A PROJECT REPORT ON

PREDICTIVE ANALYTICS & VISUALIZATIONS OF CRIME PATTERNS IN INDIA

SUBMITTED TO THE SAVITRIBAI PHULE PUNE UNIVERSITY, PUNE
IN THE PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE AWARD OF THE DEGREE

OF

BACHELOR OF ENGINEERING (COMPUTER ENGINEERING)

SUBMITTED BY

SHIVAM BHIRUD   **Exam No : B120364246**
SHEKHAR PATIL   **Exam No : B120364461**
SHRAVANI SWAMI   **Exam No : B120364537**
JYOTI SHARMA   **Exam No : B120364352**

Under The Guidance of

Prof. S.V. Dabhade

Sinhgad Institutes

DEPARTMENT OF COMPUTER ENGINEERING

**STES'S SMT. KASHIBAI NAVALE COLLEGE OF ENGINEERING
VADGAON BK, OFF SINHGAD ROAD, PUNE-411041**

SAVITRIBAI PHULE PUNE UNIVERSITY 2017 -2018

Sinhgad Institutes

CERTIFICATE

This is to certify that the Project Entitled

" PREDICTIVE ANALYTICS & VISUALIZATIONS OF CRIME PATTERNS IN INDIA"

Submitted by

SHIVAM BHIRUD	**Exam No : B120364246**
SHEKHAR PATIL	**Exam No : B120364461**
SHRAVANI SWAMI	**Exam No : B120364537**
JYOTI SHARMA	**Exam No : B120364352**

is a bonafide work carried out by them under the supervision of **Prof. S.V. Dabhade** and it is approved for the partial fulfillment of the requirement of Savitribai Phule Pune university, Pune for the award of the degree of Bachelor of Engineering (Computer Engineering).

Prof. S. V. Dabhade
Guide
Dept. of Computer Engg.

Dr. P. N. Mahalle
H.O.D
Dept. of Computer Engg.

Dr. A. V. Deshpande

Principal

Smt.Kashibai Navale College of Engineering

Signature of Internal Examiner

Signature of External Examiner

PROJECT APPROVAL SHEET

A Project Report Titled as

Predictive Analytics and Visualizations of Crime Patterns in India

Is verified for its originality in documentation, problem statement, proposed work
and implementation successfully completed by

SHIVAM BHIRUD	Exam No : B120364246
SHEKHAR PATIL	Exam No : B120364461
SHRAVANI SWAMI	Exam No : B120364537
JYOTI SHARMA	Exam No : B120364352

at

DEPARTMENT OF COMPUTER ENGINEERING

STES' SMT KASHIBAI NAVALE COLLEGE OF ENGINEERING

SAVITRIBAI PHULE PUNE UNIVERSITY,PUNE

ACADEMIC YEAR 2017-2018

Prof. S. V. Dabhade
Guide
Dept. of Computer Engg.

Dr. P. N. Mahalle
H.O.D
Dept. of Computer Engg.

Abstract

In India, there has been a drastic pattern of crime observed in the past few years giving rise to a threat to the security of common man. Considering this perilous situation, we aim to study these crime patterns and realise the changes in the overall crime based on the data obtained from the official Indian Government websites. The raw data obtained was converted to a suitable format using data mining techniques such as eliminating missing values, eliminating redundant data, data transformation, etc. This data was fed to algorithms like Linear Regression, Random Forest for performing predictions. Crime type predictions are performed, for four years, for each state as well as all the states of India using the data from year 2001-2016. These predictions are displayed using simple visualization charts. One important aspect that is used with these algorithms is that of identifying the trend changing year in order to increase the accuracy of the predictions.

Another aspect of the proposed analytics system includes the visualization of crime patterns over the past years till 2015. These visualizations are performed with the help of demographic data available at the Census India website. Each state in the visualizations are distinguished using factors like literacy and area in order to demonstrate the crime per one lakh people. While performing all these predictions and visualizations, we came across significant observations regarding the algorithms used, or certain crime types and the data related to it which has also been incorporated in one of our modules.

The entire system is established through a website - "Indian Crime Analysis". The website provides three main services - predictions, visualizations and observations. Under the predictions service, four major sectors of crime is considered - crime against women, crime against children, crimes under the Indian Penal Code (IPC) and crimes under Special and Local Laws (SLL). The visualizations service consists of the total crime against women, total crimes against children and the total IPC crimes with respect to population of an area. The observations service provides a headline of all the notable statistics and information perceived during execution of the system. This website provides a user friendly environment for all its users to take the entire benefit of studying the crime patterns across all the states of India for diverse purposes.

Acknowledgments

It gives us great pleasure in presenting the preliminary project report on

`PREDICTIVE ANALYTICS & VISUALIZATIONS OF CRIME PATTERNS IN INDIA'.

We would like to take this opportunity to thank our internal guide Prof. S. V. Dabhade for giving us all the help and guidance we needed. We are very grateful to her for her kind support. Her valuable suggestions were very helpful.

We are also grateful to Dr. P. N. Mahalle, Head of Computer Engineering Department, STES' Smt. Kashibai Navale College of Engineering for his indispensable support, suggestions.

<div align="right">

Shivam Bhirud

Shekhar Patil

Shravani Swami

Jyoti Sharma

(B.E. Computer Engg.)

</div>

Contents

List of Figures

List of Tables

CHAPTER 1

SYNOPSIS

1.1   Project Title

Predictive Analytics and Visualizations of Crime Patterns in India

1.2   Project Type

Internal project

1.3   Internal Guide

Prof. S. V. Dabhade

1.4   Technical Keywords

1) SYSTEM ORGANIZATION
   a) Predictions
   b) Visualizations
   c) Significant Observations
2) DISTRIBUTED SYSTEMS
   a) Client/Server
   b) Distributed databases
   c) Security and reliability issues in distributed applications
3) ALGORITHMS
   a) Linear Regression Algorithm
   b) Random Forest Algorithm

1.5 Problem Statement

Visualization of crime data in order to predict the rate of crime and its severity
in the forthcoming years, in different areas of the country based on the previous year
criminal records of India.

1.6 Abstract

In India, there has been a drastic pattern of crime observed in the past few years
giving rise to a threat to the security of common man. Considering this perilous situation,
we aim to study these crime patterns and realise the changes in the overall crime based on
the data obtained from the official Indian Government websites. The raw data obtained was
converted to a suitable format using data mining techniques such as eliminating missing
values, eliminating redundant data, data transformation, etc. This data was fed to
algorithms like Linear Regression, Random Forest for performing predictions. Crime type
predictions are performed, for four years, for each state as well as all the states of India
using the data from year 2001-2016. These predictions are displayed using simple
visualization charts. One important aspect that is used with these algorithms is that of
identifying the trend changing year in order to increase the accuracy of the predictions.

Another aspect of the proposed analytics system includes the visualization of crime
patterns over the past years till 2015. These visualizations are performed with the help of
demographic data available at the Census India website. Each state in the visualizations are
distinguished using factors like literacy and area in order to demonstrate the crime per one
lakh people. While performing all these predictions and visualizations, we came across
significant observations regarding the algorithms used, or certain crime types and the data
related to it which has also been incorporated in one of our modules.

The entire system is established through a website - "Indian Crime Analysis". The website
provides three main services - predictions, visualizations and observations. Under the
predictions service, four major sectors of crime is considered - crime against women, crime
against children, crimes under the Indian Penal Code (IPC) and crimes under Special and
Local Laws (SLL). The visualizations service consists of the total crime against women,
total crimes against children and the total IPC crimes with respect to population of an area.
The observations service provides a headline of all the notable statistics and information
perceived during execution of the system. This website provides a user friendly
environment for all its users to take the entire benefit of studying the crime patterns across
all the states of India for diverse purposes.

1.7 Goals and Objectives

This system is proposed for increasing transparency regarding the crime in the
society and to help police officials, Government of India and the rest concerned
authorities to take measures for minimizing the overall crime in our country.

We aim to study these crime patterns and realise the changes in the overall crime ratio based on the demographic or population ratios which shall ease the process of reaching the future crime-cases prediction. Along with this, a prediction can be made on the maximum criminal activities that need to be controlled and the amount of security measures that need to be taken to decelerate or curb the occurrence of any mild to treacherous level criminal activities.

The purpose is to help ascertain incidence and patterns of the crimes using data mining techniques and data visualization.

1.8   Relevant mathematics associated with the Project

System Description: In this system we are predicting and visualizing the various crime patterns of India. It includes a comparative analysis of various states of India based on the number of crime cases reported.

Input: Year for prediction, State name, Crime Type.

Output: Predictions of crime upto the selected year, Comparative Analysis of state based on crime using animated graph, Illustrative visualizations of the total crime per state for each sector.

Data Structure and Dataframes: Pandas dataframe, list, numpy array

Functions :

Objects
Morphisms
Overloading : Linear Regression
Functional relations

Mathematical formulation:

Linear Regression:

Success Conditions:

The success conditions of this system includes optimally accurate predictions of the desired parameters.

Failure Conditions:

The failure conditions mainly depend on the availability and consistency of the data.

1.9 Names of Conferences / Journals where papers published

"A Survey on Data Analytics, Predictions and Visualization of Crime Patterns in India",
International Journal of Innovative Research in Computer and Electronics

Engineering,

Vol.5, Issue 9, September 2017, ISSN(Online)-2320-9801.

1.10 Review of Conference/Journal Papers sup-porting Project idea

Acknowledging the fact that there is a tremendous increase in the rate of occurrence of crime, we studied various papers which enlighten us on different approaches towards crime analysis and predictions.

In [1], different methodologies for fetching raw data (criminal records) that needs preprocessing is suggested. Data classification is done by grouping of records based on states/cities, criminal profiles, etc. The explanation of the k-means algorithm for clustering the categories of crime that possess similar characteristics is provided.It is also stated that crime correlation can be performed using Pearson's correlation coefficient and crime predictions can be done with the help of linear regression.

In [2], a detailed study, analysis and distribution of numerous papers/journals published on crime data mining is performed. This paper also provides the classification of the research papers based on the techniques used; technologies used; challenges and issues addressed. An elaboration on the sources of crime data and the types of crimes is also included in this paper. Thus, it gives us a brief information of the study done in crime data mining till date.

In [3], preprocessing techniques mainly focusing on handling the missing values for state and the intensity of crime occurred by using traditional KNN with a new distance metric and LVQ (Learning Vector Quantization) is explained. Clustering using k-means and DBScan (Density-Based Spatial Clustering Application with Noise) algorithms are suggested. Prediction of crime trend can be done using C4.5 decision tree algorithm.

In [4], emphasis is laid on the computer aided crime analysis tools in India which include-

- Crime and Criminal Information System (CCIS) which assists in investigations and supervising officers and police planners to create strategies for crime control.

- Common Integrated Police Application (CIPA) to automate the workflow at police stations, to form CCIS and to enhance the accessibility, transparency and accountability in the working of police dept.

The proposed system in this paper is an Intelligent Crime Information and Analysis System (ICIAS) which involves a Crime Data Entry Module and a Crime Information Retrieval and Analysis Module to identify crime hotspots and crime zones.

In [5], a detailed study of the existing computer aided crime analysis software and tools available in the market is illustrated. The softwares such as ATAC, CrimeStat, RCAGIS (Regional Crime Analysis Geographic Information System), CrimeConnect, BRAINCEL etc. are available that allow data entry, manipulation, and analysis of the crime patterns using data mining, neural nets and deep learning techniques. Recently, Government of India's model named 'Government to Government (G2G)' model, also called as the Crime & Criminal Information System (CCIS) was designed to create computerized storage, analysis of crime and retrieval of criminal records but till date, in spite of collecting humongous data it fails to perform the crime analysis thus it is somewhat a 'standalone' system.

