Prediction and Analysis of Crime

Author:RafaTaskin A Shaikh   
*Masters In ComputerApplication   
Guide: Prof. Prini Rastogi*  
*Ajeenkya D Y Patil Unversity*Pune, India  
rafa.shaikh@adypu.edu.in

*Abstract*— Crime is one of the dominant and alarming aspect of our society. Over the past few years, the crime rate across globe has increased exponentially. So, preventing the crime from occurring is a vital task. In the recent time, it is seen that artificial intelligence has shown its importance in almost all the field and crime prediction is one of them. However, it is necessary to maintain a proper database of the crime that has occurred. The ability to predict the crime based on time, location and so on which can occur in future can help the law enforcement agencies in preventing the crime before it occurs from a strategical perspective. However, predicting the crime accurately is a challenging task because crimes are increasing at an alarming rate. Thus, the crime prediction and analysis methods are very important to detect the future crimes and reduce them. In Recent time, many researchers have conducted experiments to predict the crimes using various machine learning methods and particular inputs. For crime prediction, KNN, K-means and Random Forest and some other algorithms are used. Our system can predict regions which have high probability for crime occurrence and can visualize crime prone areas. The main purpose is to highlight the worth and effectiveness of machine learning in predicting violent crimes occurring in a particular region in such a way that it can be used by police to reduce crime rates in the society.

***Keywords***— ***Machine Learning, Artificial Intelligence, Crime Probability, Visualize* *& Reduce Crime***

# Introduction

Crime is increasing considerably day by day. Crime is among the main issues which is growing continuously in intensity and complexity. Crime patterns are changing constantly because of which it is difficult to explain behaviours in crime patterns. Crime is classified into various types like kidnapping, theft murder, rape etc. The law enforcement agencies collects the crime data information with the help of information technologies (IT). But occurrence of any crime is naturally unpredictable and from previous searches it was found that various factors like poverty, employment affects the crime rate. It is neither uniform nor random. With rapid increase in crime number, analysis of crime is also required. Crime analysis basically consists of procedures and methods that aims at reducing crime risk. It is a practical approach to identify and analyse crime patterns. But, major challenge for law enforcement agencies is to analyse escalating number of crime data efficiently and accurately. So it becomes a difficult challenge for crime analysts to analyse such voluminous crime data without any computational support. A powerful system for predicting crimes is required in place of traditional crime analysis because traditional methods cannot be applied when crime data is high dimensional and complex queries are to be processed. Therefore, a crime prediction and analysis tool were needed for identifying crime patterns effectively. This paper introduces some methodologies with the help of which it can be predicted that at what place and time which type of crime has a higher probability of occurrence. Ever Since the Coronavirus Pandemic hit the globe like a shockwave, multiple global data prediction systems have shown unexpected amount of fluctuations. Given the circumstances, it is essential to make accurate predictions based on data analysis. One of the primary advantages of big data analytics software is that it can evaluate huge quantities of data much faster than humans can, plus spot trends they’d likely miss. So, from a crime solving point of view, data analytics could help catch criminals who are trying to evade arrest.

Use the enter key to start a new paragraph. The appropriate spacing and indent are automatically applied.

## Motivation

### In the past few years, the rate of crimes be it violent crimes or crimes against the vulnerable has increased exponentially. But over the few studies done, it has been observed that some crimes follow a particular pattern which has been helpful in apprehending criminals. The major motivation behind this project is the same- to study historic data and analyze it so that patterns behind crimes can be found which will help law enforcement authorities to solve crimes easily and thus form a safe society for everyone. The results of crime scene analysis (CSA) may be used to determine the limits of the available evidence and the need for additional investigative and forensic efforts, as in a threshold assessment (discussed shortly). When sufficient behavioural evidence is available, these same results may also be used to infer offender modus operandi (MO) and signature behaviors, evidence of crime scene staging, crime scene motive, and offender characteristics, or to assist with linkage analysis efforts. Crime analysis also plays a role in devising solutions to crime problems, and formulating crime prevention strategies. Quantitative [social science](https://en.wikipedia.org/wiki/Social_science) data analysis methods are part of the crime analysis process, though qualitative methods such as examining police report narratives also play a role.

