**ACKNOWLEDGEMENT**

A project of this magnitude and nature requires kind co-operation and support from many, for successful completion . I wish to express my sincere thanks to all those who were involved in the completion.

My sincere thanks to **Honorable Founder and Chairman,Dr. JEPPIAAR, M.A., Ph.D.,** for his sincere endeavor in educating me in his premier institution.

I would like to express my deep gratitude to **Our Beloved Secretary and Correspondent, Dr. P. CHINNADURAI, M.A., Ph.D.,** for his kind words and enthusiastic motivation which inspired me a lot.

I also express my sincere thanks to **Our Dynamic Directors, Mrs. C. VIJAYA RAJESHWARI and Mr. C. SAKTHI KUMAR, M.E.,** for providing me with the necessary facilities for the completion.

I also express our appreciation and gratefulness to my **Principal, Dr. K. MANI, M.E., Ph.D.,**  who helped us in the completion of the project.

I wish to convey my thanks and gratitude to my **Head of the Department**, **Dr. M. HELDA MERCY,**  Department of Information Technology, for her support and by providing us ample time to complete our project.

I express my indebtedness and gratitude to my **staff in charge,**  **Mr. KARTHIKEYAN B,** Professor , Department of Information Technology for his guidance throughout the course of my project.

I thank my mother and friends for providing their extensive moral support and encouragement during the course of the project.

Last but never the least, I thank God Almighty for showering his abundant grace upon me so that I could complete the project successfully on time.

**DECLARATION**

I hereby declare that the project report entitled **CRIME PATTERN PREDICTION** which is being submitted in partial fulfilment of the requirement of the course leading to the award of the ‘Bachelor Of Technology in Information Technology’ in **Panimalar Engineering College, Affiliated to Anna University- Chennai** is the result of the project carried out by me under the guidance and supervision of **Mr B. KARTHIKEYAN** **, PROFESSOR in the department of Information Technology**. I further declare that I or any other person has not previously submitted this project report to any other institution/university for any other degree/ diploma or any other person.

Date:

Place**: Chennai** **(Your name )**

It is certified that this project has been prepared and submitted under my guidance.

Date: (**Your** **Guide Name** )

Place: **Chennai**  ( Designation / IT)

**TABLE OF CONTENTS**

|  |  |  |
| --- | --- | --- |
| **CHAPTER NO.** | **TITLE** | **PAGE NO** |
|  | **ABSTRACT** | **5** |
|  | **LIST OF TABLES** | **6** |
|  | **LIST OF FIGURES** | **7** |
| **1** | **INTRODUCTION** | **8** |
|  | 1.1 OVERVIEW OF THE PROJECT | 9 |
|  | 1.2 NEED FOR THE PROJECT | 9 |
|  | 1.3 OBJECTIVE OF THE PROJECT | 10 |
|  | 1.4 SCOPE OF THE PROJECT | 10 |
| **2** | **LITERATURE SURVEY** | **11** |
|  | 2.1 ReCAP (Regional Crime Analysis program) | 11 |
|  | 2.2 Clustering | 11 |
|  | 2.3 Criminal classification | 12 |
|  | 2.5 COMPARISONS DRAWN |  |
|  | 2.6 FEASIBILITY STUDY | 13 |
| **3** | **SYSTEM DESIGN** | **16** |
|  | 3.1 PROPOSED SYSTEM ARCHITECTURE DESIGN | 17 |
|  | 3.1.1 Block diagram for proposed system | 17 |
|  | 3.2 MODULE DESIGN | 20 |
|  | 3.2.1 Machine Learning | 20 |
|  | 3.2.2 Dataset | 21 |
|  | 3.2.3 Pre-processing | 22 |
|  | 3.3 IMPLEMENTATION METHODOLOGY | 24 |
| **4** | **REQUIREMENT SPECIFICATION** | **33** |
|  | 4.1 HARDWARE REQUIREMENT | 33 |
|  | 4.2 SOFTWARE REQUIREMENT | 33 |
|  | 4.2.1 Features of Python | 33 |
|  | 4.2.2 Anaconda Platform | 34 |
| **5** | **IMPLEMENTATION** | **34** |
|  | 5.1 SAMPLE CODE | 35 |
|  | 5.2 SAMPLE SCREEN SHOTS | 46 |
| **6** | **TESTING AND MAINTENANCE** | **51** |
|  | 6.1 SOFTWARE TESTING | 51 |
|  | 6.1.1 Black box testing | 52 |
|  | 6.1.2White box testing | 53 |
|  | 6.2 MAINTENANCE | 53 |
| **7** | **CONCLUSION AND FUTURE WORKS** | **54** |
|  | 7.1 CONCLUSION | 55 |
|  | 7.2 FUTURE WORKS | 55 |
|  | **REFERENCES** | 57 |

**ABSTRACT**

Crime event prediction is important to crime prevention in the society by helping the law enforcement agencies to design optimal patrol strategies. Reduction of crime events will benefit society in numerous ways. It will increase the public safety and decrease the economic loss. However, crime event prediction is a challenging task. To be better prepared to respond to criminal activity, it is important to understand patterns in crime. In our project, we analyze crime data from the city of Indore, scraped from publicly available website of Indore Police. At the outset, the task is to predict which category of crime is most likely to occur given a time and place in Indore. The use of AI and machine learning to detect crime via sound or cameras currently exists, is proven to work, and expected to continue to expand. The use of AI/ML in predicting crimes or an individual’s likelihood for committing a crime has promise but is still more of an unknown. The biggest challenge will probably be “proving” to politicians that it works. When a system is designed to stop something from happening, it is difficult to prove the negative. Companies that are directly involved in providing governments with AI tools to monitor areas or predict crime will likely benefit from a positive feedback loop. Improvements in crime prevention technology will likely spur increased total spending on this technology. We also attempt to make our classification task more meaningful by merging multiple classes into larger classes. Finally, we report and reflect on our results with different classifiers, and dwell on avenues for future work.

**LIST OF TABLES**

|  |  |  |
| --- | --- | --- |
| **S.No.** | **Table Name** | **Page Number** |
| 1 | Police Dataset | 17 |
| 2 | Dataset after preprocessing | 18 |
| 3 | Tests | 44 |

**LIST OF FIGURES**

|  |  |  |
| --- | --- | --- |
| **S.NO** | **TITLE OF THE FIGURE** | **PAGE NO** |
| 1.1 | Machine learning Process | 15 |
| 2.1 | Use case diagram | 18 |
| 2.2 | Activity diagram | 19 |
| 2.3 | Sequence diagram | 20 |
| 2.4 | System architecture | 21 |
| 3.1.1 | Principle diagram of KNN | 24 |
| 3.1.2 | Shows graphical representation of KNN | 25 |
| 3.1.3 | Distance functions | 26 |
| 3.2.1 | Decision tree | 27 |
| 4.2.2 | Decision Tree example | 27 |
| 4.3.1 | Random Forest Example | 28 |
| 4.3.2 | Decision Tree of Crime Pattern Prediction | 29 |
| 4.4.1 | Act13(Gambling vs Hour) | 29 |
| 4.4.2 | Act323(Violence vs Hour) | 30 |
| 4.4.3 | Act363(Kidnapping vs Hour) | 30 |
| 4.4.4 | Act379(Robbery vs Hour) | 31 |
| 4.4.5 | Act302(Murder vs Hour) | 31 |
| 4.4.6 | Act279(Accident vs Hour) | 32 |
| 5.1 | Predicting Surges | 46 |
| A.1 | Snapshot 1 | 46 |
| A.2 | Snapshot 2 | 47 |
| B.1 | Snapshot 3 | 47 |
| B.2 | Snapshot 4 | 48 |
| C.1 | Snapshot 5 | 48 |
| C.2 | Snapshot 6 | 49 |
| C.3 | Snapshot 7 | 49 |

1. **INTRODUCTION**

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

Many factors are relevant to the possibility that a particular type of crime event is going to occur in a region in the near future. These include demographics, the distribution of different types of services, crime history, human mobility and so on.

For a given city with R regions, we aim to identify the regions where a certain type of crime event will happen in the next time interval. Several types of crime event are studied including Theft, Unlawful Entry, Drug Offence, Traffic Related Offence, Fraud, and Assault. The criminal offence which includes the illegal taking of other persons belonging without consent to permanently or temporarily deprive the owner is defined as Theft. When a person enters a building (e.g. office, bank, shop etc.) with an intention to commit crime can be classified as Unlawful Entry. Drug Offence includes any form of sale, dealing, importing or exporting, manufacture or cultivation of illegal drugs or other substances. The offences which are related to most forms of road traffic, including pertaining to the licensing, registration, road worthiness or use of vehicles, bicycle offences, and pedestrian offences can be classified as Traffic Related Offence. All types of physical and mental harm towards a person are defined as Assault. This includes all types of physical contact with a person without their consent.

In this study, with an aim for short-term crime event prediction we partition a day into total eight intervals and each interval spans 3 hours. Crime prediction in finer temporal grain will help the police to design their patrol strategy dynamically and it will increase the probability to reduce crime rate more effectively. It is to be noted that the time span is independent of the methodology applied. User can change the interval span in different application scenarios. We derive a series of features and categorize them into four categories: historical, demographic, geographic and dynamic. We derive the historical features from the crime event records. They describe the density and trend of crime event in a region and the surrounding regions. The demographic features reflect the socio-economic conditions of residents in a region. The geographic features retain the information about the properties of venues in a region. While the crime event prediction has been studied by exploring historical, demographic, geographic features, the unique aspects of this work lies in the proposal of a series of dynamic features in crime event prediction. We extract the dynamic features from check-ins of Foursquare users. For a particular Foursquare user, the visiting history is associated with her/his habits and routines According to routine activity theory, the opportunities for crime events are optimized in places where victims and offenders come together in greater concentrations. For example, a location with visitors from diverse backgrounds in a time interval is highly correlated with some types of crime event such as Theft; monitoring the fluctuation of visitor diversity at locations provides useful information to the crime event prediction.

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

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

The output is the class of crime that is likely to have occurred. We try out multiple classification algorithms, such as KNN (K-Nearest Neighbors), Decision Trees, and Random Forests.

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

**1.1 OVERVIEW OF THE PROJECT**

Madhya Pradesh's commercial capital Indore has topped the crime record in the country in 2008 followed by Bhopal and Jaipur. Crime rate of Indore was 941.4, which is the highest in the country, according to National Crime Record Bureau's (NCRB) report - "Crime in India 2008”. With the rapid urbanization and development of big cities and towns, the graph of crimes is also on the increase. This phenomenal rise in offences and crime in cities is a matter of great concern and alarm to all of us. There are robberies, murders, rapes and what not. The frequent and repeated thefts, burglaries, robberies, murders, killings, rapes, shoplifting, pick pocketing, drug- abuse, illegal trafficking, smuggling, theft of vehicles etc., have made the common citizens to have sleepless nights and restless days. They feel very insecure and vulnerable in the presence of anti-social and evil elements. The criminals have been operating in an organized way and sometimes even have nationwide and international connections and link

**1.2 NEED FOR THE PROJECT**

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

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

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

The main aim of this research work consists of developing analytical data mining methods that can systematically address the complex problem related to various form of crime. Thus, the main focus is to develop a crime analysis tool that assists the police in • Detecting crime patterns and perform crime analysis

• Provide information to formulate strategies for crime prevention and reduction

• Identify and analyze common crime patterns to reduce further occurrences of similar incidence

• The present research work proposes the use of an amalgamation of data mining techniques that are linked with a common aim of developing such a crime analysis tool. For this purpose, the following specific objectives were formulated

• To develop a data cleaning algorithm that cleans the crime dataset, by removing unwanted data to explore and enhance clustering algorithms to identify crime patterns from historical data

• To explore and enhance classification algorithms to predict future crime behavior based on previous crime trends

• To develop anomalies detection algorithms to identify change in crime patterns.