In [6], the GDP, literacy-rate, police-rate, employment-rate and various crimes such as murder, dacoity and riots and the state in India as location data has been considered to check spatial autocorrelation between various crimes and to compare various attribute clusters and its relation. Heat maps are plotted for various clusters of murder-rate, riot-rate, dacoit- rate, overall crime-rate, police-rate, population density, GDP, etc.

In [7], crime analysis is done to identify crime patterns quickly and in an efficient manner for future crime pattern detection and action. Crime information has to be stored and analysed so as to get an accurate analysis of this growing volumes of crime data. Results of data mining could potentially be used to lessen and even prevent crime for the upcoming years. Different structures are used for recording crime data. Clustering techniques like K-means and DBscan algorithms are proposed to get the crime ratio relation between the states or districts.

In [8], ideas to support police detectives in solving crimes by data mining techniques have been suggested. A study on various classification and clustering techniques like k-means and KNN, with the help of which an attempt to classify categories of crimes at various places has been covered. Finally, reports were generated in the form of geo-spatial plots which clearly depicts the categories of crime in a particular region. As stated, the limitations of this study is, data mining can only help the detectives but cannot replace them.

In [9], a model for crime and criminal data analysis is proposed using simple k-Means algorithm for clustering the data and Apriori Algorithm for data-association. This paper clearly depicts the knowledge extraction from the datasets based on discovering patterns and trends, making forecasts, relationship analysis with most probable explanations, mapping criminal networks and identifying possible suspects. Clustering is done to discover the relation between various types of crimes. Association rules mining is based on generated rules from crime dataset based on frequent occurrence of patterns in crime for taking the preventive actions. The data was collected manually from a police department in

Libya. Analysis of this data was done with the help of WEKA mining software tool and MS-Excel was used for generating the effective visuals.

1.11   Plan of Project Execution

SR. NO.	TASK NAME	START DATE	FINISH DATE	JUNE	JULY	AUG	SEPT	OCT	NOV	DEC	JAN	FEB	MAR
1	Project Scouting	17/6/2017	25/6/2017										
2	Feasibility Study	28/6/2017	07-05-2017										
3	Literature Survey	07-03-2017	07-12-2017										
4	Project Definition	15/7/2017	19/7/2017										
5	Project Interface Definition	16/8/2017	25/8/2017										
6	Requirement Gathering	28/8/2017	09-04-2017										
7	Design	14/9/2017	26/9/2017										
8	Documentation	22/9/2017	10-02-2017										
9	Phase I Implementation	10-10-2017	20/10/2017										
10	Phase I Testing	11-02-2017	11-08-2017										
11	Phase II Implementation	30/11/2017	15/2/2018										
12	Phase II Testing	18/2/2017	03-03-2018										
13	Journal Paper Publication	03-10-2018	15/3/2018										
14	Project Report Finalising	03-12-2018	20/3/2018										
15	Final Project Implementation	03-11-2018	27/3/2018										

SDLC phases (June, 2017 to March 2018)

- Requirement gathering (8 days)
- System Specification and analysis(9 days)
- System designing(high level and low level) (12 days)
- Coding and unit testing(4 months)
- Functional and nonfunctional testing(1 month)
- Documentation and Maintenance(1.5 month)

CHAPTER 2

TECHNICAL KEYWORDS

2.1   Area of Project

Data Science

2.2   Technical Keywords

1) SYSTEM ORGANIZATION
   a) Predictions
   b) Visualizations
   c) Significant Observations
2) DISTRIBUTED SYSTEMS
   a) Client/Server
   b) Distributed databases
   c) Security and reliability issues in distributed applications
3) ALGORITHMS
   a) Linear Regression Algorithm
   b) Random Forest Algorithm
4) LIBRARIES
   a) Plotly
   b) Sklearn

CHAPTER 3

INTRODUCTION

3.1 Project Idea

The main idea of the project is to help ease the policing and security policy making and implementation by providing an elaborate view of the crime patterns that could be observed in the coming years. It also includes the idea to visualize the crime case reporting of the past years in order to identify critical areas that need to be looked upon.

3.2 Motivation of the Project

The motivation for predictive analytics of crime patterns in india aroused after the Indian Government's plan to forecast crime by 2018, since no such system yet exists that would benefit the strategizing of crime control. The plan mainly includes predicting crime considering the factors - geographical areas, population, lifestyle, race, social issues etc. Keeping the government's vision in mind, we aimed to prepare a primary model towards accomplishing it. Our system covers the factors of predicting crimes for a few upcoming years covering geographical areas (states) as one of the parameters. Also, for visualizations we have covered various demographic factors like population, literacy rates with respect to area.

3.3 Literature Survey

Ref. No.	Paper Title	Description	Limitations
1.	Crime Pattern Analysis, Visualization and Prediction using Data Mining	● Data collection which fetches raw data(criminal records) that needs preprocessing. ● Data classification, grouping of records based on states/cities, criminal profiles, etc. ● K-means algorithm for clustering. ● Correlating crimes using Pearson's correlation coefficient. ● Predicting crimes using linear regression.	● Future Crime ratio is not considered. ● Lacking area wise Police efficiency predictions.

2.	Systematic Review of Crime Data Mining	● Selection, detailed study, analysis and distribution of numerous papers/journals published on crime data mining. ● Classification of the research papers based on the techniques used; technologies used; challenges and issues addressed. ● Sources of crime data and types of crimes.	● Analysis is done regarding the categories of the possible techniques that can be used, but actual elaborated study of those techniques is missing.
3.	Enhanced Algorithm to Predict a future crime using Data Mining, International Journal of Computer Applications	● Preprocessing techniques mainly focusing on handling the missing values for state and number of crime occurred by using traditional KNN with a new distance metric and LVQ thus producing a E-KDD model. ● Clustering using k-means and DBScan algorithms. ● Prediction of crime trend using C4.5 decision tree algorithm.	● This tool is developed only for police officials and not for the common public. Public should also know the fluctuations in crime rate so that the general awareness can be spread.

4.	Design and Implementation of ICIAS based on Crime Data Mining	● Computer aided crime analysis tools in India: - CCIS assists in investigations and supervising officers and police planners to create strategies for crime control. - CIPA to automatize the workflow at police stations, to form CCIS and to enhance the accessibility, transparency and accountability in the working of police dept. ● Proposed ICIAS: - Crime Data Entry Module - Crime Information Retrieval and Analysis Module to identify crime hotspots and crime zones.	● A generalized survey of every tool used in data mining has been done, but no specific technique mentioned.
5.	A Study on the Existing Computer Aided Crime Analysis Tools to Handle Indian State of Affairs	● Detailed study of the existing computer aided crime analysis software and tools available in the market. ● Soft-wares such as ATAC, CrimeStat, RCAGIS (Regional Crime Analysis Geographic Information System), CrimeConnect, BRAINCEL etc. are available that allows data entry, manipulation, and analysis of the crime pattern using data mining, neural nets and deep learning	● A generalized survey of every tool used in data mining has been done, but no specific technique mentioned.

		techniques.	
		● Government of India's 'Government to Government (G2G)' model called the CCIS, was designed to create computerized storage, analysis and retrieval of crime criminal records but till date, in spite of collecting humongous data it fails to perform the crime analysis thus it is somewhat 'standalone' system.	
6.	Spatial Patterns of Crimes in India using Data Mining Techniques	● The GDP, literacy-rate, police-rate, employment-rate and various crimes such as murder, dacoit and riots and the state in India as location data has been considered to check spatial autocorrelation between various crimes and to compare various attribute clusters and its relation. ● Heat maps are plotted for various clusters of murder-rate, riot-rate, dacoit- rate, overall crime-rate, Police-rate, population density cluster, GDP cluster etc.	● Future Crime ratio prediction is not considered.
7.	Data mining Techniques to Analyze and Predict Crimes	● Crime analysis is done to identify crime patterns quickly for future crime pattern detection and action.	● To better understand the drawn predictions, there should be some platform for depicting the same, however nothing is mentioned

		● Different structures are used for recording crime data. Clustering techniques like K-means and DBscan algorithms are proposed to get the crime ratio relation between the states or districts.	regarding it.
8.	Pattern Detection using Data Mining	● Study of various classification and clustering techniques is done, with the help of which an attempt to classify categories of crimes at various places was made. ● Reports were generated in the form of geo-spatial plots which clearly depicts the categories of crime in a particular region	● Since the crime is subjective and unpredictable to some extent, this idea is not a complete solution for reducing crime in our society.

ms

CHAPTER 4

PROBLEM DEFINITION AND

SCOPE

4.1   Problem Statement

Visualization of crime data in order to predict the rate of crime and its severity in the forthcoming years, in different areas of the country based on the previous year criminal records of India.

4.1.1   Goals and objectives

This system is proposed for increasing transparency regarding the crime in the society and to help police officials, Government of India and the rest concerned authorities to take measures for minimizing the overall crime in our country.

We aim to study these crime patterns and realise the changes in the overall crime ratio based on the demographic or population ratios which shall ease the process of reaching the future crime-cases prediction. Along with this, a prediction can be made on the maximum criminal activities that need to be controlled and the amount of security measures that need to be taken to decelerate or curb the occurrence of any mild to treacherous level criminal activities.

The purpose is to help ascertain incidence and patterns of the crimes using data mining techniques and data visualization.

4.1.2   Statement of scope

This project scales to the scope of benefiting the following three sectors:

1. Government sector: To analyse the crime patterns to accordingly understand the severity of crime threat in different states/districts/areas of the country and implement suitable security strategies and operations.

2. Public sector: To understand the safety of residing or migrating to a particular area within the country, or for secure rehabilitation. Also to apprehend the security services available in that area for approach during emergencies.

3. Other sectors: Law and order sectors; Government servants/officials for preferential posting, etc.

1) **Input Size limitations** : It utilizes yearwise dataset of Crime types from 1990 to 2016, as the size of data will increase it would be reflected in the accuracy percentage of predictions for coming years.

2) **Input Validation** : The dataset given to the system checks for the numerical values in proper format and converted to suitable format using data mining

techniques such as eliminating missing values, eliminating redundant data, data transformation, etc.

3) **Input Dependency** : The data is obtained from the official Indian Government Websites.Algorithms for the implementation of Visualization is also an important aspect on which the project is dependent .

4) **Major Functionalities of the System** : Crime type predictions are performed, for four years, for each state as well as all the states of India using the data from year 1990-2016. These predictions are displayed using simple visualization charts.
   These visualizations would provide the comparative study between the various crimes as well as between crimes in different region.The system is not supposed to change the information without permission, this implies the system is secure and accurate information is presented on the platform.

5) **Major Inputs and Outputs** : The System requires dataset of the Crime against women and Children, Indian Penal Code and Special and Local Laws Crimes.The proposed system would present the supplied data in some visuals such as Animate and barcharts as output with the prediction of the crime for coming four years with accuracy percentage.

4.2   Major Constraints

1) **Data Dependency** : The Data is fetched from the Trusted third party ,if some values are missing then the preferable steps are performed to minimise the inaccuracy for Visualization and prediction of the proper datasets.

2) **Workload** : DB system can easily handle and read transactions but cannot update transactions frequently and only administrators can update the Database .

3) **Scalability** : The system is Scalable as the Database increases, the variation in the Visualization increases. As the increase in the system's workload, the system would be able to process and so the system is Scalable to an apex.The scalability required is often driven by the lifespan and the maturity of the system. In this case, the new (and hence immature) proposed system could suffer an unexpected growth in popularity and suffer from a significant increase in workload as it becomes popular with new users. More mature systems which represent improvements on older systems are likely to have more accurately defined workloads and thus be less likely to suffer in this respect. Initially,the proposed system may face some issues but this could be handled.

