**­­­Crime Scene : Crime Prediction and Analysis**

**Abstract**

India is the second most populous country in the world with the Rapid increase in population and Rapid Urbanization of cities. It is not new that people hear of crimes happening quite often the crime rate has also been on increase which can threat to the safety of citizens of India in order to avoid the unfortunate, this research proposes a model that is able to predict the future crime based on the location pattern and time which can serve as a valuable source of knowledge for law enforcement or police department either from strategic or tactical perspectives to do so the random forest algorithm is used to increase the accuracy of the result.

**Index**

|  |  |  |
| --- | --- | --- |
| **Chapter No.** | **Particulars** | **Pg. No.** |
| 1 | Introduction | 1 |
|  | 1.1. Problem statement | 1 |
|  | 1.2. Aim and objective | 2 |
|  | 1.3. Motivation | 2 |
| 2 | Literature review | 3 |
|  | 2.1. Existing system | 3 |
|  | 2.2. Proposed System | 5 |
| 3 | Project description | 6 |
|  | 3.1. Modules | 7 |
|  | 3.1.1. Prediction | 7 |
|  | 3.1.2. Daily News | 8 |
|  | 3.1.3. Emergency Call | 11 |
|  | 3.2. Project Schedule | 14 |
|  | 3.3. Analysis. | 15 |
|  | 3.3.1. Functional Requirements | 15 |
|  | 3.3.2. Non-Functional Requirements | 16 |
|  | 3.4. H/W and S/W Requirements | 18 |
| 4 | System design | 19 |
|  | 4.1. UML Diagram | 20 |
|  | 4.2. DFD diagram etc. | 21 |
| 5 | Implementation Methodology | 24 |
| 6 | Results | 27 |
| 6 | Conclusion | 29 |
| 7 | References | 30 |

**Introduction**

**1.1 Problem Statement**

Crimes are the significant threat to the humankind. This project aims towards common people or public to help them in their lives by notifying them the current situation of their surroundings. Developing an android application that sets forth the numerous crime hotspots all over India to the user when the location is specified accompanied with latest newsletter and effortless crime reporting. Crime analysis and prediction is a systematic approach for identifying patterns and trends in crime. The crime rate prediction strategies can be applied on historical data available in the police records by examining the data at various angles like reason of crime, frequency of similar kind of crimes at specific location with other parameters**.**

**1.2 Aim and Objective**

* Displaying various crime hotspots all over India or on specified location with crime prediction and analysis and using google maps
* Daily crime news popups
* Crime reporting through visual or text media directly to nearby stations or government enterprises related to crime investigations
* Famous crime documentaries for the users who wish to gain more knowledge about the criminal laws and conduct.
* User Authentication and authorization for the safety of the user as well as for the convenience of keeping record of the crime being reported.

