

**CRIME ANALYSIS AND PREDICTION USING DATA MINING**

**BY**

**JANNATUL FARIA**

**ID: 151-15-5183**

This Report Presented in Partial Fulfillment of the Requirements for the Degree of  
Bachelor of Science in Computer Science and Engineering

Supervised By

**Mr. Md. Riazur Rahman**

**Senior lecturer**

Department of CSE

Daffodil International University

Co-Supervised By

**Ms. Zerín Nasrín Tumpa**

**Lecturer**

Department of CSE

Daffodil International University



**DAFFODIL INTERNATIONAL UNIVERSITY**

**DHAKA, BANGLADESH**

**NOVEMBER 2018**

## **APPROVAL**

This Project titled “**CRIME ANALYSIS AND PREDICTION USING DATA MINING**”, submitted by **Jannatul Faria**, ID No: **151-15-5183** to the Department of Computer Science and Engineering, Daffodil International University, has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering (BSc) and approved as to its style and contents. The presentation has been held on 10<sup>th</sup> November, 2018.

## **BOARD OF EXAMINERS**

---

**Dr. Syed Akhter Hossain**  
**Professor and Head**

Department of Computer Science and Engineering  
Faculty of Science & Information Technology  
Daffodil International University

**Chairman**

---

**Dr. Sheak Rashed Haider Noori**

**Associate Professor & Associate Head**

Department of Computer Science and Engineering  
Faculty of Science & Information Technology  
Daffodil International University

**Internal Examiner**

---

**Md. Zahid Hasan**

**Assistant Professor**

Department of Computer Science and Engineering  
Faculty of Science & Information Technology  
Daffodil International University

**Internal Examiner**

---

**Dr. Mohammad Shorif Uddin**

**Professor**

Department of Computer Science and Engineering  
Jahangirnagar University

**External Examiner**

## DECLARATION

We hereby declare that, this project has been done by us under the supervision of **Mr. Md. Riazur Rahman, Senior lecturer Department of CSE** Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

### Supervised by:

---

**Mr. Md. Riazur Rahman**  
**Senior lecturer**  
Department of CSE  
Daffodil International University

### Co-Supervised by:

---

**Ms. Zerine Nasrin Tumpa**  
**Lecturer**  
Designation  
Department of CSE  
Daffodil International University

### Submitted by:

---

**Jannatul Faria**  
**ID: 151-15-5183**  
Department of CSE  
Daffodil International University

## ACKNOWLEDGEMENT

First, we express our heartiest thanks and gratefulness to almighty God for His divine blessing makes us possible to complete the final year project/internship successfully.

We really grateful and wish our profound our indebtedness to **Mr. Md. Riazur Rahman, Senior Lecturer**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of “*Field name*” to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts and correcting them at all stage have made it possible to complete this project.

We would like to express our heartiest gratitude to **Prof. Dr. Syed Akhter Hossain, Head, Department of CSE**, for his kind help to finish our project and also to other faculty member and the staff of CSE department of Daffodil International University.

We would like to thank our entire course mate in Daffodil International University, who took part in this discuss while completing the course work.

Finally, we must acknowledge with due respect the constant support and patients of our parents.

## **DEDICATION**

This book is dedicated to my caring Mother and Loving Father, for their kindness and devotion, and also to my husband for his endless support when I needed most; their selflessness will always be remembered.

## **ABSTRACT**

Crime analysis and prediction is systematic way for detecting crime, analyzing patterns of crime and predicting crime trends. Data mining, it is suitable method to apply on large quantity of crime dataset and information obtained from data mining technique is very effective that support law enforcement officers. Our system can detect and predict areas which are crime prone and have high probability for crime occurrence and can visualize crime prone areas in radius. The results of this method can be used to increase people's awareness related to the locations that are dangerous and to assist police force to predict crimes that will occur in the future in a target area within a target time. With the rapid increase and usages of computerized and informational systems, data analysts of crime can assist the police forces to fasten up the process of crime solving in our community. It is interesting to know that about 15% of the criminals commits 40% of the total occurrence. Even though it is not possible to estimate who may be the victims of a crime but we can estimate the place that has high chance for crime occurrence. K-means algorithm, is popular for cluster analysis in data mining, is done by splitting data into different subset based on their different means. Expectation - maximization algorithm is an extension where it is possible to split the data based on their available parameters. Data mining framework which is easy to implement, works with the geospatial plot of crime and criminal activities and assists to increase the effectiveness of the police forces and other law enforcement agencies. Our developed system can also be applied for the Bangladesh crime departments, by doing this the Bangladesh police can reduce the crime and solve the crimes with minimum time.

## TABLE OF CONTENTS

<b>CONTENS</b>	<b>PAGE</b>
Board of examiners	i
Declaration	iii
Acknowledgements	iv
Dedication	v
Abstract	vi
<b>CHAPTER</b>	
<b>CHAPTER 1: INTRODUCTION</b>	<b>1-6</b>
1.1 Introduction	1
1.2 Motivation	2
1.3 Rationale of the Study	3
1.4 Research Questions	5
1.5 Expected Output	5
1.6 Report Layout	6
<b>CHAPTER 2: BACKGROUND</b>	<b>7 - 13</b>
2.1 Introduction	7
2.2 Related Works	7
2.3 Research Summary	10
2.4 Scope of the Problem	11
2.5 Challenges	12

<b>CHAPTER 3: RESEARCH METHODOLOGY</b>	<b>14-20</b>
3.1 Introduction	14
3.1 Research Subject and Instrumentation	14
3.1 Data Collection Procedure	16
3.1 Statistical Analysis	17
3.5 Implementation Requirements	19
<b>CHAPTER 4: EXPERIMENTAL RESULTS AND DISCUSSION</b>	<b>21-28</b>
4.1 Introduction	21
4.2 Experimental Results	21
4.3 Descriptive Analysis	23
4.4 Implementation and Testing	25
4.5 Summary	28
<b>CHAPTER 5: SUMMARY, CONCLUSION, RECOMMENDATION AND IMPLICATION FOR FUTURE RESEARCH</b>	<b>29-32</b>
5.1 Summary of the Study	29
5.2 Conclusions	29
5.3 Recommendations	31
5.4 Implication for Further Study	32

## **APPENDIX**

**33-34**

## **REFERENCES**

**35-37**

## **LIST OF FIGURES**

<b>FIGURES</b>	<b>PAGE NO</b>
FIG: 1.1: Boston Police department Crime report	3
FIG: 1.2: Homicide Rates in Big Cities	6
FIG: 2.1: Boston Police Department Crime Map	8
FIG: 3.1: Crime Analysis Steps	15
FIG: 3.2: Boston Police Department Crime Dataset in xls format	17
FIG: 3.3: Attribute Sets	18
FIG: 3.4: WEKA, Test data file import	19
FIG: 3.5: Procedure in Data Model	20
FIG: 4.1: Naïve Bayes Classification Algorithm Result	22
FIG: 4.2: Decision Table in WEKA	23
FIG: 4.3: Crime Report Graphs I	24
FIG: 4.4: Crime Report Graphs II	24
FIG: 4.5: Home Page	25
FIG: 4.6: Search Interface of Crimes	26
FIG: 4.7: Details of Crime Report	26
FIG: 4.8: Google map generated by predicted data	27
FIG: 4.9: Planned Mobile App	27
FIG: 4.10: Interactive Google Map of Crime Prone Area	28

# CHAPTER 1

## Introduction

### 1.1 Introduction:

Crimes are a reasonably common social drawback that touching the standard of everyone's life and it additionally affects the economic process of any society [1]. it's thought of a very important driving issue that determines whether or not individuals move to a brand-new town and what places ought to be avoided once they travel specifically [2]. With the rise of crimes in our society, law enforcement agencies are requesting to demand advanced geographic data systems and new data processing techniques to boost crime analytics and better fight their criminal to protect their communities [3].

In spite of the fact that violations could occur in anyplace and everyplace, it's very basic that hoodlums take a shot at possibilities and openings they discover simple and face in most familiar regions for them [4]. By giving an information mining method to spot and locate the principal criminal inclined hotspots and understand the sort, area and time of any carried-out wrongdoings, we will in general want to increate individuals' mindfulness identifying with the unsafe regions in any bound eras. In this manner, our arranged goals will most likely encourage people in network stay away from the areas at a specific time. moreover, having this sort of information would encourage people to improve their living spot choices in an exceptionally town. On the contrary hand, police powers will utilize this goal to broaden the measure of wrongdoing forecast and look at to stop before it may happen. It will encourage inside the circulation of police powers at the most apparently wrongdoing places for some random time, to concede conservative and greatly productive utilization of police assets [5]. By having the majority of this information out there related with wrongdoings, we tend to plan to shape our locale more secure for the people living there and moreover for other people, who can go there. Criminals will think twice before committing crime they know there are system in work which can predict his motive/action even before he/she commits it. It is a great tool of modern data and information revolution to fight and combat criminals of modern times.

## **1.2 Motivation:**

Security, is seen as an important aspect, in recent times by all government and political parties around the world targeting to lower the crime incidence. Criminals and crime threat for the society in all parts of world for a long time now since rise of human civilization and precautions are required to remove crimes from our community. Crime and criminal activity prevention application to keep ordinary public safe is our mission. Current strategies of police, which is tradition and work by searching the crime and criminals, generally after the crime has occurred. But, with the help of informational and technological advancement, we can use history of crime data to create crime a usable pattern and use these created patterns to predict crimes before they actually take place.

In current situation, hoodlums and guilty parties are ending up mechanically progressed, and carrying out wrongdoing effortlessly. The test included by analyst and police organizations is that it's issue to research monster volume of learning associated in wrongdoing and psychological militant exercises so law requirements must be constrained to realize new system to avert wrongdoing and catch criminal, remain ahead inside the unending race between the offenders and furthermore the police. to understand this, appropriate procedure needs to select to perform wrongdoing examination bolstered data found. As information handling alludes to extricating or mining data from goliath measures of data, is utilized here on huge volume wrongdoing dataset and information acquired from information preparing systems is advantageous and bolster law implementations.

Out of all the grouping calculation that accessible, K-implies is that the least complex and most normally utilized bunch algorithmic program in logical and mechanical code. it's few advantages over elective bunch algorithmic program for instance, its use less machine unpredictability and suitable for group mammoth data sets. In that capacity, it's been with progress used in various reason and examines capacities, and additionally showcase division, workstation vision, geostatistics, urbanology and agribusiness. it's regularly utilized as a preprocessing venture for elective calculations that out there, for instance to look out a starting setup.

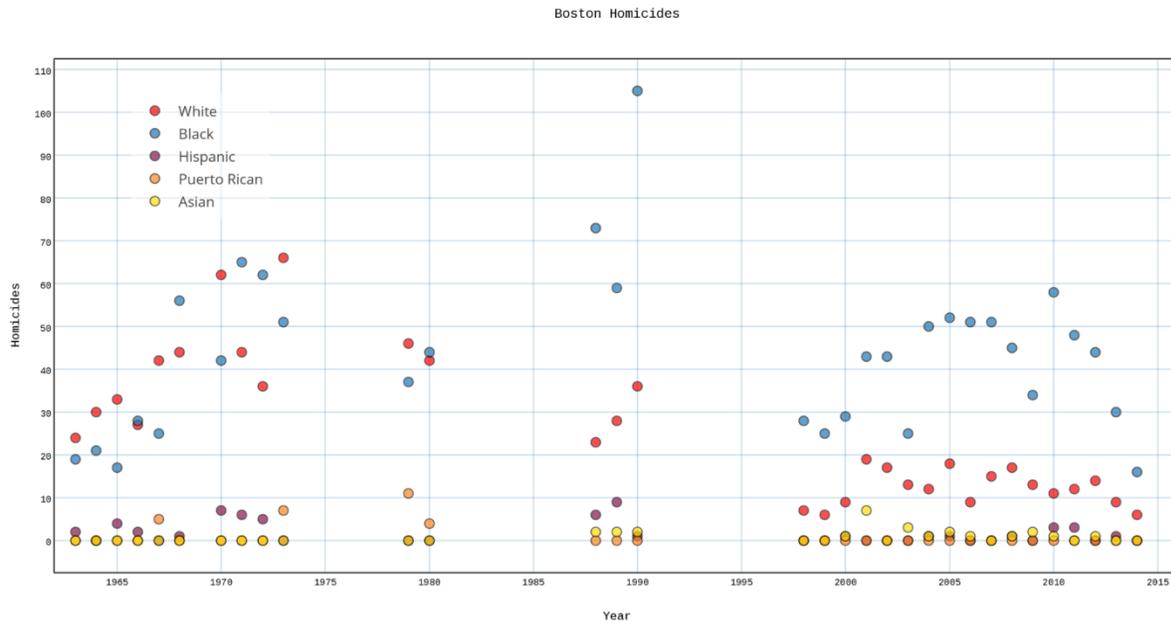


FIG: 1.1: Boston Police Department Crime report

We chose group algorithmic guideline over the other algorithmic principle like order because of wrongdoings differ in nature hugely and wrongdoing data square measure normally packed with uncertain violations. In this way, arrangement algorithmic principle that may have certainty the present and renowned settled wrongdoings, won't give sensible prognosticative factors to future violations.