4) **Platform** : The system development requires various generic products to be used, for example, Tableau, Flask ,etc.

5) **IDE** : Visual Studio for development.

6) **Programming Language** : Python, HTML5, CSS, Javascript (Javascript libraries-Chart.js, Plotly.js).

7) Above given Requirements are Average minimum hardware requirement for Web Browser System.

4.3   Methodologies of Problem solving and efficiency issues

- The system uses Linear regression Algorithm in major section i.e. Predicting the Crime Patterns with minimum accuracy rates as factor. Linear Regression is the suitable Algorithm to predict pattern of this data. We had tried some other algorithms for testing purpose such as Random Forest Algorithm, and found Linear regression as the best fit for this proposed system with minimal accuracy in the results.
- DB system can easily handle and read transactions but cannot update transactions frequently and only administrators can update the Database .
- The system is Scalable as the Database increases, the variation in the Visualization increases. As the increase in the system's workload, the system would be able to process and so the system is Scalable to an apex.

4.4   Outcome

The System "Indian Crime Analysis" is designed specifically for crime analysis have been created to perform functions that are not available in other existing software.Thus, it can be understood that, even though several solutions to solve the problem has been proposed it can be seen that a perfect solution to each city, state and country is developed for all type of users.

The System is precise and would present the analysis in the form of animate visual and predict the crime ratio precisely, if the system is unable to provide accurate results then it would notify about the unavailability of data or the proximate cause.

4.5   Applications

- Crime investigation and interrogation of large database of information held by law enforcement departments and criminal justice agencies.
- Anyone can access the Predicted and Visualized data for the reference.

- Used by the Government to transfer the official who can handle the vulnerable area effectively that helps to reduce the criminal activities.
- Used by the Migrants to decide which region would be the best suitable for them.
- Used by the Non-governmental organizations or nonprofit organisation to spread awareness in society.
- Law Students can use the data to study and analyze the areas under crime and investigation.
- Researchers and analysts can use the website to extract information in bulk as it contains large amount of data of criminal activities across the country.

4.6   Hardware Resources Required

Sr. No.	Parameter	Minimum Requirement
1	CPU Speed	1 GHz
2	RAM	128 MB
3	Hard Disk Space	120 MB

Table 3.1 : Hardware requirements

Above given Requirements are Average minimum hardware requirement for Web Browser System

4.7   Software Resources Required

Platform : Flask, MS-Excel, Tableau, Ipython Notebook, Browser like chrome, firefox, etc for viewing website

1. Operating System: Linux, Windows

2. IDE: Visual Studio

3. Programming Language: Python, HTML5, CSS, Javascript (Javascript libraries- Chart.js, Plotly.js)

CHAPTER 5

PROJECT PLAN

5.1 Project Estimates

Use Waterfall model and associated streams derived from assignments 1,2, 3, 4 and 5(Annex A and B) for estimation.

5.1.1 Reconciled Estimates

5.1.1.1 Cost Estimate

5.1.1.2 Time Estimates

5.1.2 Project Resources

Project resources [People, Hardware, Software, Tools and other resources] based on Memory Sharing, IPC, and Concurrency derived using annexures to be referred.

5.2 Risk Management w.r.t. NP Hard anal-ysis

This section discusses Project risks and the approach to managing them.

5.2.1 Risk Identi cation

For risks identi cation, review of scope document, requirements speci cations and schedule is done. Answers to questionnaire revealed some risks. Each risk is categorized as per the categories mentioned in [?]. Please refer table 5.1 for all the risks. You can refereed following risk identi cation questionnaire.

1. Have top software and customer managers formally committed to sup-port the project?

2. Are end-users enthusiastically committed to the project and the sys-tem/product to be built?

3. Are requirements fully understood by the software engineering team and its customers?

4. Have customers been involved fully in the de nition of requirements?

5. Do end-users have realistic expectations?

6. Does the software engineering team have the right mix of skills?

7. Are project requirements stable?

8. Is the number of people on the project team adequate to do the job?

9. Do all customer/user constituencies agree on the importance of the project and on the requirements for the system/product to be built?

5.2.2   Risk Analysis

The risks for the Project can be analyzed within the constraints of time and quality

ID	Risk Description	Probability	Impact		
			Schedule	Quality	Overall
1	Description 1	Low	Low	High	High
2	Description 2	Low	Low	High	High

Table 5.1: Risk Table

Probability	Value	Description
High	Probability of occurrence is	> 75%
Medium	Probability of occurrence is	26 75%
Low	Probability of occurrence is	< 25%

Table 5.2: Risk Probability de nitions [?]

5.2.3   Overview of Risk Mitigation, Monitoring, Management

Following are the details for each risk.

Impact	Value	Description
Very high	> 10%	Schedule impact or Unacceptable quality
High	5 10%	Schedule impact or Some parts of the project have low quality
Medium	< 5%	Schedule impact or Barely noticeable degradation in quality Low Impact on schedule or Quality can be incorporated

Table 5.3: Risk Impact de nitions [?]

Risk ID	1
Risk Description	Description 1
Category	Development Environment.
Source	Software requirement Speci cation document.
Probability	Low
Impact	High
Response	Mitigate
Strategy	Strategy
Risk Status	Occurred

Risk ID	2
Risk Description	Description 2
Category	Requirements
Source	Software Design Speci cation documentation review.
Probability	Low
Impact	High
Response	Mitigate
Strategy	Better testing will resolve this issue.
Risk Status	Identi ed

Risk ID	3
Risk Description	Description 3
Category	Technology
Source	This was identi ed during early development and testing.
Probability	Low
Impact	Very High
Response	Accept
Strategy	Example Running Service Registry behind proxy balancer
Risk Status	Identi ed

5.3 Project Schedule

5.3.1 Project task set

Major Tasks in the Project stages are:

Task 1:

Task 2:

Task 3:

Task 4:

Task 5:

5.3.2 Task network

Project tasks and their dependencies are noted in this diagrammatic form.

5.3.3 Timeline Chart

Please refer Annexure C for the planner

5.4 Team Organization

The manner in which sta is organized and the mechanisms for reporting are noted.

5.4.1 Team structure

The team structure for the project is identi ed. Roles are de ned.

5.4.2 Management reporting and communication

Mechanisms for progress reporting and inter/intra team communication are identi ed as per assessment sheet and lab time table.

CHAPTER 6

SOFTWARE REQUIREMENT

SPECIFICATION

6.1 Introduction

6.1.1 Purpose and Scope of Document

The purpose of Software Requirement Specification is to guide the users about the various sections of the proposed System "Indian Crime Analysis" , scope and limitation of the system.

The proposed system contains three major modules with their functionalities :

1) Visualization : The proposed system is supposed to Visualize the the Crime data after classifying and clustering the datasets using matplotlib: bar graphs, line graphs, Animates, scatter plots, and choropleth (map) plots. It will also create figures with multiple sub-figures,and customize labels, colors, error bars, etc.

2) Prediction : The system uses statistics to predict outcomes to detect crimes and identify suspects, after the crime has taken place using the dataset. Nearly any regression model can be used for prediction purposes by making specific assumptions with regard to one or more of the parameters.

3) Some basic modules to improve the nature of Interaction :
- Statistical Reports or Observations
- Dashboards
- Geographical Analysis tools
- Notifications

6.1.2 Overview of responsibilities of Developer

Activities carried out by the developers :

- Information gathering : Data extraction of various Crime Report against women, children, senior citizens, Indian Penal Code and Special and local laws crimes.

- Data Cleaning : Data is converted to suitable format using data mining techniques such as eliminating missing values, eliminating redundant data, data transformation, etc.

- Algorithm Implementation : The extracted data is fed to the suitable algorithm. We have implemented some of the algorithms and found Linear Regression as the best fit for the dataset for data science domain mainly for Crime Analysis.

● Web Designing : Some of the major prerequisites are HTML5, CSS, javascript, Python, Ruby on Rails, Tableau and functionalities of various Libraries such as Chart.js, plotly, sklearn.

● Mentioned the statistics of the Dataset, Predictions and Visualizations and provide the specific observations for the users.

6.2   Usage Scenario

This section provides various usage scenarios for the system to be developed which describes a real-world example of how one or more people or organizations interact with a system.

The system "Indian Crime Analysis" will be used by various organizations such as Non governmental Organizations, Policy makers, Law Enforcements and other Private Agencies and Business Representatives which provides Comparative study for various purpose of Business benefits and social workings.

The system is also useful for the individuals such as Law students, government servants, Researchers and Analysts, Advocates and Migrants for studying crime patterns at individual or organizational levels.

6.2.1   User profiles

● **Researchers or Analysts :** Researchers can use the proposed system to predict some categories of crime if correct analysis is applied. For example, Hidden in the Twittersphere are nuggets of information that could prove useful to crime fighters -- even before a crime has been committed.Thus the system would make crime prediction more flexible.

● **Government Servants :** The Center can survey the police departments to determine how they can use analytic mapping in policing and began developing training programs to enhance departments' ability to use spatial maps and data sets. Thus, the project's statistical analysis and analysis of spatial data to lead to:
- More effective deployment of police.
- Better use of public safety resources.
- Stronger crime policies.
- Greater understanding of crime.

● **Migrants :** Geography has a major influence on crime. The features and characteristics of cityscapes and rural landscapes can make it easier or more difficult for crime to occur. Combining geographic data with police report data and then displaying the information on a map is an effective way to analyze where, how and why crime occurs. Thus , helpful in choosing the most appropriate and preferable place for the person to stay.

- **Law Students :** The Students can gather information for many fields to study various kinds of crime and the probable reason behind them. For example, students interested in studying drug use among college students on campus or the other statistics of crime in the society.

- **Non governmental Organisation :** NGOs can combat crimes in a number of different ways, such as:
  - Working with governments to improve legislation;
  - Monitoring and reporting incidents;
  - Acting as a voice for victims of crimes, especially by serving as intermediaries with the authorities.
  - Providing practical assistance to victims of hate crimes, such as legal advice, counselling and other services;
  - Raising awareness about the existence of discrimination,intolerance and hate crimes;
  - Campaigning for action to meet the challenge of hate crimes.

- **Law Enforcement and other Private Agencies :** Crime data are also a critical source of information for program and policy evaluations by various agencies in government, academia, and the public and private sectors.

- **Advocates :** The data and visualizations of the same are also used by advocates of particular issues and by the public, and are often seen as measures of accountability.

- **Policy makers :** Policy makers at the local, state, and federal levels need accurate and timely data on crime to inform budgetary decisions about the amount of resources.

- **Business representatives :** The statistical information can be used by the business representatives to establish their business near the appropriate region and use the business tactics to develop the strategy and thus increase the revenue.

6.2.2   Use-cases

All use-cases for the software are presented. Description of all main Use cases using use case template is to be provided.

Table 6.1: Use Cases

Sr No.	Use Cases	Description	Actors	Assumptions
1.	User Credentials	User is required to fill some of the basic personal information for further communication in any case.	User (Primary Actor)	User would provide the proper credentials, as this is to ensure that user is using the

				system for legal purpose.
2.	Data Acquisition	Data is extracted from the trusted third party.As per the updations in dataset the results of the System would be more accurate.	Datastore (secondary Actor)	Data provided by the Indian Government Source is assumed to be proper.
3.	Data Preprocessing	Data is converted to suitable format using data mining techniques such as eliminating missing values, eliminating redundant data, data transformation, etc.	Datastore (secondary Actor)	Data preprocessing techniques used are accurate.
4.	Visualizations	Visualizations are used by various types of users to represent crime pattern analysis.	User (primary Actor)	The initial module of this system would presents only some of the Visuals.
5.	Results	This module will be presenting the observations to the user, those can be generalized observation or specific to the dataset.	User (primary Actor)	The results are assumed to be accurate.