**1.3 OBJECTIVE OF THE PROJECT**

The objective of our work is to:

* + - Predicting crime before it takes place.
    - Predicting hotspots of crime.
    - Understanding crime pattern.
    - Classify crime based on location.
    - Analysis of crime in Indore

**1.4 SCOPE OF PROJECT**

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

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

1. **LITERATURE SURVEY:**

### 2.1 ReCAP(Regional Crime Analysis Program)

### D.E. Brown constructed a software framework called ReCAP(Regional Crime Analysis Program) for mining data in order to catch professional criminals using data mining and data fusion techniques. In 2009, Li Ding et al.[11] propose an integrated system called PerSearch that takes a given description of a crime, including its location, type, and the physical description of suspects(personal characteristics or vehicles) as input. To detect suspects, the system will process these inputs through four integrated components: geographic profiling ,social network analysis, crime profile, and physical matching. Essentially, geographic profiling determines where" the suspects are, while other components determine the suspects. De Bruin et. al. (2006) introduced a framework for crime trends using a new distance measure for comparing all individuals based on their profiles and then clustering them accordingly .This method also provided a visual clustering of criminal career sand identification of classes of criminals .From the literature study, it could be concluded that crime details increasing to very large quantities running into zota bytes(1024bytes). This in turn is increasing the need for advanced and efficient techniques for analysis. Data mining as an analysis and knowledge discovery tool has immense potential for crime data analysis. As is the case with any other new technology, the requirement of such tool changes, which is further augmented by the new and advanced technologies used by criminals. All these facts confirm that the field is not yet mature and needs further investigations.

2.2 CLUSTERING

Bruin et al. proposed a technique which is used to determine the clustering of criminals based on the criminal careers. The criminal profile per offense per year is extracted from the database and a profile distance is calculated. After that, the distance matrix in profile per year is created. The distance matrix including the frequency value is made to form clusters by using naïve clustering algorithm. They made a criminal profile which is established in a way of representing the crime profile of an offender for a single year. With this information, the large group of criminals is easily analyzed and they predicted the future behavior of individual suspects. It will be useful for establishing the clear picture on different existing types of criminal careers They tested the tool on actual Dutch National Criminal Record Database for extracting the factors for identifying the criminal careers of a person.

2.3 CRIMINAL CLASSIFICATION

Sheehy et al. [10] came up with a research idea which was geared towards the treatment of the mentally ill people inside the prison. According to their work, the mentally ill criminals are identified using their Social Security Number (SSN) with all the criminal personal records and their crime career records attached. As the outcome, the Criminals are classified into “high”, “medium” and “low” levels of recidivism risk potential according

Their objective was to describe and classify the criminals into a misdemeanor and a felony which can be referred and not referred based on the mental health of the criminals. Their ill activities are monitored and data collection is continuous. By these, the criminals can be separated from other criminals who are hazardous and those who can cause damage to other inmates along with them. Further, their study also involves the classification of the mental health of the criminals into two categories i.e. “referred” and “not-referred”. This helps the guards to identify the prisoners who are referred for the mental health check-up. The research work they had undergone will provide a summary of the inmates who are seriously mentally ill and those who are to be separated from the other inmates.

**2.6 FEASIBILITY STUDY**

Feasibility study is of prime importance in any project management scenario especially as per the Software Engineering practices. It takes into concern a number of factors such as the technical, operational and economic factors. This kind of study is important to find out whether the system so developed is feasible to use or not. The feasibility study is done at the earlier stages so as to reduce the cost involved in reconfiguring the entire system during testing.

**Steps in Feasibility study**

* Form a project team and appoint a project leader.
* Prepare system flowcharts.
* Enumerate potential candidate system.
* Describe and identify characteristics of candidate systems.
* Form a project team and appointing a project leader.
* Prepare system flowcharts.
* Enumerate potential candidate system.
* Describe and identify characteristics of candidate systems.
* Determine and evaluate performance and cost effectiveness of each candidate system.
* Weight system performance and cost data.
* Select the best candidate system.
* Prepare and report the final project directive to management.

The document provide the feasibility of the project that is being designed and lists various areas that were considered very carefully during the feasibility study of this project such as Economic feasibility, Technical feasibility and Operational feasibility.

**Economic Feasibility**

This study is carried out to check the economic impact that the system will have on the organization. The amount of fund that the company can pour into the research and development of the system is limited. The expenditures must be justified. Thus the developed system as well within the budget and this was achieved because most of the technologies used are freely available. Only the customized products had to be purchased.

The fund required by the company for the research and development of the system. The expenditures faced by the company. For the developed system to be within the budget, the technologies which are freely available and the customized products are to be used.

The following are some of the important financial questions asked during preliminary investigation:

* The costs conduct a full system investigation.
* The cost of the hardware and software.
* The benefits in the form of reduced costs or fewer costly errors.

The system is economically feasible because it considers all the pros and cons of the construction and the implementation of the project and hence it is considered as feasible economically.

**Technical Feasibility**

The system must be evaluated from the technical point of view first. The assessment of this feasibility must be based on an outline design of the system requirement in the terms of input, output, programs and procedures. Having identified an outline system, the investigation must go on to suggest the type of equipment, required method developing the system, of running the system once it has been designed. Technical issues raised during the investigation are:

* Does the existing technology sufficient for the suggested one?
* Can the system expand if developed?

The project should be developed such that the necessary functions and performance are achieved within the constraints. The project is developed within latest technology. Through the technology may become obsolete after some period of time, due to the fact that never version of same software supports older versions, the system may still be used. So there are minimal constraints involved with this project. As the system has been developed using Java the project is technically feasible.

This study is carried out to check the technical feasibility, that is, the technical requirements of the system. Any system developed must not have a high demand on the available technical resources. This will lead to high demands being placed on the client. The developed system must have modest requirements, as only minimal or null changes are required for implementing this system.

**Social Feasibility**

This study is to check the level of acceptance of the system by the user. This includes the process of training the user to use the system efficiently. The level of acceptance by the users solely depends on the methods that are employed to educate the user about the system and to make him familiar with it. The training process for the users becomes so easier while this system is into the use and so the acceptance level also is higher and the system will be familiar and so this project is also socially feasible.

**Operational Feasibility**

The aspect of study is to check the level of acceptance of the system by the user. This includes the process of training the user to use the system efficiently. The user must not feel threatened by the system, instead must accept it as a necessity .the level of acceptance by the users solely depends on the methods that are employed to educate the user about the system and to make him familiar with it. His level of confidence must be raised so that he is also able to make some constructive criticism, which is welcomed, as he is the final user of the system.

**3.2.1 Machine learning**

Arthur Samuel, a pioneer in machine learning and artificial intelligence defined machine learning as a field of study that gives computers the ability to learn without being explicitly programmed . In essence, machine learning is a computer system's method of learning by way of examples. There are many machine learning algorithms available to users that can be implemented on datasets. However, there are two major types of learning algorithms: supervised learning and unsupervised learning algorithms. Supervised learning algorithms work by inferring information or "the right answer" from labeled training data. The algorithms are given a particular attribute or set of attributes to predict. Unsupervised learning algorithms, however, aim to find hidden structures in unlabeled class data. In essence, the algorithms learn more about the dataset as it is given more examples to be implemented on. There are five types of machine learning algorithms that are used to conduct analysis in the field of data mining: (i) Classification Analysis Algorithms - These algorithms use the attributes in the dataset to predict values for one or more variables that take discrete values. (ii) Regression Analysis Algorithms - These algorithms use the attributes of a dataset to predict values for one or more variables that take continuous values (e.g., profit/loss). It a statistical tool used in the process of investigating the relationships between variables. (iii) Segmentation Analysis Algorithms - Divide data into groups or clusters of items that have similar properties. (iv) Association Analysis Algorithms - Find correlations between different attributes in a dataset. Typical application of such type of algorithms involves creation of association rules, which can be used in market basket analysis. (v) Sequence Analysis Algorithms

The term machine learning refers to the automated detection of meaningful patterns in data. In the past couple of decades it has become a common tool in almost any task that requires information extraction from large data sets. We are surrounded by a machine learning based technology: search engines learn how to bring us the best results (while placing portable ads), anti-spam software learns to filter our email messages, and credit card transactions are secured by a software that learns how to detect frauds. Digital cameras learn to detect faces and intelligent personal assistance applications on smart-phones learn to recognize voice commands. Cars are equipped with accident prevention systems that are built using machine learning algorithms.

Machine learning is also widely used in scientific applications such as bioinformatics, medicine, and astronomy. One common feature of all of these applications is that, in contrast to more traditional uses of computers, in these cases, due to the complexity of the patterns that need to be detected, a human programmer cannot provide an explicit, fine detailed specification of how such tasks should be executed. Taking example from intelligent beings, many of our skills are acquired or reined through learning from our experience (rather than following explicit instructions given to us). Machine learning tools are concerned with endowing programs with the ability to learn and adapt

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

    longitude), class of crime

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

The output is the class of crime that is likely to have occurred. We try out multiple classification algorithms, such as KNN (K-Nearest Neighbors), Decision Trees, and Random Forests.

## 

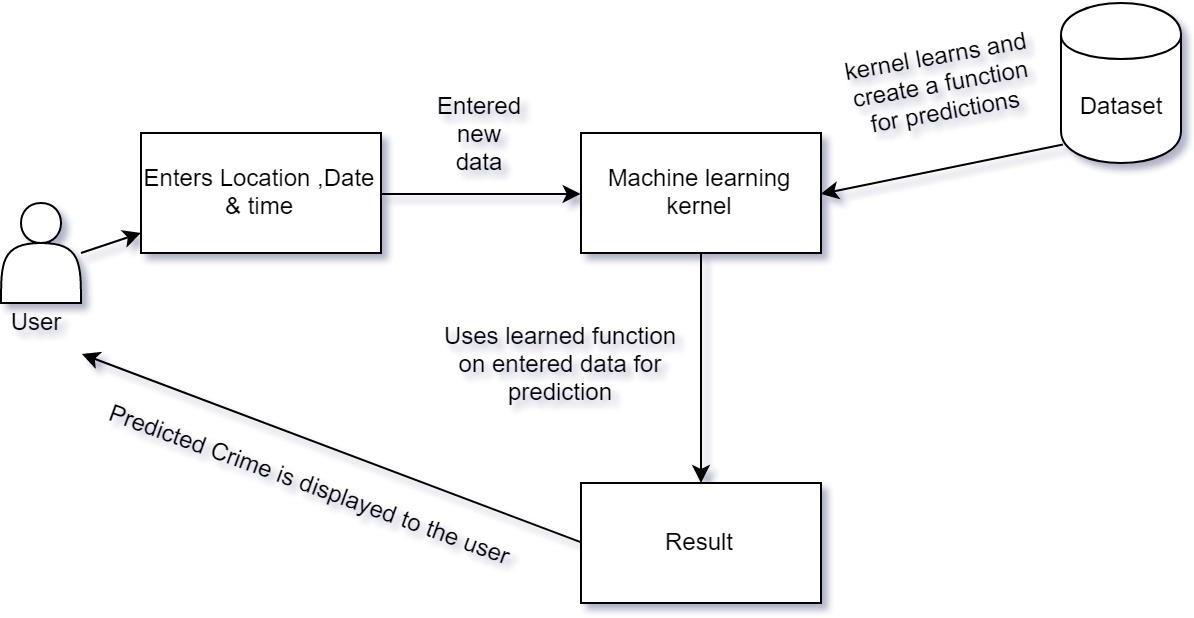
1. **SYSTEM DESIGN**

**3.1 PROPOSED SYSTEM ARCHITECTURE DESIGN**

The system architectural design is the design process for identifying the subsystems

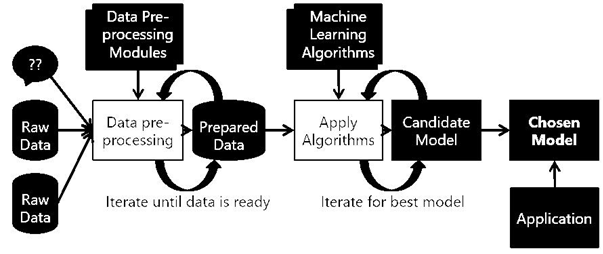
making up the system and framework for subsystem control and communication. The

goal of the architectural design is to establish the overall structure of software system.