### 1.3 Objectives:

Fitting information mining approach should be chosen to perform wrongdoing investigation and as grouping is relate approach of information mining. bunch groups a gathering of articles in particular implies that question inside a similar gathering region unit a great deal of comparable than those in elective groups and concerned various calculations that take issue significantly in their idea of what portrays a bunch and the best approach to with effectiveness acknowledge them. K implies that group approach information mining is utilized amid this examination to remove supportive learning from the high-volume wrongdoing dataset. To translate the dataset that help law authorizations in recognizing and break down wrongdoing examples to anticipate wrongdoing and cut back events of wrongdoing occurrence. This moreover offer data to downsize the wrongdoing inside the future occurrence by foreseeing wrongdoing. K mean group amid this

investigation is upheld by abuse open supply information preparing apparatus that territory unit diagnostic devices utilized for examining data. Among the changed offered open supply information handling suite like KNIME, WEKA, R, ORANGE, quick excavator, WEKA is picked. K implies that bunch is done with the help of fast mineworker device that is relate open supply connected science and information preparing PC code written in Java with various flexible information handling bolster decisions. Likewise, for wrongdoing investigation, a dataset is utilized recorded by the police in USA and space time from 1990 to 2011-12. amid this investigation, executing that is wrongdoing dedicated by human by slaughtering another human, is being broke down.

Grouping calculations are being utilized to investigation and anticipate wrongdoing event in wrongdoing hazardous territories. There are a few grouping calculations out there and proper to bunch the significant data into wanted bunches for examination. The monstrous volumes of wrongdoing data set likewise in light of the fact that the unpredictability of connections between these sorts of information have made human science a worthy field for applying and testing information handling strategies. Human science might be a term that centers the logical investigation of wrongdoing, criminal conduct, implementation. it's a way that expects to search out wrongdoing and criminal's attributes. It's one in everything about principal essential fields wherever the utilizations of learning mining approach will turn out fundamental imperative outcomes. recognizing wrongdoing qualities is that the activity for this technique to foresee violations. The data and learning picked up from information handling method might be a frightfully valuable instrument which may encourage and bolster law requirements. This bunching approach takes dataset and it changes over it to groups that are extra analyzed for discovering wrongdoing unsafe zones. These bunches outwardly indicate territories of violations overlaid on guide for law requirements. Groups store area of wrongdoings along the edge of various subtle elements of wrongdoing characteristics like sort and time. These bunches are arranged and ordered on the preface of their characteristic individuals. Wrongdoing inclined region unit as are thickly possessed groups, though less bunch individuals are unnoticed. Measures are taken to avert and utilized and upheld predictable with wrongdoing assortments bolstered wrongdoing hazardous zones.

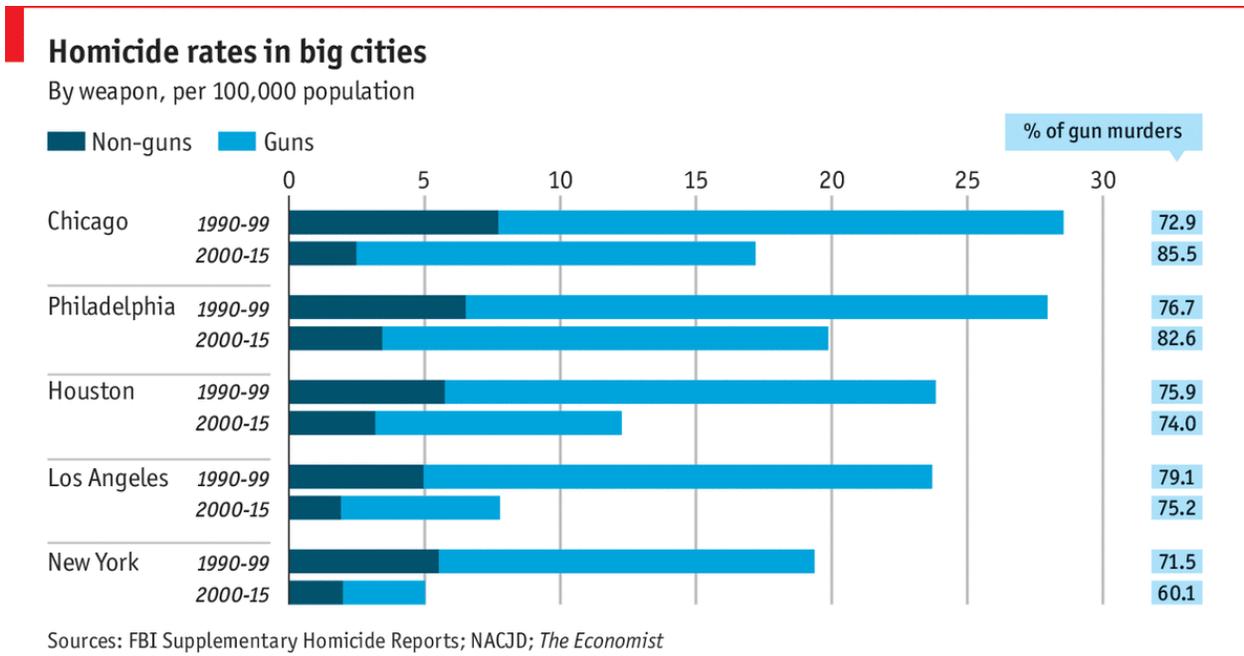
## **1.4 Research Questions & Objectives:**

The objectives of the research that is summarized in this report are to:

1. Determine and examine analysis and different study that solve the long run of crime, with a spotlight on the first two decades of the twenty first Century.
2. Determine people and community in Bangladesh and abroad that have developed crime predictions system, examine their goal and works, the character of their analysis, and therefore the strategies, instruments, and tools they used for analytics.
3. Strengths and weaknesses is analyzed and the present analysis and forecasts done by them.
4. Examine, analysis the organizations in North American country, USA, Australia so as to relatively analyze this capacities and practices in Bangladesh against different countries.
5. Provide suggestions to the Police Force on further study and measurement to be undertaken in Bangladesh.
6. Create a prediction model and prediction criminal activities before they occur using data mining tools.

## **1.5 Expected Outcome:**

Wrongdoing rate is expanding impressively step by step. the issue with wrongdoing is that it can't be prognosticated since it's neither orderly nor irregular. Additionally, the in vogue advancements and innovative methodologies encourage offenders in accomplishing their objectives effortlessly. with regards to Crime Records Bureau violations like robbery, torching, lawful offense and so forth are diminished though wrongdoings like sex misuse, murder, assault and so on. are expanded. regardless of whether it's intolerable to foresee who all could likewise be the casualties of wrongdoing be that as it may, we will anticipate the place that has chance wrongdoing predominance. the foreseen outcomes can't be 100 percent precision anyway the outcomes demonstrate that our application encourages in diminishing rate to a correct degree by giving security in wrongdoing inclined zones. The framework can be utilized by law authorization offices and they will ready to oil the territories they find most wrongdoing inclined all the more as often as possible to stop a wrongdoing that was anticipated utilizing our framework.



Economist.com

FIG: 1.2: Homicide Rates in Big Cities

The wrongdoing inclined regions where the wrongdoing event is anticipated utilizing our framework can be spoken to graphically utilizing a warmth outline demonstrates movement level, normally low action is appeared by lighter shading and high action is shown by darker shading.

### 1.6 Report Layout:

This report consists of total of five chapters. In the first, we have provided the summary and introduction of our project, it's goal and objective of doing the project. We have provided detailed explanation and motive behind choosing this project and validated it by providing crime data related to incidents.

In chapter two, we have discussed about the backgrounds of this research, the past related works that has done on the similar field and how it was successful on implementation. We have also stated the problem faced during conducting the research and how they were overcome.

In chapter three, detailed research methodology was described, various techniques that was used and implemented in our research are stated with diagrams and pictures.

Chapter four contains the results obtained from our research and discussions about it. Finally, chapter five states the conclusions of our research and how it can be improved in the future.

## **CHAPTER 2**

### **Background**

#### **2.1 Introduction:**

Crime analysis is an analytical process which gives pertinent data with respect to wrongdoing examples and pattern connections to help law implementations in arranging the arrangement of assets for the counteractive action and forecast of criminal exercises. It is critical to break down wrongdoing because of following reasons:

1. Analyze wrongdoing to educate police constrain about everyday exercises and particular wrongdoing patterns in way.
2. Analyze wrongdoing to exploit a lot of data existing in equity framework open space.

Wrongdoing rates are quickly changing worldwide and enhanced examination instruments or model empower us to discover shrouded examples of wrongdoing, assuming any, with no express earlier learning of these examples.

The fundamental goals of wrongdoing examination recorded underneath:

1. Extraction of examples from wrongdoing dataset by investigation of accessible wrongdoing and information.
2. Prediction of wrongdoing dependent on order model and circulation of existing information expectation of wrongdoing rate utilizing distinctive information mining methodologies and strategies.
3. Detection of wrongdoing.

#### **2.2 Related Works:**

Information mining in the investigation and examination of criminology can be arranged into principle regions, wrongdoing control and wrongdoing concealment. [6] Introduced a system for wrongdoing patterns utilizing another separation measure for looking at all people dependent on their profiles and afterward bunching them as needs be. [7].