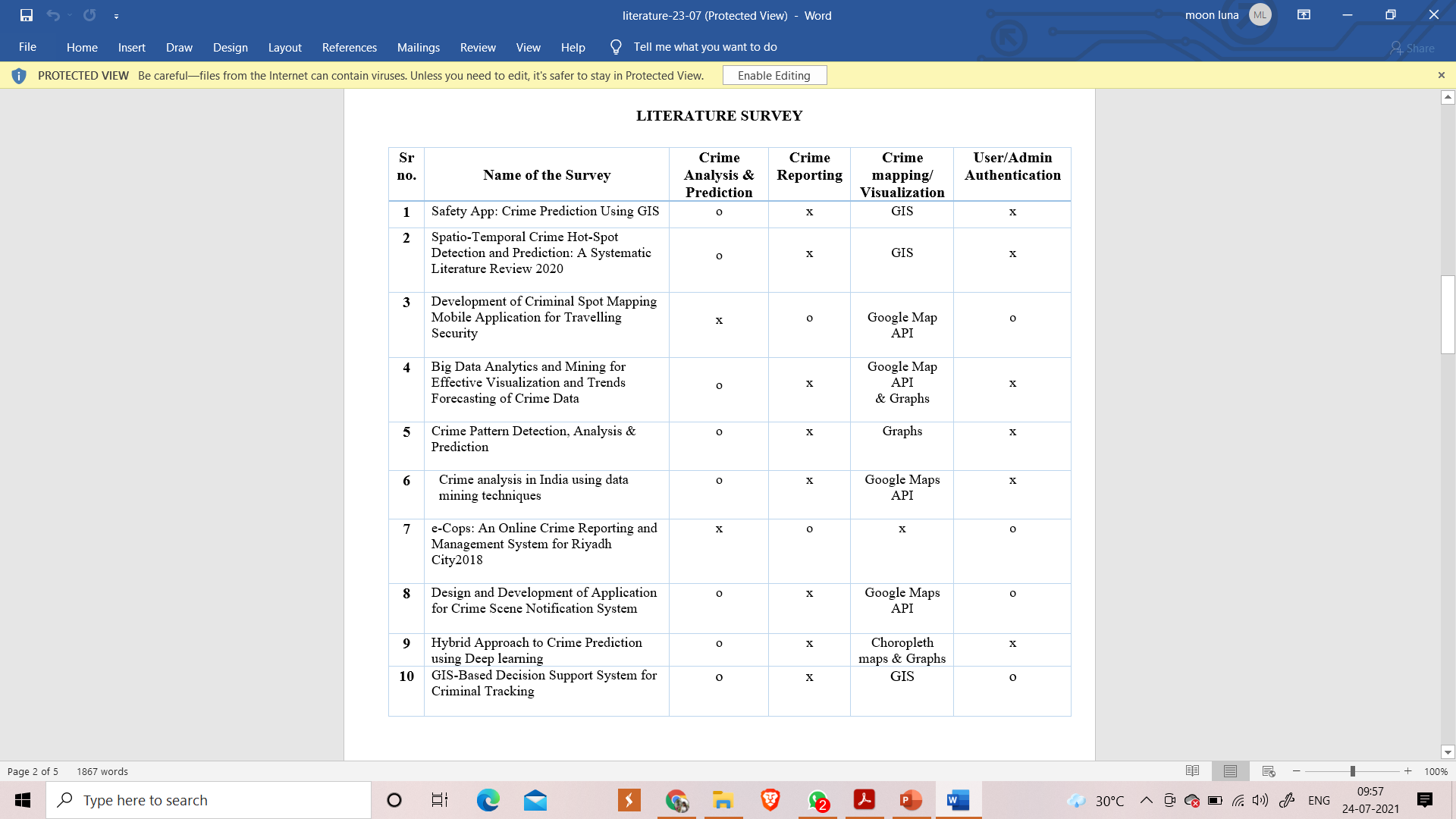
**1.3 Motivation**

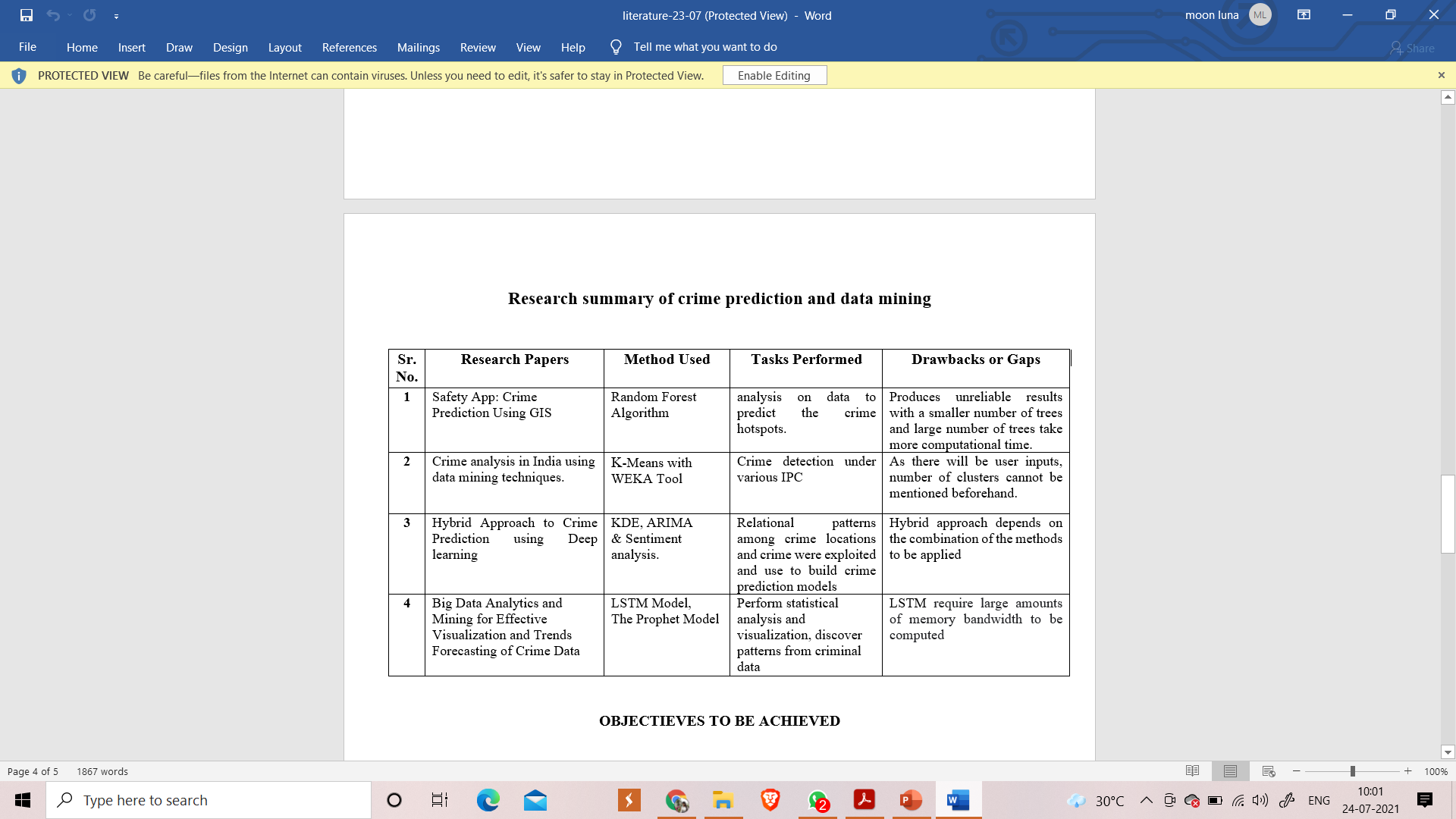
* Crime and Criminal activities are very sensitive topics to work on and this brought us more interest into building a app that can help us fight crimes
* When we started reading about crime we came across many sites and apps that were attracting us towards building something different.
* In today’s world crimes are increasing and so it was a trending need to perform a task that would help us and our country
* We came up with the idea of Crime Scene through our previous crime analysis and prediction project in 6th sem. We live in a world which isn’t safe for us and our family. This project is just a way to help you out to stay updated with different crimes taking place.

­**Literature review**

**2.1. Existing system**

* As to build Crime Scene, we came across an [1] IEEE paper **“Safety App: Crime prediction using GIS”.** This paper talks about the analysis done on different crimes which are then represented in the application in form of google maps. The application mentioned on this paper focuses only on the crimes taking place in Mumbai.
* **“Spatio-Temporal Crime Hot Spot Detection and Prediction”** IEEE paper describes hotspots of crime. A wide range of literature has been reported for crime hotspot detection and prediction. The primary objective is to investigate which methods are superior as compared to others in Spatio-temporal crime hotspot identification and predicting. One crucial point is to study the impact of Spatio-temporal datasets as compared to other datasets presented in the literature for crime hotspot identification and predicting.
* After analysing a couple of IEEE papers, we explored some crime-based applications. We have reviewed a number of safety applications that have been published on the Android Play Store/IOS AppStore such as “**Raksha, A Women’s Safety App**” and “The **Nirbhaya App**”. We discovered that these applications majorly focused on sending out an SOS alert to the guardians of the users in the times of difficulties and also track the user’s location, which in case of the Nirbhaya app could be triggered using a panic button.
* An android-based application is developed to monitor the safety of user while traveling. With this application, user’s inner circle can monitor their trip location and status therefore increasing the cautiousness. It provides up-to-date information about the security level of a certain place that can be estimated by application users by viewing vulnerability information, the ability to trigger panic mode automatically hen the user passes certain boundaries while doing a trip, reporting events, commenting and voting against vulnerable points allow users to be more aware of the environment.
* Use of big data analytics techniques and different types of visualization is described. Analysis on crime data in US cities (San-Francisco, Chicago and Philadelphia.) is done. For finding a pattern prophet model and LSTM algorithm is used. According to the author, the Prophet model and LSTM algorithm perform better than conventional neural networks. For prediction mechanism RMSE and spearman correlation is used. With the rise of data mining, developers are looking for a model that can remember past information for a longer time than LSTMs. In technical terms, they need high memory-bandwidth because of linear layers present in each cell which the system usually fails to provide for. Thus, hardware-wise, LSTMs become quite inefficient.