**Fig 3.1-System architecture of Crime Pattern Prediction**

**3.1.1 Block Diagram for Proposed System**

****

**3.2 Module Design**

3.1.1 Use case diagram

Use case diagram represent the overall scenario of the system. A scenario is nothing but

a sequence of steps describing an interaction between a user and a system.

Thus use case is a set of scenario tied together by some goal. The use case diagram are

drawn for exposing the functionalities of the system

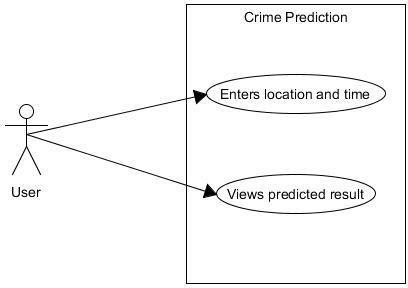


Fig 3.2-Use case diagram of Crime Pattern Prediction

3.1.2 Activity diagram

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

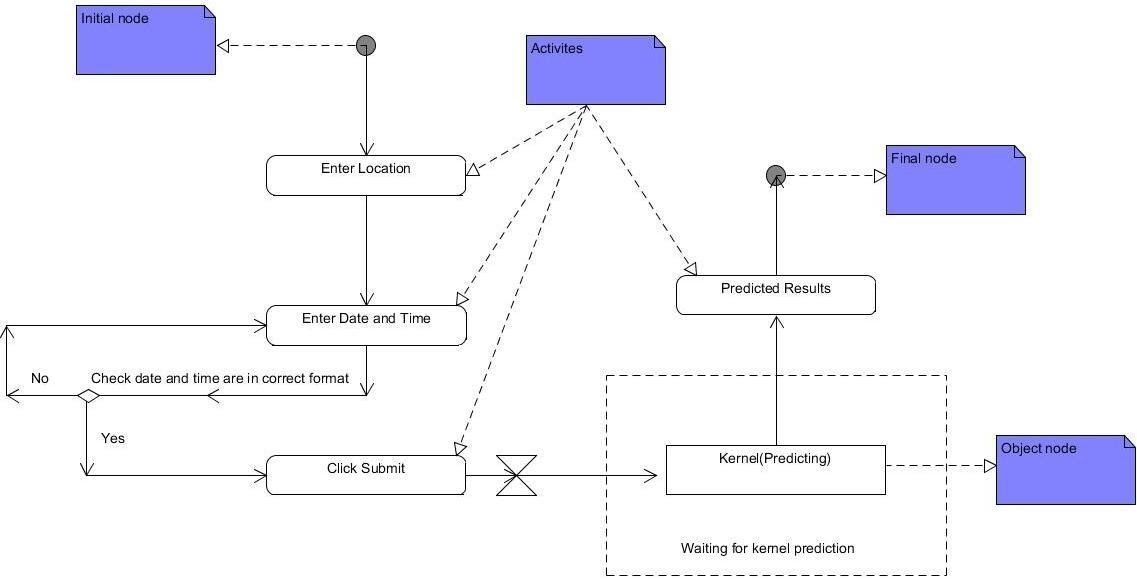


Fig 3.3-Activity diagram of Crime Pattern Prediction

3.1.3 Sequence diagram

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

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

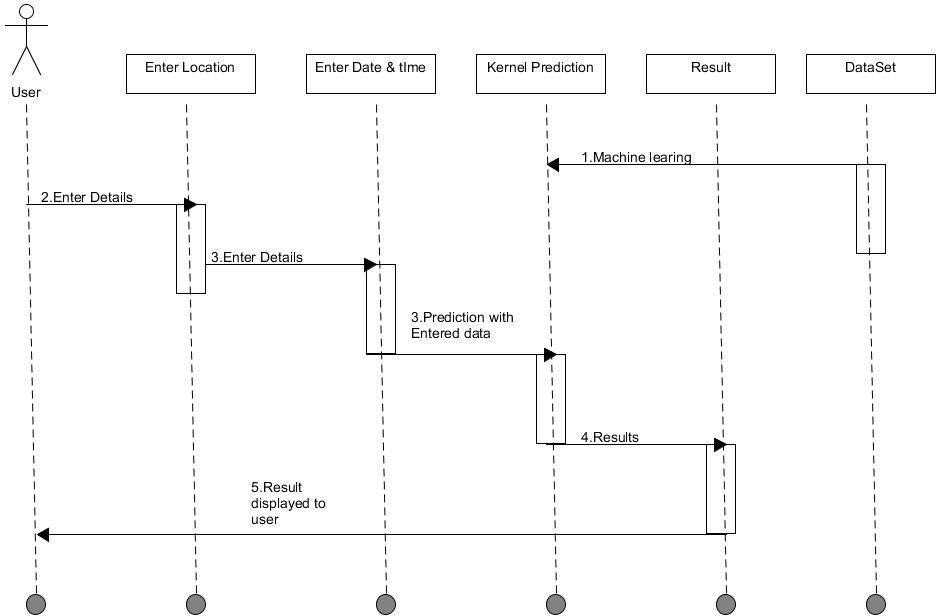


Fig 3.4-Sequence diagram of Crime Pattern Prediction

## 3.2.2 Our Dataset

Dataset which we are using is scraped daily from website of Indore police which is publically available.

But the dataset is Hindi and in order to perform machine learning this data cannot be used as it is.

Hence the data needs to be processed Features of this dataset

* + - * थाना : Police Station
      * थाना अपराध/मगग क्रमाांक : Police Station identification number
      * धारा : I.P.C. act number
      * फररयादी का नाम एवां पता : Complainant name & address
      * आरोपी का नाम एवां पता : Accused name & address
      * घटना स्थल : Incident place
      * घटना ददनाांक व समय : Incident date & time
      * कायमी ददनाांक व समय : Reporting date & time
      * दवलांब से कायमी का कारण : Reason of Time delay in reporting to police
      * घटना के  कारण सदित दववरण :  Incident information in brief

**Table 3.1** **POLICE DATASET**



**3.2.3 Preprocessing**

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

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

,Accused name & address

* + - * Dropping features such as Resolution, Description and Address: The resolution and description of a crime are only known once the crime has occurred, and have limited significance in a practical, real-world scenario where one is trying to predict what kind of crime has occurred, and so, these were omitted. The address was dropped because we had information about the latitude and longitude, and, in that context, the address did not add much marginal value.
      * The timestamp contained the year, date and time of occurrence of each crime. This was decomposed into five features: Year (2018), Month (1-12), Date (1-31), Hour (0-

23) and Minute (0-59).

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

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

**Table 1.2: Dataset after Preprocessing**

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

* KNN( K- Nearest neighbors)
* Decision Tree
* Random Forests

**3.3 IMPLEMENTATION METHODOLOGY**

**3.3.1 IMPLEMENTATION DETAILS**

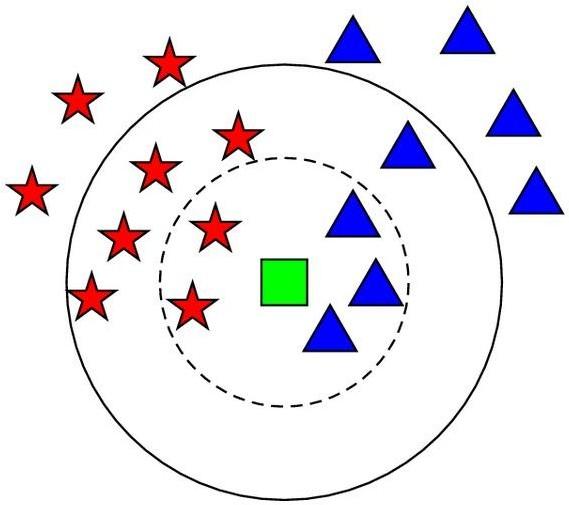
For the purpose of proper implementation and functioning several Algorithms and techniques were used. Following are the algorithms used:

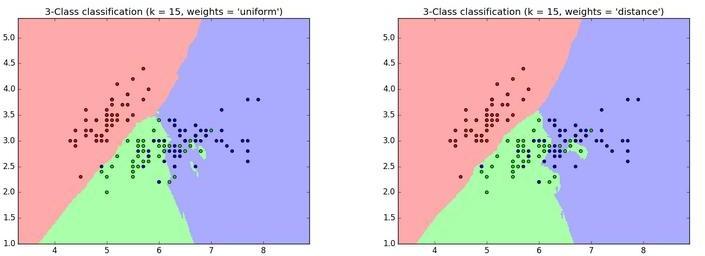
**3.3.1.2 KNN (K-Nearest neighbors)**

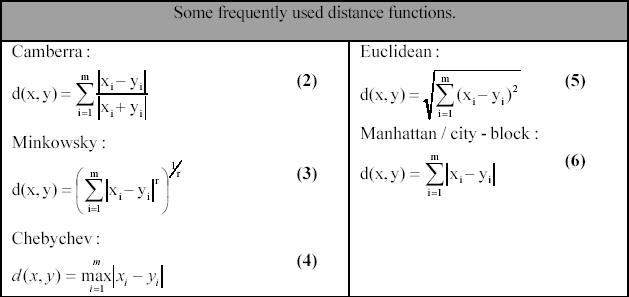
A powerful classification algorithm used in pattern recognition K nearest neighbors stores all available cases and classifies new cases based on a similarity measure (e.g. distance function). One of the top data mining algorithms used today. A non-parametric lazy learning algorithm (An Instance based Learning method).