6.2.3   Use Case View
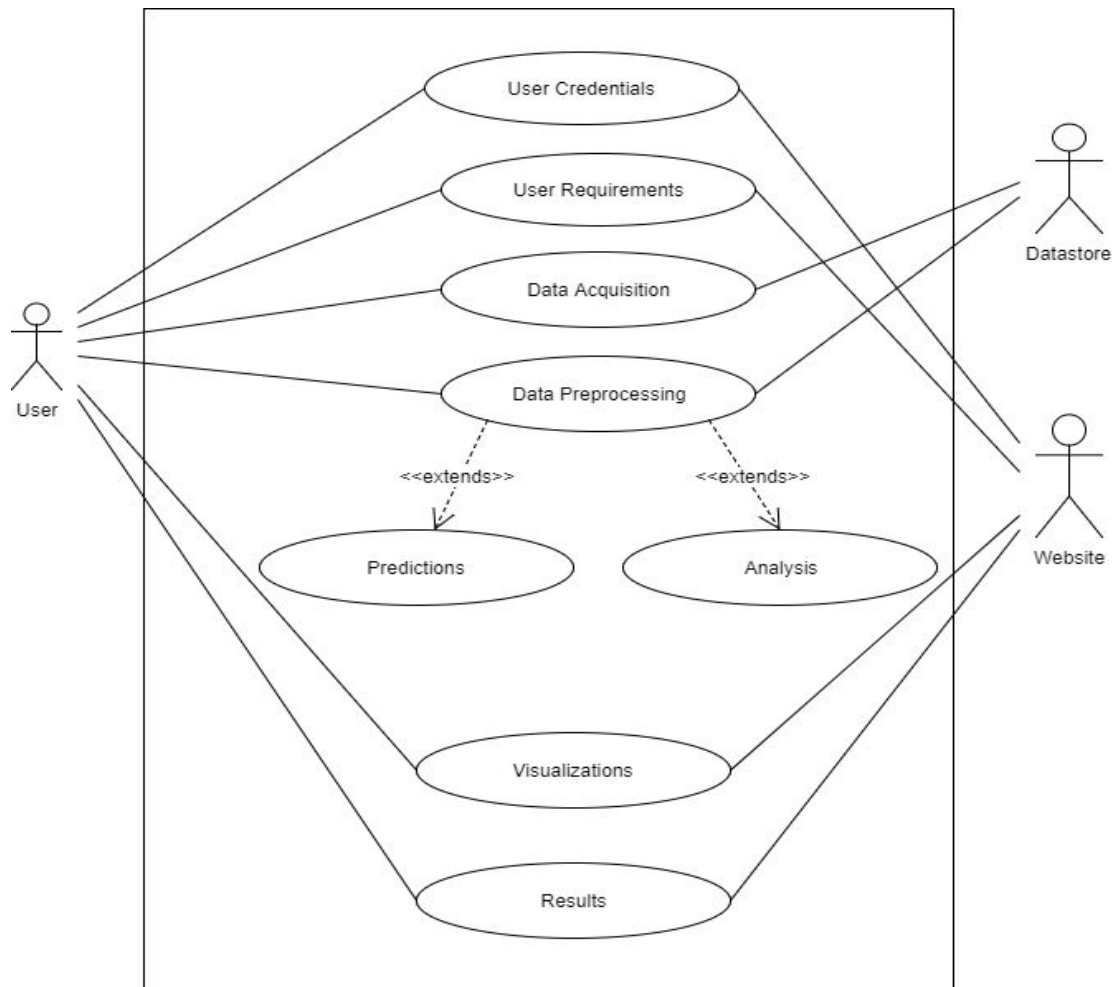
Figure 6.1: Use case diagram

6.3 Data Model and Description

6.3.1 Data Description

India's Crime Dataset from National Crime Records Bureau (NCRB) and Open Government Data website:
- State wise Cases Reported Crimes committed against Women from 1990 to 2016.
- State wise Cases Reported Crimes committed against Children from 1994 to 2016.
- State wise cases reported under IPC Crimes from 2001 to 2016.
- State wise cases reported under SLL Crimes from 2001 to 2016.

India's Demographic data from the Census India website:
- State wise Population from year 1991 to 2011.

6.3.2 Data objects and Relationships

Data objects and their major attributes and relationships among data objects are described using an Entity Relationship Diagram .
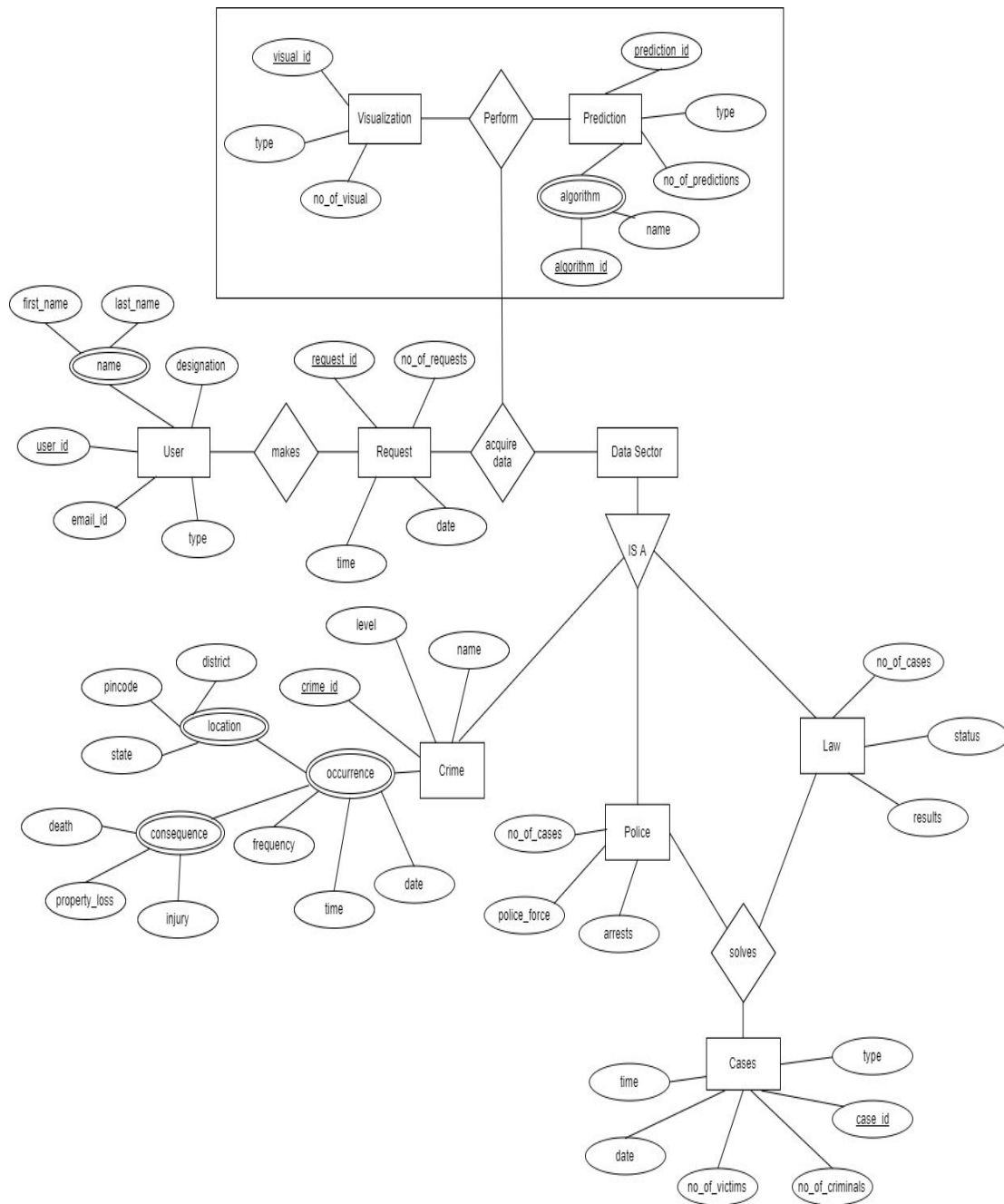
Fig 3.9 : Entity Relationship Diagram

6.4 Functional Model and Description

A description of each major software function, along with dataflow (structured analysis) or class hierarchy (Analysis Class diagram with class description for object oriented system) is presented.
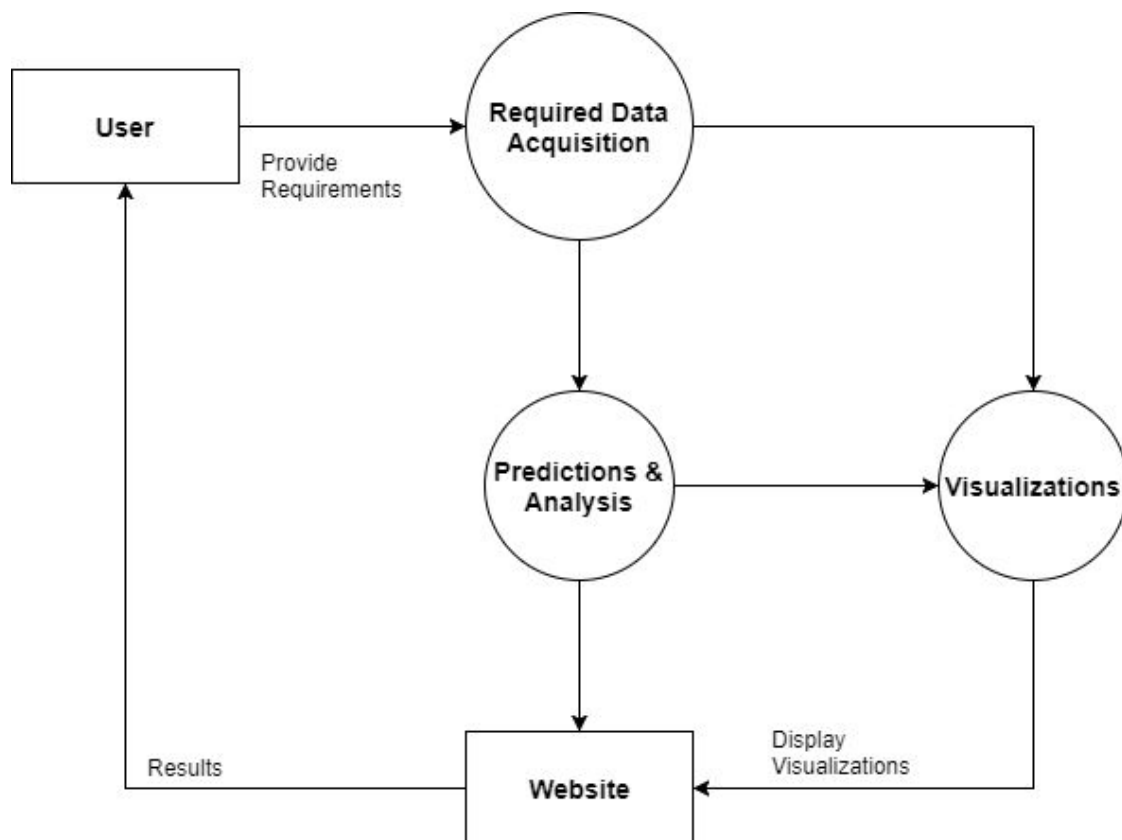
6.4.1 Data Flow Diagram

In this case, we are using data flow diagrams to show the generalized process to achieve the goal of presenting a system which shows the Visualized and Predicted information from the gathered Crime information.

