in planning and deployment of man power and other resources.

Chapter–2 SYSTEM REQUIREMENT SPECIFICATION

2.1 Overall Description

This section and its subsections contain the description of the project compoents such as interfaces, performance requirements, design constraints, assumptions and dependencies etc.

2.1.1 Product Perspective

The application will be a Windows / Linux based, self contained and independent product.

2.1.1.1 System Interfaces

- **Client on Internet**
  Web Browser and Operating Software (any)
- **Web Server**
  Django
- **Database Server**
  SQLite Database
- **Development End**
  IDE – Jupyter notebook

  OS, Chrome Browser

2.1.1.2 User Interfaces

The web application will have a user friendly and menu based interface. Following screens will be provided:

a. A Login screen to enter the email id and password.

b. A registeration page for applying for new users.

c. Notifications of new insights and forecasts.

2.1.1.3 Hardware Interfaces

- Screen resolution of at least 800 x 600 pixels is required for proper and complete viewing of screens. Higher resolutions in wide-screen mode will be better for a better view.
- Support for printer (dot-matrix / deskjet / inkjet / laser) is required. This implies that appropriate drivers should be installed and printer device should be connected for printing of reports and marksheets.
- A network connection (internet / intranet) is required to make the web service accessible on other systems connected over the network.
- Other hardware interface specifications are as follows

Table 2.1 – Client Side Hardware Interfaces

HARDWARE INTERFACES - CLIENT SIDE (Minimum)	
HARDWARE	**SPECIFICATIONS**
Intel Pentium 4 and Higher Processor	
QWERTY Keyboard (U.S. Design)	
USB 2.0 / PS2 Mouse	

Table 2.2 – Server Side Hardware Interfaces

HARDWARE INTERFACES - SERVER SIDE (Minimum)		
HARDWARE	**RAM**	**DISK SPACE**
Intel Core i3 / i5 / i7 2.27 GHz and higher Or AMD 4XXX and higher	2048 MB	20 GB

Table 2.3 – Client Side Hardware Interfaces

HARDWARE INTERFACES - CLIENT SIDE (Recommended)		
HARDWARE	**RAM**	**DISK SPACE**
Intel Core i3 / i5 / i7 2.27 GHz and higher Or AMD 4XXX and higher	1024 MB	1 GB
QWERTY Keyboard (U.S. Design)		
USB 2.0 Optical Mouse		

Table 2.4 – Server Side Hardware Interfaces

HARDWARE INTERFACES - SERVER SIDE (Recommended)		
HARDWARE	**RAM**	**DISK SPACE**
Intel Xeon higher Or AMD equivalent	4096 MB	40 GB

2.1.1.4 Software Interfaces

- Any Microsoft Windows 7 and higher (Windows 7 / 8 / 8.1 / 10) or equivalent Linux based operating system with minimum kernel support 3.X.
- Crystal Reports 8 for generation and viewing of reports

Table 2.5 – Minimum Software Interfaces

SOFTWARE INTERFACES (Minimum)		
Software Tool	**Version**	**Purpose of Use**
Operating system	Windows 7 and higher or Linux with Kernel 3.x and higher	Installation and operational platform
Web Browser	Firefox web browser	Access to the web application
Web Server	Django	Running the web application over internet / intranet
Database	SQLite3	Running and linking the database over internet / intranet to the online web application

Table 2.6 – Recommended Software Interfaces

SOFTWARE INTERFACES (Recommended)		
Software Tool	**Version**	**Purpose of Use**
Operating system	Windows 8 & higher or Linux with Kernel 4.x & higher	Installation and operational platform
Web Browser	Firefox web browser compatible	Access to the web application
Web Server	Django	Running the web application over internet / intranet
Database	SQLite3	Running and linking the DB on internet / intranet

2.1.1.5 Communiation Interfaces

- Client (customer) on Internet will be using HTTP/HTTPS protocol.
- Client (system user) on Internet will be using HTTP/HTTPS protocol.

2.1.1.6 Memory Constraints

- At least 256 MB of RAM and 2 GB of space on hard disk will be required for running the application on client end.
- Similarily, a minimum of 2048 MB of RAM and 20 GB of space on hard disk will be required for running the application on server end.

2.1.1.7 Operations

- This product release will not cover any automated housekeeping aspects of the database.
- The DBA at the client side will be assumed responsible for manually deleting or archieving obsolete or non-required data from the database as per clients requirements.
- This will include database backup and recovery options also.
- The Django webserver will hosted and maintained on a remote server addressed by a URL based address.
- The URL address may be intranet or internet based as per clients requirements.
- The 'SYSTEM RESET' function is provided that after confirmation from the administrator, will delete all the selective or complete data from the system.

2.1.1.8 Site Adaption Requirements

The computing terminals conneted to network (internet / intranet) at the client end will be required to support the hardware and software interfaces specified in above sections.

2.1.2 Project Functions

The applicant/ user will be allowed to register and login to our website, so that they can view the insights and forecasts generated and request for new insighs, edit and delete etc. The visualization will automatically be created when data is supplied and other parameters are defined.

A summary of the major functions that the software will perform:

a. A Login facility for enabling only authorized access to the system.

b. Register facility for enabling users to register themselves by entering following credentials: Full Name, Post, Department, Email, Mobile Number, Age, Aadhaar Number, work ID, Password.

c. They can view insights and forecasts after the successful login.

d. Get the notifications regarding new insights.

e. Get the notifications regarding queries and troubleshoot.

2.1.3 User Characteristics

- **User** – A person who is authorized to access the website and belongs to a government authority.
- **Administrator** – He/ She has the priviledges to add or remove a user, manage website content and address any queries.

2.1.4 Constraints

- The applicants require a computer to submit their information.
- Although the security is given high importance, there is always a chance of intrusion in the web world which requires constant monitoring.
- The user has to be careful while submitting the information. Much care is required.
- Minimum hardware configuration must be present.
- Internet Connection should be available every time.
- ID proof is required to sign up and login.

2.1.5 Assumptions & Dependencies

- The user should provide all the valid details.

2.2 Specific Requirements

This section presents the software requirements to a level of detail sufficiency to enable designers to design and testers to test the system.