## Problem Definition

## Our study aims to find spatial and temporal criminal hotspots using a set of real-world datasets of crimes. We will try to locate the most likely crime locations and their frequent occurrence time. In addition, we will predict what type of crime might occur next in a specific location within a particular time. Finally, we intend to provide an analysis study by combining our findings of a particular crimes dataset with its demographics information. Some instances where crime analysis has been able to overcome challenges faced by criminal justice departments are:

## University of Tennessee at Knoxville even has a Murder Accountability Project, where a group developed an algorithm to help with serial homicide cases.

## Information collected by data analytics tools can help prevent crime, too. Police in India are among those using data analytics like that. They incorporate criminal behaviour patterns, crime anniversaries, dates and conditions and more as they rely on data to help them detect useful patterns that could keep people safer.

## Call detail records (CDRs) encompass one example of the kind of data they could look at during an investigation. A CDR includes information including an identification code, the duration of the call and the specific cell towers that routed the communications for both the receiver and caller.

## Objectives

The main objective of the proposed concept is to develop a system which might help in predicting the crime rate and therefore help in apprehending the criminals.

The second major objective of this proposed concept is to prevent crime. The data collected and processed will provide a mechanism to prevent crimes.

To study the pattern of crime and trends observed so far so that a common factor can be established.

To create awareness about increasing crime rate, so that civilians will always be cautious.

To determine areas which are vulnerable to a particular crime.

# Methodology

## Steps Involved:

### Architecture: Predictive modeling was used for making predictions since it has the method which is able to build a model and has the capability to make predictions.This method consists of different algorithms of Machine Learning that can study properties from the data used for training which is used for producing.

### Data collection:

### Crime dataset from kaggle is use in CSV format.

### Data Preprocessing:

Data pre-processing basically involves methods to remove the infinite or null values from data which might affect the performance of the model. In this step the data set is converted into the understandable format which can be fed into machine learning models. The categorical attributes

(Location, Block, Crime Type, Community Area) are converted into numeric using Label Encoder. The date attribute is splitted into new attributes like month and hour which can be used as feature for the model.

### Feature selection:

Features selection is done which can be used to build the model. The attributes used for feature selection are Offense ID,District Id,Location, X coordinate , Y coordinate, Latitude , Longitude, Hour and month.

### Building and TraIning Model:

After feature selection location and month attribute are used for training. The dataset is divided into pair of xtrain ,y train and xtest, y test. The algorithms model is imported from sklearn. Building model is done using model. Fit (xtrain, ytrain).

### Prediction:

After the model is build using the above process, prediction is done using model.predict(xtest). The accuracy is calculated using accuracy\_score imported from metrics.

### Visualization:

Using matplotlib library from sklearn. Analysis of the crime dataset is done by plotting various graphs

# Design

## Software Requirement Specification

### User class and characteristcs

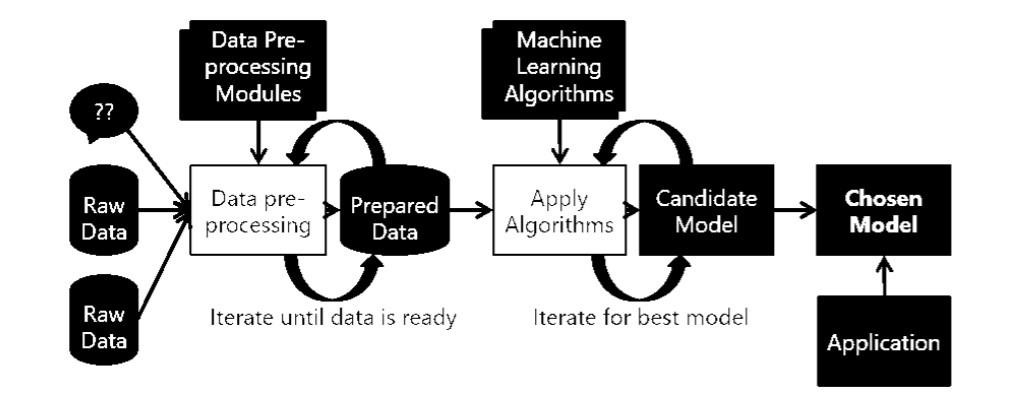
Using the data training process, data will be trained and the model will then predict and User of this system has to first login into system (web page made using flask) and then user will select the .csv file which will contain attributes such as Geo(longitude, latitude), month, year, date, etc and will predict the crime.