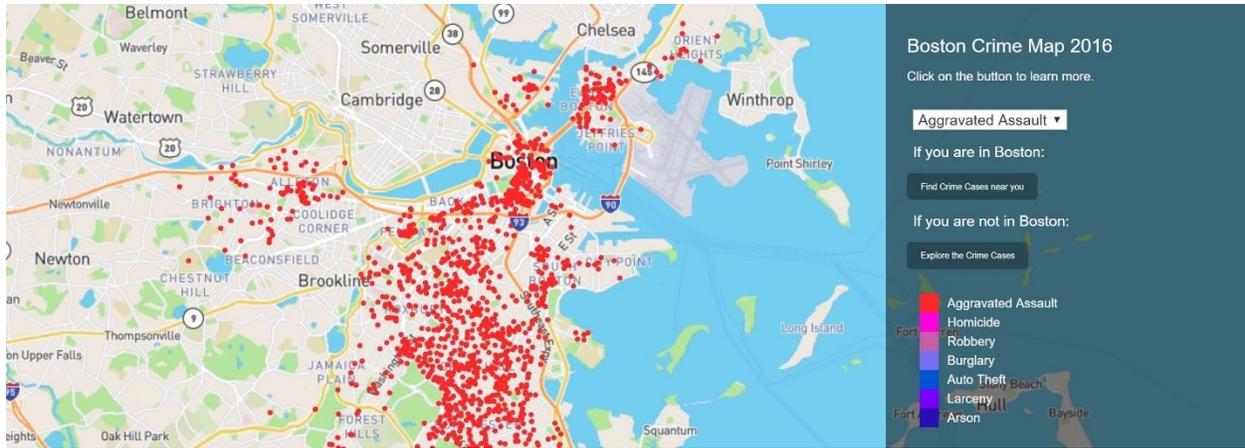


FIG: 2.1: Boston Police Department Crime Map

In nations like England, Tokyo or Sydney, Police Department have completed an investigation named Series Finder for finding the examples in thievery. For accomplishing this they utilized the business as usual of criminal and they extricated a few examples of wrongdoing which were trailed by wrongdoer. The calculation builds usual way of doing things of the criminal. The M.O. is an arrangement of propensities for a wrongdoer and is a sort of conduct used to describe an example. The information included different methods for passage (front entryway, window, and so forth.) by a criminal, day of the week they pick, attributes of the property (condo, house) they choose, and geographic vicinity to another break-ins. Utilizing nine known wrongdoing arrangement of thefts Series Finder recouped the majority of the violations inside these discovered examples and furthermore recognized nine extra violations. The anticipated outcome demonstrated over 85% exactness. Along these lines, a similar idea we are applying here in this examination, for instance: discover obscure examples from known information and dataset [8]. It's the primary numerically principled way to deal with the expectation display and robotized learning of wrongdoing arrangement. Features the current frameworks utilized by Indian police as e-administration activities and furthermore proposes an intelligent question-based web interface as wrongdoing examination device to help law implementation officers in their day by day exercises. In the interface which is utilized to remove helpful data of wrongdoings from the immense wrongdoing database kept up by National Crime Record organization and discover wrongdoing problem areas utilizing different information mining systems, for example, grouping and so on. The adequacy of the UI has been outlined on wrongdoing records. [9] examine on an improvement of Visual Interactive Malaysia Crime News Retrieval System (I-JEN) and portray the methodology,

arranged and client contemplates, the engineering of the framework and future arrangement. Their fundamental targets were to model and build wrongdoing based occasion; research the utilization of wrongdoing based occasion in enhancing the order and grouping; build up an intelligent recovery framework news identified with wrongdoing; picture news identified with violations in a powerful and intuitive way; incorporate them into a usable and strong and dependable framework and assess the ease of use and execution of the framework and the investigation will add to the better comprehension of the information identified with wrongdoing and utilization in the Malaysian setting and additionally the framework created with the perception highlights to address wrongdoing information and the inevitable objective of battling the violations. [10]

Inspects the utilization of group investigation in the given space, especially inconsistency discovery in review. The reason for his examination is to inspect the utilization of bunching innovation to robotize misrepresentation sifting and find through the identified with submitting extortion amid a review. He utilized bunch investigation to enable evaluators to center their endeavors while assessing their gathering life coverage claims which is suspicious. [11] He took a gander at the utilization of missing quality and bunching calculation for an information mining way to deal with help anticipate the wrongdoings examples and accelerate the way toward unraveling wrongdoing in a network. [12] By utilizing a bunching/group-based model to foresee wrongdoing patterns, the information mining procedures are utilized to investigate the city wrongdoing information from Police Department. The after effects of this information mining could adequately be utilized to diminish and even forestall wrongdoing for the coming a long time in future. [13] Research work concentrated on building up a wrongdoing examination instrument for Indian situation utilizing distinctive information mining approach that can help police division to viably and effectively handle wrongdoing examination. The apparatus empowers law requirement officers to effortlessly and monetarily perfect, portray and investigate wrongdoing information to distinguish conspicuous examples. [14] It introduces a structure for the wrongdoing and criminal information examination and recognition utilizing Algorithms like choice tree for information order and Simple K Means calculation for information bunching. The paper will in general help analysts and masters in finding examples and patterns, making expectations, discovering connections and conceivable clarifications of wrongdoing, mapping criminal systems and distinguishing conceivable suspects. [15] Apply distinctive arrangements of devices on wrongdoing informational collections to dig for data and to pick up learning that is escaped human

perception. With the assistance of best in class perception methods, we present the examples found through our calculations in an available, slick and natural way that empowers police divisions to all the more likely use their assets as needs be. [16] Examine the likelihood of bunching innovation and utilization of this innovation for evaluating. Computerizing misrepresentation sifting can be of a magnificent device to incentive to persistent reviews. The target of their examination is to look at the utilization of bunch investigation as an option and inventive irregularity identification strategy in the wire exchange framework. [17] Tried to catch long periods of human experience and data into PC models by means of information mining and by structuring a usable reproduction show.

### **2.3 Research Summary:**

This area outlines and looks at the examination directed and discoveries that have model of forecasts of wrongdoing for the underlying long stretches of the 21st Century. Additionally, specific rundown has been put on how the extension and nature of wrongdoing will change: Crime is relied upon to change in both quantitative and subjective terms. While this outline examinations how future wrongdoing rates may transform from those previously, its majority centers around how the idea of wrongdoing is relied upon to change over the time later on. This segment starts by condensing the key factors that exploration shows will impact the extension and nature of wrongdoing later on specifically statistic, full scale financial, and innovative variables. A short outline of research that has made particular expectations on future wrongdoing rates is then displayed. This is trailed by a survey of expectations identified with the idea of wrongdoing, with a specific spotlight on future patterns and foreseen improvements that will recognize wrongdoing in the 21st Century. This examination target and recognizes items and administrations that will be focused by culprits later on, particular sorts of new wrongdoings that are relied upon to increment later on, the degree to which wrongdoing later on will affect society, and a profile of exploited people and guilty parties of things to come.

#### **Factors Driving Future Crime Trends:**

There are various elements that impact the kind of wrongdoing, in spite of the fact that there is no particular or authoritative clarification about why the wrongdoing rate has fallen over the most recent couple of years or why it started to drastically increment in the circumstances such as 1960s.

In any case, the two most essential main thrusts that have impacted wrongdoing rates before and are required to be profoundly persuasive on future rates are full scale financial variables like quality of the economy, buyer spending, joblessness rates levels and statistic factors specifically the quantity of guys in the wrongdoing inclined age gathering. The one intriguing variable that may have had the most critical effect on wrongdoing rates as of late is innovation. It will, most likely, keep on impacting the idea of wrongdoing significantly.

### **Technology:**

While statistic and large-scale financial components are fundamental to determining the future reasons of wrongdoing, another variable that is probably going to intensely impact both the nature and extent of current and future wrongdoing is innovation. The impact of innovation on the eventual fate of wrongdoing and criminal exercises can be portrayed into three general classifications:

1. Advances in innovation will keep on giving culprits and guilty parties the apparatuses to encourage the commission of customary wrongdoings (misrepresentation, burglary, tax evasion, and falsifying).
2. Technology itself will be the objective of criminal offenses and criminal exercises (burglary of broadcast communications administrations and the spread of infections).
3. New innovation will be utilized to counteract or change criminal assault.

One of the key explanations behind the expansion in falsifying in recent years is the progression made in such advances as PCs, shading laser printing, scanners, printers and work area distributing programming. The business accessibility and also the less expenses of these innovations have expanded criminal's openness, which thus has given a more prominent number of individuals, beginners and experts alike with the chance to carry out various misrepresentation related violations in past years, which were at one time the space of exceedingly talented falsifiers or forgers who required particular hardware and aptitude.

## **2.4 Scope of the Problem:**

The inspiring inquiry for the majority of this report is basically expressed as, "What may be 'wrongdoing'?" This segment of our exploration paper is positively determined by that question, as well, however it is additionally propelled by a double inquiry, "What are 'wrongdoing insights'?"

that imparts to the principal question the with a property that it appears to be basic yet is extremely mind boggling to answer when we attempt. The simple answer is that has two essential hotspots for broadly arranged measurements on the frequency of wrongdoing: the information assembled by the Uniform Crime Reporting office and the aftereffects of the continuous National Crime Victimization Survey done by outsider association. The previous information is started on the deliberate commitment of data from nearby law authorization offices and the last are gotten from a noteworthy example review supported by the Bureau of Justice Statistics that straightforwardly talks with individuals and family units on their encounters with wrongdoing and brutality.

In spite of the fact that there is some motivation to surmise that the dimension of wrongdoings and wrongdoing rates has expanded throughout the most recent too many years of the twentieth century, it stays hard to see if the wrongdoing issue has really exacerbated or showed signs of improvement. Amid the 1980s, open worry over the dimension of inclination persuaded violations in the United States climbed definitely. Such reasons and the resulting authorization of wrongdoing rates over the United States most likely expanded, in any event to some degree, from an evident compounding of the general wrongdoing issue. Measurements from both free and legislative information gathering associations guarantees the end that wrongdoing expanded through the span of the 1980s and, to a vast degree, leveled off amid the 1990s. These insights, doesn't make a difference how frightful be that as it may, stay conflicting and deficient. In addition, the measurements accumulated toward the finish of the 1980s and all through the ahead of schedule to mid-1990s reflected not just a development in the inclination wrongdoing issue, yet in addition a development in authoritative and managerial attention to the issue.

## **2.5 Challenges:**

The most tedious part of crime and crime rate forecasting is that it is a new and hi-tech innovation which is still in its development stages. At current, it has been used in the UK, USA, China and some parts of Europe. However, the results have not been always promising. There are couple of reasons behind it. Algorithms are not developed and matured yet and prone to inaccuracies. Crime patterns cannot be predicted correctly, if the crime-related data is inconsistent. This eventually leads to poor crime forecasting and predictions.

Experts are currently of the view that predictive crime is not an all accepted and go-to crime prevention method yet. It has to be used with a real-time human intervention that starts with crime prevention at a community level itself.

Since data science and artificial intelligence are already being used in defense of many countries, it's only a matter of time before they will become an important part of law enforcement in general. As of now, crime prediction and forecasting have a lot of dependencies. If perfected, it can become one of the most useful tools for the safety of mankind at large without any doubt.

## **CHAPTER 3**

### **Research Methodology**

#### **3.1 Introduction:**

In crime rate prediction model, the primary difficulties we are confronting are:

1. Increase in wrongdoing data that must be put away and examined.
2. Analysis of information is troublesome since information is fragmented and conflicting.
3. Limitation in getting wrongdoing information records from Law Enforcement division.
4. Accuracy of the program relies upon exactness of the preparation set.

Finding the examples and patterns in wrongdoing and outrages is a testing factor. To recognize an example, wrongdoing experts takes a considerable measure of time by doing manual inquiry, examining and going through information to discover whether a specific wrongdoing fits into a known example. On the off chance that it doesn't fit into a current example, the information must be named another example. Subsequent to recognizing an example, it very well may be utilized to foresee, envision and avoid wrongdoing.

Before this grouping calculations have been utilized for wrongdoing investigation. For example, one wrongdoing it is uncovered that suspect has dark hair and from next site/witness it is uncovered that suspect is youth and from third one uncovers that the guilty party has tattoo to his left side arm and so on. By portraying the guilty party subtle elements, it gives an entire picture from various wrongdoing occurrences. Today its greater part is physically finished with the assistance of numerous reports that the criminologists generally get from the PC information examiners and their very own wrongdoing logs. Nature of violations change after some time, so with the end goal to have the capacity to distinguish more current and obscure examples in future, grouping systems work better.