2.2.1 User Interface Requirements

Following screens will be provided by the system:

Login Screen:

This will allow the user/ applicant to enter into his dashboard. User need to enter 2 things.

- User ID: The user id of the applicant, he used at the time of registeration.
- Password: The password of the applicant, he used at the time of registeration.

Apply for User Registration:

User can apply for driving test by submitting the below informations.

- Full Name: User full name as per Aadhaar Card.
- Designation: His position in the office for clearance.
- Email: User's email id.
- Mobile No.: User's mobile number.
- DOB: As mentioned in the ID.
- ID : A valid id given by authorized government agency.
- Password: User's Password.

2.2.2 System Product Features

2.2.2.1 Security

The application will be password protected. Users will have to enter correct username, password and role in order to access the application modules allowed to their privilage.

2.2.2.2 Maintainability

The application will be designed in a manner to make it easy to incorporate new requirements in individual modules such as source management, user info, insights and forecasts, report generation and user accounts activation / deactivation.

2.2.2.3 Portabilty

The application will be easily portable among any windows or linux based systems that have Django, Python, SQLite database installed.

Chapter-3 SYSTEM DESIGN SPECIFICATION

3.1 System Architecture

System architecture presents the schematic view of the complete system along with its major components and their connectivities. The overall architecture of the proposed system will be as follows.

3.2 Module Decomposition

The proposed system can be decomposed into following major modules :

1. Website: The user need to register on the website to see insights and forecasts provided admin approves the request.
2. Data Import, cleansing, transformation: Data will be imported, cleaned and made fit visualizing.
3. Gathering insights and forecasts: Using Data Science and Machine Learning techniques visualizations will be created and forecasts for the future events will be generated.

Module - 1: The user first need to register or login to the panel by filling up the required credentials. After that he can see insights and forecasts and make other requests.

Module - 2: In this module data will be imported from central database and then Data Science techniques will be used to clean and transform the data according to the need of models.

Module - 3: In this module we will generate visualizations and gather deep insights from the data also machine learning will be used to forecast future criminal events.