### Assumptions and dependencies

### We assume that the dataset selected by the user will have data of the necessary information to predict. No extra or less data is present. No null values are there. And the prediction accuracy will depend on the dataset chosen while training.

# Modeling

## System Architecture



|  |
| --- |
| Prediction |

1. This is a figure of the System’s Architecture

1. *STEPS TO BE FOLLOWED:*

Dataset will be collected

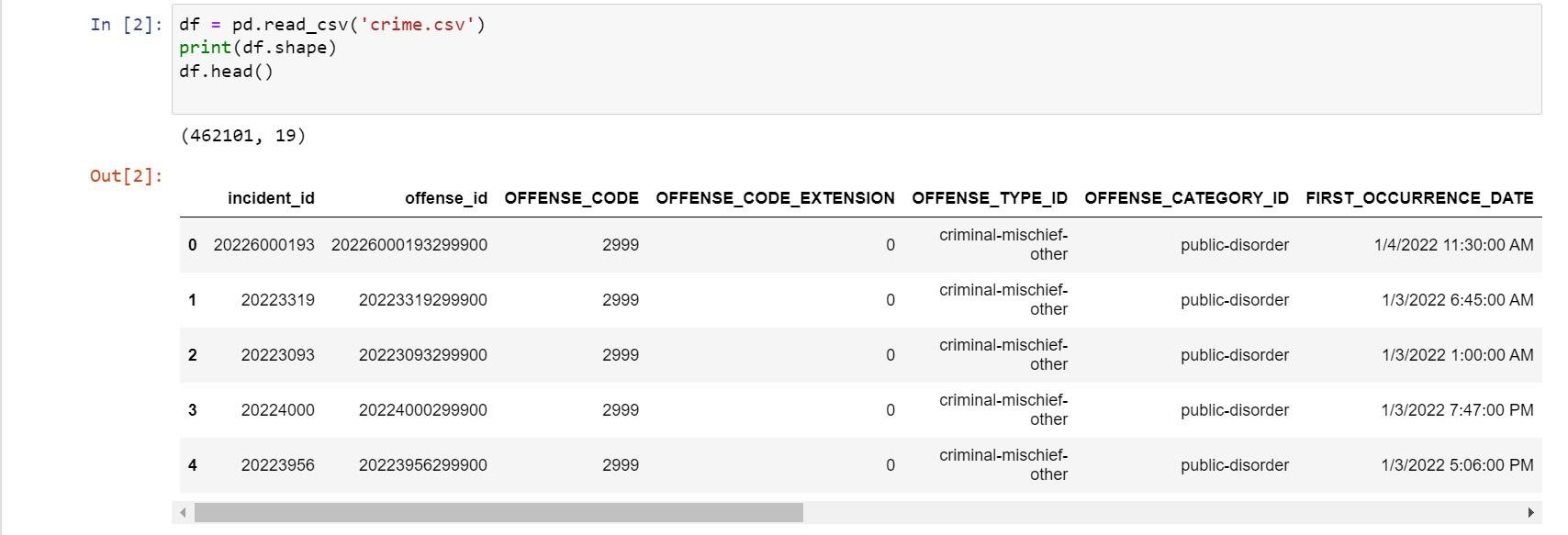
The raw data will undergo pre-processing in which it will made consistent, meaningful and will be fit for being used in any machine learning algorithms.