#### **3.2 Research Subject and Instrumentation:**

There are steps in doing Crime Analysis:

- 1) Data Collection
- 2) Classification
- 3) Pattern Identification

- 4) Prediction
- 5) Visualization

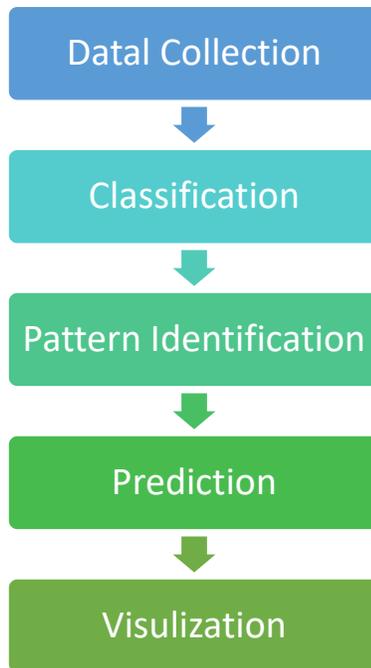


FIG: 3.1: Crime Analysis Steps

**Classification:**

For arrangement we are utilizing a calculation called Naïve Bayes which is a managed learning technique and also a measurable strategy for grouping. Guileless Bayes classifier is a probabilistic classifier which when given an info gives a solitary yield as opposed to giving a likelihood dispersion of set everything being equal. The calculation arranges a news article into a wrongdoing type to which it fits the best. From arrangement what we get is "What is the likelihood that a wrongdoing report D has a place with a given class C?" [18]. The benefit of utilizing Naive Bayes Classifier is that it is basic. Contrasted with different calculations like SVM (Support Vector Machine) which is asset hungry and takes parcel of memory and elite makes it not the same as different calculations. Additionally, if there should be an occurrence of SVM as size of preparing set expands the speed of execution is less. Utilizing Naive Bayes calculation, we make a model via preparing wrongdoing information identified with homicide, vandalism, theft, thievery, assault, sex misuse, fire related crime, equipped burglary, grabbing, a rip-off and so on. Via preparing implies we need to show them on specific data sources with the end goal that we can

test them for obscure information sources. For testing the exactness of the model, we apply test information. Not at all like SVM as the span of preparing information builds exactness of test set likewise increments. Another favorable position of Naïve Bayes is that it functions admirably for any little measure of preparing to ascertain the arrangement parameters. Additionally, it settles the Zero-recurrence issue i.e. while evaluating likelihood once in a while checking a likelihood  $P(A) * P(B/D) * P(C/D) * P(E/D)$  where  $P(C/D) = 0$ . In this way, the assessed likelihood results dependably give zero which prompts vulnerability in a few outcomes. To stay away from this sort condition, we add +1 to the tally of each zero esteem classes to accomplish uniform circulation. Test outcomes demonstrates that Naive Bayes indicates over 90% exactness since its classifications each word as tokens and expelling successive words like "the", "and", "of" and so forth which enhances precision. A word is naturally ended in the event that it happened less occasions or under multiple times. Figure demonstrates an example pseudo code of Naïve Bayes calculation. Thus, by utilizing this idea in wrongdoing article we can get more subtle elements identified with wrongdoing like unfortunate casualty and wrongdoer names, area of wrongdoing, date, time and so on [20].

### **3.3 Data Collection Procedure:**

In information accumulation step we are gathering information from various sites like news locales, online journals, internet-based life, RSS channels and so on. The gathered information is put away into database for further process. Since the gathered information is unstructured information, we utilize mySQL. Wrongdoing information is an unstructured information since the no of field, substance, and size of the report can contrast starting with one record then onto the next the better alternative is to have a diagram less database. Likewise, the nonappearance of joins diminishes the intricacy. Different advantages of utilizing an unstructured database is that:

- Large volumes of organized, semi-organized, and unstructured information.
- Object-situated programming that is anything but difficult to utilize and adaptable.

The upside of NoSQL database over SQL database is that it permits inclusion of information without a predefined blueprint. Dissimilar to SQL database it not has to realize what we are putting away ahead of time, indicate its size and so forth.

For this research, dataset was obtained from online data source <https://www.crimereports.com/>

Here Boston Police Department Crime report was downloaded and used for research purpose. Total of 300 entries was used in our test project.

INCIDENT	OFFENSE	OFFENSE_CODE_GROUP	OFFENSE_DESCRIPTION	DISTRICT	REPORTIN SHOOTING	OCCURRED_ON_YEAR	MONTH	DAY_OF	HOURL	UCR_PART	STREET	Lat	Long	Location
18204237	3006 Medical Assistance	SICK/INJURED/MEDICAL - PERSON	E5	740	01,06,18 20:49	2018	6	Friday	20	Part Three VFW PKW	42.29112	-71.1669	(42.29112036, -71.16691153)	
18204237	1402 Vandalism	VANDALISM	D4	165	01,06,18 20:30	2018	6	Friday	20	Part Two HARRISON	42.34285	-71.0652	(42.34285014, -71.06516235)	
18204237	3501 Missing Person Reported	MISSING PERSON	C11	257	01,06,18 20:08	2018	6	Friday	20	Part Three CUSHING	42.31364	-71.0632	(42.31364183, -71.06316493)	
18204237	3006 Medical Assistance	SICK/INJURED/MEDICAL - PERSON	D4	165	01,06,18 20:58	2018	6	Friday	20	Part Three ALBANY S	42.34256	-71.0627	(42.34255919, -71.06270824)	
18204236	3115 Investigate Person	INVESTIGATE PERSON	D4	144	01,06,18 20:37	2018	6	Friday	20	Part Three SAINT GEF	42.34572	-71.0856	(42.34571916, -71.08562464)	
18204236	3301 Verbal Disputes	VERBAL DISPUTE	C11	361	01,06,18 20:02	2018	6	Friday	20	Part Three DORCHES	42.29557	-71.0615	(42.29557185, -71.06152074)	
18204236	1402 Vandalism	VANDALISM	B3	430	01,06,18 19:58	2018	6	Friday	19	Part Two WOODRO	42.28576	-71.0903	(42.28576084, -71.09030349)	
18204236	3115 Investigate Person	INVESTIGATE PERSON	D14	749	01,06,18 20:17	2018	6	Friday	20	Part Three COMMON	42.33906	-71.1564	(42.33906470, -71.15642795)	
18204235	802 Simple Assault	ASSAULT SIMPLE - BATTERY	C11	490	01,06,18 18:45	2018	6	Friday	18	Part Two WASHINGTON	42.30512	-71.0802	(42.30511791, -71.08019708)	
18204235	724 Auto Theft	AUTO THEFT	E18	527	01,06,18 19:30	2018	6	Friday	19	Part One HOWE PAR	42.23631	-71.1323	(42.23631051, -71.13231083)	
18204235	3201 Property Lost	PROPERTY - LOST	B2	287	10,04,18 10:00	2018	4	Tuesday	10	Part Three WASHINGTON	42.33095	-71.0829	(42.33094752, -71.08285104)	
18204235	3831 Motor Vehicle Accident Response	M/V - LEAVING SCENE - PROPERTY DAMAGE			01,06,18 8:20	2018	6	Friday	8	Part Three	42.37053	-71.061	(42.37052953, -71.06101929)	
18204235	3802 Motor Vehicle Accident Response	M/V ACCIDENT - PROPERTY DAMAGE			01,06,18 19:18	2018	6	Friday	19	Part Three	42.30811	-71.0583	(42.30811068, -71.05830858)	
18204235	2647 Other	THREATS TO DO BODILY HARM	B3	429	01,06,18 20:04	2018	6	Friday	19	Part Two THEODOR	42.28407	-71.0882	(42.28406985, -71.08824920)	
18204235	1402 Vandalism	VANDALISM	B3	458	01,06,18 20:01	2018	6	Friday	20	Part Two GLENWAY	42.30131	-71.0844	(42.30131086, -71.08443841)	
18204235	2647 Other	THREATS TO DO BODILY HARM	B3	458	01,06,18 20:01	2018	6	Friday	20	Part Two GLENWAY	42.30131	-71.0844	(42.30131086, -71.08443841)	
18204234	802 Simple Assault	ASSAULT SIMPLE - BATTERY	A15	45	01,06,18 18:40	2018	6	Friday	18	Part Two MYSTIC ST	42.38003	-71.0625	(42.38003093, -71.06248252)	
18204234	614 Larceny From Motor Vehicle	LARCENY THEFT FROM MV - NON-ACCESSORY	A1	117	01,06,18 18:30	2018	6	Friday	18	Part One BEACH ST	42.35153	-71.0625	(42.35152810, -71.06248241)	
18204234	3115 Investigate Person	INVESTIGATE PERSON	B2	265	01,06,18 19:19	2018	6	Friday	19	Part Three PERRIN ST	42.32189	-71.0791	(42.32188540, -71.07906272)	
18204234	2647 Other	THREATS TO DO BODILY HARM	B2	276	01,06,18 19:39	2018	6	Friday	19	Part Two WILLIAM	42.33238	-71.0824	(42.33237733, -71.08237779)	
18204234	3501 Missing Person Reported	MISSING PERSON			31,05,18 7:30	2018	5	Thursday	7	Part Three NORTH PC	-1	-1	(-1.00000000, -1.00000000)	
18204234	802 Simple Assault	ASSAULT SIMPLE - BATTERY	D4	167	01,06,18 18:56	2018	6	Friday	18	Part Two E CANTON	42.33789	-71.0692	(42.33789065, -71.06919427)	
18204234	3115 Investigate Person	INVESTIGATE PERSON	E18	490	01,06,18 18:00	2018	6	Friday	18	Part Three AMERICA	42.27531	-71.1159	(42.27531111, -71.11537798)	
18204234	3831 Motor Vehicle Accident Response	M/V - LEAVING SCENE - PROPERTY DAMAGE	B3	476	01,06,18 14:30	2018	6	Friday	14	Part Three BLUE HILL	42.28817	-71.0939	(42.28817397, -71.09387677)	
18204233	3301 Verbal Disputes	VERBAL DISPUTE	B2	309	01,06,18 19:04	2018	6	Friday	19	Part Three N CHARLA	42.31967	-71.0847	(42.31966899, -71.08474412)	
18204233	3820 Motor Vehicle Accident Response	M/V ACCIDENT INVOLVING PEDESTRIAN - INI	B3	427	01,06,18 17:40	2018	6	Friday	17	Part Three ESSENDE	42.27856	-71.092	(42.27855984, -71.09199697)	
18204233	3831 Motor Vehicle Accident Response	M/V - LEAVING SCENE - PROPERTY DAMAGE	D4	156	01,06,18 18:30	2018	6	Friday	18	Part Three CONCOR	42.34052	-71.0781	(42.34051771, -71.07806475)	
18204233	3803 Motor Vehicle Accident Response	M/V ACCIDENT - PERSONAL INJURY	E13	651	01,06,18 18:05	2018	6	Friday	18	Part Three ROSEMAR	42.30423	-71.1147	(42.30422988, -71.11464399)	
18204233	3301 Verbal Disputes	VERBAL DISPUTE	D14	778	01,06,18 18:21	2018	6	Friday	18	Part Three SPARHAW	42.35148	-71.1514	(42.35147715, -71.15143461)	
18204233	3125 Warrant Arrests	WARRANT ARREST	A1	116	01,06,18 17:52	2018	6	Friday	17	Part Three HUDSON	42.35074	-71.0601	(42.35074197, -71.06006135)	
18204233	724 Auto Theft	AUTO THEFT	C11	393	30,05,18 23:30	2018	5	Wednesday	23	Part One DRACUT S	42.28742	-71.0619	(42.28741838, -71.06190780)	
18204233	3802 Motor Vehicle Accident Response	M/V ACCIDENT - PROPERTY DAMAGE	D14	812	01,06,18 17:31	2018	6	Friday	17	Part Three N HARVA	42.36669	-71.1253	(42.36668653, -71.12529557)	
18204232	2610 Other	TRESPASSING	C11	257	01,06,18 16:31	2018	6	Friday	16	Part Two STOUGHTR	42.31662	-71.0641	(42.31661750, -71.06407827)	
18204232	3115 Investigate Person	INVESTIGATE PERSON	C11	462	01,06,18 17:51	2018	6	Friday	17	Part Three BULLARD	42.30235	-71.0732	(42.30235109, -71.07319116)	
18204232	3109 Police Service Incidents	SERVICE TO OTHER PD INSIDE OF MA.	A7	5	01,06,18 18:10	2018	6	Friday	18	Part Three SARATOG	42.3849	-71.0019	(42.38490217, -71.00186562)	
18204232	619 Larceny	LARCENY ALL OTHERS	A15	56	31,05,18 14:30	2018	5	Thursday	14	Part One GREEN ST	42.3785	-71.062	(42.37850462, -71.06198887)	
18204232	1402 Vandalism	VANDALISM	E13	649	31,05,18 5:00	2018	5	Thursday	5	Part Three SAINT JOS	42.30667	-71.1161	(42.30667342, -71.11610570)	
18204232	3115 Investigate Person	INVESTIGATE PERSON	C11	450	01,06,18 17:24	2018	6	Friday	17	Part Three WASHING	42.29361	-71.0718	(42.29360585, -71.07188650)	