3.3 High Level Design Diagrams

3.3.1 Usecase Diagrams

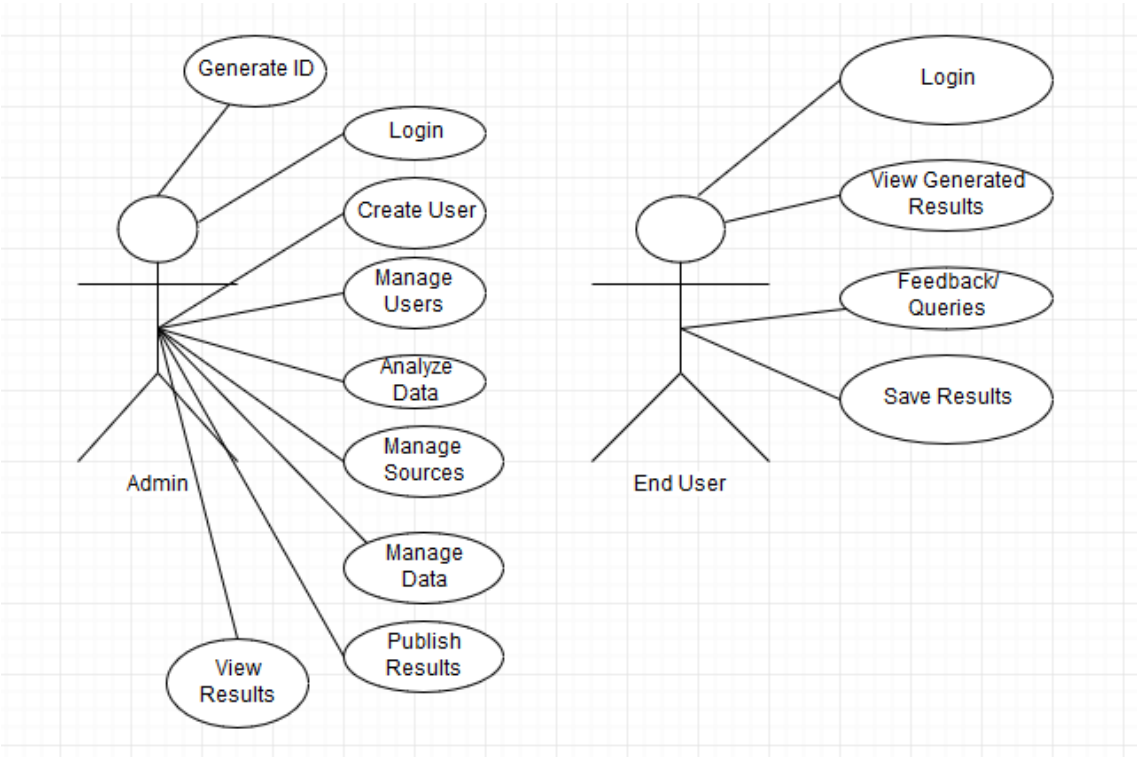

Figure 3.1 – Usecase Diagram

3.3.2 Sequence Diagram

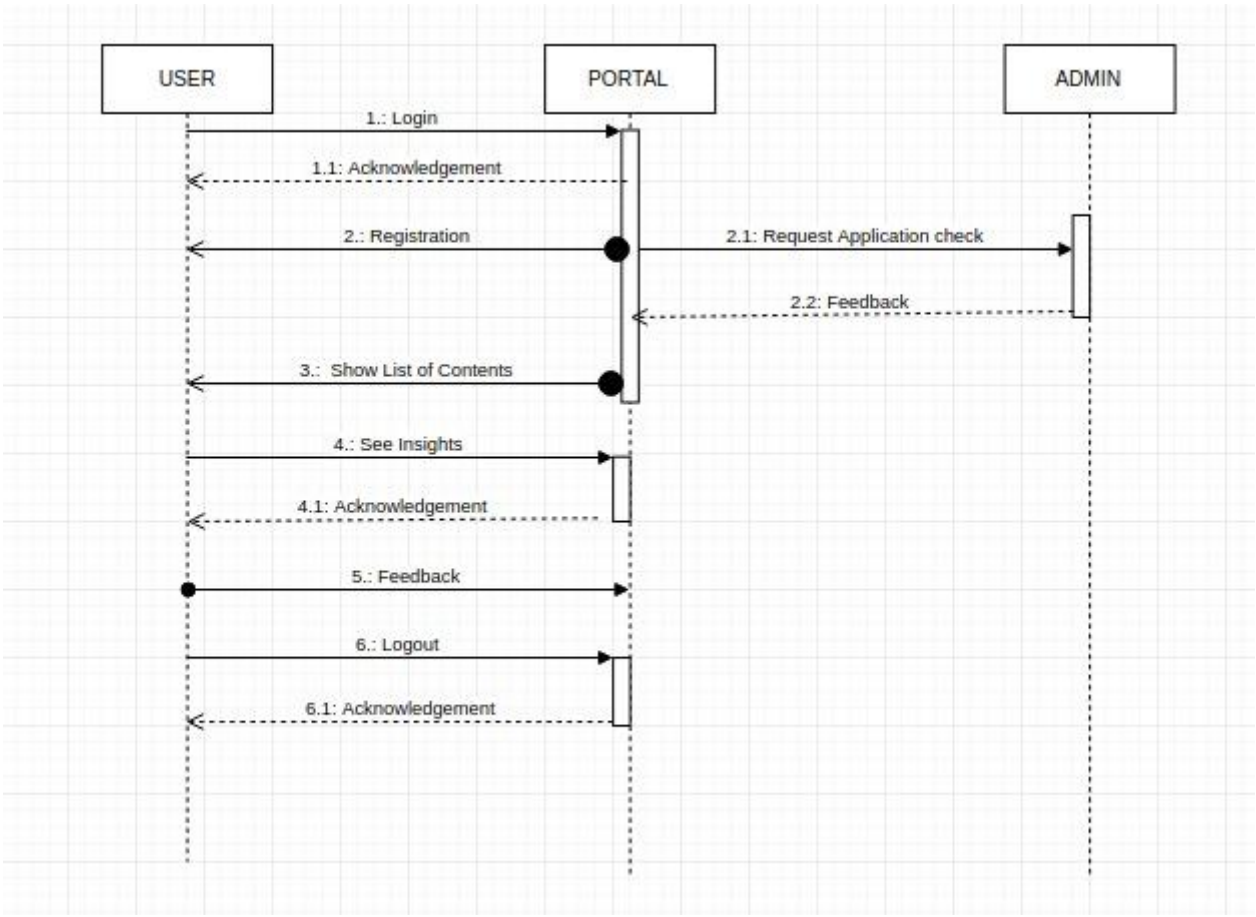

Figure 3.3 – Sequence Diagram

3.3.3 Class Diagram

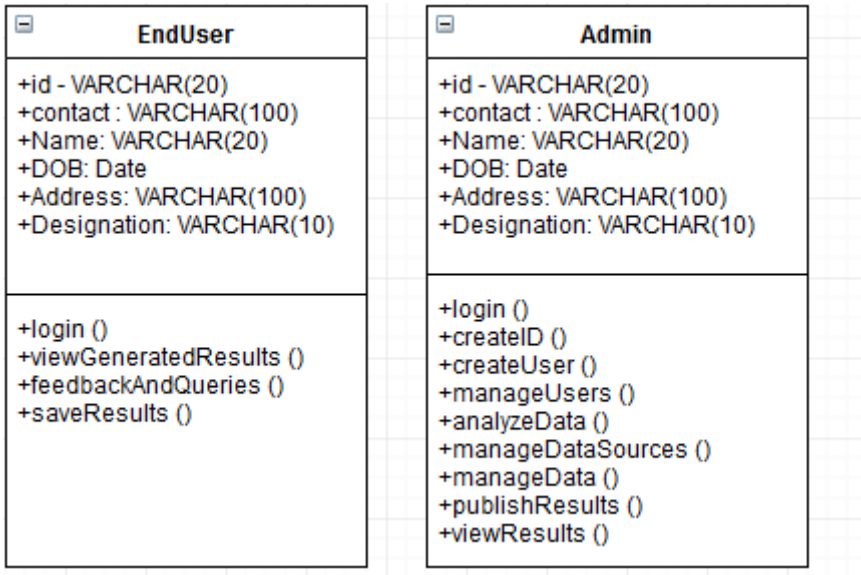

Figure 3.6 – Class Diagram

3.3.4 Entity Relationship Diagram

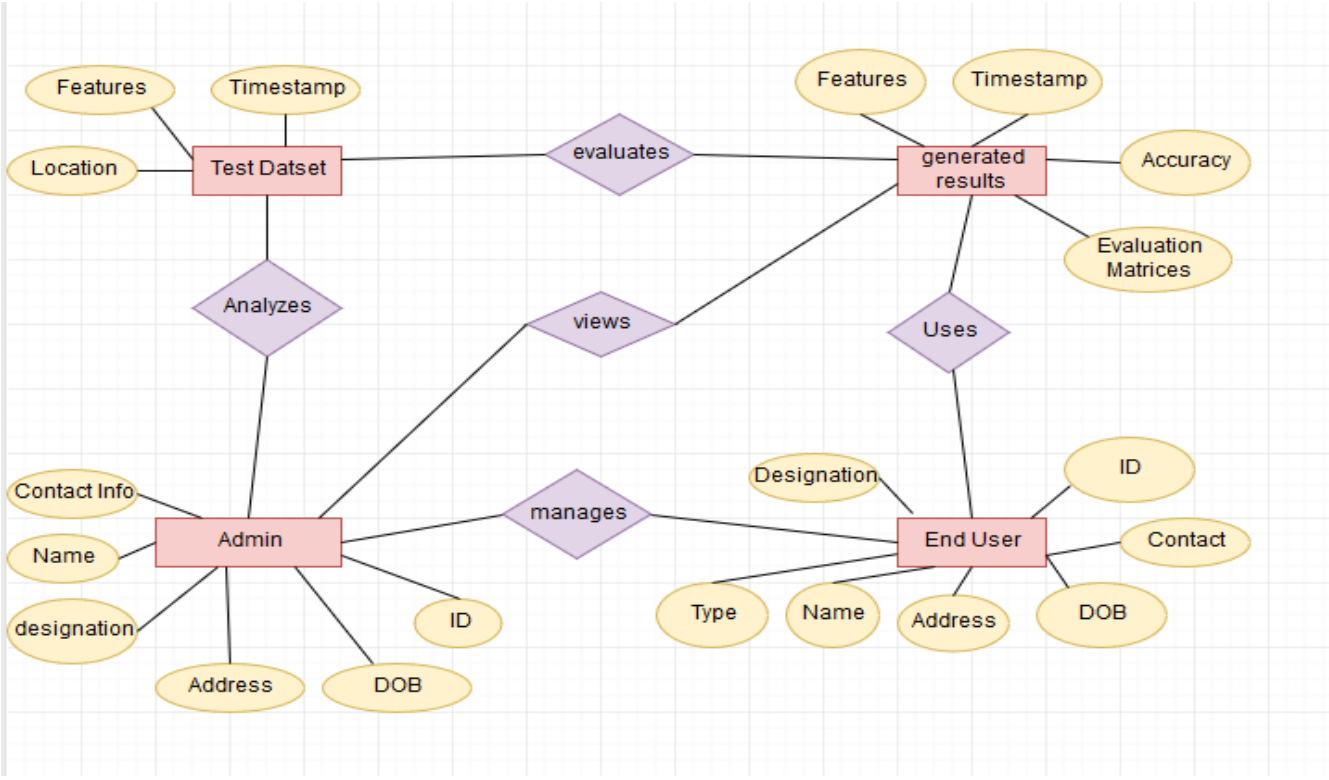

Figure 3.7 – Entity Relationship Diagram

Chapter–4 METHODOLOGY & TEAM

4.1 Introduction to CRISP-DM Methodology:

CRISP-DM stands for cross-industry process for data mining. The CRISP-DM methodology provides a structured approach to planning a data mining project. It is a robust and well proven methodology. Its very powerful, flexible and its very useful in solving thorny business problem using data science, data analytics, machine learning.

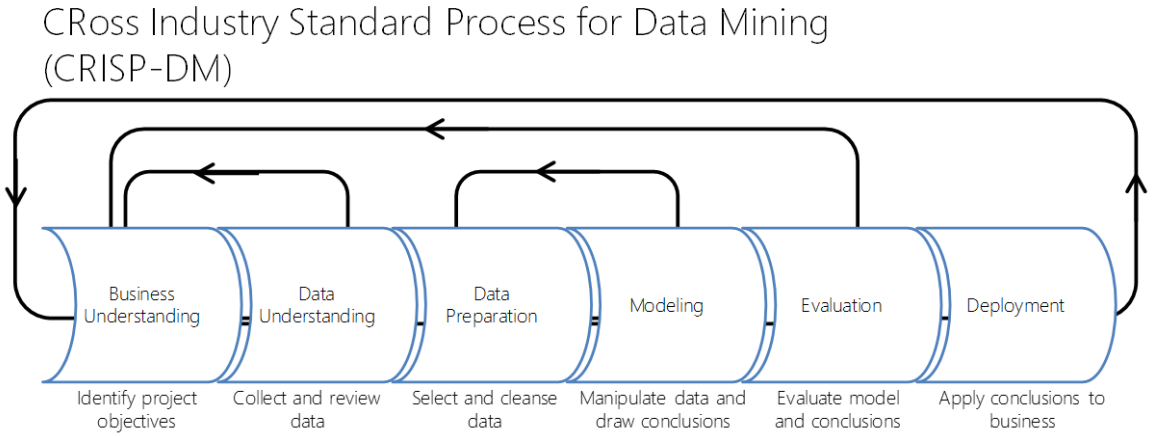

Figure 4.1 – CRISP-DM Methodology

This model is idealized sequence of events. In practice many of the tasks can be performed in a different order and it will often be necessary to backtrack to previous tasks and repeat certain actions. The model does not try to capture all the possible routes through the data mining process.

The phases for CRISP-DM method are:

1. **Business Understanding** – Understand what you accomplish from a business perspective.
2. **Data Understanding** – Acquire the data listed in project resources.
3. **Data Preparation** – Data Cleaning, Transformation, Slicing and Dicing.
4. **Modeling** – Generate visualizations, gather insights and apply models.
5. **Evaluation** – Quality assessment of models.
6. **Deployment** – Deployment of model to predict some actual business data.

CRISP-DM Model Pros & Cons

Advantage

- Responding to change
- Accepting uncertainty
- Faster review cycles
- Less up-front work

Disadvantage

- Lack of understanding
- Flexibility can lead to bad behaviors
- Culture fit
- Lack of predictability
- Challenges at scale

4.2 Team Members, Roles & Responsibilities

Table 4.1 – Roles and responsibilities