The pre-processed and cleaned data (i.e. data free from noisy values) will be trained and tested with suitable machine learning algorithms.

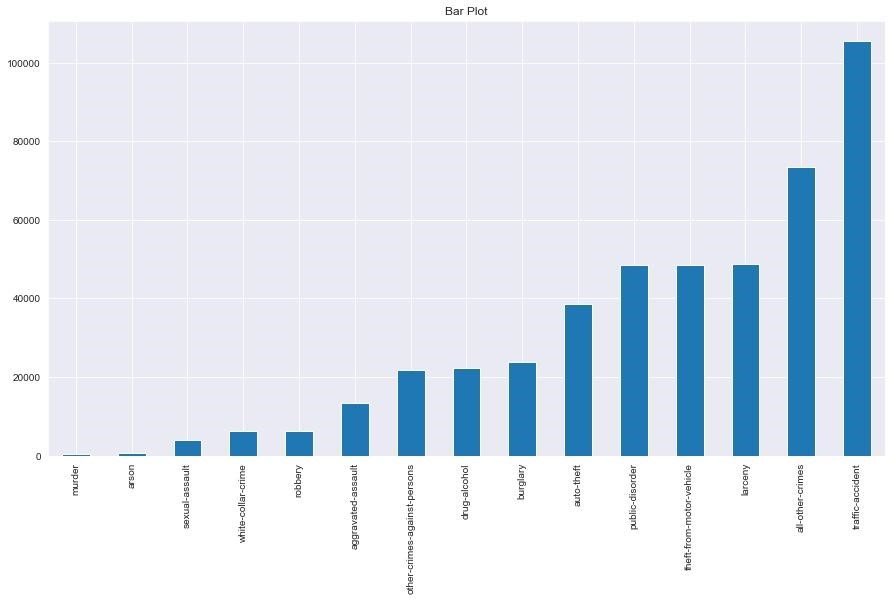
Once the best machine learning model with good accuracy fits the data, it will be used for prediction.