KNN: Classification Approach

* + - * An object (a new instance) is classified by a majority votes for its neighbor classes.
      * The object is assigned to the most common class amongst its K nearest neighbors.(measured by distance function)

**Fig 3.2.1 Principle diagram of KNN** 

**Fig 3.2.2 Shows graphical representation of KNN**



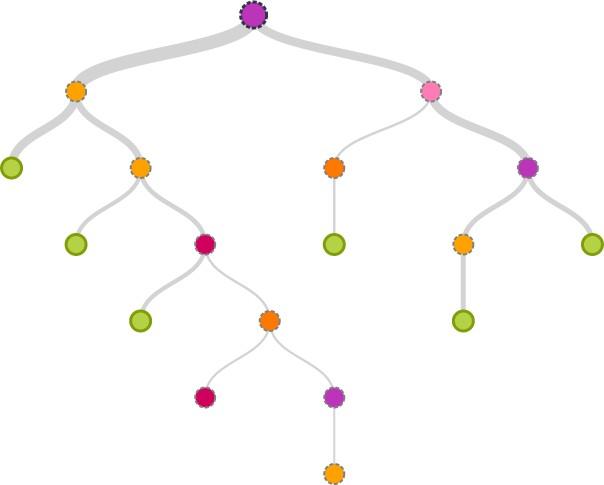
**Fig 3.2.3 Distance functions**

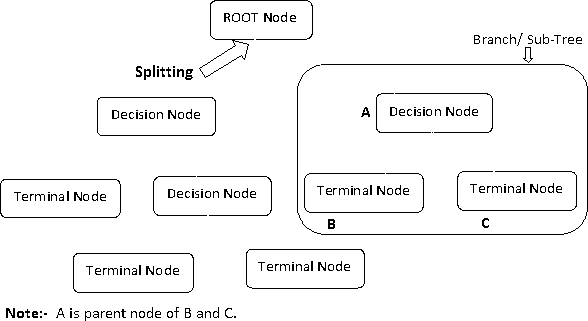
**3.3.1.3 Decision Tree**

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

* It is different from others because it works intuitively i.e., taking decisions one-by-one.
* Non-parametric: Fast and efficient.

It consists of nodes which have parent-child relationship:

**Fig 3.3.1 Decision tree**



**Fig 3.3.2 Decision Tree example**

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

## 3.3.1.4 Random forest

Random Forests is a very popular ensemble learning method which builds a number of

classiﬁers on the training data and combines all their outputs to make the best predictions

on the test data.

Thus, the Random Forests algorithm is a variance minimizing algorithm that uses

randomness when making split decision to help avoid overﬁtting on the .training data.

A random forests classiﬁer is an ensemble classiﬁer, which aggregates a family of

classiﬁers h(x|θ1),h(x|θ2),..h(x|θk). Each member of the family, h(x|θ), is a classiﬁcation

tree and k is the number of trees chosen from a model random vector.

Also, each θk is a randomly chosen parameter vector. If D(x,y) denotes the training

dataset, each classiﬁcation tree in the ensemble is built using a different subset

Dθk(x,y) ⊂ D(x,y) of the training dataset.

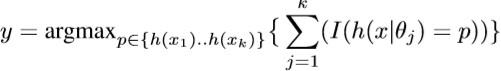
Thus, h(x|θk) is the kth classiﬁcation tree which uses a subset of features xθk ⊂ x to

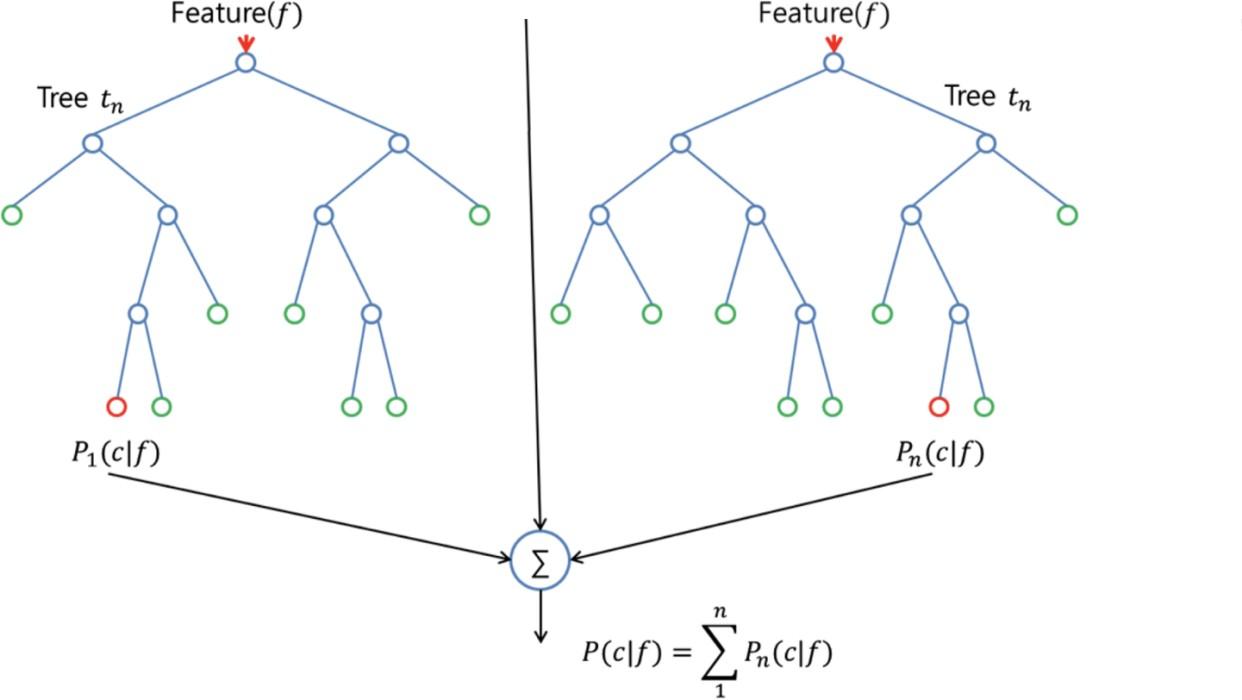
build a classiﬁcation model. Each tree then works like regular decision trees: it V K

partitions the data based on the value of a particular feature (which is selected

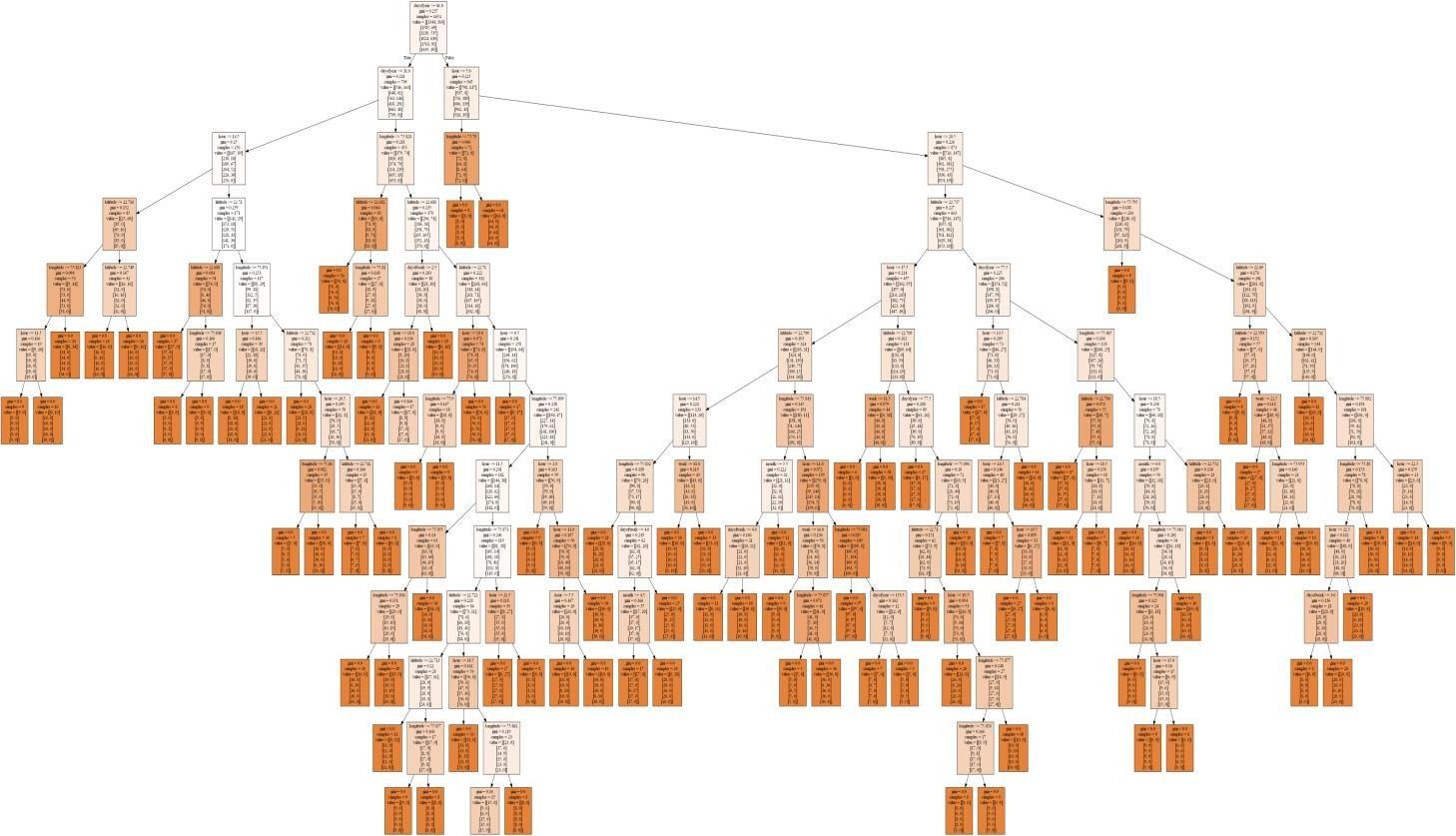
randomly from the subset), until the data is fully partitioned, or the maximum allowed

depth is reached. The ﬁnal output y is obtained by aggregating the results thus:

where I denotes the indicator function.