As to build Crime Scene, many papers were taken into consideration. “Safety App: Crime prediction using GIS”. This paper talks about the analysis done on different crimes which are then represented in the application in form of google maps. The application mentioned on this paper focuses only on the crimes taking place in Mumbai. On further research we came across few more papers. [2] “Spatio-Temporal Crime Hot Spot Detection and Prediction” IEEE paper describes hotspots of crime. A wide range of literature has been reported for crime hotspot detection and prediction. The primary objective is to investigate which methods are superior as compared to others in Spatio-temporal crime hotspot identification and predicting. One crucial point is to study the impact of Spatio-temporal datasets as compared to other datasets presented in the literature for crime hotspot identification and predicting. The paper is fully focused on GSI technique. [10] This paper aims to develop a GIS-based decision support system to help in improving the police personnel response to crimes. The system aids the police personnel to get critical spatial and non-spatial information. The system can identify the nearest police stations to the crime location, and also determine the shortest route from the selected police station to the crime location. Below are some research papers which were considered while building this particular application. [3] An android-based application is developed to monitor the safety of user while traveling. With this application, user’s inner circle can monitor their trip location and status therefore increasing the cautiousness. It provides up-to-date information about the security level of a certain place that can be estimated by application users by viewing vulnerability information, the ability to trigger panic mode automatically hen the user passes certain boundaries while doing a trip, reporting events, commenting and voting against vulnerable points allow users to be more aware of the environment.

[4] Use of big data analytics techniques and different types of visualization is described. Analysis on crime data in US cities (San-Francisco, Chicago and Philadelphia.) is done. For finding a pattern prophet model and LSTM algorithm is used. According to the author, the Prophet model and LSTM algorithm perform better than conventional neural networks. For prediction mechanism RMSE and spearman correlation is used. With the rise of data mining, developers are looking for a model that can remember past information for a longer time than LSTMs. In technical terms, they need high memory-bandwidth because of linear layers present in each cell which the system usually fails to provide for. Thus, hardware-wise, LSTMs become quite inefficient.

[5] Clustering of crime and crime profiling has been done. According to the author, visualization of clusters on boxplot and prediction on the same would be beneficial for the police department to make optimal decisions as well as it would be useful for strategy building. Dataset from Chicago police station is used for data processing. K-means for clustering and SVM for prediction is used

 [6] Author has used k-means clustering for crime identification and for crime classification neural networks as well as Random Forest algorithms are used. Google maps is utilized for visualization purposes in which crime spots are marked on the Indian map. For verification of the mining process WEKA tool was used. K means cannot handle noisy data and outliers and is not suitable to identify clusters with non-convex shapes as crime is neither systematic nor entirely random.

[7]We found about online crime reporting. A website or an application exclusively for crime reporting the process of registration and form filling makes it very tedious to file a complaint during emergencies. Hence in our application we are considering to incorporate crime reporting where the user can upload a visual or text media of the crime witnessed.

[8]This research proposes a development of a notification system for the crime scene investigation on android devices and web application. This system facilitates form-filling, file-recording, saving multimedia files, geographical search, navigation, and push notification message.

[9] Use of sentiment analysis to calculate the crime rate in certain areas considering the various posts that have been uploaded on social media. The social media often circulate erroneous news which can give the user wrong data and may cause inconvenience. Hence in our project, we will be considering the crimes that have been registered under the Indian Penal Code which the official criminal code of conduct in India along with sentiment analysis.

By considering flaws in the papers’, we are going to design our system to overcome those flaws. We also incorporate multivariate visualization, graph mining techniques and trends within these datasets. Machine learning is similar to data mining, concepts of machine learning can be used for better prediction. The data privacy, reliability, accuracy can be improved for enhanced prediction. The knowledge which is acquired from the data mining techniques will help in reducing crimes as it helps in finding the culprits faster and also the areas that are most affected by crime.

**2.2. Proposed System**

The different crimes happening around the country were taken from the Indian government site: <https://data.gov.in/> This data needs to be first processed and cleaned for classification. Prediction will be performed district wise. After processing the data will be following the ensemble learning method which helps increasing the accuracy of prediction data. The application will be capable of prediction the crimes in particular area and will also provide analysis. There are different algorithms to predict data from which random forest was is to selected for predict crime. We are also considering of having more features in our application which are given below

Emergency call:  [Emergency](https://en.wikipedia.org/wiki/Emergency) [location-based service](https://en.wikipedia.org/wiki/Location-based_service) (LBS) available on [smartphones](https://en.wikipedia.org/wiki/Smartphone) that, when a caller dials the local (in country) short dial [emergency telephone number](https://en.wikipedia.org/wiki/Emergency_telephone_number), sends the best available [geolocation](https://en.wikipedia.org/wiki/Geolocation) of the caller to a dedicated end-point (like family, friends, police)

Crime news Daily: Will provide the user with only crime related news. This will help the user to be updated by the crimes that are occurring in their vicinity or in their city etc.

Report the crime: We all have seen crime related videos getting viral on internet. Instead of making them viral on internet we can provide user to upload videos or images or tweets against the crime they must have come across. These videos and images can then be used as evidence for that particular crime investigation.

**Project description**

**3.1. Modules :**

**3.1.1 Prediction**

The ability to predict the crime which can occur in future can help the law enforcement agencies in preventing the crime before it occurs. the crime prediction and analysis methods are very important to detect the future crimes and reduce them for crime prediction, Random Forest , Decision trees algorithms are used.

**3.1.2 Daily News**

This module gives the news in India There's also a list of Headlines, which is a mix of news across various categories, such as sports, tech and business.