FIG: 3.2: Boston Police Department Crime Dataset in xls format

### 3.4 Statistical Analysis:

#### Pattern Identification:

Third stage is the example distinguishing proof stage where we need to recognize possibilities and designs in wrongdoing. For discovering wrongdoing design that happens every now and again we are utilizing Apriori calculation. Apriori can be utilized to decide affiliation rules which feature standard and general patterns in the database. The consequence of this stage is the wrongdoing design for a specific place. Here relating to every area, we take the traits of that put, climate characteristics, zone affectability, nearness of criminal gatherings, eminent occasion and so forth. In the wake of getting a general wrongdoing design for a place, when another case arrives on the off chance that it pursues the comparable wrongdoing design, we can state that the region has a possibility for wrongdoing event. Data with respect to designs lay authorization authorities to encourage assets in a successful way. They can keep away from wrongdoing event by giving security/watching in wrongdoing inclined territories, settling criminal cautions/CCTV and so

forth. Take an example rundown of 100 news for a place and apply Apriori calculation in it. It will mine the regular wrongdoing designs for a place. In this way, in the event that there is an example in which wrongdoing happened, we accept that in the event that again that design happens in a place, there is likelihood for wrongdoing event in that put. We are thinking about properties for wrongdoing design discovery. E.g.: For a place Meerut the example subsequent to mining will be:

- attribute 1, trait 2, characteristic 3, quality 4
- attribute 1, trait 3, characteristic 4, quality 5

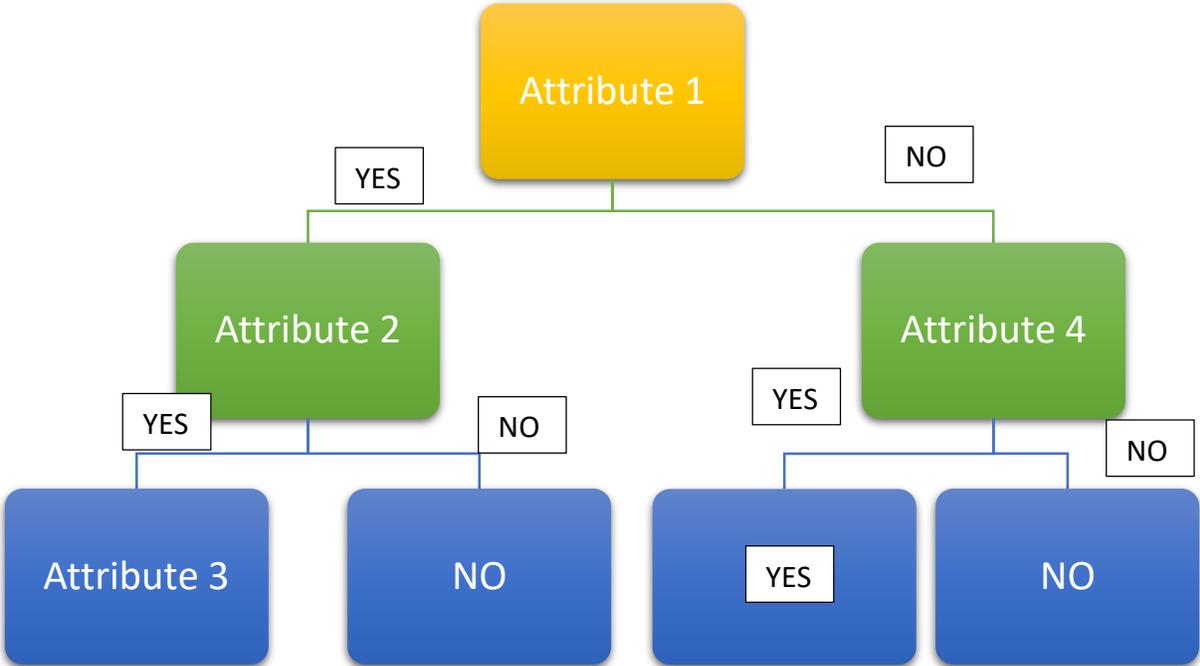


FIG: 3.3: Attribute Sets

Along these lines, the above will be the wrongdoing design for a city. Thus, wrongdoing happens just if the above examples happen on multi day or more. On the off chance that any of these examples happen, no one but, we can state that there is likelihood for wrongdoing event.

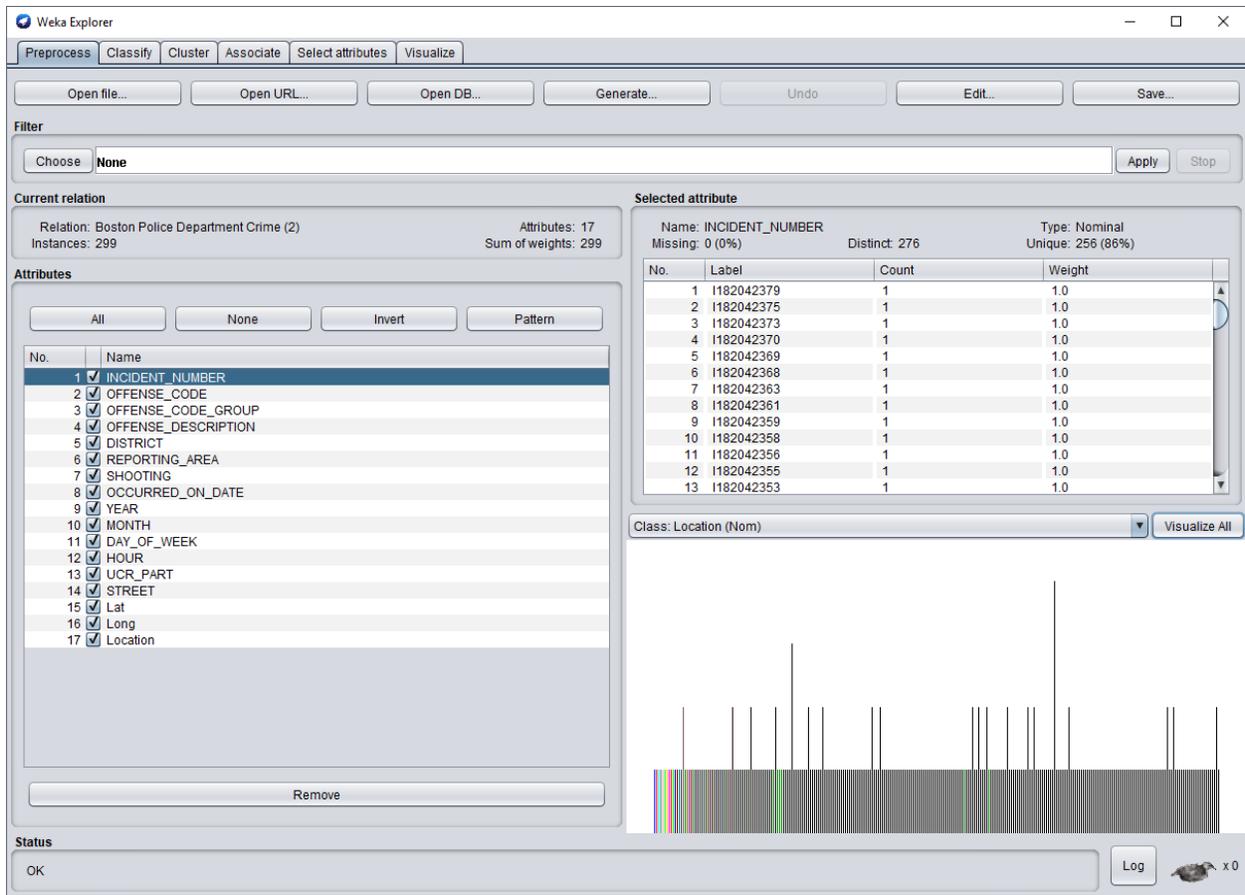


FIG: 3.4: WEKA, Test data file import

### 3.5 Implementation Requirements:

An open source data mining tool is used for analysis purpose and it can be setup very easily. Here crime analysis is conducted on dataset obtained by applying k means clustering algorithm and by using rapid miner tool.

The procedure is given below:

1. In the first stage, crime dataset is taken
2. Create a new database and apply filtering algorithm in dataset according to requirement and matching attributes according to analysis to be done
3. Read excel file of crime dataset in rapid miner tool and apply “Replace Missing value operator” on it and execute operation
4. Execute operation and perform “Normalize operator” on resultant dataset

5. Perform k means clustering on resultant dataset formed after normalization and execute operation
6. From plot view of result plot data between crimes and get required cluster
7. Cluster is formed and analysis can be conducted now

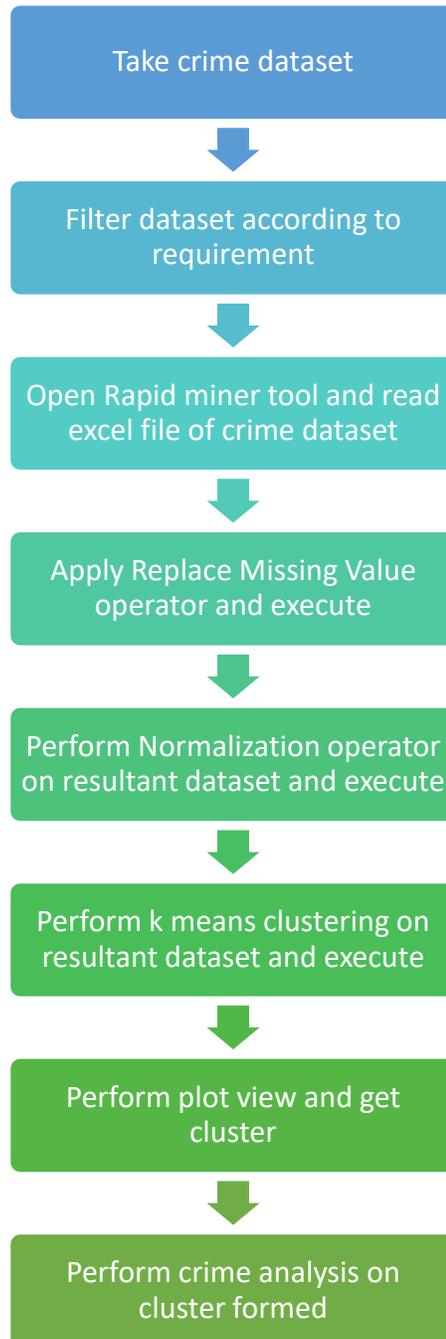


FIG: 3.5: Procedure in Data Model

## CHAPTER 4

### Experimental Results and Discussion

#### 4.1 Introduction:

For forecast we are utilizing the choice tree idea. A choice tree is like a diagram in which interior gesture speaks to test on a characteristic, and each branch speaks to result of a test. The fundamental favorable position of utilizing choice tree is that it is easy to comprehend and translate. Alternate favorable circumstances incorporate its vigorous nature and furthermore it functions admirably with substantial informational collections. This component encourages the calculations to settle on better choices about factors. Relating to each place we manufacture a model. In this way, for getting the wrongdoing inclined territories we pass current date and current qualities into the expectation programming. The outcome is demonstrated utilizing some representation components.

#### 4.2 Experimental Results:

The working of choice tree is by all accounts small befuddling yet it's extremely simple. Think about an assortment of plant animal categories. We arrange them as per arrange, sort, species and so on. Rather we need to order them into a typical classification as bushes and trees. On the off chance that another species is distinguished, we need to arrange this into any of the two classes. Fundamentally, we sort it dependent on its attributes i.e. we have an arrangement of inquiries to check whether it fulfills the conditions. In the event that first condition is fulfilled, we check the following case and on the off chance that the principal condition itself isn't fulfilled, there is no compelling reason to check the rest. Along these lines, the arrangement of inquiries and their answers can be sorted out as a choice tree. The tree has three sorts of hubs:

1. A Root hub, that has approaching edges and at least zero active edges.
2. Internal hubs, every one of which makes them approaching edge and at least two active edges.
3. Leaf hub or end hub, every one of which has precisely one approaching edge and no friendly edges.