Team Member	Project Role	Responsibilities
Ishan Kalra	**Data Analysis**	**Analysing Data**
Chandra Prakash	**Web Developer**	**Backend**
Bhumika Saxena	**Web Design**	**Frontend**

UNIT– 5 SYSTEM TESTING

The designed system has been testing through following test parameters.

5.1 Functinality Testing

In testing the functionality of the web sites the following features were tested:

- **Links**

a) Internal Links:

All internal links of the website were checked by clicking each link individually and providing the appropriate input to reach the other links within.

b) External Links

Till now no external links are provided in our website but in future we do plan to host links of other government agencies websites.

c) Mail Links

No mail links are provided in our website till this stage but this is also a future enhancement of our website to trigger mails to people for keeping them updated about the latest news letters, updates, insights and forecasts.

d) Broken Links

Broken link are those links which so not divert the page to specified page or any page at all. By testing the links on our website there was no link found on clicking which we did not find any page.

- **Forms**

a) Field validation

Check on Posts of the user is applied: If a user does not have required for clearance then he / she will not be able to log in to website.

b) Error message for wrong input

Error messages have been displayed as and when we enter the wrong details (eg. Dates), and when we do not enter any detail in the mandatory fields. For example: when we enter wrong password we get error message for acknowledging us that we have entered

it wrong and when we do not enter the username and/or password we get the messages displaying the respective errors.

c) Optional and Mandatory fields

All the mandatory fields have been marked with a red asterisk (*) and apart from that there is a display of error messages when we do not enter the mandatory fields. For

- **Database**

Testing is done on the database connectivity.