6.4.1.1   Level 0 Data Flow Diagram

Fig 3.2 : DFD level 0

6.4.1.2   Level 1 Data Flow Diagram

Fig 3.3 : DFD level 1
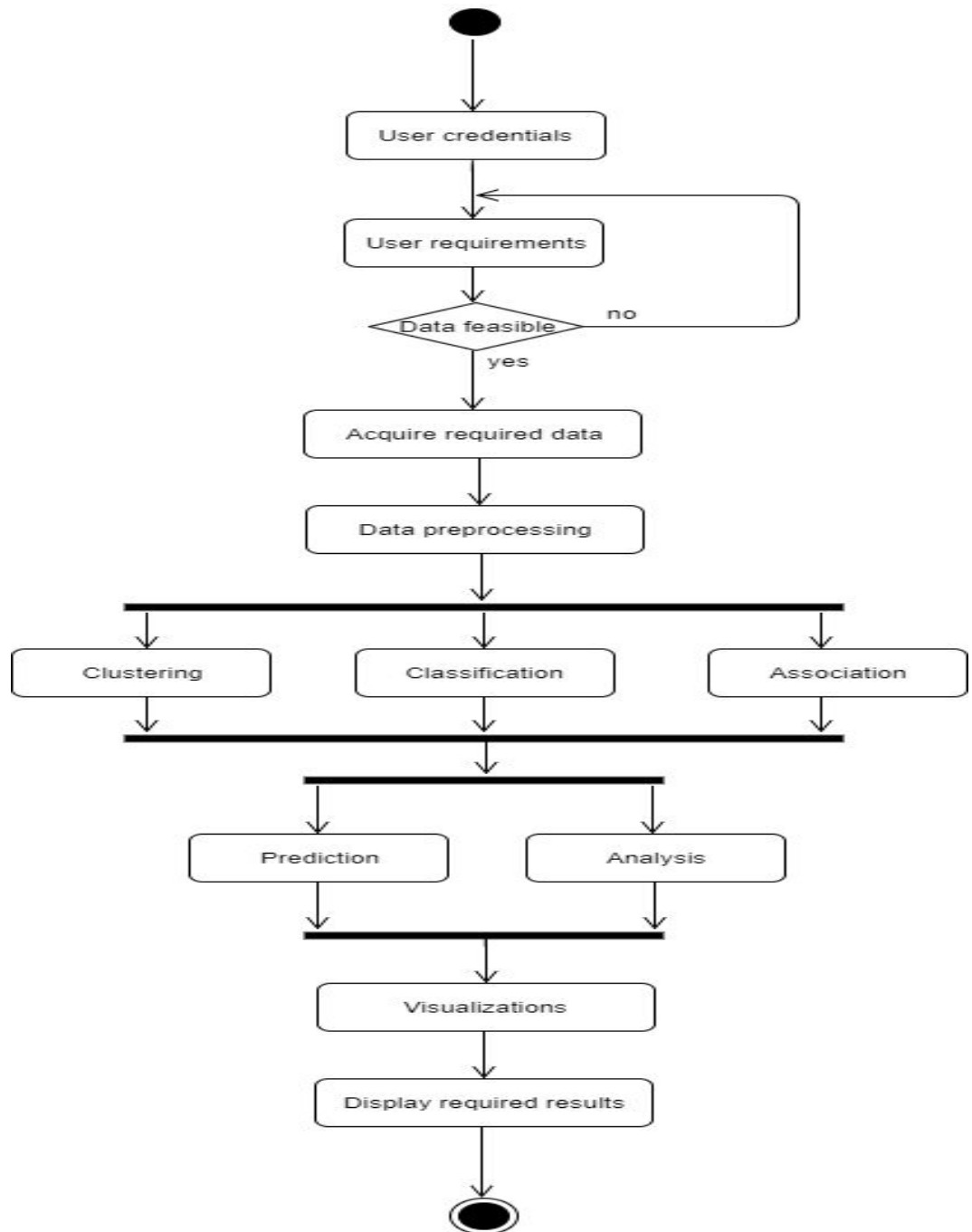
6.4.1.3   Level 2 Data Flow Diagram
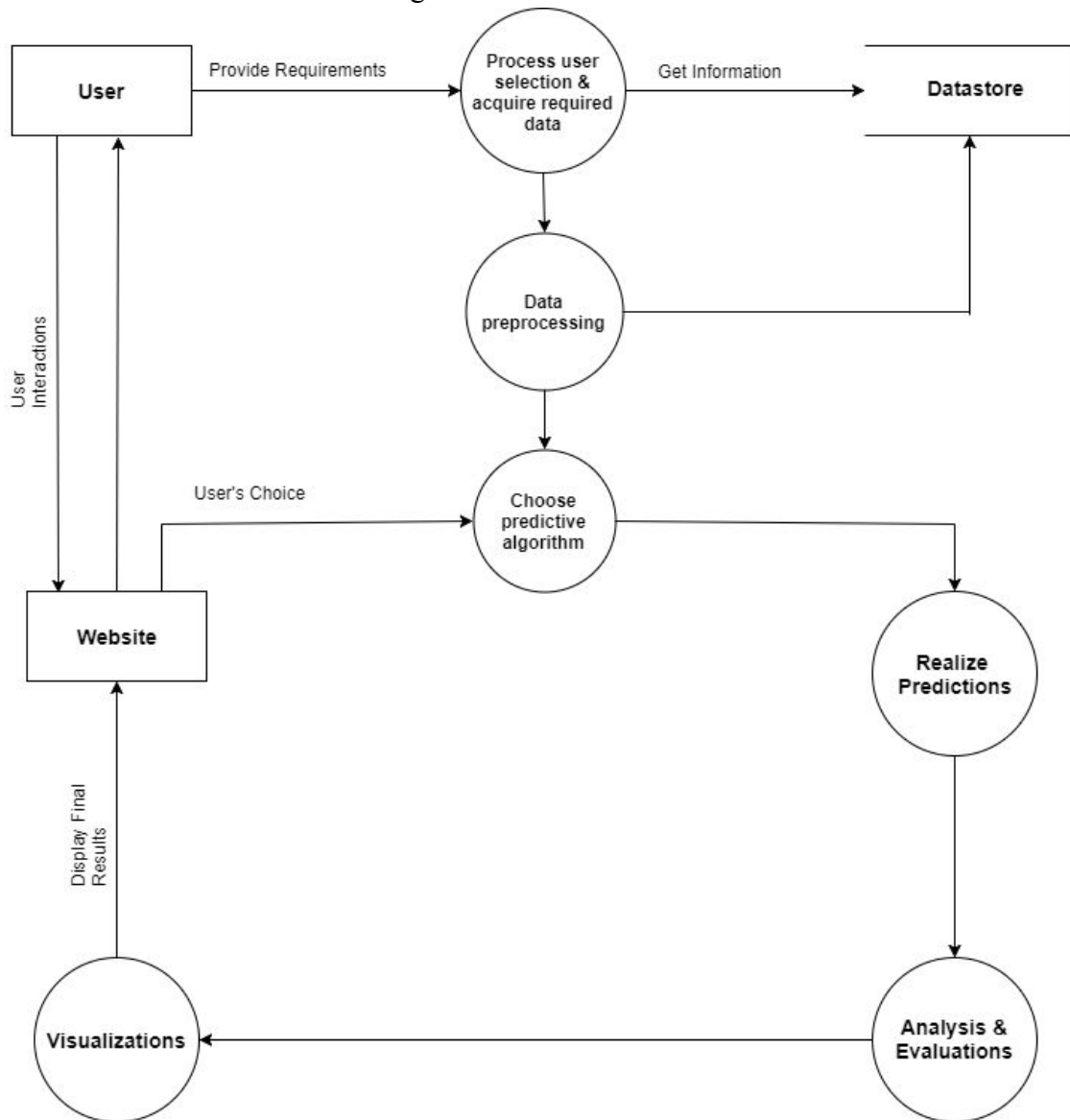
Fig 3.4 : DFD level 2

6.4.2   Activity Diagram:

Fig 3.5 : Activity Diagram

6.4.3   Non Functional Requirements:

Interface Requirements for the System : The user just require Browser like chrome, firefox, etc for viewing website.
For development the required Scripting/Programming languages :
- Python Programming language.
- Ruby on Rails for website development.

Performance requirement of the proposed System :

- **Response Time** : 0.1 second will be response time for the system, the limit so that the user feel that the system is reacting instantaneously as no special feedback is necessary except to display the result. Python programming language is used in the backend, however it is well known for its computational speed as far as data mining is concerned. For systems that have to support significant numbers of users the cost of response times delays can actually be measured in monetary terms and therefore can form part of trade-off studies between different architectures providing different levels of performance.
- **Workload** : DB system can easily handle and read transactions but cannot update transactions frequently and only administrators can update the Database .
- **Scalability** : The system is Scalable as the Database increases, the variation in the Visualization increases. As the increase in the system's workload, the system would be able to process and so the system is Scalable to an apex.
  The scalability required is often driven by the lifespan and the maturity of the system. In this case, the new (and hence immature) proposed system could suffer an unexpected growth in popularity and suffer from a significant increase in workload as it becomes popular with new users. More mature systems which represent improvements on older systems are likely to have more accurately defined workloads and thus be less likely to suffer in this respect. Initially,the proposed system may face some issues but this could be handled.
- **Platform** : The system will require various generic products to be used, for example, Tableau, Weka, Ipython Notebook, Spyder,etc.

Software Quality Attributes

Application will satisfy following software quality attributes:
- Correctness: Most accurate algorithm will be considered for prediction of crime. Thus the results will be the most accurate.
- Reusable: This system can be reused if there is a proper track of crime records maintained by the government in forthcoming years.
- Portable: As software is Web based, investigators can access or request for evidences anytime he/she needs regardless of place. It automatically introduces portability.
- Also, system satisfies other quality attributes such as reliability, interoperability, maintainability.

6.4.4 State Diagram:

State Transition Diagram

Fig.6.2 example shows the state transition diagram of Cloud SDK. The states are represented in ovals and state of system gets changed when certain events occur. The transitions from one state to the other are represented by arrows. The Figure shows important states and events that occur while creating new project.
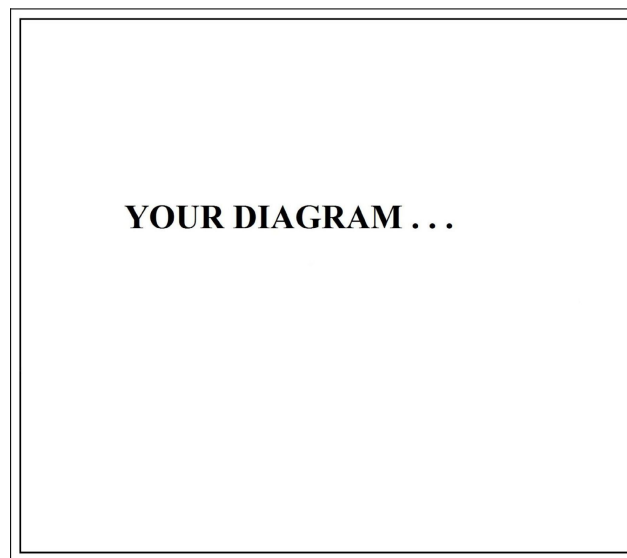
YOUR DIAGRAM . . .

Figure 6.2: State transition diagram

6.4.5   Design Constraints

The Algorithms used in the implementation would show result in predicted Crime patterns. As per our testing results the system would show best results using Linear Regression as compared to Random Forest Algorithm.