Further patterns will also be analyzed

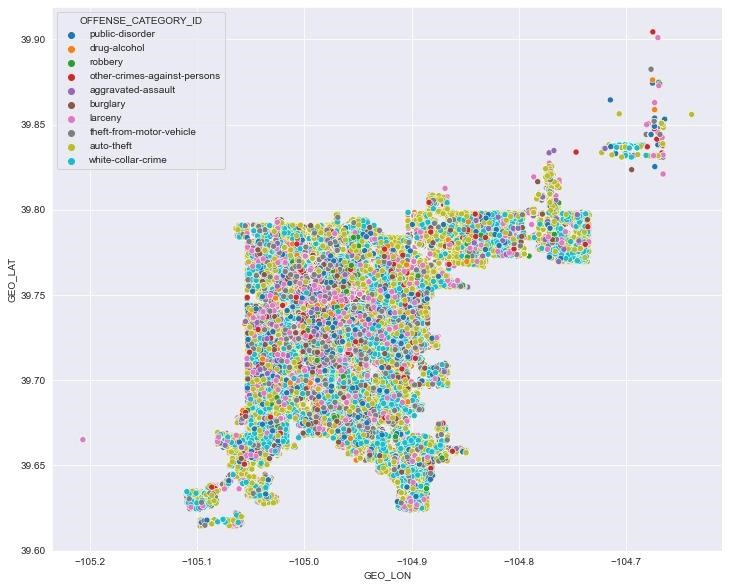
# TEST DATASET, ANALYSIS & RESULT



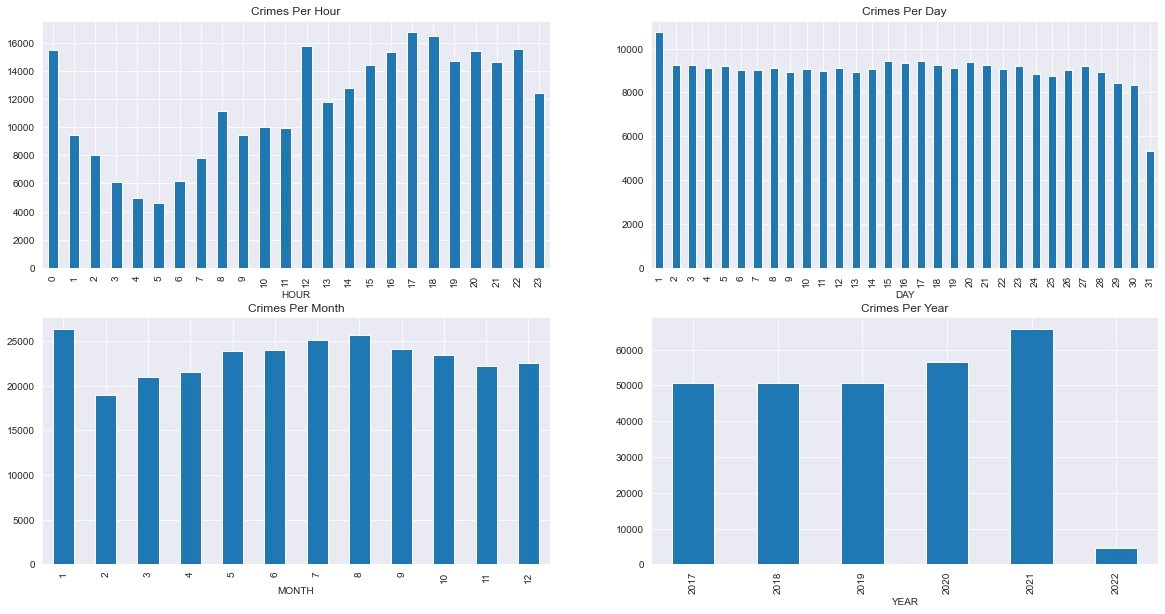
1. This is a figure of the test data set



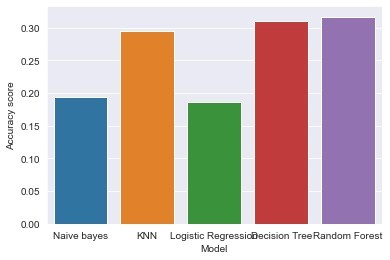
1. This is a figure of the offence category vs frequency id