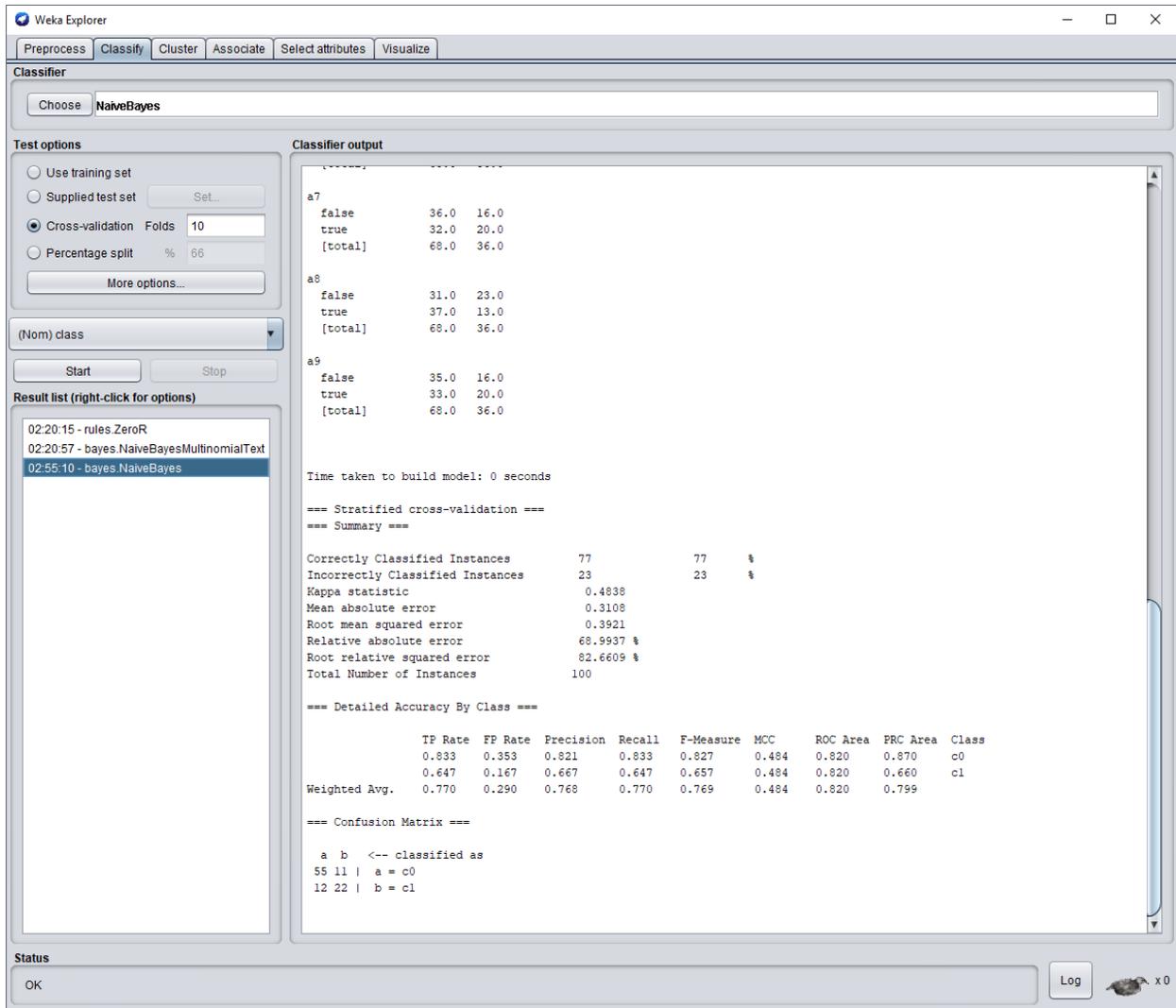


FIG: 4.1: Naïve Bayes Classification Algorithm Result

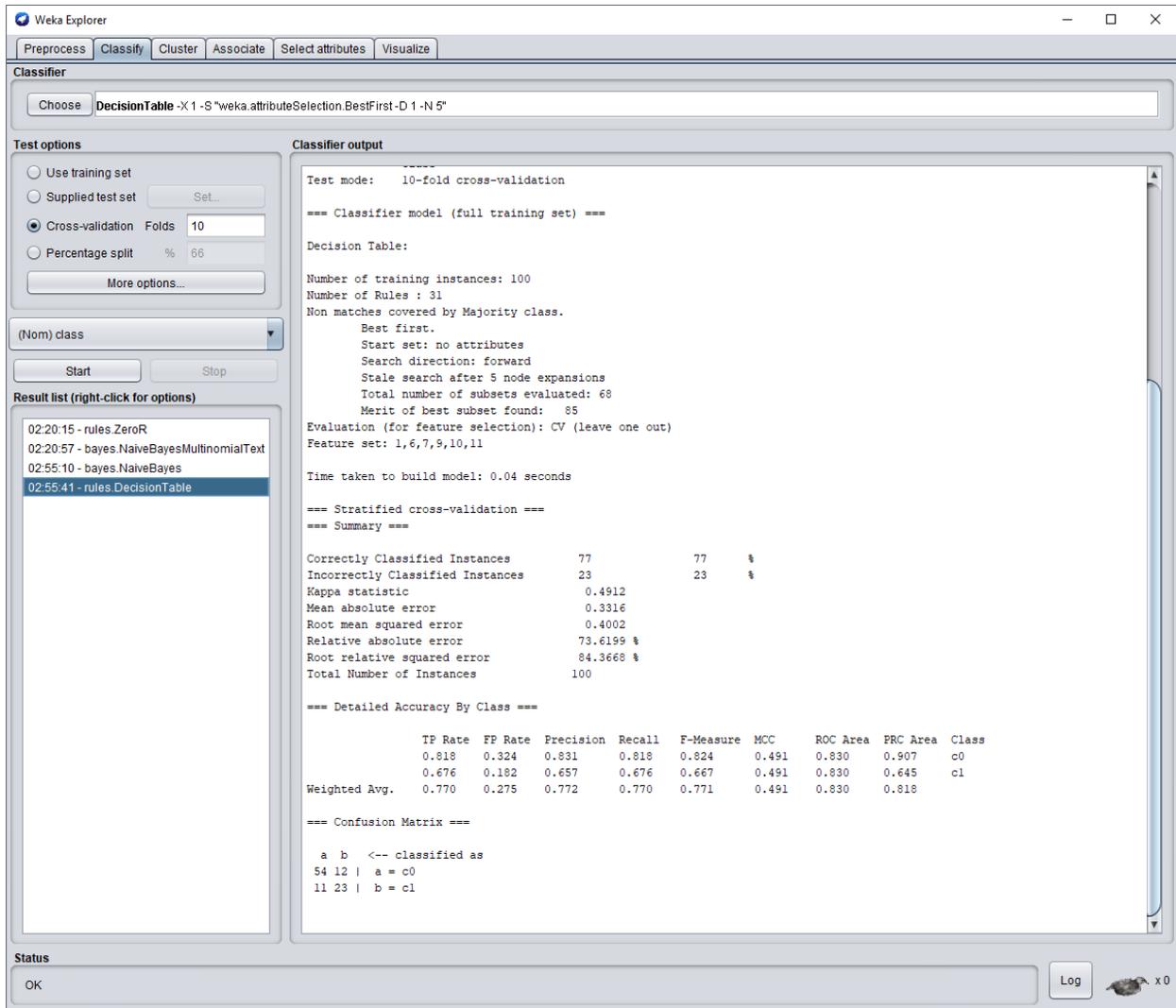


FIG: 4.2: Decision Table in WEKA

### 4.3 Descriptive Analysis:

This directed machine learning method fabricates a choice tree from an arrangement of class marked preparing tests and by utilizing this tree, tests the new examples. It is a prescient model which utilizes an arrangement of parallel standards to compute the class esteem. The tree decides:

1. Which variable to part at a hub.
2. Decision to stop or split.
3. Assign terminal hubs



#### 4.4 Implementation and Testing:

I have designed and developed a web system by using PHP, Python and MySQL to demonstrate the findings and prediction results in an interactive google map. In this system, python library 'apriori.py' analysis the data and generate the list of google map coordination and php displays that data in user interface by using html and bootstrap. Below are couple of screenshots of that system:

Home Page: It consists of 2 main function. One is to search crime reports from crime data table, stored in MySQL database. Another one is to upload crime data file to generate crime prediction.