In the database testing we included following cases:

a) Entries in database through frontend and checked for the same in the tables.
b) Checked for the data types.
c) Checked for range of each type of data.
d) If entries made in one table are affecting other tables then we have checked those entries also.
e) We not just added details in the database by default but also did the same using the frontend.

5.2 Performance Testing

Performance testing can be applied to understand the website's scalability, or to benchmark the performance in the environment of third party products such as servers and middleware for potential purchase. This can only be done once it is put into use on the actual internet server and tested by the users.

Till now it is done using the null modem on two systems.

The system load includes:

a) What is the number of users per time?
b) Checking for peak loads and how system behaves.
c) Amount of data accessed by user.

This is done using only 2 systems for now so cannot be tested for load unless we deploy it on a real server machine.

5.3 Usability Testing

Usability testing is the process by which the human-computer interaction characteristics of a system are measured, and weaknesses are identified for correction.

a) Ease of learning
b) Navigation
c) Subjective user satisfaction
d) General appearance

As system is not put into the real time use so it's not yet tested for usability.

5.4 Server Side Interfacing

In this we tested the server side interface. This was done by verifying that communication is done properly. Also the compatibility of server with software, hardware, network and database was tested.

5.5 Client Side Compatibility

The client side compatibility is also tested using various browsers like Google Chrome, Mozilla Firefox and Internet Explorer.

Chapter-6 TEST EXECUTION SUMMARY

Execution Test Summary Report is an overall view of Testing Process from start to end. Test Plan comes at the starting of project while Test Summary Report comes at the end of testing process. This report is given to the client for his understanding purpose.

The Test Summary Report contents are:

1. Test Case ID generated =
2. Total number of resources consumed =
3. Passed Test Cases =
4. Failed Test Cases =
5. Status of Test Cases =

Table 6.1 – Test Case Summary

S.No	Test Case ID	Test Case Description	Expected Outcome	Test Case Status	No. of Resources Consumed
1.	PR01	Data Visualization	OK	PASS	
2.	PR02	Data Visualization	OK	PASS	
3.	PR03	Web Testing	OK	PASS	
4.	PR04	Web Testing	OK	PASS	
5.	PR05	Web Testing	OK	PASS	
6.	PR06	Data Visualization	OK	PASS	
7.	PR07	Data Visualization	OK	PASS	
8.	PR08	Data Visualization	OK	PASS	
9.	PR09	Data Visualization	OK	PASS	
10.	PR10	Web Testing	OK	PASS	