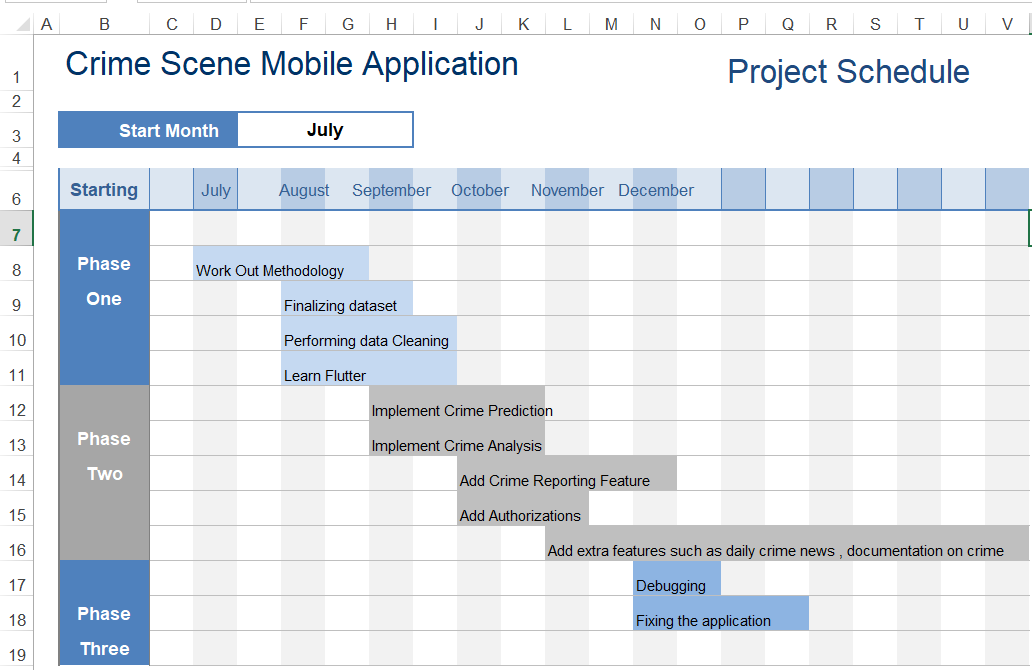
**3.1.3 Emergency call**

It notifies to your family and authorities in the case of an emergency. Phone numbers and messages are pre-stored in the app and triggered by pressing one button

**3.1.3 Crime Report**

Help you to post evidence of crime in your area. Example photos videos etc.;

**3.2. Project Schedule**



**3.3. Analysis**

**3.3.1. Functional Requirements :**

A Functional Requirement (FR) is a description of the service that the software must offer. It describes a software system or its component. It can be a calculation, data manipulation, business process, user interaction, or any other specific functionality which defines what function a system is likely to perform.

* Predict crime
* Analysis data correctly
* Working feature – news section should be able to give updated news etc.

**3.3.2 Non-functional Requirements :**

Non function requirement are those requirement of the system which are not directly concerned with specific functional delivered by the system. They may be related to emergent properties such as reliability, extendibility, usability and so on.

* To provide prediction of crime.
* To provide maximum accuracy.
* Provide visualized analysis.
* Ease of use.
* Availability
* Reliability
* Maintainability

**3.4. H/W and S/W Requirements**

Hardware Details

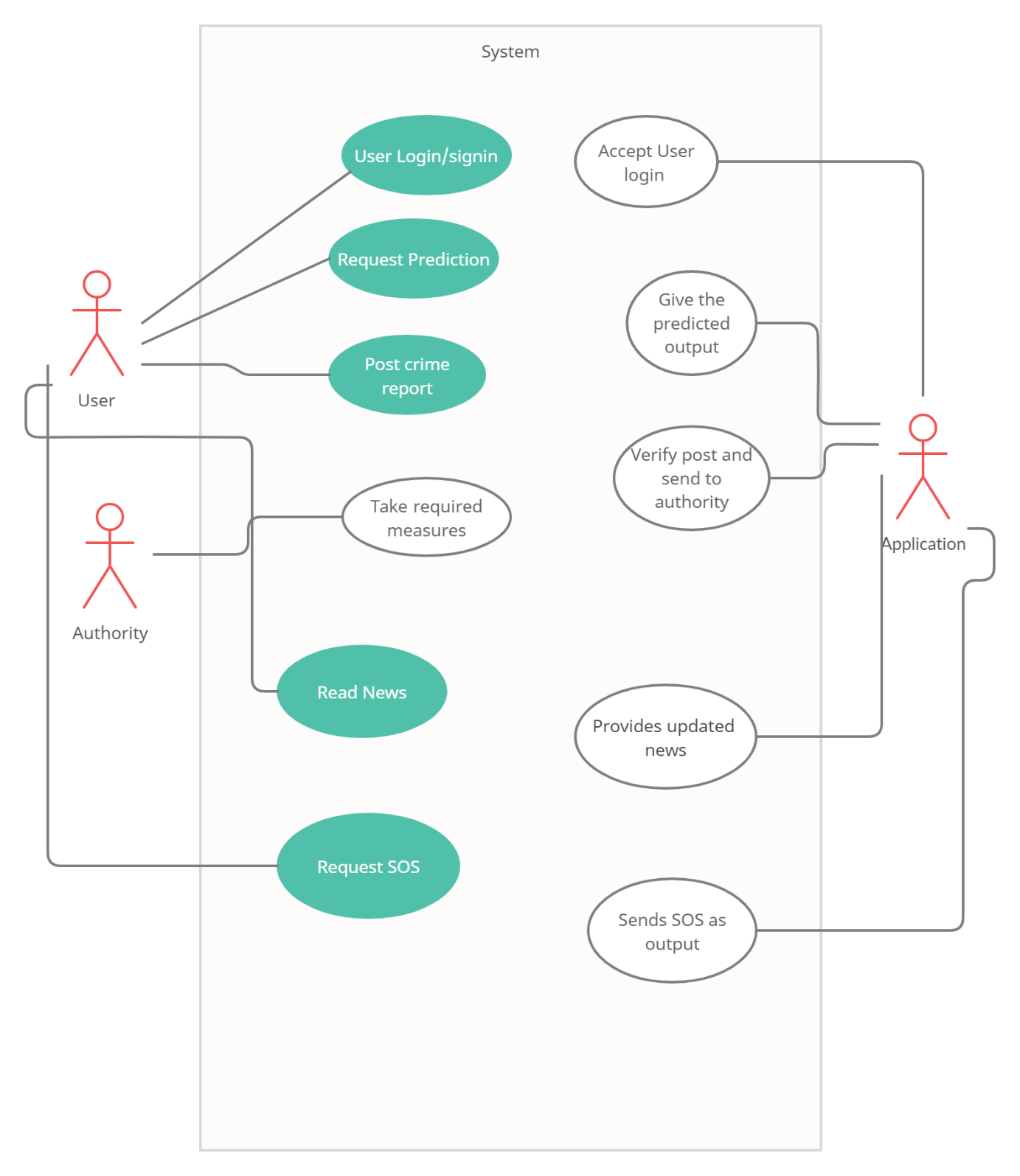
* Operating system: Windows 7 or newer, 64-bit macOS 10.9+, or Linux.
* System architecture: 64-bit x86, 32-bit x86 with Windows or Linux.
* CPU: Intel Core 2 Quad CPU Q6600 @ 2.40GHz or greater.
* RAM: 4 GB or greater.