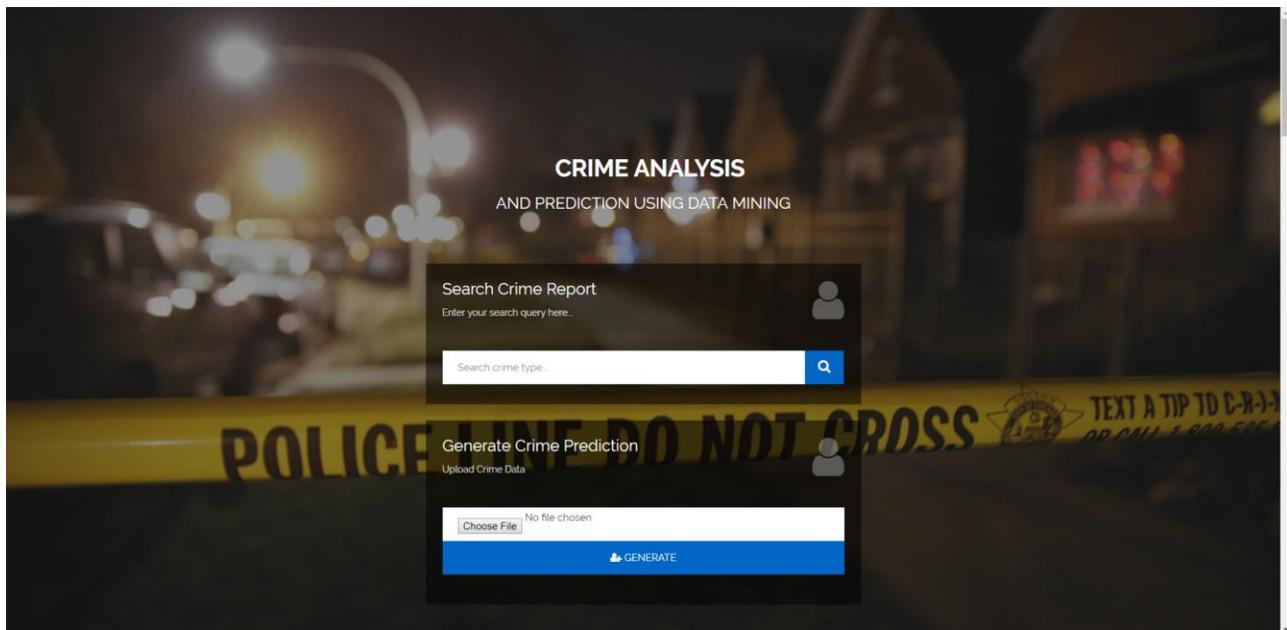


FIG: 4.5: Home Page

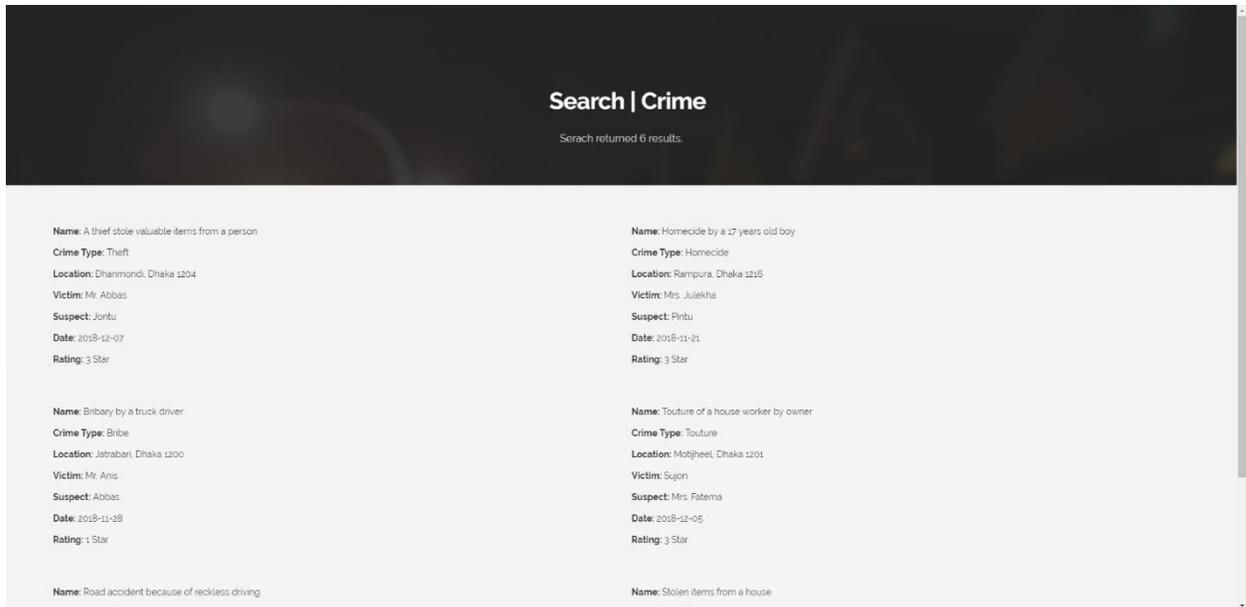


FIG: 4.6: Search Interface of Crimes

The Search page shows the list of crimes with given search keyword.

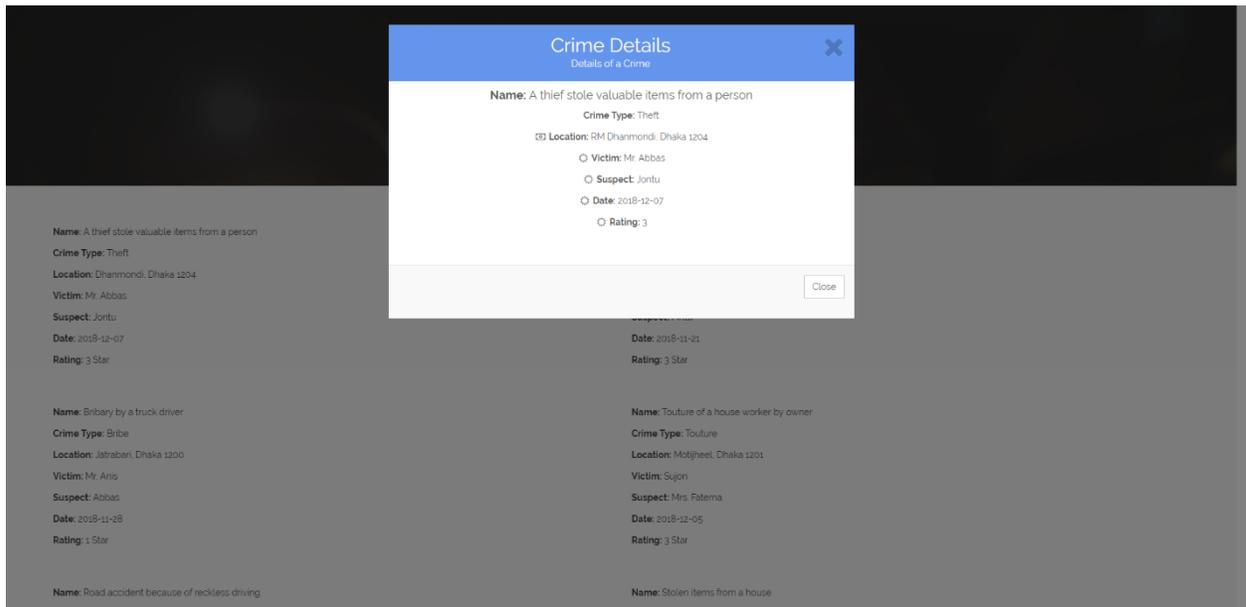


FIG: 4.7: Details of Crime Report

Upon clicking on a crime, it will display the details of that crime in a modal.

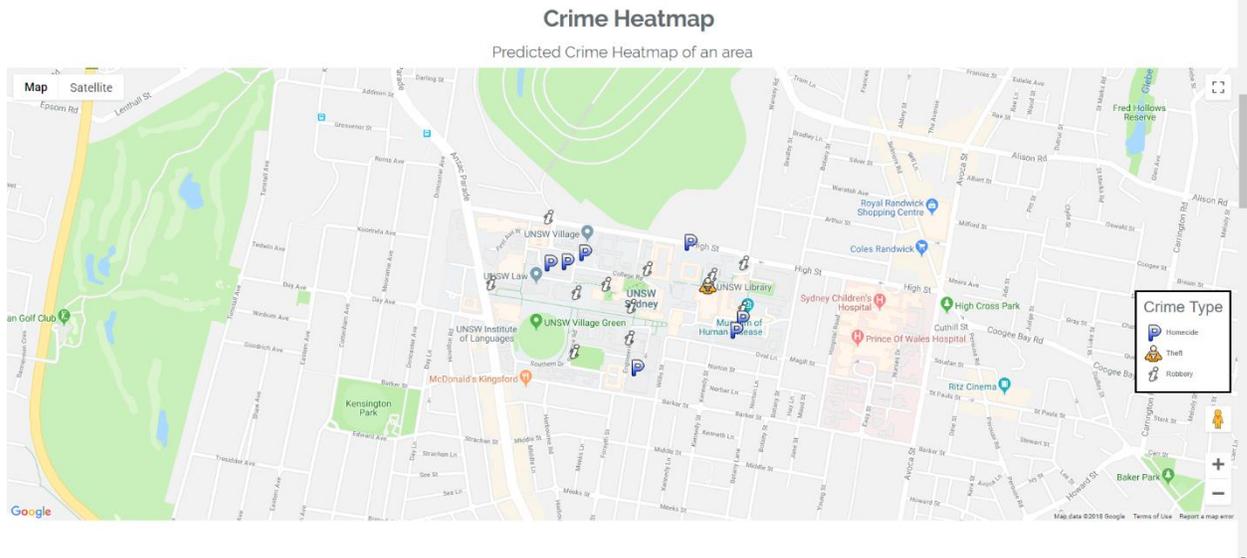


FIG: 4.8: Google map generated by predicted data

This google map is generated by analyzing given data in csv or xls file format. It displays the legend of crime and the approximate location of that crime.

FIG: 4.9: Planned Mobile App

A mobile app is planned where use will be able to upload crime that happened nearby them, find and search for crime that occurred nearby and alert others. It will also have chat options to contact real-time with the nearest police helpline.

## 4.5 Summary:

The wrongdoing inclined territories can be graphically spoken to utilizing a warmth outline demonstrates dimension of movement, generally darker hues to show low action and more splendid hues to show high action. The upsides of utilizing heat maps over other illustrative instruments are:

1. Numeric and classification based shading pictures.
2. Gradient shading range.
3. Analyze just the information we need.
4. Out of range information is consequently disposed of.

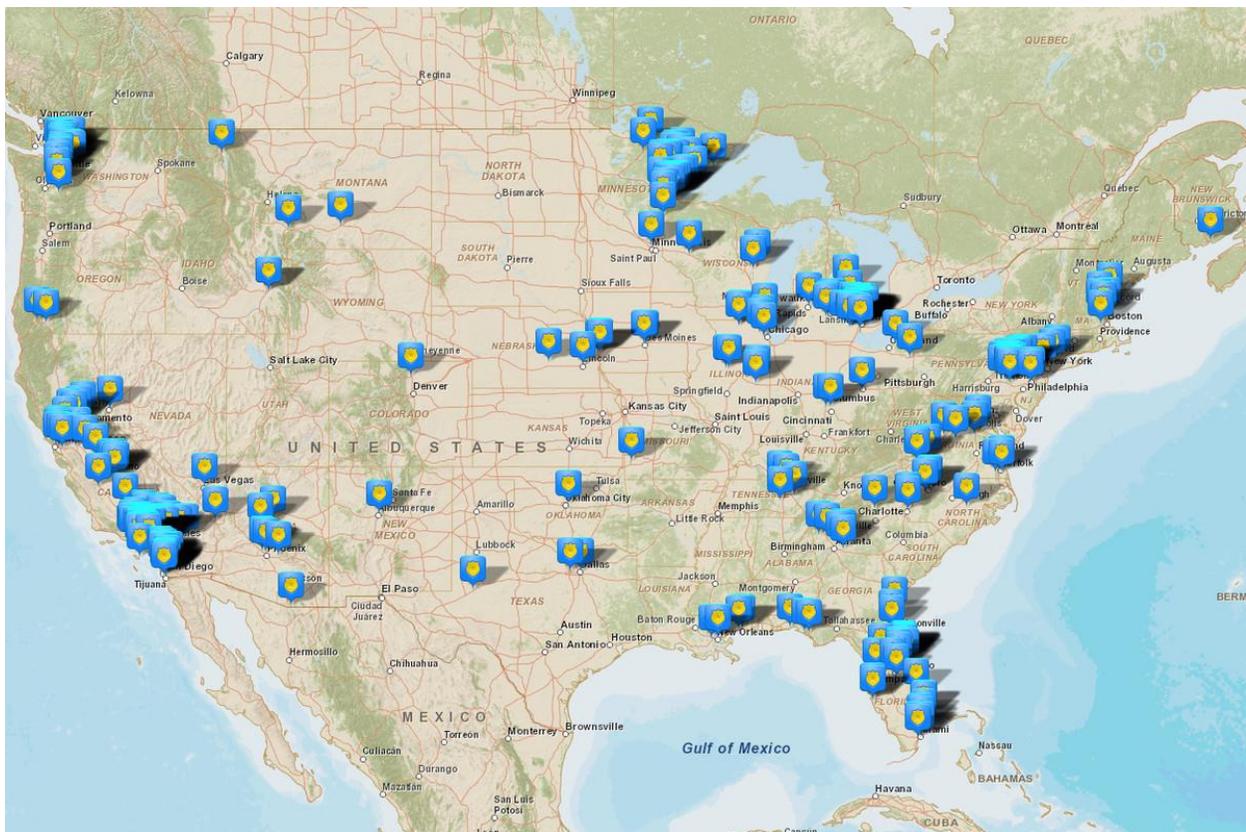


FIG: 4.10: Interactive Google Map of Crime Prone Area

Along these lines, by thinking about the plausible areas we can counteract wrongdoings by taking preventive instruments like night watching, settling thief cautions, settling CCTV camera and so on.

## CHAPTER 5

### Summary, Conclusion, Recommendation and Implication for Future Research

#### 5.1 Summary of the Study:

In this examination we have tried the precision of arrangement and forecast dependent on various test informational collections. Characterization is led dependent on the Bayes hypothesis which demonstrated over 85% exactness. Utilizing this arrangement calculation, we prepared various wrongdoing dataset and manufacture a model. For testing we embedded some test information into the model which indicates better outcomes. Our framework takes factors/characteristics of a place and Apriori calculation gives the continuous examples of that put. The example is utilized for building a model for choice tree. We created numerous diagrams and graphs and discovered intriguing insights that demonstrated the benchmark to comprehend different urban communities' violations datasets. At that point, we connected Apriori calculation to discover wrongdoing designs that are visit in these urban areas. From that point onward, we connected Decision Tree and Naïve Bayesian classifiers to help and anticipating future violations in an objective area inside a given time. We accomplished 51% of forecast precision in one city and 54% expectation exactness in the other. At long last, we gave an examination of concentrate by joining our discoveries of urban areas wrongdoing dataset with its statistic's and land data. We planned to see more about our model's discoveries and to examination and catch the components that may influence the security of a network.