Chapter-7 PROJECT SCREENSHOTS

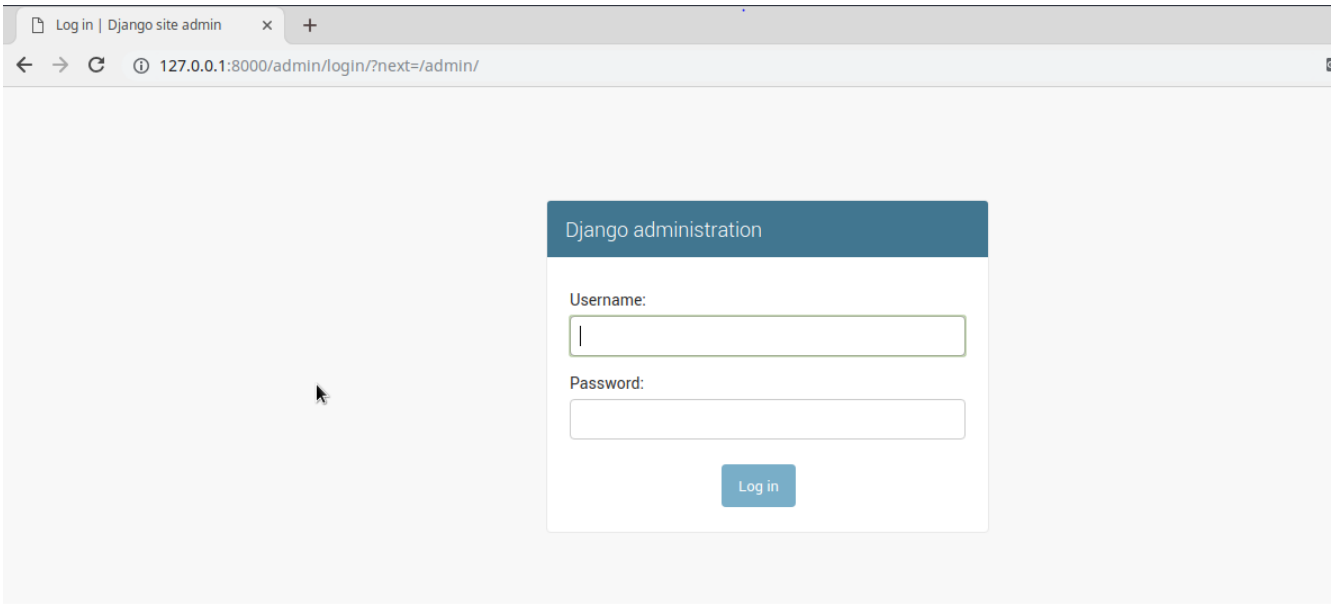

Figure 7.1 – Admin Login

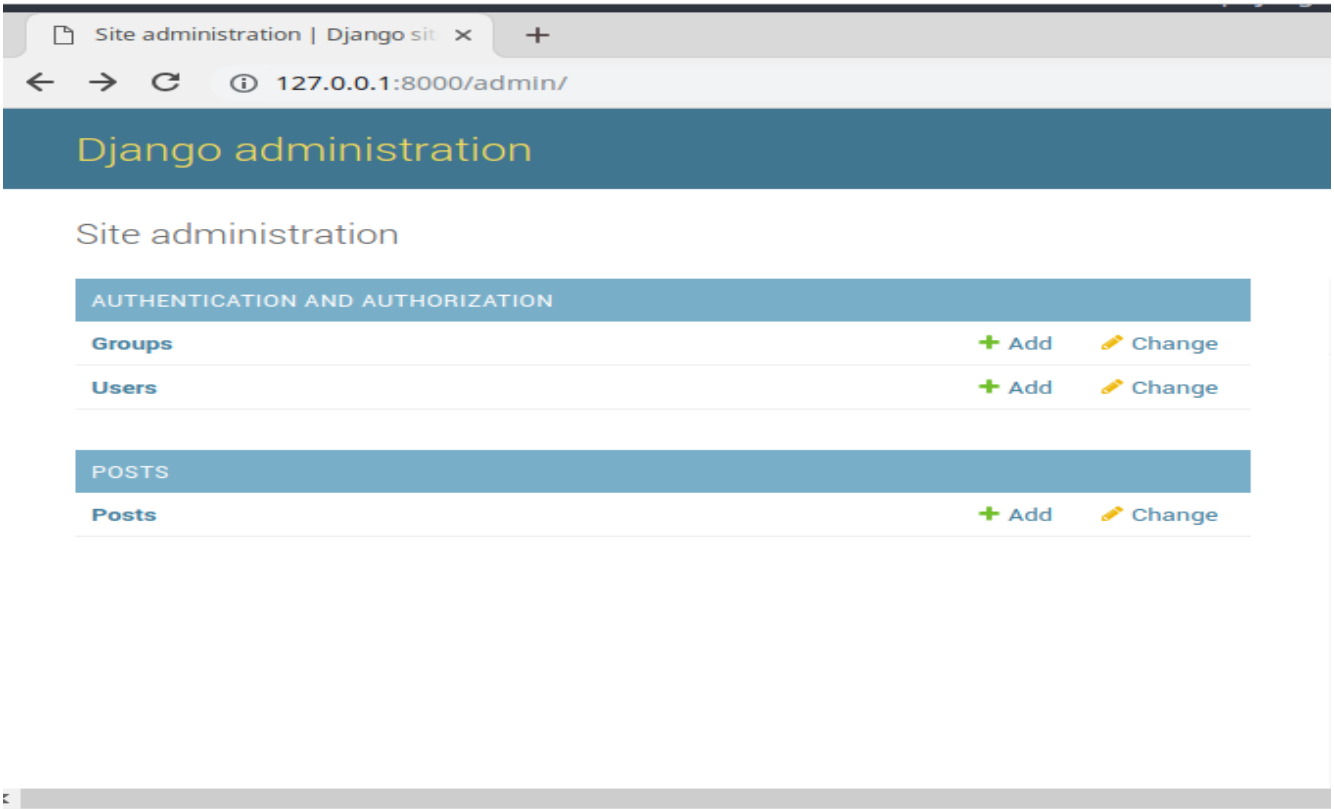

Figure 7.2 – Admin Control Panel

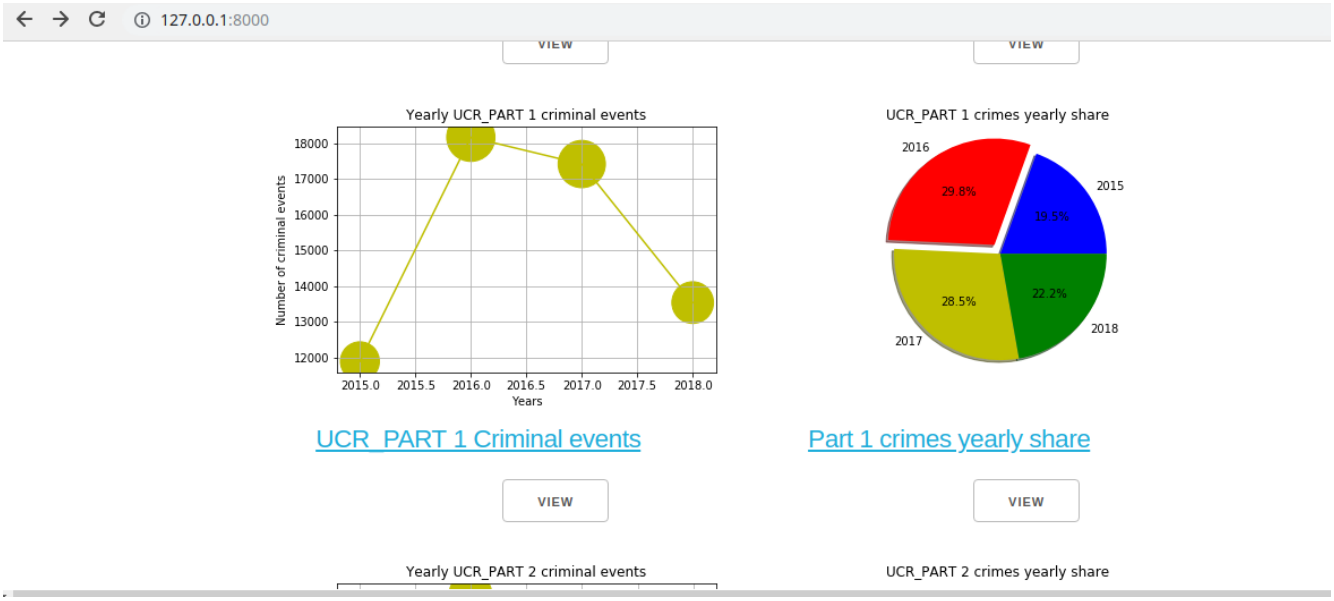

Figure 7.3 – Home Page

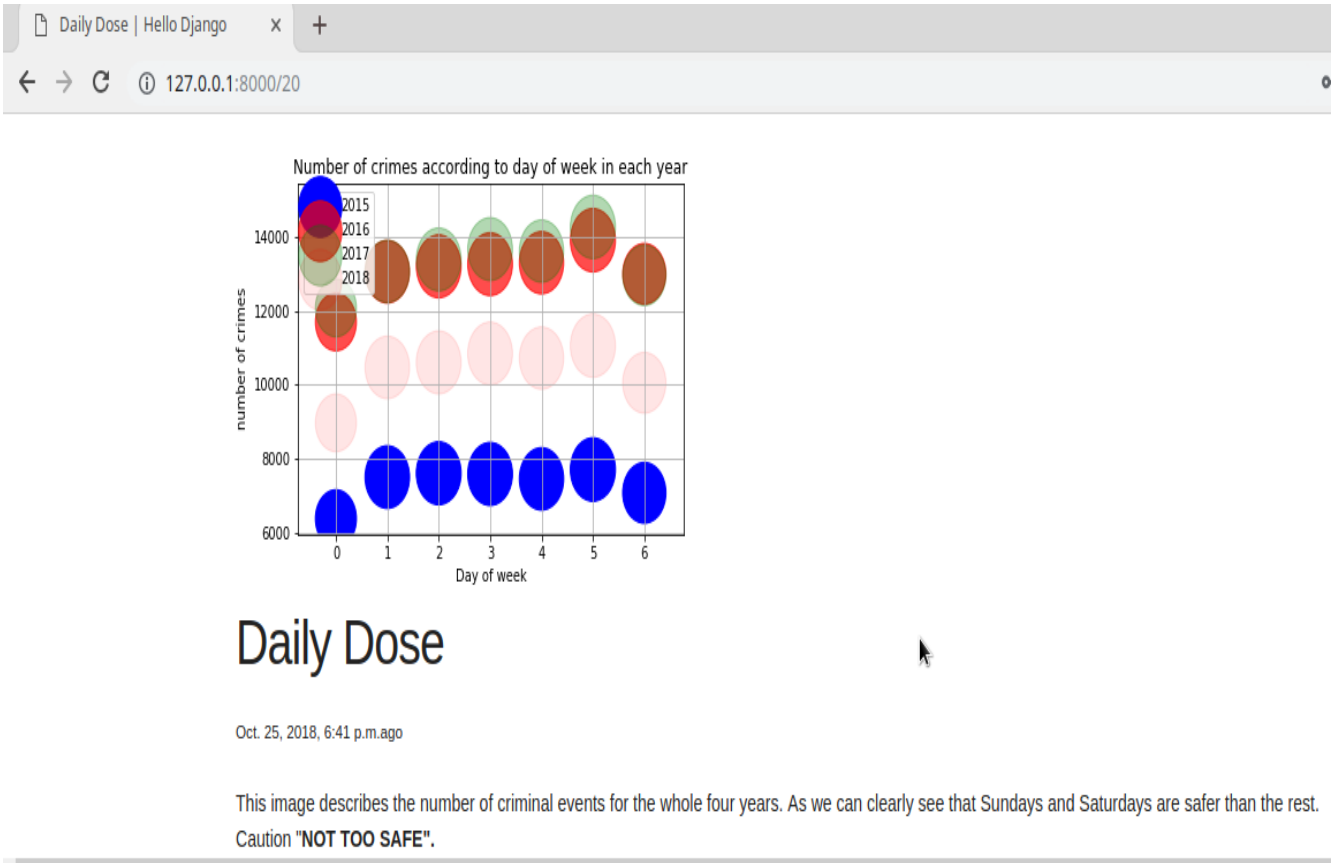

Figure 7.4 – View an insight

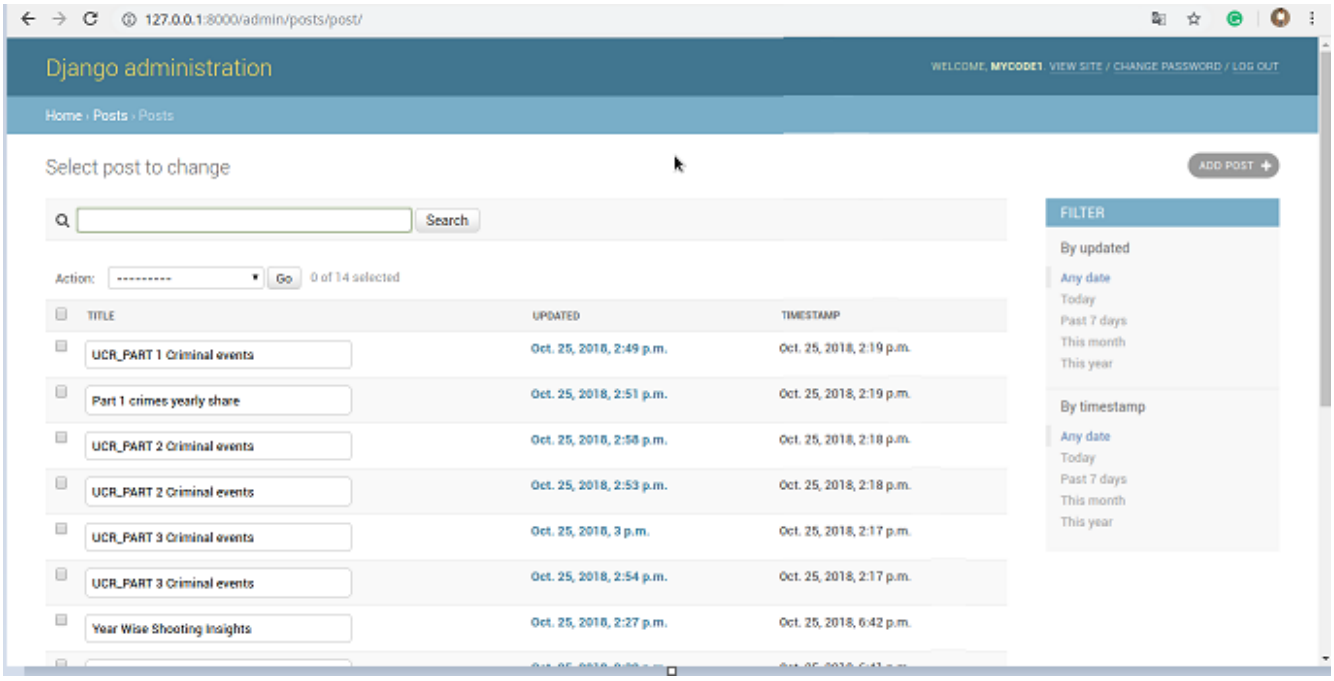

Figure 7.5 – Manage Posts

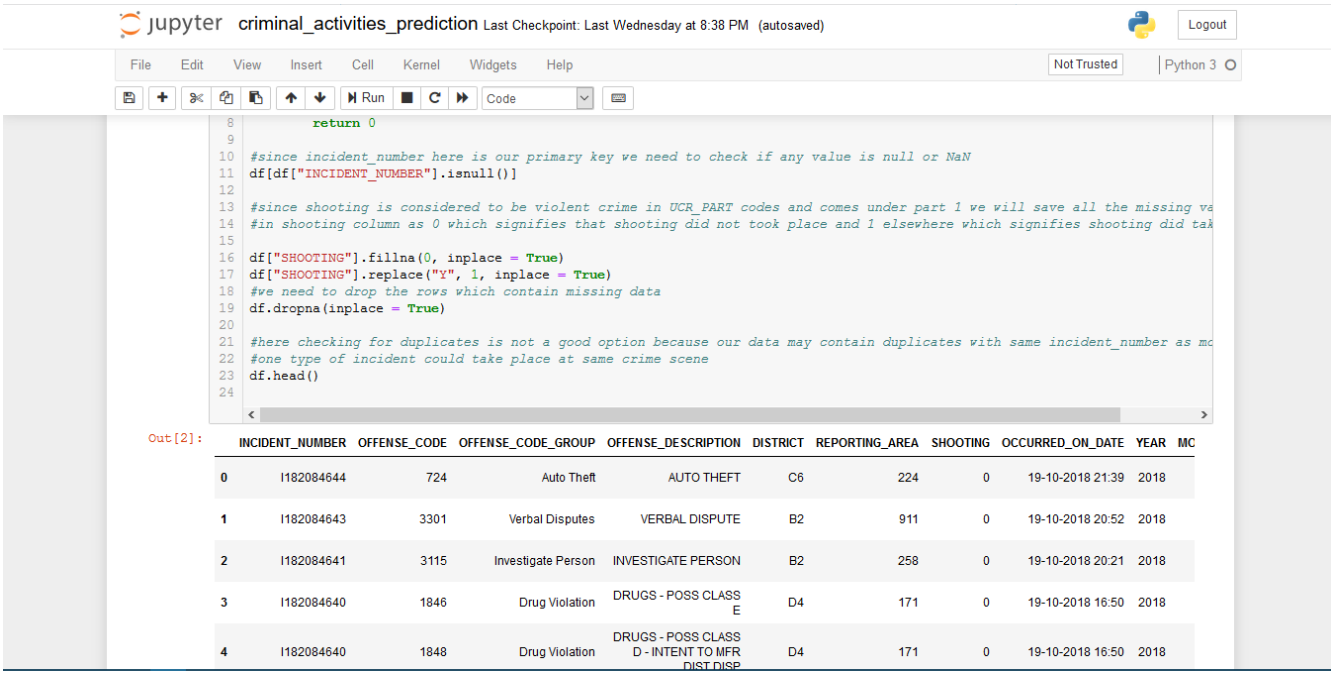

Fig 7.6 – Jupyter notebook home

	INCIDENT_NUMBER	OFFENSE_CODE	OFFENSE_CODE_GROUP	OFFENSE_DESCRIPTION	DISTRICT	REPORTING_AREA	SHOOTING	OCCURRED_ON_DATE	YEAR	MC
0	I182084644	724	Auto Theft	AUTO THEFT	C6	224	NaN	19-10-2018 21:39	2018	
1	I182084643	3301	Verbal Disputes	VERBAL DISPUTE	B2	911	NaN	19-10-2018 20:52	2018	
2	I182084641	3115	Investigate Person	INVESTIGATE PERSON	B2	258	NaN	19-10-2018 20:21	2018	