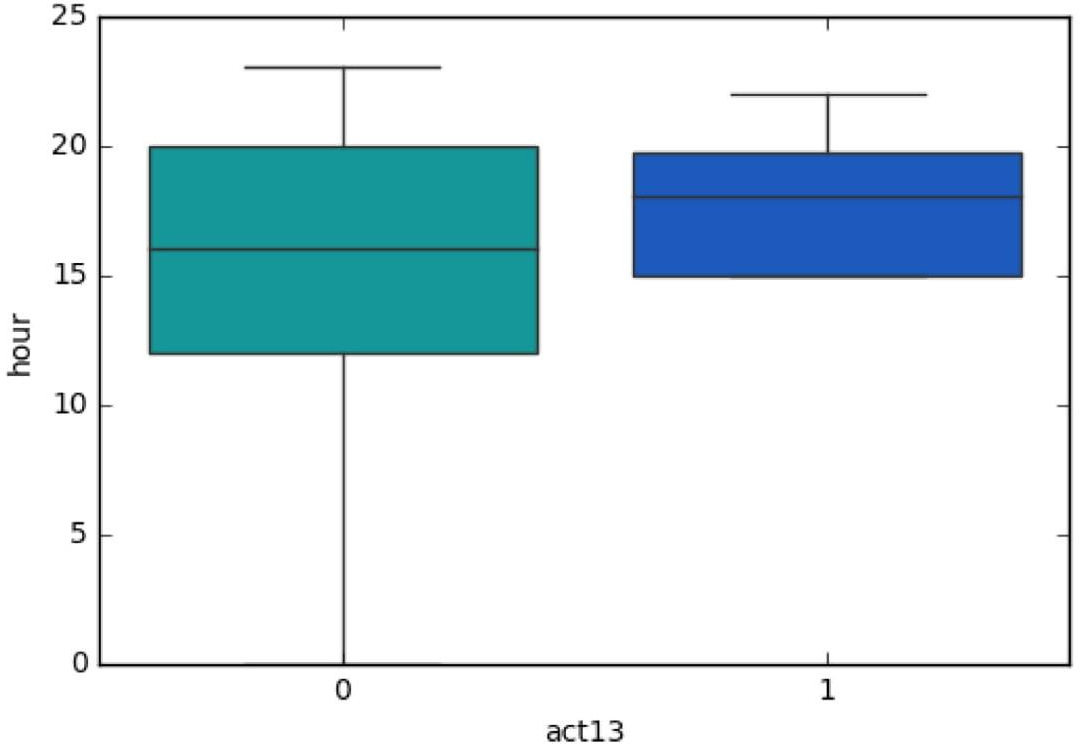


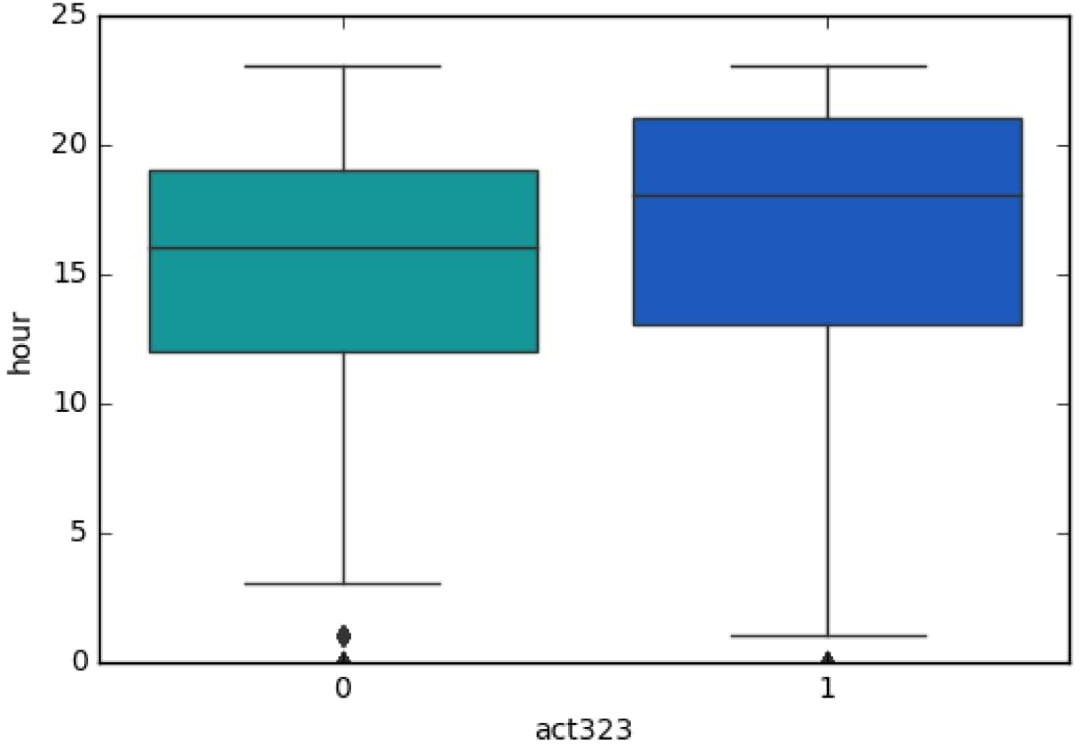
**Fig 3.4.1 Random Forest Example**



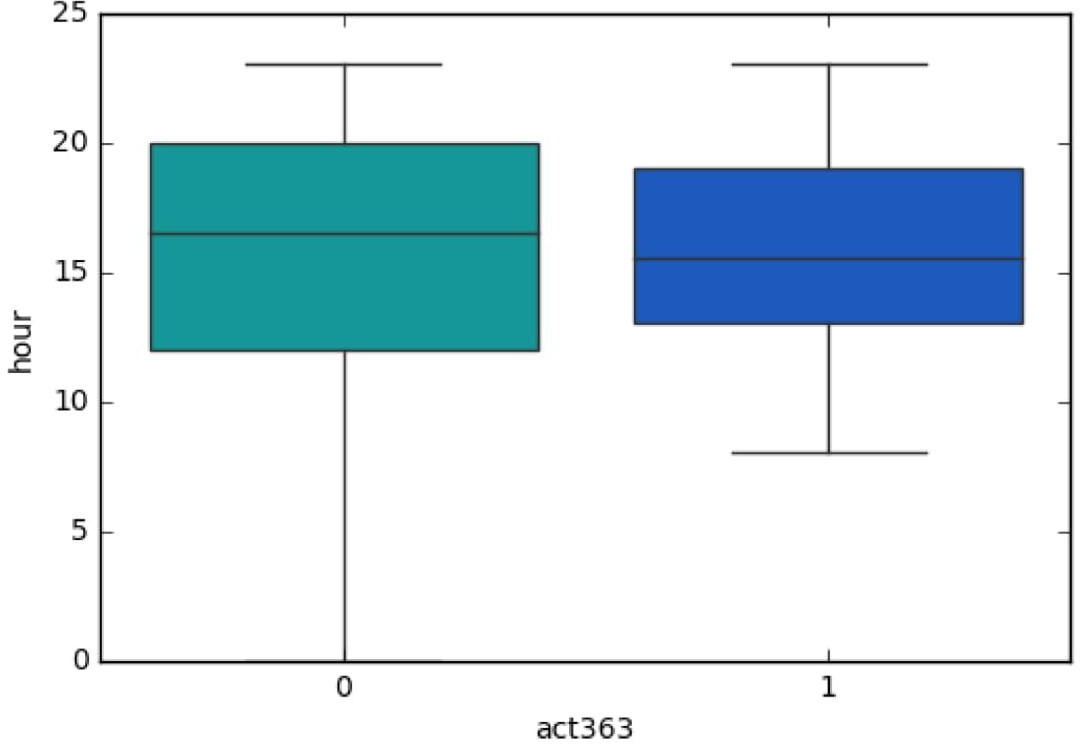
**Fig 3.4.2 Decision Tree of Crime Patter Prediction**

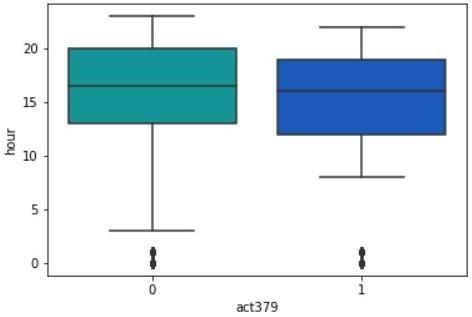
**3.3.1.5 Data Visualization**

**Fig 3.5.1 Act13(Gambling vs Hour)**

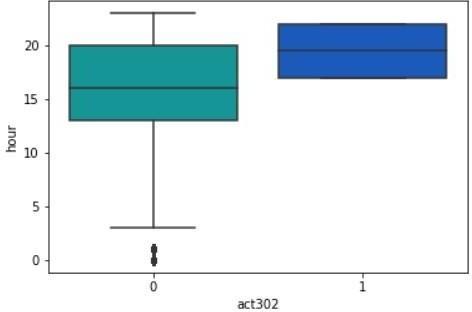


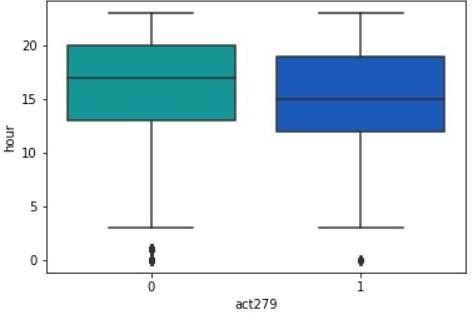
**Fig 3.5.2 Act323(Violence vs Hour)**

**Fig 3.5.3 Act363(Kidnapping vs Hour)**



**Fig 3.5.4 Act379(Robbery vs Hour)**

**Fig 3.5.5 Act302(Murder vs Hour)**



**Fig3.5.6 Act279(Accident vs Hour)**

**4 REQUIREMENT SPECIFICATION**

**4.1 HARDWARE REQUIREMENT**

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

**4.2 SOFTWARE REQUIREMENT**

* + Anaconda Distribution (v5.1)
  + Python (3.6.5)

Packages Used:

oFlask (0.12.2)

* + - * Pandas (0.22.1)

oNumpy (1.14.2)

* + - * Sklearn (0.19.1)

oGeopy (1.13.0)

HTML 5

CSS 3

Bootstrap 4

Java Script 1.8

**4.2.1 Features of PYTHON**

* Python is easy to learn and use. It is developer-friendly and high level programming language.
* Python language is more expressive means that it is more understandable and readable.
* Python can run equally on different platforms such as Windows, Linux, Unix and Macintosh etc. So, we can say that Python is a portable language
* Python supports object oriented language and concepts of classes and objects come into existence.
* Graphical user interfaces can be developed using Python.

**4.2.2 ANACONDA platform**

**Anaconda** (Python distribution) **Anaconda** is a free and open-source distribution of the Python and R programming languages for scientific computing (data science, machine learning applications, large-scale data processing, predictive analytics, etc.), that aims to simplify package management and deployment.

Directly from the platform and without involving DevOps, data scientists can develop and deploy AI and machine learning models rapidly into production. Anaconda provides the tools needed to easily:

* Collect data from files, databases, and data lakes
* Manage environments with Conda (all package dependencies are taken care of at the time of download)
* Share, collaborate on, and reproduce projects
* Deploy projects into production with the single click of a button

**5. IMPLEMENTATION**

The implementation of the project is done with the help of python language. To be particular, for the purpose of machine learning Anaconda is being used.

Anaconda is one of several Python distributions. Anaconda is a new distribution of the Python. It was formerly known as Continuum Analytics. Anaconda has more than 100 new packages. Anaconda is used for scientific computing, data science, statistical analysis, and machine learning.

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

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

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

The entries was done just to make the machine learn what all it has to do with the data and what actually the output is being demanded. As soon as the machine learnt the algorithms and the process, accuracy of different algorithms were measured & the algorithm with the most accuracy is used for the prediction kernel i.e. Random forest.