Software Details

* Android Studio
* Flutter SDK

**System design**

**4.1. UML Diagram**



**Implementation Methodology**

The Ensemble Learning:

Ensemble learning is the process by which multiple models, such as classifiers, are strategically generated and combined to solve a particular computational intelligence problem.

Some researchers suggest that the performance of the ensembles depends on two properties, which are the individual success of the base classifiers of the ensemble and the independence of the base classifier’s results from each other, while some other researcher suggests that the accuracy of individual models, diversity among the individual models, decision making strategy, and number of base classifiers used for constructing an ensemble are among the factors responsible for the success of an ensemble. An ensemble is itself a supervised learning algorithm because it can be trained and then used to make predictions.

Ensemble methods are techniques that aim at improving the accuracy of results in models by combining multiple models instead of using a single model. The combined models increase the accuracy of the results significantly. This has boosted the popularity of ensemble methods in machine learning. Bagging, the short form for bootstrap aggregating, is mainly applied in classification and regression. It increases the accuracy of models through decision trees, which reduces variance to a large extent. Bagging is classified into two types:

Bootstrapping:

is a sampling technique where samples are derived from the dataset using the replacement procedure. The sampling with replacement method helps make the selection procedure randomized. The base learning algorithm such as decision tree is run on the samples to complete the procedure.

Aggregation:

in bagging is done to incorporate all possible outcomes of the prediction and randomize the outcome. Without aggregation, predictions will not be accurate because all outcomes are not put into consideration. Therefore, the aggregation is based on the probability bootstrapping procedures or on the basis of all outcomes of the predictive models.

Bagging is advantageous since weak base learners are combined to form a single strong learner that is more stable than single learners. It also eliminates any variance, thereby reducing the overfitting of models. One limitation of bagging is that it is computationally expensive. Thus, it can lead to more bias in models when the proper procedure of bagging is ignored

Variance Reduction:

Ensemble methods are ideal for reducing the variance in models, thereby increasing the accuracy of predictions. The variance is eliminated when multiple models are combined to form a single prediction that is chosen from all other possible predictions from the combined models. An ensemble of models combines various models to ensure that the resulting prediction is the best possible, based on the consideration of all predictions.

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.

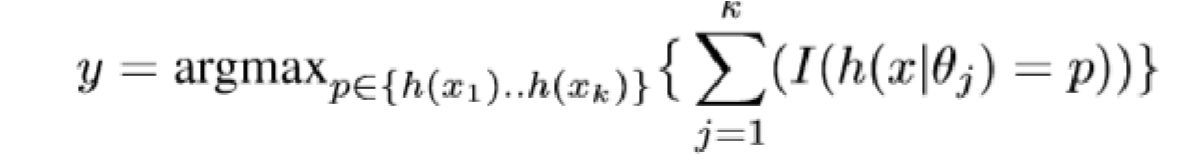
Non-parametric: Fast and efficient. It is a graphical representation for getting all the possible solutions to a problem/decision based on given conditions

Decision tree algorithm:

* Step-1: Begin the tree with the root node, says S, which contains the complete dataset.
* Step-2: Find the best attribute in the dataset using Attribute Selection Measure (ASM)
* Step-3: Divide the S into subsets that contains possible values for the best attributes.
* Step-4: Generate the decision tree node, which contains the best attribute.
* Step-5: Recursively make new decision trees using the subsets of the dataset created in step -3. Continue this process until a stage is reached where you cannot further classify the nodes and called the final node as a leaf node.

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. Random forest is considered as a highly accurate and robust method because of the number of decision trees participating in the process. It does not suffer from the overfitting problem. The main reason is that it takes the average of all the predictions, which cancels out the biases. Random forest 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 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.

Random forest algorithm:

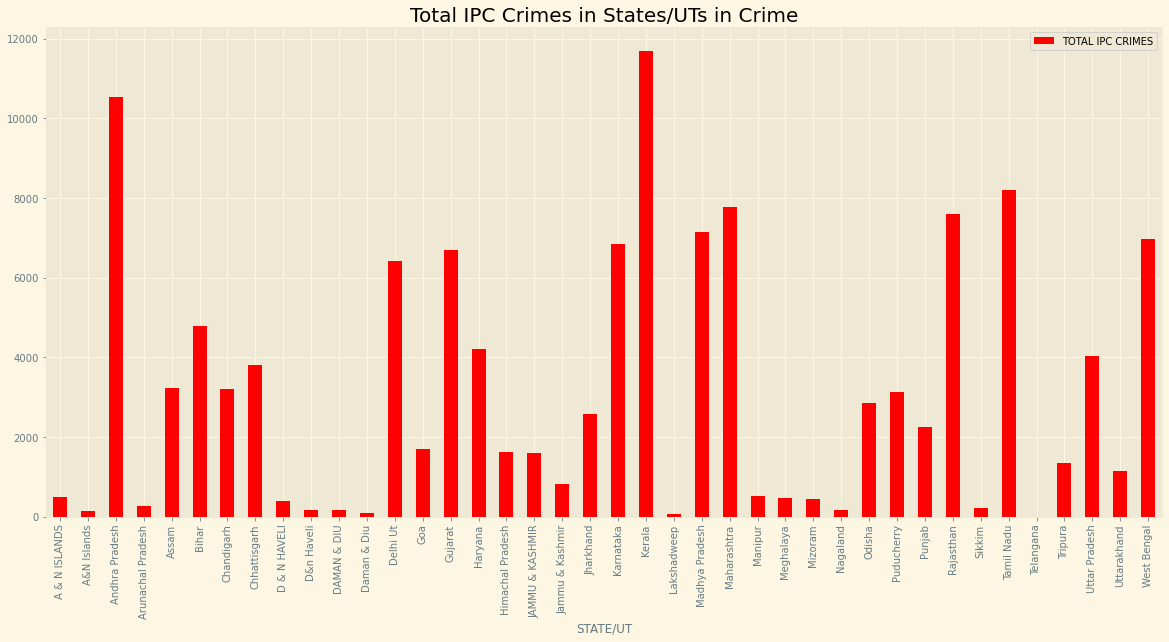
* Step-1: Select random samples from a given dataset.
* Step-2: Construct a decision tree for each sample and get a prediction result from each decision tree.
* Step-3: Perform a vote for each predicted result.
* Step-4: Select the prediction result with the most votes as the final prediction.