#### 5.2 Conclusions:

Our structured framework predicts wrongdoing hazardous districts in a specific nation. It will be more exact in the event that we think about a specific state/area inside a given date. This investigation centers around wrongdoing examination by actualizing some bunching calculation on a given wrongdoing dataset utilizing quick digger apparatus, WEKA and here we do wrongdoing examination by thinking about wrongdoing manslaughter and plotting it with target years. From our finding and bunched results, it is anything but difficult to distinguish wrongdoing

pattern over years and can be utilized to structure more successful safeguard strategies for future as forecast show.

Identified with every region we fabricate an information show via preparing on these successive examples. Examples of wrongdoing can't be static since examples change are inclined to change after some time. Via preparing implies we are encouraging the framework dependent on some specific sources of info and datasets. Thus, the framework consequently takes in the example change in wrongdoing by assessing the wrongdoing designs. Wrongdoing factors changes after some time as well. By going through the wrongdoing information, we need to recognize new wrongdoing factors that prompt individuals carrying out wrongdoing. For showing signs of improvement results in forecast, we need to discover more wrongdoing characteristics of spots as opposed to changing a few traits. Till now we figured out how to prepared our framework utilizing a few properties yet we are wanting to incorporate more model to enhance precision later on.

### **5.3 Recommendations:**

Our structured framework predicts wrongdoing hazardous districts in a specific nation. It will be more exact in the event that we think about a specific state/area inside a given date. This investigation centers around wrongdoing examination by actualizing some bunching calculation on a given wrongdoing dataset utilizing quick digger apparatus, WEKA and here we do wrongdoing examination by thinking about wrongdoing manslaughter and plotting it with target years. From our finding and bunched results, it is anything but difficult to distinguish wrongdoing pattern over years and can be utilized to structure more successful safeguard strategies for future as forecast show.

Identified with every region we fabricate an information show via preparing on these successive examples. Examples of wrongdoing can't be static since examples change are inclined to change after some time. Via preparing implies we are encouraging the framework dependent on some specific sources of info and datasets. Thus, the framework consequently takes in the example change in wrongdoing by assessing the wrongdoing designs. Wrongdoing factors changes after some time as well. By going through the wrongdoing information, we need to recognize new wrongdoing factors that prompt individuals carrying out wrongdoing. For showing signs of improvement results in forecast, we need to discover more wrongdoing characteristics of spots as

opposed to changing a few traits. Till now we figured out how to prepared our framework utilizing a few properties yet we are wanting to incorporate more model to enhance precision later on.

#### **5.4 Implication for Further Study:**

As a future augmentation of this undertaking, we intend to apply more characterization models to expand expectation exactness and to improve the general execution of the framework. It is likewise a supportive expansion for our examination to consider the salary data of a network with the end goal to check whether there are connections between a territory's pay level and their wrongdoing rate. Additionally, we plan to investigate numerous different urban areas and socioeconomics topographical data with its wrongdoing discoveries. Besides, we need to think about different violations datasets from new urban communities alongside their statistic datasets.

From the empowering results from this undertaking, we feel that wrongdoing information mining has a promising future for expanding the adequacy and proficiency of criminal and insight investigation. Visual and natural criminal and knowledge examination procedures can be produced for wrongdoing designs. As we have tried and connected different bunching strategies for information digging for wrongdoing examination, we can likewise perform different procedures of information mining. Additionally, we can perform investigation on different dataset, for example, wide scale study dataset, business dataset, destitution dataset, help viability dataset, and so on.

Notwithstanding this another idea called Criminal profiling which helps the wrongdoing specialists and police power to record the attributes of offenders in a specific city or network. It is an extremely precise device for profiling the attributes or points of interest of culprits or guilty parties. It is a conduct and insightful apparatus that is planned to help law authorization officers to precisely foresee and profile the qualities of obscure hoodlums or guilty parties. Criminal profiling's principle objective is that:

- to furnish law implementation with a social and mental appraisal of the criminal;
- to assess things found in the ownership of the criminal.

For doing this we need to break down the foundation of criminal and their records for gathering the most extreme related information. In this way, the most extreme points of interest of every guilty party are gathered from wrongdoing records. At the point when wrongdoings like murder happen in a specific place then from reports like FIR, we get the criminal's subtle elements and

their method of task. Subsequent to getting these subtle elements we can think about the guilty parties with these practices. Along these lines, filtering going through every wrongdoing record after a specific wrongdoing event is weighty assignment, rather we will ready to utilize some perception systems to speak to the wrongdoer's subtle elements in a shape that justifiable by human.

## APPENDIX

How much crime can there be within the Bangladesh within the next five or ten years? Is it possible that the crime rate will go up or go down, that totally depends on many factors? Can crime rate stay regarding the same? Can the amount of juvenile serious violent offenders/homicide perpetrators increase? what is going to be the ensuing demands on the juvenile and criminal justice systems? Over the past 3 decades, criminologists have created variety of tries to handle these and connected queries. These typically have taken the shape of efforts to clarify past variations or to project future levels of crime by applying techniques of demographic and applied mathematics analysis. These techniques generally consist of:

1. the application of demographic age standardization strategies to mix comparatively correct estimates of the age structure of the yank population with age-specific arrest rates for numerous styles of crimes and classes of the population to calculate expected numbers of criminal offenses or crime rates or
2. the construction of AN instructive time-series regression or structural equation models to elucidate or predict variations in crime rates over time.

Criminal killing (murder and nonnegligent manslaughter): The will-full (nonnegligent) killing of 1 person by another. Excluded square measure deaths caused by negligence, tries to kill, assaults to kill, suicides, accidental deaths, traffic fatalities, and excusable homicides. excusable homicides square measure restricted to the killing of a felon by an enforcement officer within the line of duty and also the killing of a felon, throughout the commission of a law-breaking, by a non-public national.

The purposes of this paper are twofold. First, we tend to review variety of living demographic projections of crime rates and offenses that are created for the us over the past few decades, with a special specialize in projections of juvenile crime rates and offenses. we start within the next section with a quick outline of demographic analyses of the rise within the Sixties supported the approaching older of the baby boomers. this can be followed by a review of projections of

downturns in crime rates within the Eighties supported the smaller “baby bust” birth cohorts. a lot of recently, following the increase in delinquent and criminal offenses by adolescents and teenagers within the 1985-1993 amount, criminologists have made some shivery projections, that we tend to next describe, of accelerating numbers of violent criminal offenses expected within the amount 1995-2005, because the “echo boomers” enter their young years.

## Reference:

- [1] A. Bogomolov, B. Lepri, J. Staiano, N. Oliver, F. Pianesi and A. Pentland, 'Once Upon a Crime: Towards Crime Prediction from Demographics and Mobile Data', CoRR, vol. 14092983, 2014.
- [2] R. Arulanandam, B. Savarimuthu and M. Purvis, 'Extracting Crime Information from Online Newspaper Articles', in Proceedings of the Second Australasian Web Conference - Volume 155, Auckland, New Zealand, 2014, pp. 31-38.
- [3] A. Buczak and C. Gifford, 'Fuzzy association rule mining for community crime pattern discovery', in ACM SIGKDD Workshop on Intelligence and Security Informatics, Washington, D.C., 2010, pp. 1-10.
- [4] M. Tayebi, F. Richard and G. Uwe, 'Understanding the Link Between Social and Spatial Distance in the Crime World', in Proceedings of the 20th International Conference on Advances in Geographic Information Systems (SIGSPATIAL '12), Redondo Beach, California, 2012, pp. 550-553.
- [5] S. Nath, 'Crime Pattern Detection Using Data Mining', in Web Intelligence and Intelligent Agent Technology Workshops, 2006. WI-IAT 2006 Workshops. 2006 IEEE/WIC/ACM International Conference on, 2006, pp. 41,44.
- [6] Tong Wang, Cynthia Rudin, Daniel Wagner, and Rich Sevieri. Detecting patterns of crime with series finder. In Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECMLPKDD 2013), 2013.

- [7] De Bruin ,J.S.,Cocx,T.K,Kosters,W.A.,Laros,J. and Kok,J.N(2006) Data mining approaches to criminal career analysis ,”in Proceedings of the Sixth International Conference on Data Mining (ICDM’06) ,Pp. 171-177
- [8] Manish Gupta<sup>1\*</sup>, B.Chandra<sup>1</sup> and M. P. Gupta<sup>1</sup>,2007 Crime Data Mining for Indian Police Information System
- [9] Nazlena Mohamad Ali<sup>1</sup>, Masnizah Mohd<sup>2</sup>, Hyowon Lee<sup>3</sup>, Alan F. Smeaton<sup>3</sup>, Fabio Crestani<sup>4</sup> and Shahrul Azman Mohd Noah<sup>2</sup> ,2010 Visual Interactive Malaysia Crime News Retrieval System
- [10] Sutapat Thirprungsri Rutgers University .USA ,2011 Cluster Analysis of Anomaly Detection in Accounting Data : An Audit Approach 1
- [11] A.Malathi ,Dr.S.Santhosh Baboo. D.G. Vaishnav College,Chennai ,2011 Algorithmic Crime Prediction Model Based on the Analysis of Crime Clusters.
- [12] Malathi.A <sup>1</sup> ,Dr.S.Santhosh Baboo <sup>2</sup> and Anbarasi . A <sup>31</sup> Assistant professor ,Department of Computer Science ,Govt Arts College ,Coimbatore , India . <sup>2</sup> Readers , Department of Computer science , D.G. Vaishnav Collge ,Chennai , India , 2011 An intelligent Analysis of a city Crime Data Using Data Mining
- [13] Malathi , A; Santhosh Baboo , S, 2011 An Enhanced Algorithm to Predict a Future Crime using Data Mining
- [14] Kadhim B.Swadi al-Janabi . Department of Computer Science . Faculty of Mathematics and Computer Science .University of Kufa/Iraq , 2011 A Proposed Framework for Analyzing Crime DataSet using Decision Tree and Simple K-means Mining Algorithms.
- [15] Aravindan Mahendiran, Michael Shuffett, Sathappan Muthiah, Rimmy Malla, Gaoqiang Zhang,2011 Forecasting Crime Incidents using Cluster Analysis and Bayesian Belief Networks

[16] Sutapat Thiprungsri,2012 Cluster Analysis for Anomaly Detection in Accounting Data : An Audit Approach1

[17] K. Zakir Hussain, M. Durairaj and G. Rabia Jahani Farzana ,2012 Application of Data Mining Techniques for Analyzing Violent Criminal Behavior by Simulation Model

[18] [http://www.cs.sunysb.edu/~cse634/lecture\\_notes/07apriori.pdf](http://www.cs.sunysb.edu/~cse634/lecture_notes/07apriori.pdf). Last accessed:12-April-2014, 10: AM

[19] <http://www-users.cs.umn.edu/~kumar/dmbook/ch4.pdf>. Last accessed:14-April-2014, 1:00 PM.

[20] Ned Horning . Introduction to Decision trees and Random Forests, American Museum of Natural History's Center for Biodiversity and Conservation.

[21] Lafferty, McCallum, and Pereira (2001); Sutton and McCallum(2010).“<http://aliasi.com/lingpipe/demos/tutorial/classify/read-me.html> [ 2010].

[22] <http://nlp.stanford.edu/software/jenny-ner-2007.pdf>. Last accessed :24-Feb-2014.

[23] Wikipedia contributors.(9 July 2013 ), Stanford NLP. [Online].Available :<http://www-nlp.stanford.edu/software/dcoref.shtml>. Last accessed: 24- Feb-2014, 10:00 AM.

[24] Wikipedia contributors.(12 May 2014 at 19:05.), Series Finder. [Online].Available:[http://en.wikipedia.org/wiki/Crime\\_analysis](http://en.wikipedia.org/wiki/Crime_analysis), Lastaccessed: 12-Feb-2014, 12:00 PM.