**5.1 SAMPLE CODE**

def predict():

rfc = joblib.load('model/rf\_model')

print('model loaded')

if request.method == 'POST':

address = request.form['Location']

geolocator = Nominatim()

location = geolocator.geocode(address)

print(location.address)

lat=[location.latitude]

log=[location.longitude]

latlong=pd.DataFrame({'latitude':lat,'longitude':log})

print(latlong)

DT= request.form['timestamp']

latlong['timestamp']=DT

data=latlong

my\_prediction = rfc.predict(data)

return render\_template('result.html', prediction = my\_prediction)

**5.1.1 PREDICTION CODE**

**#Imports library like panda numpy**

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

%matplotlib inline

dataset=pd.read\_csv('data.csv')

data=pd.read\_csv('data.csv')

dataset.head()

for col in data:

print (type(data[col][1]))

data['timestamp'] = pd.to\_datetime(data['timestamp'], coerce=True)

data['timestamp'] = pd.to\_datetime(data['timestamp'], format = '%d/%m/%Y %H:%M:%S')

data['timestamp']

# DATE TIME STAMP FUNCTION

column\_1 = data.ix[:,0]

#To extract Date, Time & Location

db=pd.DataFrame({"year": column\_1.dt.year,

"month": column\_1.dt.month,

"day": column\_1.dt.day,

"hour": column\_1.dt.hour,

"dayofyear": column\_1.dt.dayofyear,

"week": column\_1.dt.week,

"weekofyear": column\_1.dt.weekofyear,

"dayofweek": column\_1.dt.dayofweek,

"weekday": column\_1.dt.weekday,

"quarter": column\_1.dt.quarter,

})

dataset1=dataset.drop('timestamp',axis=1)

data1=pd.concat([db,dataset1],axis=1)

data1.info()

data1.dropna(inplace=True)

data1.head()

sns.pairplot(data1,hue='act363')

sns.boxplot(x='act379' ,y='hour' ,data=data1, palette='winter\_r')

sns.boxplot(x='act13' ,y='hour' ,data=data1 , palette='winter\_r')

sns.boxplot(x='act323' ,y='hour' ,data=data1, palette='winter\_r')

sns.boxplot(x='act363' ,y='hour' ,data=data1, palette='winter\_r')

df = pd.DataFrame(data=data1, columns=['act13', 'hour', 'day'])

df.plot.hexbin(x='act13',y='hour',gridsize=25)

df.plot(legend=False)

df1 = pd.DataFrame(data=data1, columns=['act13', 'act323', 'act379'])

df1.plot.kde()

X=data1.iloc[:,[1,2,3,4,6,16,17]].values

X

y=data1.iloc[:,[10,11,12,13,14,15]].values

## Splitting the data

from sklearn.cross\_validation import train\_test\_split

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.20, random\_state=50)

#Creating and Training KNN Model:

from sklearn.neighbors import KNeighborsClassifier

knn = KNeighborsClassifier(n\_neighbors=10)

knn.fit(X\_train,y\_train)

knn.score(X\_test,y\_test)

Elbow Method For optimum value of K:

error\_rate = []

for i in range(1,140):

knn = KNeighborsClassifier(n\_neighbors=i)

knn.fit(X\_train,y\_train)

pred\_i = knn.predict(X\_test)

error\_rate.append(np.mean(pred\_i != y\_test))

plt.figure(figsize=(10,6))

plt.plot(range(1,140),error\_rate,color='blue', linestyle='dashed', marker='o',

markerfacecolor='red', markersize=5)

plt.title('Error Rate vs. K Value')

plt.xlabel('K')

plt.ylabel('Error Rate')

#Creating & Training Decision Tree Model:

from sklearn.tree import DecisionTreeClassifier

dtree = DecisionTreeClassifier(max\_depth=500, random\_state=300)

dtree.fit(X\_train,y\_train)

y\_pred=dtree.predict(X\_test)

dtree.score(X\_test,y\_test)

dtree.score(X\_train,y\_train)

y\_pred

treefeatures=dtree.feature\_importances\_

indices = np.argsort(treefeatures)

features = data1.iloc[:,[1,2,3,4,6,16,17]]

plt.figure(1)

plt.title('Feature Importances')

plt.barh(range(len(indices)), treefeatures[indices], color='b', align='center')

plt.yticks(range(len(indices)), features[indices])

plt.xlabel('Relative Importance')

Tree Visualization:

feature\_names=[ 'dayofweek', 'dayofyear', 'hour', 'month', 'week','latitude', 'longitude']

from IPython.display import Image

from sklearn.externals.six import StringIO

from sklearn.tree import export\_graphviz

import pydot

import os

os.environ["PATH"] += os.pathsep + 'C:/Program Files (x86)/Graphviz2.38/bin/'

dot\_data = StringIO()

export\_graphviz(dtree, out\_file=dot\_data,feature\_names=feature\_names,filled=True)

graph = pydot.graph\_from\_dot\_data(dot\_data.getvalue())

Image(graph[0].create\_png())

Creating & Training Random Tree Model:

from sklearn.ensemble import RandomForestClassifier

rfc = RandomForestClassifier(n\_estimators=100)

rfc.fit(X\_train, y\_train)

y\_pred=rfc.predict(X\_test)

rfc.score(X\_test,y\_test)

rfc.score(X\_train,y\_train)

om=rfc.feature\_importances\_

indices = np.argsort(om)

om

features = data1.columns

plt.figure(1)

plt.title('Feature Importances')

plt.barh(range(len(indices)), om[indices], color='b', align='center')

plt.yticks(range(len(indices)), features[indices])

plt.xlabel('Relative Importance')

!ipython nbconvert --to slides CrimePrediction.ipynb

!ipython nbconvert --to pdf CrimePrediction.ipynb

**5.1.2 PYTHON FLASK WEB APPLICATION**

**import os, sys, shutil, time**

**from flask import Flask, request, jsonify, render\_template,send\_from\_directory**

**import pandas as pd**

**from sklearn.externals import joblib**

**from sklearn.ensemble import RandomForestClassifier**

**import numpy as np**

**import urllib.request**

**import json**

**from geopy.geocoders import Nominatim**

**app = Flask(\_\_name\_\_)**

**@app.route('/')**

**def root():**

**return render\_template('index.html')**

**@app.route('/images/<Paasbaan>')**

**def download\_file(Paasbaan):**

**return send\_from\_directory(app.config['images'], Paasbaan)**

**@app.route('/index.html')**

**def index():**

**return render\_template('index.html')**

**@app.route('/work.html')**

**def work():**

**return render\_template('work.html')**

**@app.route('/about.html')**

**def about():**

**return render\_template('about.html')**

**@app.route('/contact.html')**

**def contact():**

**return render\_template('contact.html')**

**@app.route('/result.html', methods = ['POST'])**

**def predict():**

**rfc = joblib.load('model/rf\_model')**

**print('model loaded')**

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

**address = request.form['Location']**

**geolocator = Nominatim()**

**location = geolocator.geocode(address,timeout=None)**

**print(location.address)**

**lat=[location.latitude]**

**log=[location.longitude]**

**latlong=pd.DataFrame({'latitude':lat,'longitude':log})**

**print(latlong)**

**DT= request.form['timestamp']**

**latlong['timestamp']=DT**

**data=latlong**

**cols = data.columns.tolist()**

**cols = cols[-1:] + cols[:-1]**

**data = data[cols]**

**data['timestamp'] = pd.to\_datetime(data['timestamp'].astype(str), errors='coerce')**

**data['timestamp'] = pd.to\_datetime(data['timestamp'], format = '%d/%m/%Y %H:%M:%S')**

**column\_1 = data.ix[:,0]**

**DT=pd.DataFrame({"year": column\_1.dt.year,**

**"month": column\_1.dt.month,**

**"day": column\_1.dt.day,**

**"hour": column\_1.dt.hour,**

**"dayofyear": column\_1.dt.dayofyear,**

**"week": column\_1.dt.week,**

**"weekofyear": column\_1.dt.weekofyear,**

**"dayofweek": column\_1.dt.dayofweek,**

**"weekday": column\_1.dt.weekday,**

**"quarter": column\_1.dt.quarter,**

**})**

**data=data.drop('timestamp',axis=1)**

**final=pd.concat([DT,data],axis=1)**

**X=final.iloc[:,[1,2,3,4,6,10,11]].values**

**my\_prediction = rfc.predict(X)**

**if my\_prediction[0][0] == 1:**

**my\_prediction='Predicted crime : Act 379-Robbery'**

**elif my\_prediction[0][1] == 1:**

**my\_prediction='Predicted crime : Act 13-Gambling'**

**elif my\_prediction[0][2] == 1:**

**my\_prediction='Predicted crime : Act 279-Accident'**

**elif my\_prediction[0][3] == 1:**

**my\_prediction='Predicted crime : Act 323-Violence'**

**elif my\_prediction[0][4] == 1:**

**my\_prediction='Predicted crime : Act 302-Murder'**

**elif my\_prediction[0][5] == 1:**

**my\_prediction='Predicted crime : Act 363-kidnapping'**

**else:**

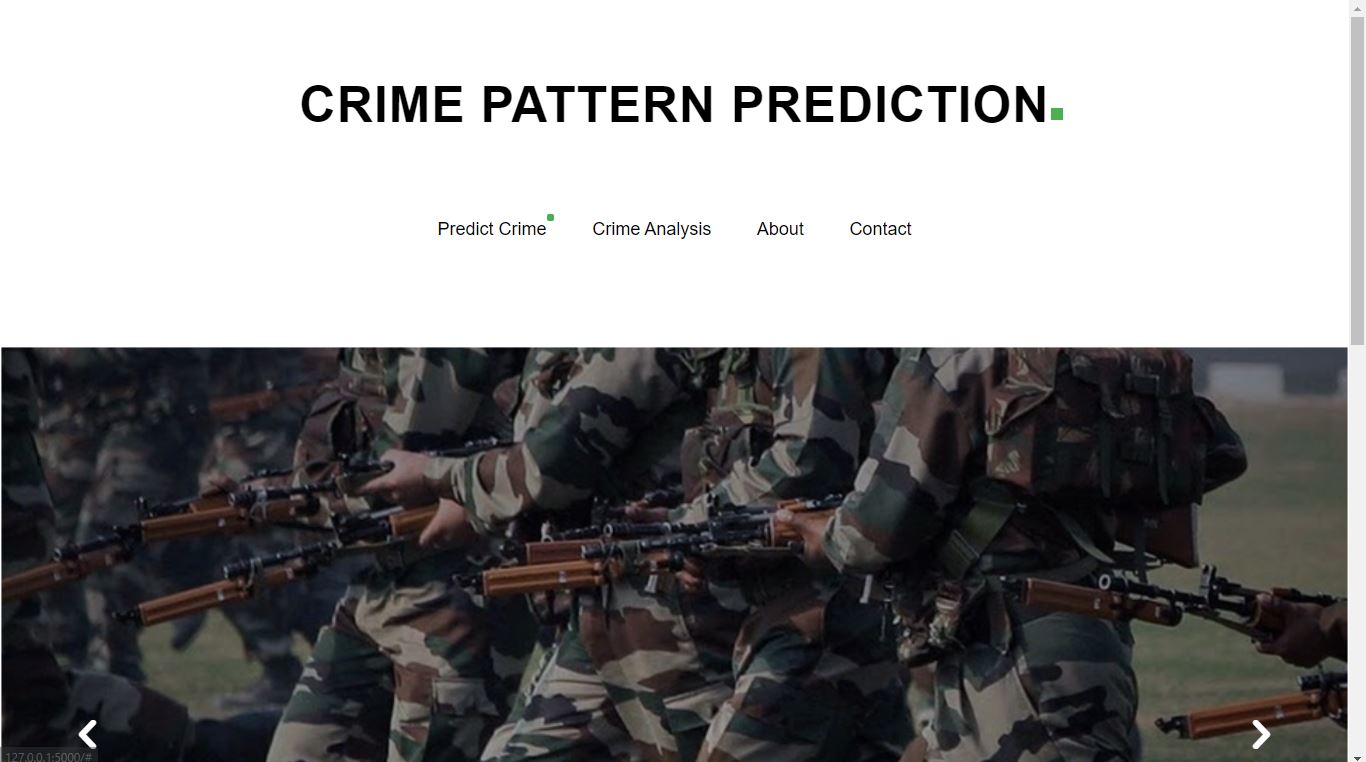
**my\_prediction='Place is safe no crime expected at that timestamp.'**