6.4.6   Software Interface Description

The resources it provides or requires, and the nature of its interactions are as shown below :

- Statistical Reports or Observations
- Dashboards
- Graphical Analysis tools
- Login for the administrator to discover knowledge from the Dataset.
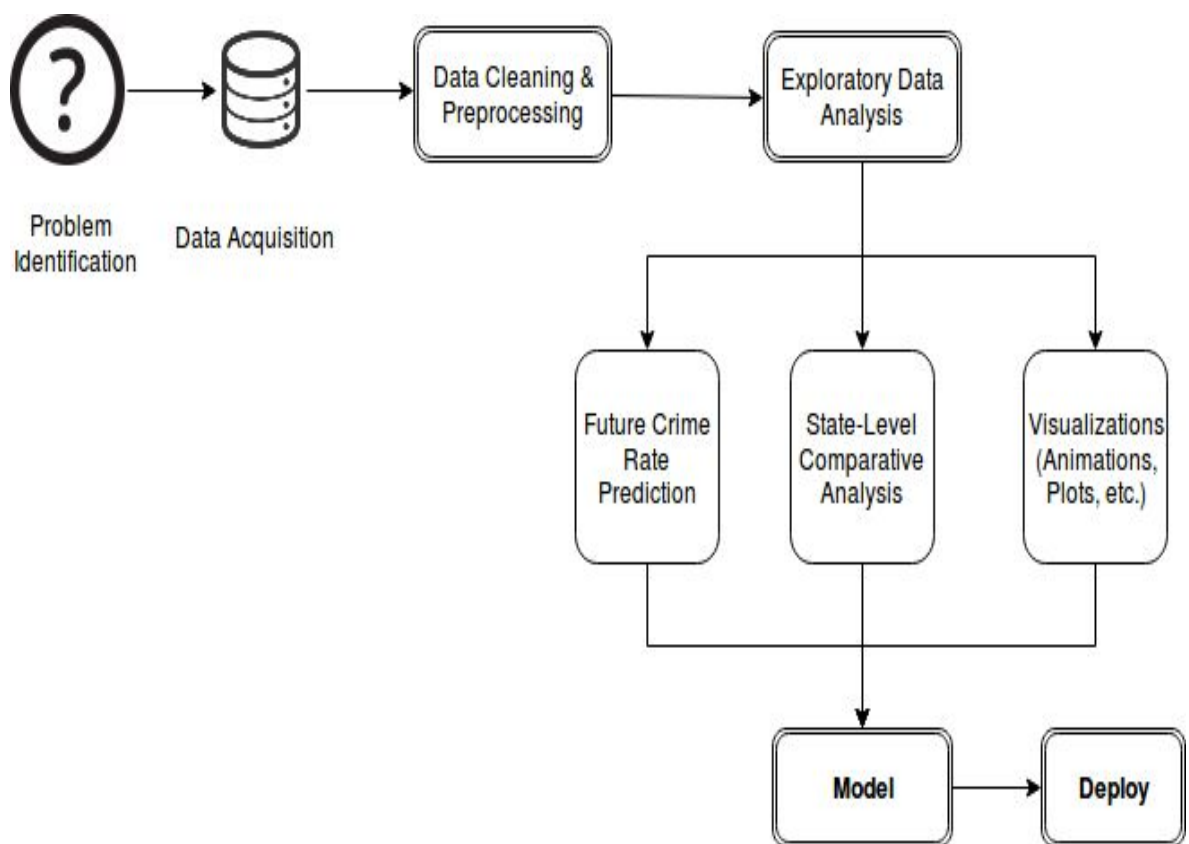- Notifications.
- Responsive themes.

CHAPTER 7

DETAILED DESIGN DOCUMENT USING ANNEXURE A

7.1   Introduction

According to our problem statement, our system is supposed to present the predictions and visualizations to the user for crime analysis. The design used to solve this problem can be as discussed in detail below.

7.2   Architectural Design

Fig 4.1 : System Architecture

Above shown diagram is the system architecture that clearly depicts every module of our system. The problem is invented and appropriate data is collected from the India's National Crime Records Bureau. Collected data then undergoes preprocessing, cleaning, wrangling, munging and removing nan values. Analysis of this data is done according to various aspects of crime patterns such as clustering states and districts based in the crime

intensity, predictions of crime ratio in future, district level comparative analysis of crime etc. All the predictions are visualized on a dynamic website.

- Problem Identification : Visualization of crime data in order to predict the rate of crime and its severity in the forthcoming years, in different areas of the country based on the previous year criminal records of India.

- Data Acquisition : Acquisition of data from the official government websites
  - National Crime Records Bureau (NCRB)
  - Open Government Data, Census India.

- Data Cleaning and Preprocessing :
  - Eliminating redundant values,
  - Handling missing/NA values,
  - Data transformation

- Exploratory Data Analysis :
  - Future Crime Rate Prediction : This module of the architecture would predict the crime patterns for coming four years.
  - State-level Comparative Analysis : This module would present the comparative observations for all type of users, including the region wise comparison and crime type wise comparison.
  - Visualizations : This module is for representing comparative information in Visuals for making the study easier and interesting and this will grab user's interest.

- Model : This section represent complete model of the system with all the subsections and working algorithm.

- Deploy : The system is deployed and user's can easily access the Website to analyse and check the predictions for various regions and crimes.

7.3   Data design

Database description

India's Crime Dataset from National Crime Records Bureau (NCRB) and Open Government Data website:

- State wise Cases Reported Crimes committed against Women from 1990 to 2016.
  - Crime attributes against women:
    - Rape
    - Kidnapping & Abduction
    - Dowry Death
    - Assault on women with intent to outrage her modesty
    - Insult to the modesty of women
    - Cruelty by husband or relatives
    - Total crime against women
- State wise Cases Reported Crimes committed against Children from 1994 to 2016.
  - Crime attributes against children:
    - Infanticide
    - Rape
    - Kidnapping & Abduction
    - Foeticide
    - Abetment of suicide
    - Exposure & Abandonment
    - Procuration of minor girls
    - Buying of girls for prostitution
    - Selling of girls for prostitution
    - Prohibition of Child Marriage Act, 2006
    - Total crimes against children
- State wise cases reported under IPC Crimes from 2001 to 2016.
  - Crime attributes for IPC crimes:
    - Murder
    - Attempt to commit murder
    - Culpable Homicide not amounting to murder
    - Rape
    - Kidnapping & Abduction
    - Dacoity
    - Preparation & Assembly for dacoity
    - Robbery
    - Burglary
    - Theft

- - - Riots
    - Criminal breach of trust
    - Cheating
    - Counterfeiting
    - Arson
    - Hurt/Grievous hurt
    - Dowry Death
    - Assault on women with intent to outrage her modesty
    - Insult to the modesty of women
    - Cruelty by husband or relatives
    - Causing death by negligence
    - Total IPC crimes
- State wise cases reported under SLL Crimes from 2001 to 2016.
  - Crime attributes for SLL crimes:
    - Arms Act, 1959
    - Narcotic Drugs & Psychotropic Substances Act, 1985
    - Gambling Act, 1867
    - Excise Act, 1944
    - Prohibition Act
    - Explosives & EXplosive Substances Act
    - Immoral Traffic (Prevention) Act, 1956
    - Indian Railways Act, 1989
    - Registration of Foreigners Act, 1939
    - Protection of Civil Rights Act, 1955
    - Indian Passport Act, 1967
    - Essential Commodities Act, 1955
    - Antiquities & Art Treasures Act, 1972
    - Dowry Prohibition Act, 1961
    - Prohibition of Child Marriage Act, 2006
    - Indecent Representation of Women (Prevention) Act, 1986
    - Copyright Act, 1957
    - SC/ST (Prevention of Atrocities) Act, 1989
    - Forest Act, 1927
    - Total SLL crimes

India's Demographic data from the Census India website:
- State wise Population from year 1991 to 2011.

7.4   Component Design
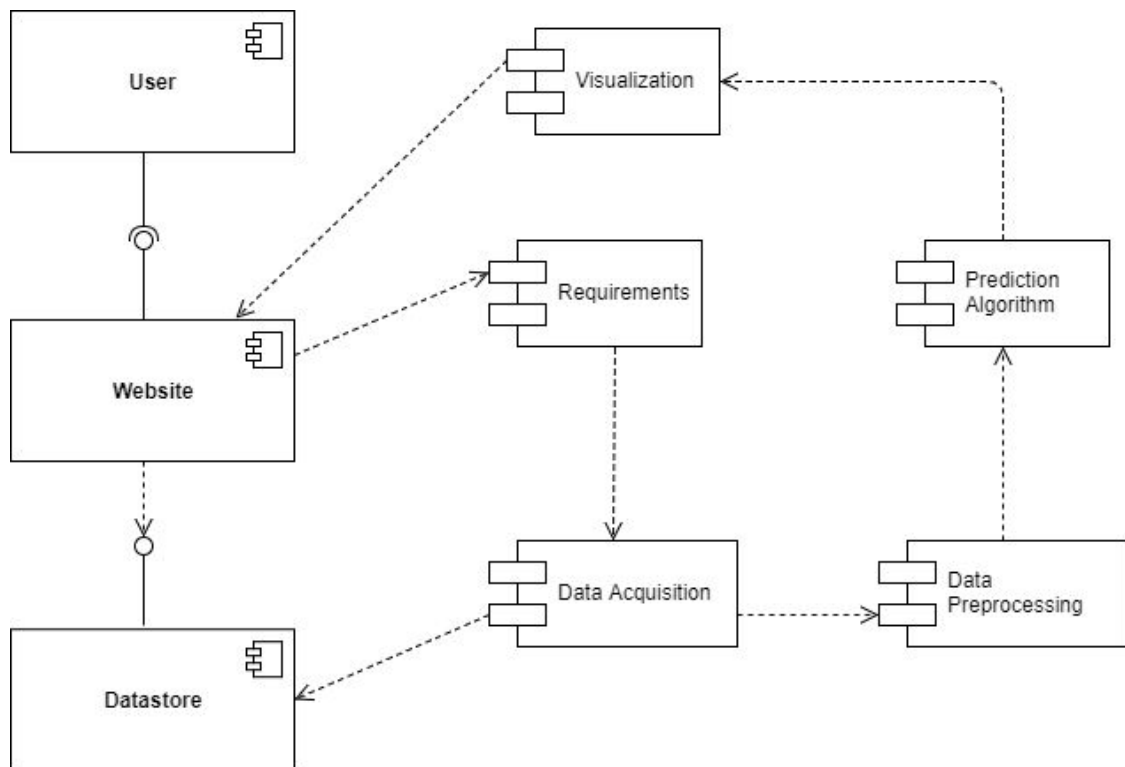
Fig 3.8 : Component Diagram

CHAPTER 8

PROJECT IMPLEMENTATION

8.1 Introduction

8.2 Tools and Technologies Used

Tools used:
MS Excel/LibreOffice Calc
Tableau

Tecnologies:
Python programming - for performing predictions
Chart.js - for displaying predictions
Plotly.js - for visualizations

8.3 Methodologies/Algorithm Details

8.3.1 Algorithm 1/Pseudo Code

LINEAR REGRESSION

8.3.2 Algorithm 2/Pseudo Code

RANDOM FOREST

8.4 Verification and Validation for Acceptance

CHAPTER 9

SOFTWARE TESTING

9.1   Type of Testing Used

UNIT TESTING : This model does not attempt to start with a full specification of requirements. Instead, development begins by specifying and implementing just part of the software, which can then be reviewed in order to identify further requirements. This process is then repeated, producing a new version of the software for each cycle of the model.

In Iterative model we create rough product or product piece in one iteration, then review it and improve and test it in next iteration and so on until it's finished. Hence, in this model the whole product is developed and tested step by step.