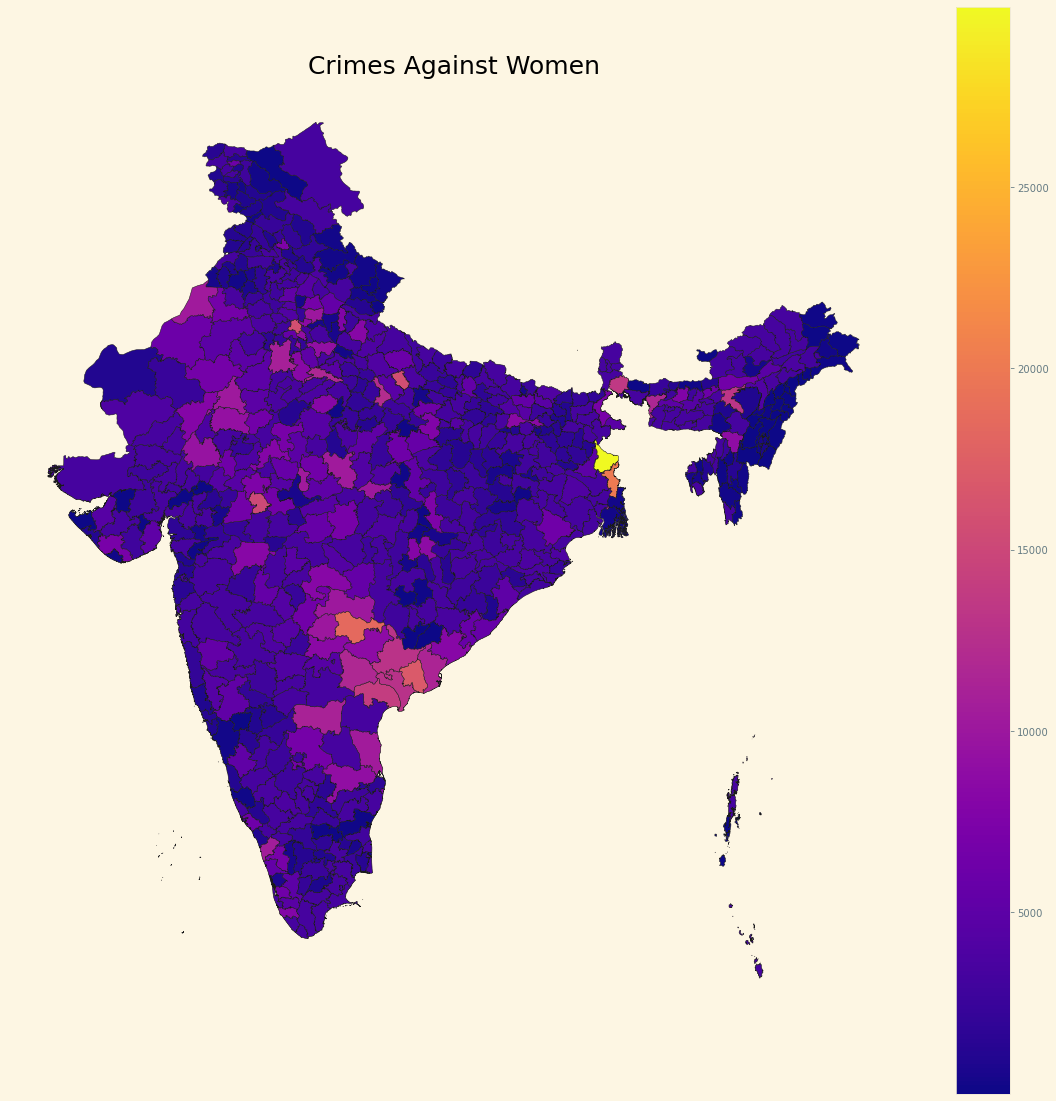
**Results**

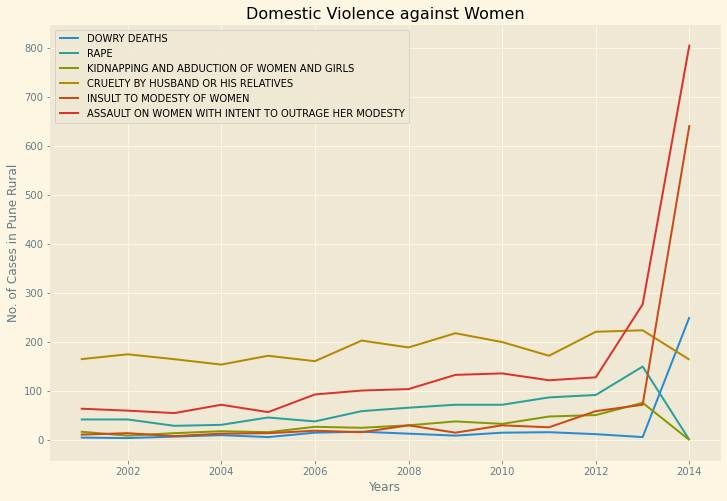
A word-cloud to explain the dataset

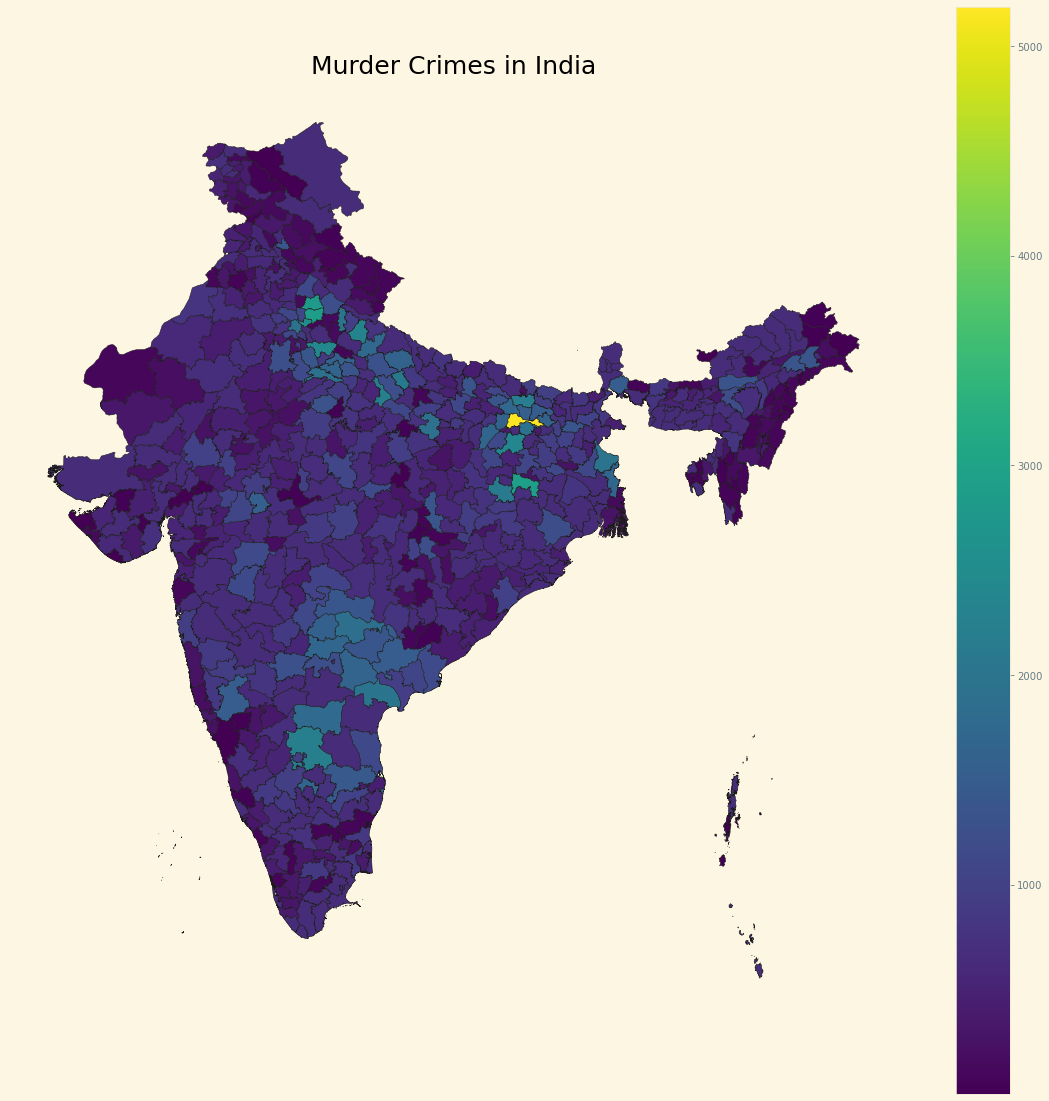


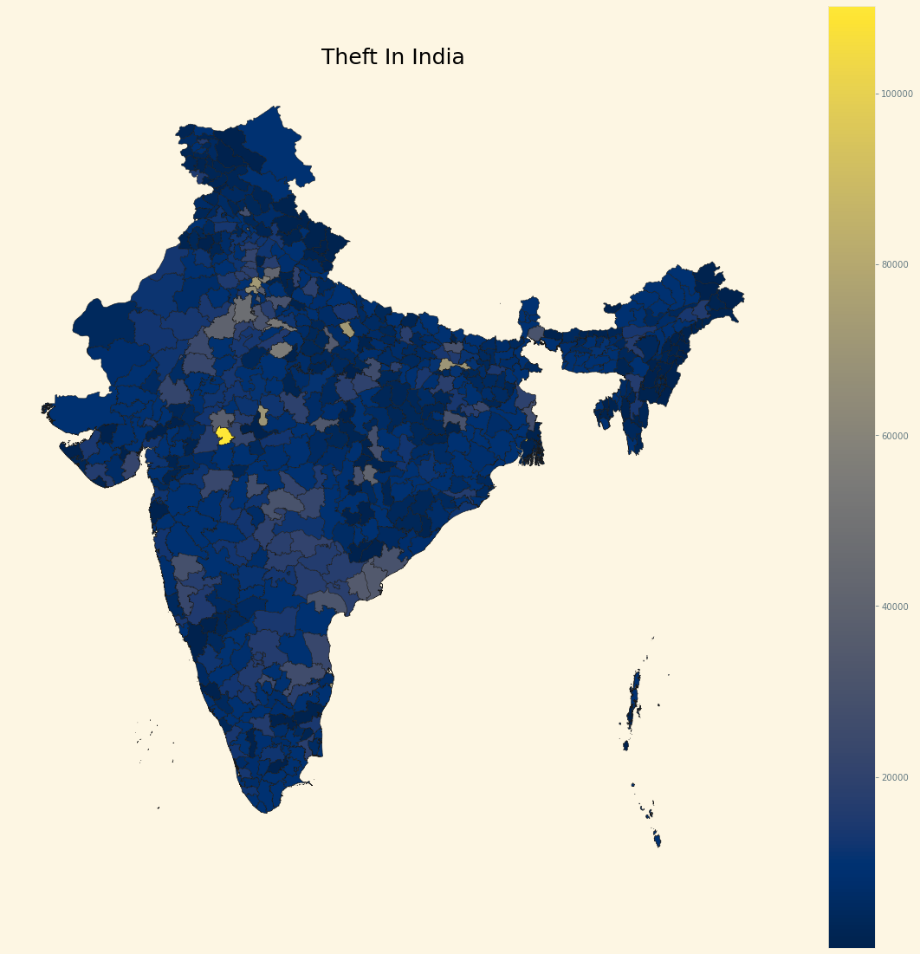
Total IPC Crimes from year 2001-14 in all the State/UTs