1. This is a figure District-wise distribution of crimes

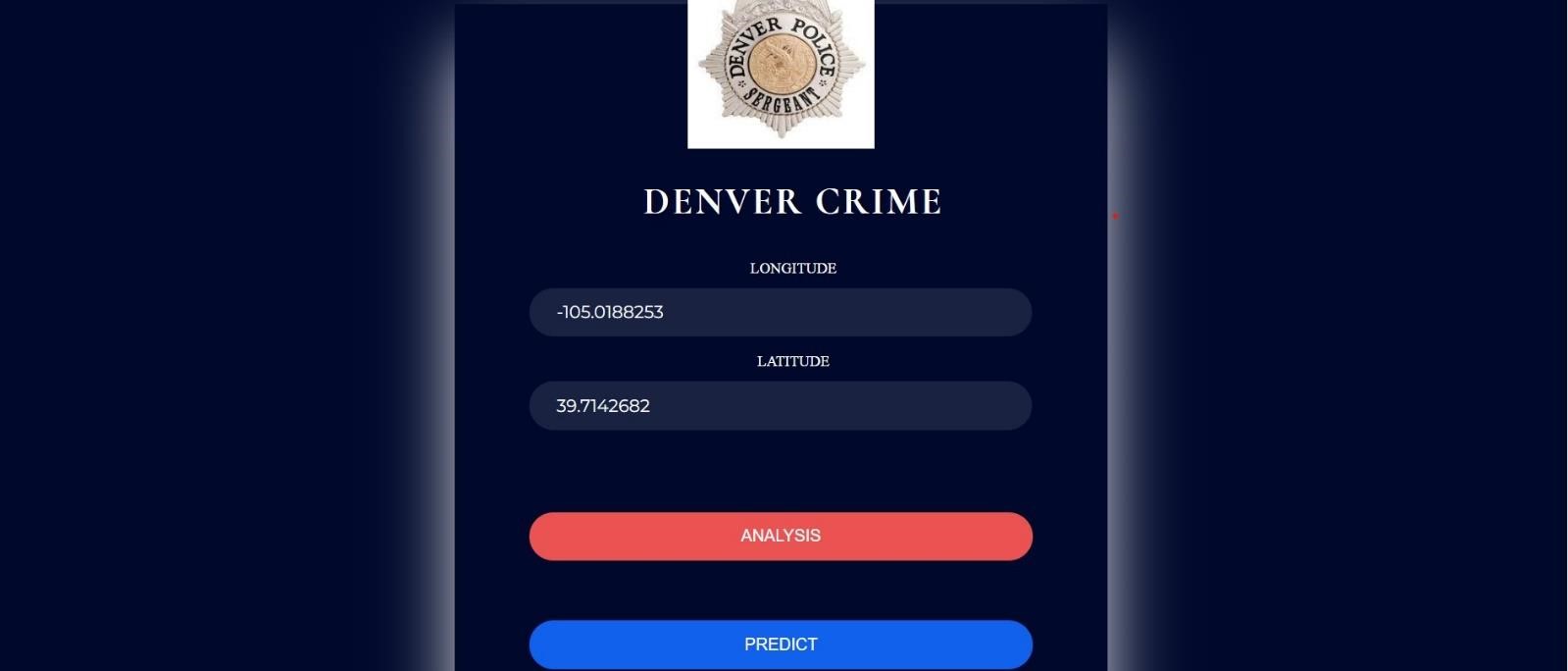


1. This is a figure of the crime analysis for per hour, day, month, year

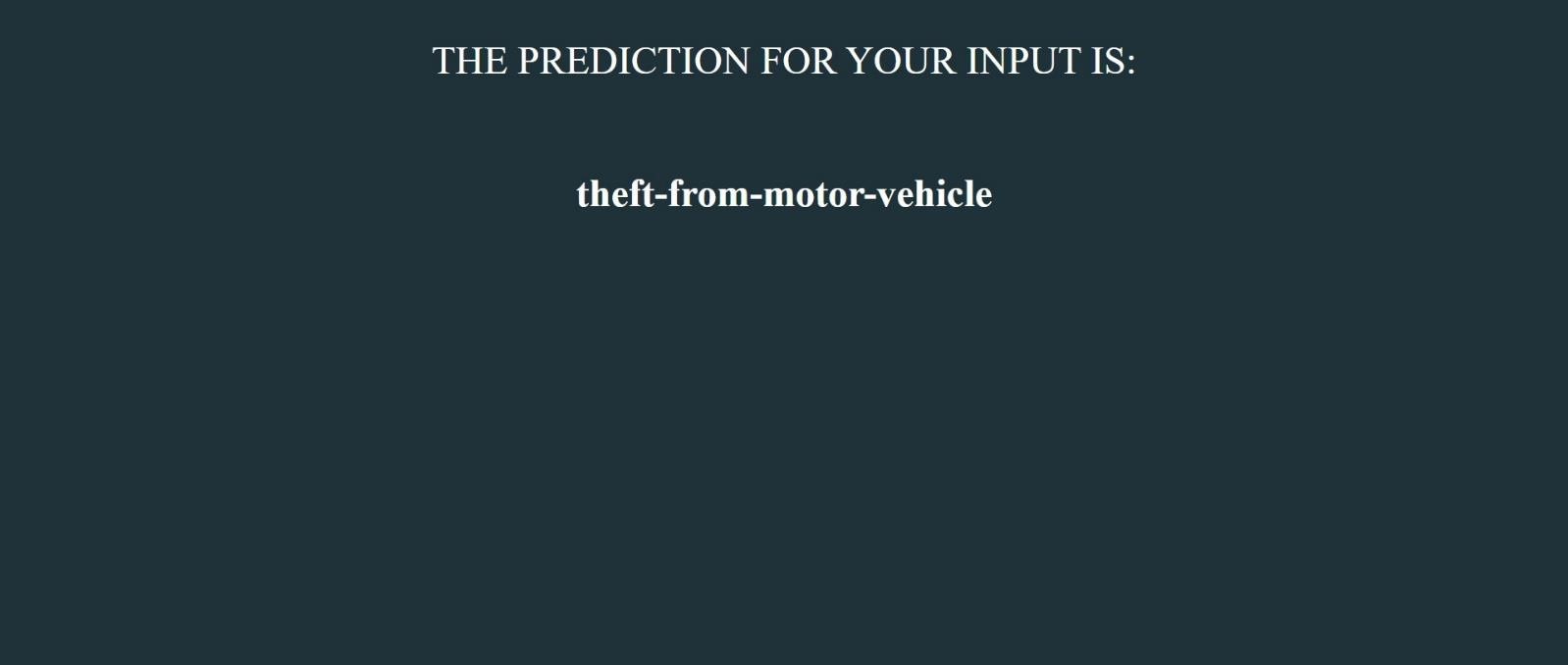


1. This is a figure of the Results achived by five algorithms used

# Prediction and outcome



1. This is the fgure of the login page



1. Result of the Crime when input coordnates are given

# Conclusion

Crime prediction is one the current trends in the society. Crime prediction intends to reduce crime occurrences. It does this by predicting which type of crime may occur in future. Here, analysis of crime and prediction are performed with the help of various approaches some of which are KNN, K means Clustering & Random Forest. However which model will work best is totally dependant on the dataset that is being used. This research work offers a way to foresee and predict crimes and frauds within a city. It focuses on having a crime prediction tool that can be helpful to law enforcement. This paper is aimed at increasing the prediction accuracy as much as possible. As compared to the previous work, this work was successful in achieving the highest accuracy in prediction. The KNN system helps law implementing agencies for improved and exact crime analysis. The result of the optimized k-means algorithm is efficient and provides improved accuracy of the final cluster reduced the number of iterations. We know that the prediction accuracy of random forest model can be