**return render\_template('result.html', prediction = my\_prediction)**

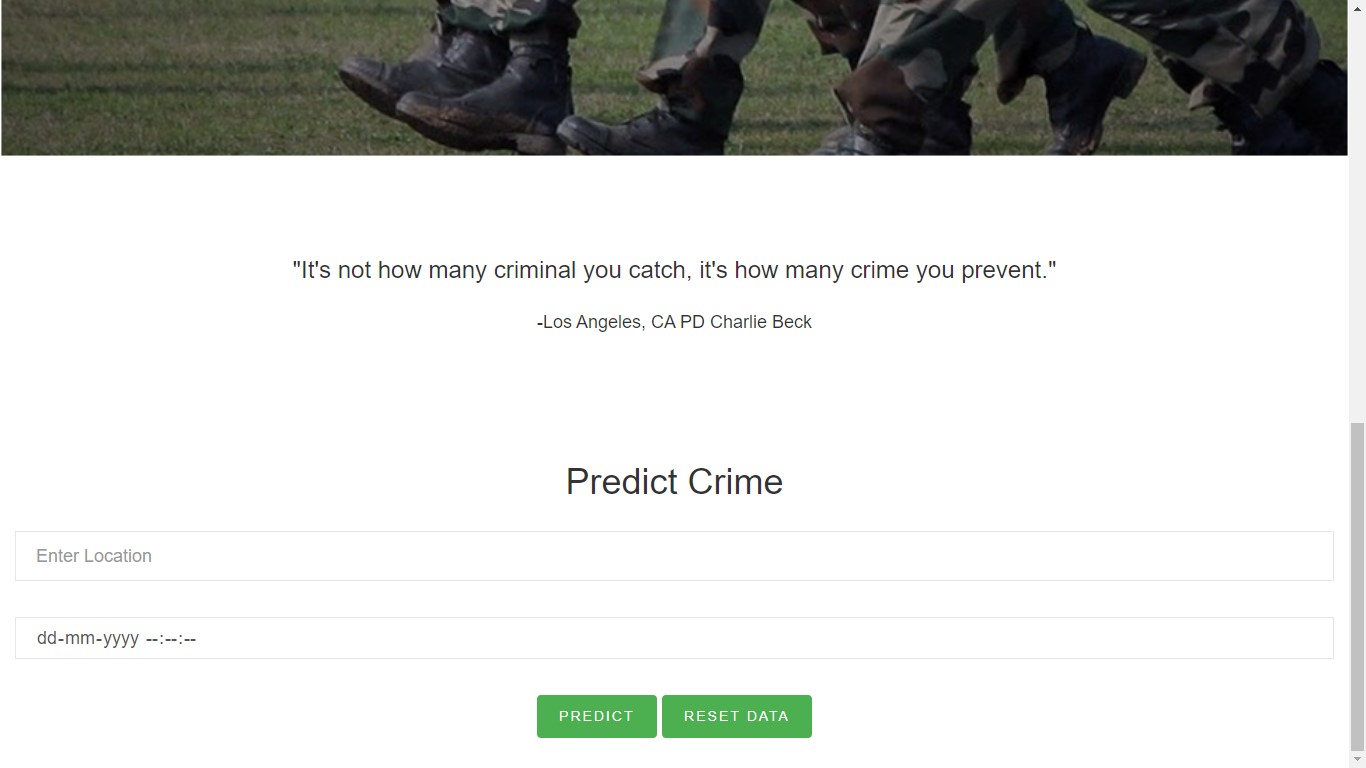
**if \_\_name\_\_ == '\_\_main\_\_':**

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

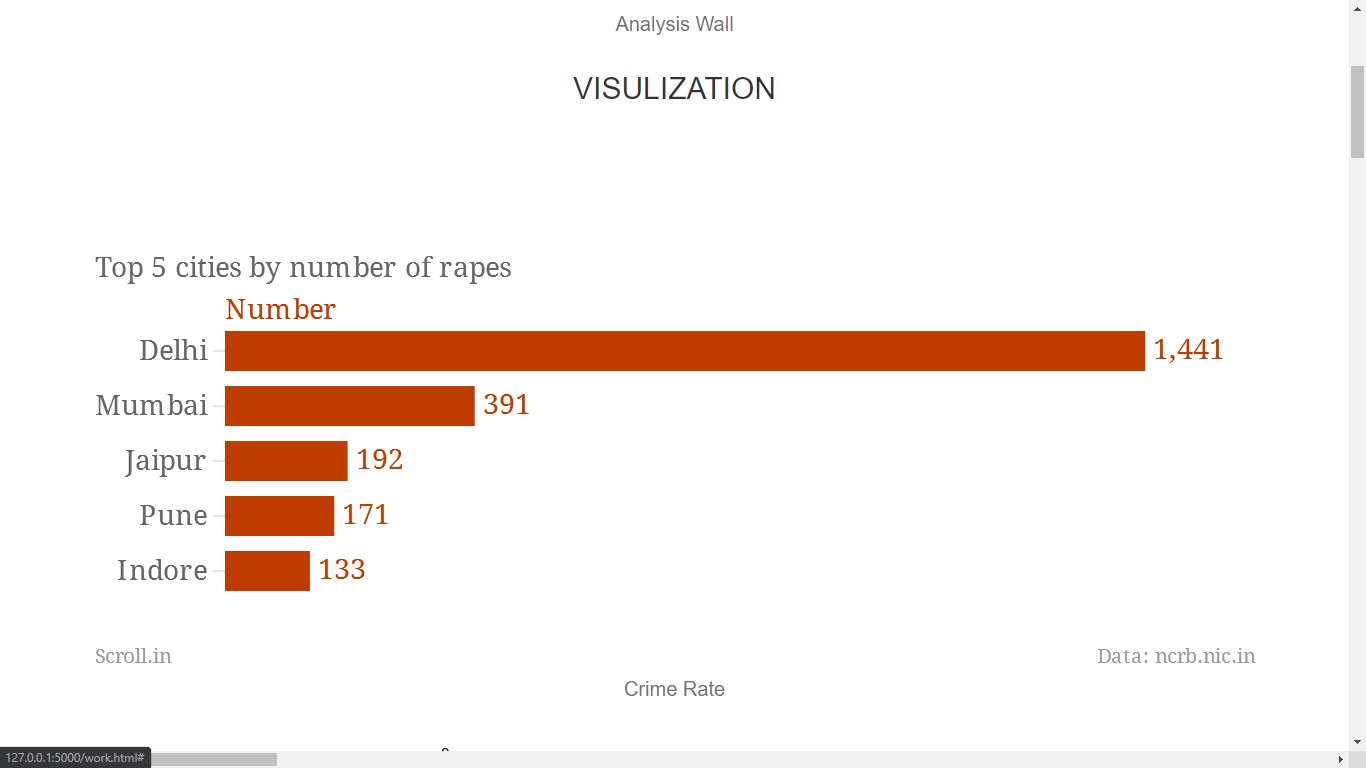
**5.2 SAMPLE SCREEN SHOTS**

****

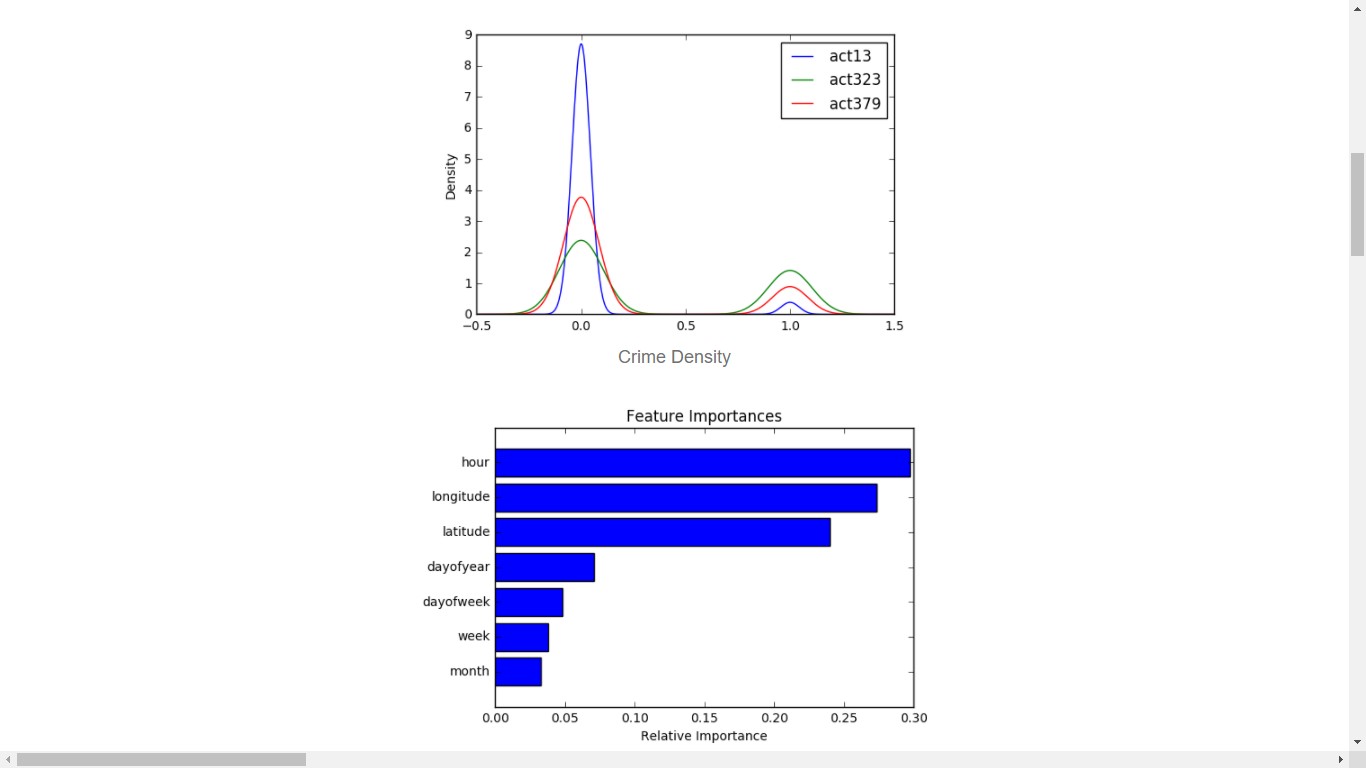
**Fig5.1**  **Snapshot 1**

****

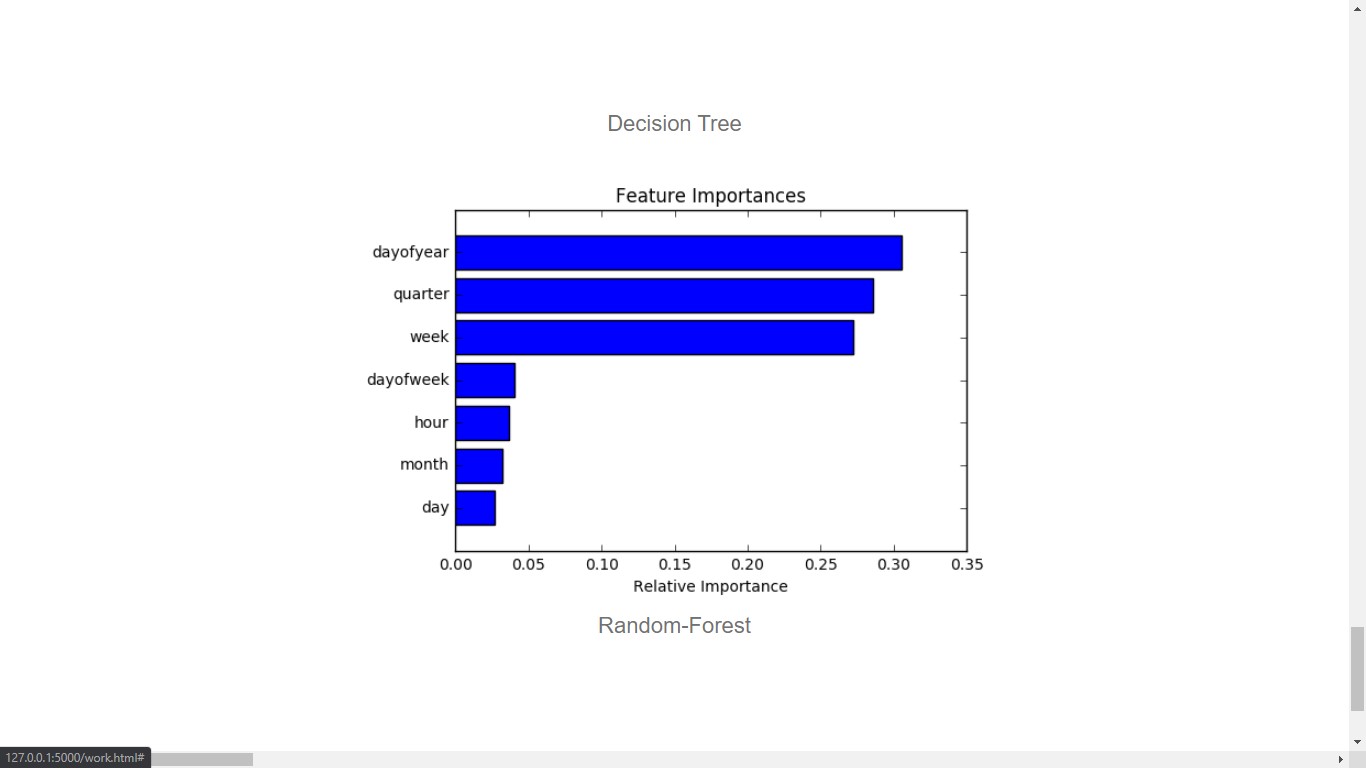
**Fig5.2**  **Snapshot 2**

****

**Fig5.3**  **Snapshot 3**

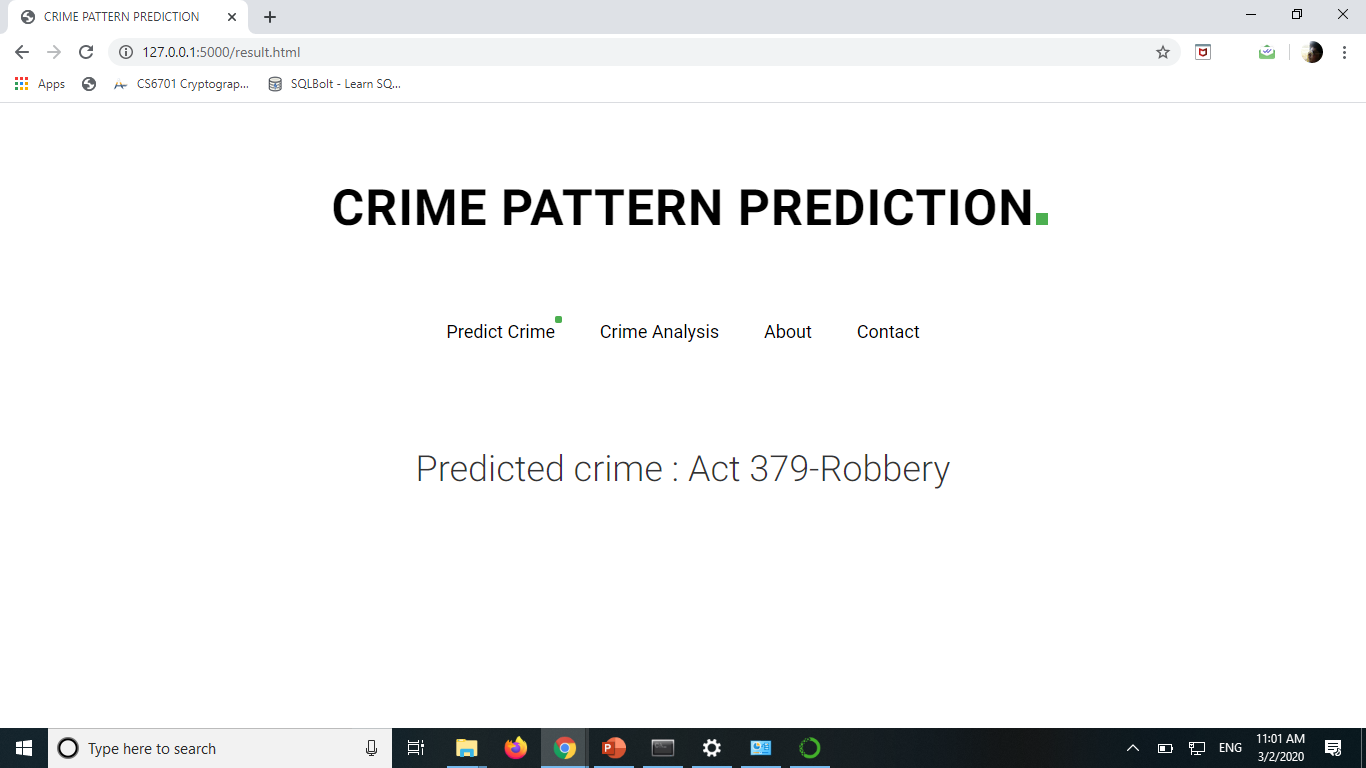
****

**Fig5.4**  **Snapshot 4**

****

**Fig5.5**  **Snapshot 5**

**RESULT SCREENSHOTS**

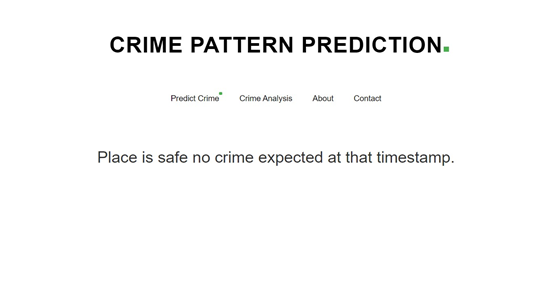
****

**Fig5.6**  **Snapshot 6**

****

**Fig5.7**  **Snapshot 7**

**POSITIVE RESULT**

****

**Fug 5.8 Snapshot 8**

**6 TESTING AND MAINTENANCE**

**6.1 SOFTWARE TESTING**

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

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

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

**6.1.5 Black box testing**

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

**UNIT TESTING:**

UNIT TESTING is a level of software testing where individual units/ components of a software are tested. The purpose is to validate that each unit of the software performs as designed. A unit is the smallest testable part of any software. It usually has one or a few inputs and usually a single output.

**6.1.6 White box testing**

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

**Table 6.1 Tests**

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

## 

**7 CONCLUSION AND FUTURE WORKS**

**7.1 CONCLUSION**

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

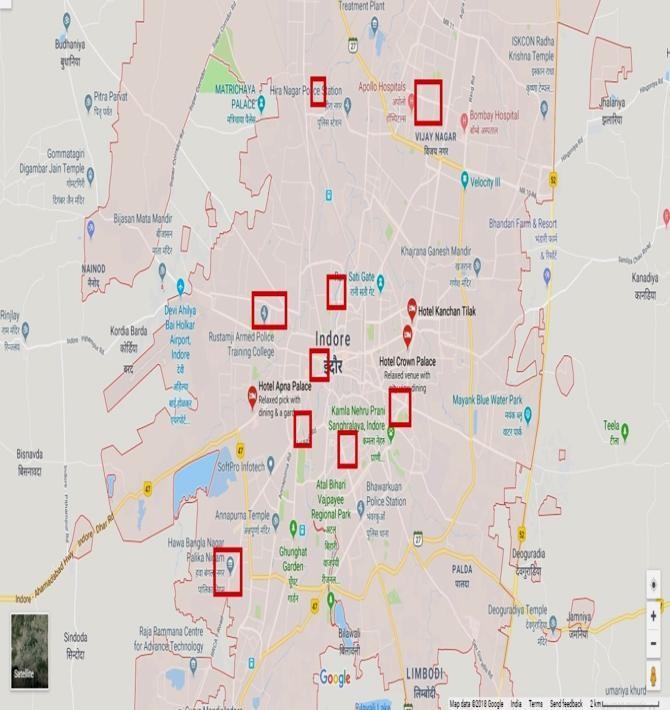
Possible avenues through which to extend this work include time-series modeling of the data to understand temporal correlations in it, which can then be used to predict surges in different categories of crime. It would also be interesting to explore relationships between surges in different categories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

**7.2 FUTURE WORKS**

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

* **Predicting Future Crime Spots:**

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



**Fig 7.1 Predicting Surges**

* **Predicting Who Will Commit a Crime**:

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