3	I182084640	1846	Drug Violation	DRUGS - POSS CLASS E	D4	171	NaN	19-10-2018 16:50	2018	
4	I182084640	1848	Drug Violation	DRUGS - POSS CLASS D - INTENT TO MFR DIST DISP	D4	171	NaN	19-10-2018 16:50	2018	

fig 7.7 – Input dataset

```
MONTH
1       22213
2       20089
3       22304
4       22446
5       24280
6       28557
7       32151
8       32528
9       31674
10      28275
11      22105
12      22056
Name:  INCIDENT_NUMBER,  dtype:  int64
```

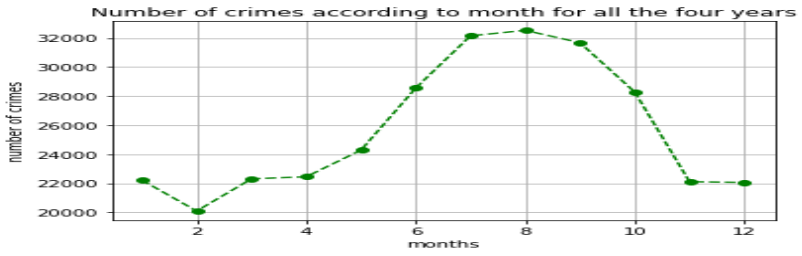

Fig 7.7 – Gathered Insight

Chapter-8 PROJECT SUMMARY AND CONCLUSIONS

The object of this project is to harness the power of Machine Learning, Data Science for our practical and potential one. This report explains to extensively cover this concept and plant a seed of inquisitiveness in the mind of users.

Through the use of this project authorities will be able to plan and deploy their resources effectively and it will help in reduction of crime rates.

Our Project is currently on track with 2 Stages of the project are completed and rest are to be completed in near future.

Further research into the project is required to enhance predictive features of the project.

Chapter-9 FUTURE SCOPE

The possible future scope of this project are:

1. Designing of mobile application for this project on Android / iPhone platforms.

2. Generalization of this project for other cities.

UNIT– 10 REFERENCES

BOOKS:

1. Python For Data Analysis – Wes McKinney

2. Sebastian Raschka, Python Machine Learning

WEBSITES:

1. http://www.stackoverflow.com/

2. http://developers.google.com/

3. https://djangoproject.com

4. https://getbootstrap.com/docs/4.1/getting-started/introduction/

5. https://www.python.org/doc/