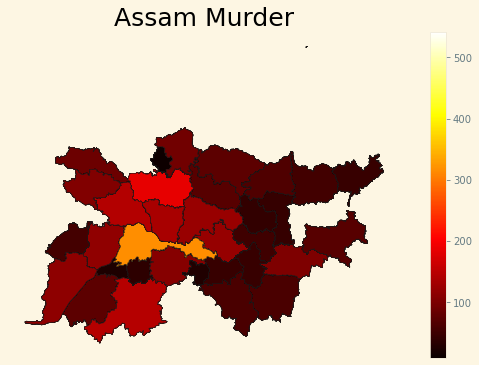


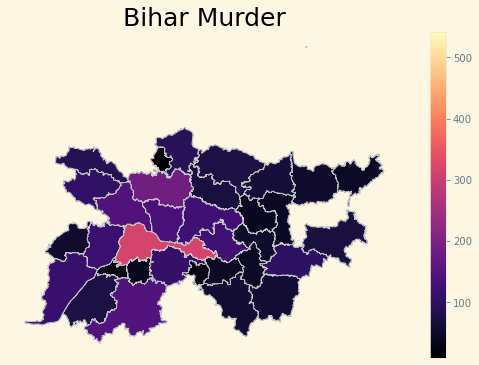


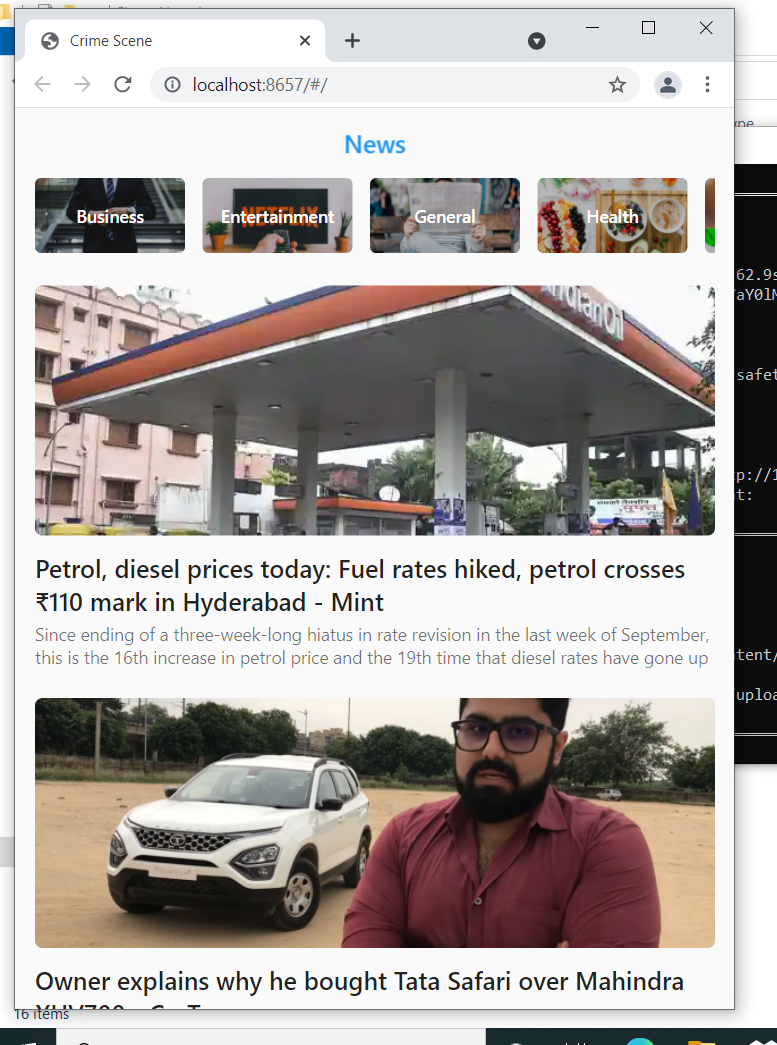












**Conclusion**

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. This paper identifies mobile phone technology as a crime fighting platform in society, as its use can effectively bridge the communication gap between the police and the general public in crime fighting. allowing both the police and the general public to interact more effectively with the help of a mobile application. In this paper, the complete analysis, classification, and prediction of the crime in India is done using Random forest algorithm with the data provided by data.gov the algorithm involves multiple decision tress on various subsets of the given dataset and combines to final prediction The stated results in this paper show that Bagging method works best for predicting crimes

**References**

* 1.Safety App: Crime Prediction Using GIS

<https://ieeexplore.ieee.org/document/9137772>

* 2.Spatio-Temporal Crime Hot-Spot Detection and Prediction: A Systematic Literature Review

<https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9187772>

* 3.Development of criminal spot mapping mobile application for travelling security

<https://ieeexplore.ieee.org/document/8273541>

* .Crime pattern detection, analysis & prediction

<https://ieeexplore.ieee.org/document/8203676>

* 6.Crime analysis and prediction using data mining

<https://ieeexplore.ieee.org/document/6906719>

* 7.e-Cops: An Online Crime Reporting and Management System for Riyadh City

<https://ieeexplore.ieee.org/document/8441987>

* 8.Design and development of application for crime scene notification system

<https://ieeexplore.ieee.org/document/8074103>

* 9.Hybrid approach to crime prediction using deep learning

<https://ieeexplore.ieee.org/document/7275858>

* 10.GIS-based decision support system for criminal tracking

<https://ieeexplore.ieee.org/document/6523543>