* **Pretrial Release and Parole:**

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

* **RESULT AND FUTURE SCOPE:**

Result of this research will be to analyse, correlate and predict the crimes from huge data set 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 displayed using various diagrams pie charts, heat maps, spikes and graphs.

**7.2.1 Expected Outcome**

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

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

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

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

**REFERENCES**

1. Bogomolov, Andrey and Lepri, Bruno and Staiano, Jacopo and Oliver, Nuria and Pianesi, Fabio and Pentland, Alex.2014. Once upon a crime: Towards crime prediction from demographics and mobile data, Proceedings of the 16th International Conference on Multimodal Interaction.
2. Yu, Chung-Hsien and Ward, Max W and Morabito, Melissa and Ding, Wei.2011. Crime forecasting using data mining techniques, pages 779-786, IEEE 11th International Conference on Data Mining Workshops (ICDMW)
3. Kianmehr, Keivan and Alhajj, Reda. 2008. Effectiveness of support vector machine for crime hot-spots prediction, pages 433-458, Applied Artificial Intelligence, volume 22, number 5.
4. Toole, Jameson L and Eagle, Nathan and Plotkin, Joshua B. 2011 (TIST), volume 2, number 4, pages 38, ACM Transactions on Intelligent Systems and Technology
5. Wang, Tong and Rudin, Cynthia and Wagner, Daniel and Sevieri, Rich. 2013. pages 515- 530, Machine Learning and Knowledge Discovery in Databases
6. Friedman, Jerome H. ”Stochastic gradient boosting.” Computational Statistics and Data Analysis 38.4 (2002): 367-378.sts
7. Leo Breiman, Random Forests, Machine Learning, 2001,Volume 45, Number 1, Page 5
8. Shyam Varan Nath Florida, “Crime Pattern Detection Using Data Mining”, Proceedings

of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT 2006 Workshops) (WI-IATW06).

1. Priyanka Gera, Rajan Vohra, “City Crime Profiling Using Cluster Analysis” ,International Journal of Computer Science and Information Technologies, Vol. 5 (4) , 2014.
2. Malathi. A and Dr. S. Santhosh Baboo, “Enhanced Algorithms to Identify Change in Crime Patterns”, International Journal of Combinatorial Optimization Problems and Informatics, Vol. 2, No.3, Sep-Dec 2011, pp. 3238, ISSN: 200