9.2   Test Cases and Test Results

For testing purpose, the System assumes the threshold value for accuracy percentage as 65% for prediction for all states.
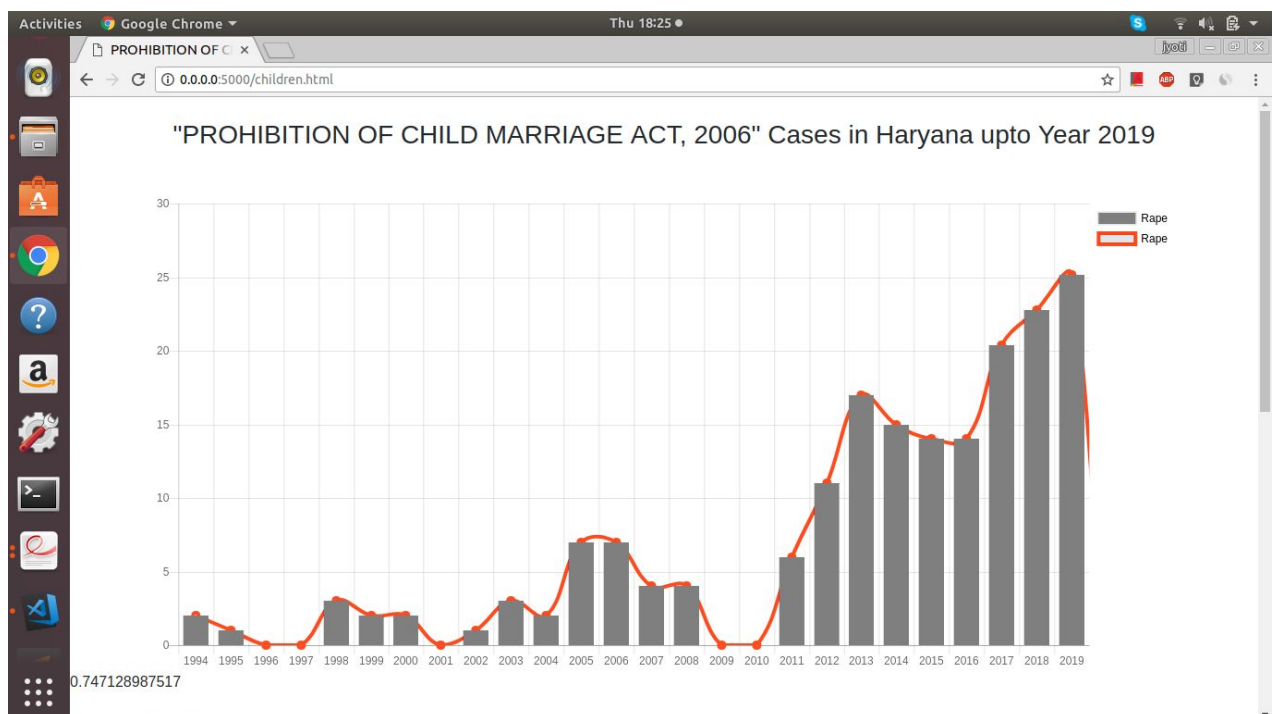
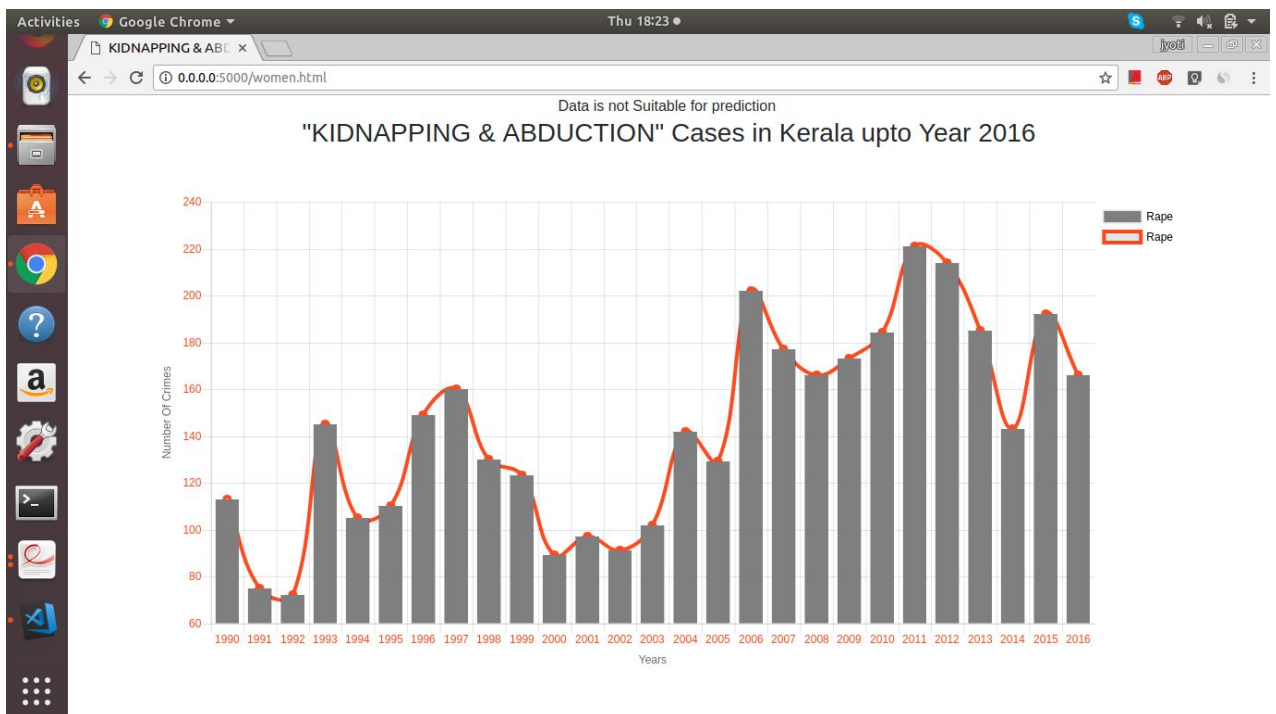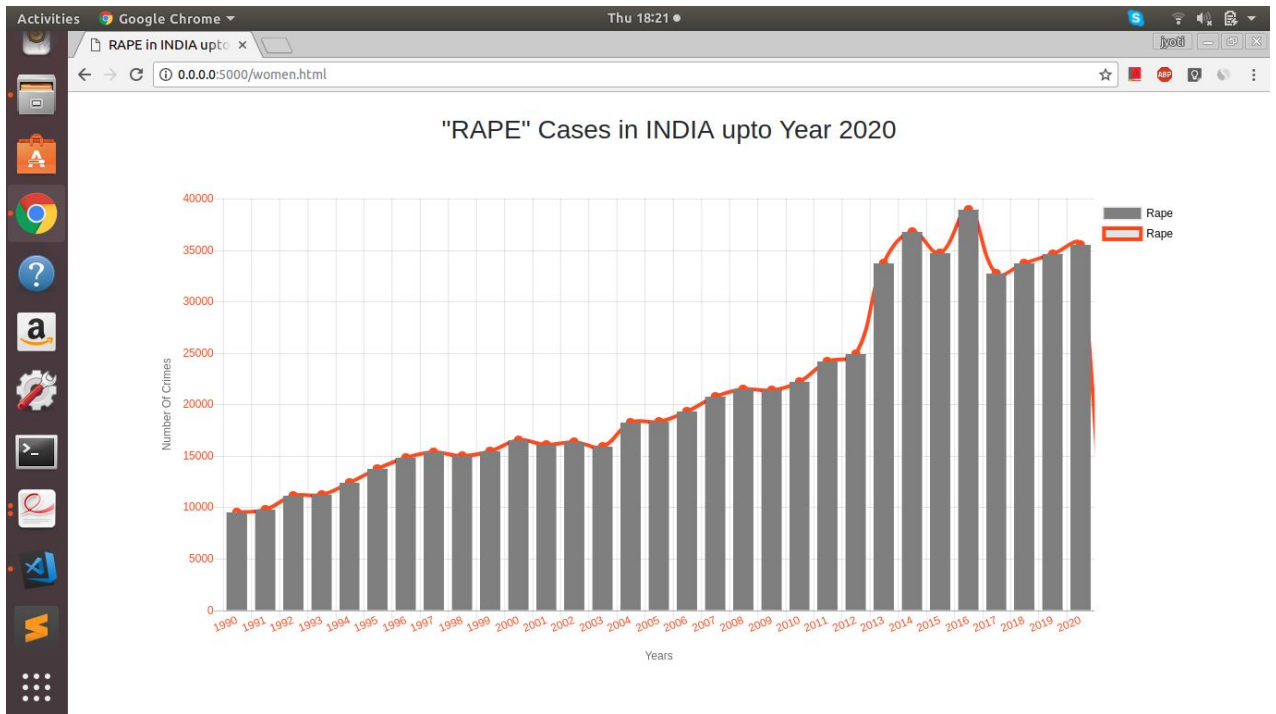Test Case 1 : Prediction for
 Crime Type : Prohibition of Child Marriage Act, 2006
 State : Haryana
 Year Chosen : 2019
Observation : Test case is cleared with accuracy percentage more than 65%.

"RAPE" Cases in INDIA upto Year 2020

Data is not Suitable for prediction

"KIDNAPPING & ABDUCTION" Cases in Kerala upto Year 2016

CHAPTER 10

RESULTS

10.1   Screen shots

Outputs / Snap shots of the results

10.2   Outputs

Outputs / Snap shots of the results

CHAPTER 11

DEPLOYMENT AND

MAINTENANCE

11.1 Installation and uninstallation

No requirement of any installation and installation since the system is implemented on a website, freely available all user classes.

11.2 User help

The website built for this system is extremely user friendly. Any user using the site will receive the desired results in just at a few clicks. On accessing the services section of the website, the user can select a particular service on a single click. After reaching the selected service, for

Predictions:
- Click on the desired crime sector displayed in the menu.
- Select the year upto which the predictions are required.
- Select the crime type for which predictions are required.
- Select a particular state/all states(India in total)
- Submit to reach the results.

Visualizations:
- Click on the desired crime sector to obtain the animated visualization of the selected crime sector for the years 2001-2015.

CHAPTER 12

CONCLUSION AND FUTURE

SCOPE

12.1 Conclusion

Result of this research will be to present a system which will analyse, correlate and predict the crimes from huge data available. Results will be in the form of correlation between various crime and location of crime i.e. state/city. Crime can also be correlated on the basis of age group, location of crime & type of crime. Prediction of the crime will be presented using various techniques and Algorithm.

Another point noted is that the crime rate is increasing and crime prevention has become an upheaval task. The legal force departments around the world are required to remain ahead in the eternal race between lawbreakers and law enforcers. So we are presenting the system which maintains, predict and visualize the crime records and the people who are involved whether the Government official or the Police or it may be guilty.

This System is designed specifically for crime analysis have been created to perform functions that are not available in other existing software.Thus, it can be understood
that, even though several solutions to solve the problem has been proposed it can be seen that a perfect solution to each city, state and country is still elusive.

12.2 Future Scope

Future Scope of the proposed system :

1. More Secure access to the Website by linking to Aadhaar Number under the Unique Identification Authority of India.
2. Frequent and dynamic extraction of the Dataset from the Government Authorities after having the Professional deal with Government Authorities as the Datasets are not fully accessible to the Unauthorized users.

REFERENCES

[1]   Tushar Sonawanev1, Shirin Shaikh2, Shaista Shaikh3, Rahul Shinde4, Asif Sayyad5, Crime Pattern Analysis, Visualization and Prediction using Data Mining, IJARIIE International Journal of Advanced Research and Innovative Ideas in Education, *Vol-1 Issue-4 2015, IJARIIE-ISSN(O)-2395-4396* Methodology for crime pattern analysis.

[2]   Sarpreet Kaur1 and Dr. Williamjeet Singh2, Systematic Review of Crime Data Mining, IJARCS International Journal of Advanced Research in Computer Science, Volume 8, No. 5, May-June 2017, ISSN No. 0976-5697.