improved based on historical crime data and covariates (POI data and demographic information). From the overall prediction results, the prediction model of random forest with covariates has better performance compared with Naive Bayes model and logistic regression model. As a combinatorial classification model, random forest overcomes the limitations of single decision tree classification and can effectively avoid the problem of overfitting. Among the model evaluation indexes selected in this paper, the random forest prediction model with covariates is better than other models.

# References

1. Pratibha, A. Gahalot, Uprant, S. Dhiman and L. Chouhan, "Crime Prediction and Analysis," 2nd International Conference on Data, Engineering and Applications (IDEA), 2020, pp. 1-6, doi: 10.1109/IDEA49133.2020.9170731.
2. S. Yao et al., "Prediction of Crime Hotspots based on Spatial Factors of Random Forest," 2020 15th International Conference on Computer Science & Education (ICCSE), 2020, pp. 811-815, doi:

10.1109/ICCSE49874.2020.9201899.

1. S. G. Krishnendu, P. P. Lakshmi and L. Nitha, "Crime Analysis and Prediction using Optimized KMeans Algorithm," 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC), 2020, pp. 915-918, doi: 10.1109/ICCMC48092.2020.ICCMC-000169.
2. A. Kumar, A. Verma, G. Shinde, Y. Sukhdeve and N. Lal, "Crime Prediction Using K-Nearest Neighboring Algorithm," 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE), 2020, pp. 1-4, doi: 10.1109/ic-ETITE47903.2020.155.