[3]   Malathi. A 1, Dr. S. Santhosh Baboo 2 and Anbarasi. A 3; 1 Assistant professor Department of Computer Science, Govt Arts College, Coimbatore, India., 2 Readers Department of Computer Science, D. G. Vaishnav College, Chennai, India. 3 Associative Professor Department of Civil Engineering, Kumaraguru college of Technology, Coimbatore, India; Enhanced Algorithm to Predict a future crime using Data Mining, International Journal of Computer Applications (0975 – 8887), Volume 21– No.1, May 2011.

[4]   Neeru Mago, Design and Implementation of Intelligent Crime Information and Analysis System (ICIAS) based on Crime Data Mining,IJARCSSE International Journal of Advanced Research in Computer Science and Software Engineering, Volume 6, Issue 1, January 2016 ISSN: 2277 128X.

[5]   D.Gnana Rajesh1 and Dr. M. Punithavalli2, A Study on the Existing Computer Aided Crime Analysis Tools to Handle Indian State of Affairs, IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 2, No 3, March 2012 ISSN (Online): 1694-0814.

[6]   Ahamed Shafeeq B M1, Dr. Binu V S2, Spatial Patterns of Crimes in India using Data Mining Techniques, Certified International Journal of Engineering and Innovative Technology (IJEIT) Volume 3, Issue 11, ISSN: 2277-3754   ISO 9001:2008, May 2014.

[7]   1S.Yamuna, 2N.Sudha Bhuvaneswari , 1 M.Phil (CS) Research Scholar School of IT Science, Dr.G.R.Damodaran College of science Coimbatore, 2 Associate Professor MCA, Mphil (cs), (PhD) School of IT Science Dr.G.R.Damodaran College of science Coimbatore, Data mining Techniques to Analyze and Predict Crimes, The International Journal of Engineering And Science (IJES), Volume 1, Issue 2, Pages 243-247 2012, ISSN: 2319 – 1813, ISBN: 2319 – 1805.

[8]   P.DHAKSHINAMOORTHY, T.KALAISELVAN , Assistant Professor, Erode Builder Educational Trust's Group of Institutions, Erode, TamilNadu, CRIME PATTERN DETECTION USING DATA MINING, International Journal of Advanced Research in Computer Science and Applications Vol.1, Issue 1, July 2013, ISSN 2321 -872X.

[9]   Dr. Zakaria Suliman Zubi1, Ayman Altaher Mahmmud2, 1Sirte University, Faculty Of Science, Computer Science Department Sirte, P.O Box 727, Libya, 2The Libyan Academy, Information Technology Department, Tripoli, Libya, Using

Data Mining Techniques to Analyze Crime patterns in the Libyan National Crime Data, ISBN: 978-960-474-350-6.

[10]   Dr. A.Bharathi1, R. Shilpa2, 1 Professor Dept. of IT Bannari Amman Institute of Technology Sathyamangalam – India, 2 Research Scholar Dept. of IT Bannari Amman Institute of Technology Sathyamangalam – India, A Survey on Crime Data Analysis of Data Mining Using Clustering Techniques, International Journal of Advance Research in Computer Science and Management Studies, Volume 2, Issue 8, August 2014, ISSN: 232 7782.

[11]   Vineet Pande, Viraj Samant, Sindhu Nair, Crime Detection using Data Mining, International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181, Vol. 5 Issue 01, January-2016.

[12]   Lakshmi Iyer1, Anandi Mani 2, Prachi Mishra3, Petia Topalova4, 1 Harvard Business School 2 University of Warwick, 3 Research Department, IMF, 4 Research Department, IMF, Political Representation and Crime: Evidence from India November 2010(Work in progress).

[13]   Shiju Sathyadevan1, Devan M.S2, Surya Gangadharan. S3, 1 2 Amrita Center for Cyber Security Amrita Vishwa Vidyapeetham, Amritapuri, Kerala, Amrita Vishwa Vidyapeetham Amritapuri, Kerala, India, 2014 First International Conference on Networks & Soft Computing, 978-1-4799-3486-7 114 2014 IEEE.

[14]   http://blog.hackerearth.com/beginners-tutorial-apriori-algorithm-data-mining-r-implementation.

[15]   India's National Crime Record Bureau for data collection, http://ncrb.nic.in/.

ANNEXURE A

LABORATORY

ASSIGNMENTS.REFER B.E.

COMPUTER ENGINEERING 2012

COURSE TERM II SYLLABUS

- Feasibility justification of the system using concepts of knowledge canvas and IDEAMatrix.
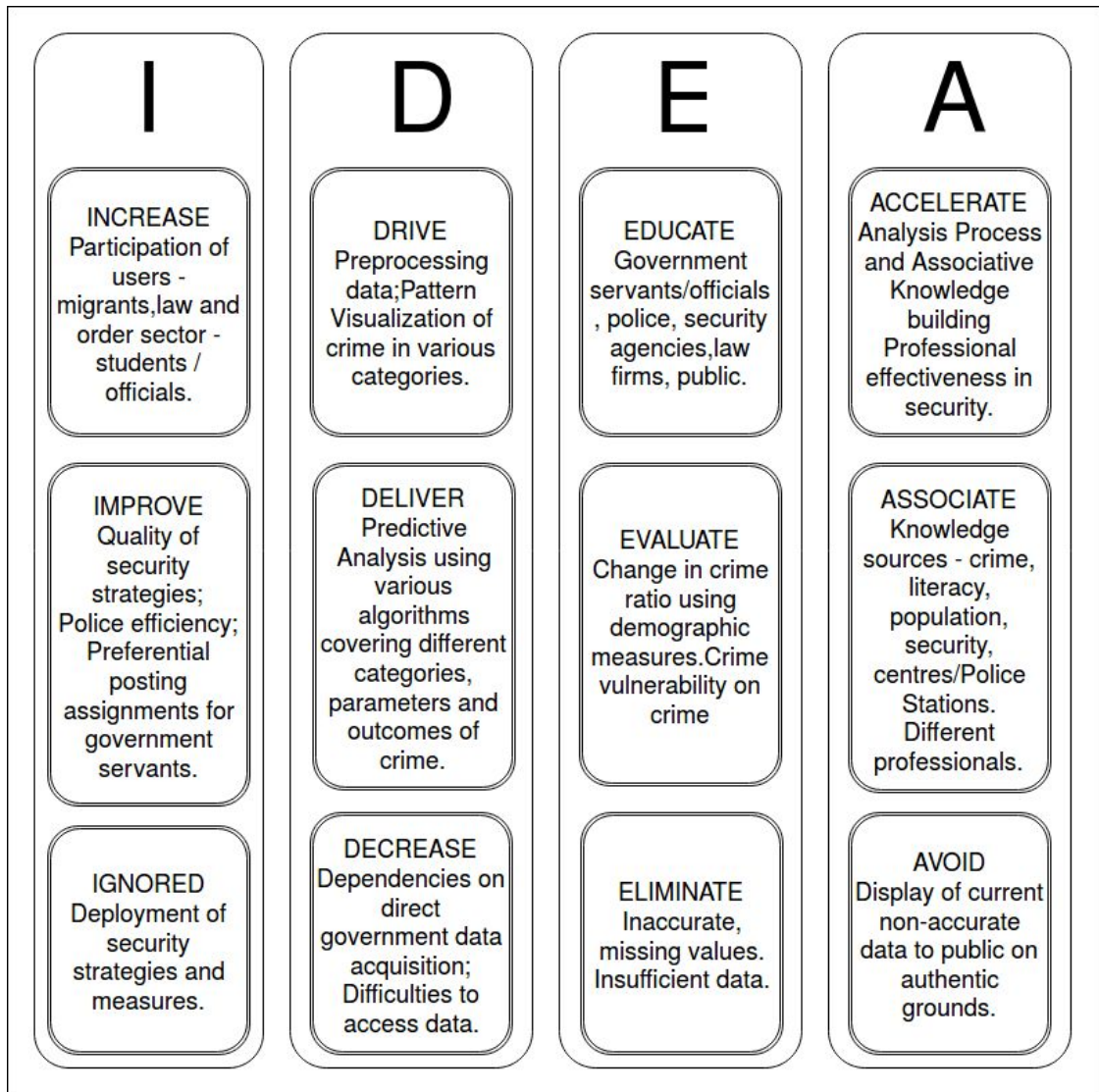
Fig 8.1 : Idea Matrix

Fig 8.2 : Knowledge Canvas of the proposed System.

- Project problem statement feasibility assessment using NP-Hard, NP Complete or satisfiability issues using modern algebra and/or relevant mathematical models.

NP-Complete:

The problem would be NP-complete as it is in the set of NP problems and the solution to the decision problem can be verified in polynomial time, and also in the set of NP-hard problems. NP is the set of yes/no problems with the following property: If the answer is yes,then there is a proof of this fact that can be checked in polynomial time.Intuitively, NP is the set of problems where we can verify a yes answer quickly if we have the solution in front of us. In this case, we are going to use the algorithm which will consider some values as testing set(generally 20% of the data) and remaining values as the training set(80% of the data). If the assumed and tested values are same or nearer to the assumed value then it is NP problem, as we have the solution to the problem.

In the system we can use the dataset of the criminal activities from 2001 to 2012 as training and can take the dataset of 2013-14 as the testing dataset.

ANNEXURE B

PROJECT PLANNER

SR. NO.	TASK NAME	START DATE	FINISH DATE	JUNE	JULY	AUG	SEPT	OCT	NOV	DEC	JAN	FEB	MAR
1	Project Scouting	17/6/2017	25/6/2017										
2	Feasibility Study	28/6/2017	07-05-2017										
3	Literature Survey	07-03-2017	07-12-2017										
4	Project Definition	15/7/2017	19/7/2017										
5	Project Interface Definition	16/8/2017	25/8/2017										
6	Requirement Gathering	28/8/2017	09-04-2017										
7	Design	14/9/2017	26/9/2017										
8	Documentation	22/9/2017	10-02-2017										
9	Phase I Implementation	10-10-2017	20/10/2017										
10	Phase I Testing	11-02-2017	11-08-2017										
11	Phase II Implementation	30/11/2017	15/2/2018										
12	Phase II Testing	18/2/2017	03-03-2018										
13	Journal Paper Publication	03-10-2018	15/3/2018										
14	Project Report Finalising	03-12-2018	20/3/2018										
15	Final Project Implementation	03-11-2018	27/3/2018										

SDLC phases (June, 2017 to March 2018)

- Requirement gathering (8 days)
- System Specification and analysis(9 days)
- System designing(high level and low level) (12 days)
- Coding and unit testing(4 months)
- Functional and nonfunctional testing(1 month)
- Documentation and Maintenance(1.5 month)

ANNEXURE C

REVIEWERS COMMENTS OF

PAPER SUBMITTED

(At-least one technical paper must be submitted in Term-I on the project de-sign in the conferences/workshops in IITs, Central Universities or UoP Con-ferences or equivalent International Conferences Sponsored by IEEE/ACM)

1. Paper Title: A Survey on Data Analytics, Predictions and Visualization of Crime Patterns in India

2. Name of the Journal where paper submitted: International Journal of Innovative Research in Computer and Electronics Engineering, Vol.5, Issue 9, September 2017, ISSN(Online)-2320-9801.

3. Paper accepted/rejected: Accepted

4. Review comments by reviewer:

5. Corrective actions if any: No corrections

ANNEXURE D

PLAGIARISM REPORT

Plagiarism report

ANNEXURE E

INFORMATION OF PROJECT

GROUP MEMBERS

```
PHOTOGRAPH

(3.5 X 4.5 aspect
ratio)
```

1. Name :

2. Date of Birth :

3. Gender :

4. Permanent Address :

5. E-Mail :

6. Mobile/Contact No. :

7. Placement Details :

8. Paper Published :

s