EXPLORE INNOVATION EXCEL

# SOFTWARE ENGINEERING PROJECT
# TOPIC : CRIME RATE PREDICTION AND ANALYSIS
# SRS DOCUMENT
# SEMESTER : 05


| S.NO. | NAME | SID |
|---|---|---|
| 1 | JASKIRAT SINGH | 20103048 |
| 2 | TANVI PUNDIR | 20103093 |
| 3 | RAHUL KERLA | 20103096 |
| 4 | BHARAT | 20103101 |

## PROBLEM STATEMENT:-

Crimes now a days are increasing day by day and with different level of intensity and versatility. The result is a great loss to society in terms of monitory loss, social loss and further it enhances the level of threat against the smooth livelihood in the society. To overcome this problem, the computing era can help to reduce the crime or even may be helpful in predicting the crime so that sufficient measures can be taken to minimize the loss to property and life. 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 to prepare the model crime prediction. It is a major challenge to understand the versatile data available with us, then model it to predict the future incidence with acceptable accuracy and further to reduce the crime rate.

Crimes now a days are increasing day by day and with different level of intensity and versatility. The result is a great loss to society in terms of monitory loss, social loss and further it enhances the level of threat against the smooth livelihood in the society. To overcome this problem, the computing era can help to reduce the crime or even may be helpful in predicting the crime so that sufficient measures can be taken to minimize the loss to property and life. 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 to prepare the model crime prediction. It is a major challenge to understand the versatile data available with us, then model it to predict the future incidence with acceptable accuracy and further to reduce the crime rate.

# EXISTING SYSTEM:-

k-NN, RF, SVM and Bayes models are existing methods Although studies have been done in the medical field with an advanced data exploration using machine learning algorithms, orthopaedic disease prediction is still a relatively new area and must be explored further for the accurate prevention and cure. it mines the double layers of hidden states of vehicle historical trajectories, and then selects the parameters of Hidden Markov Model(HMM) by the historical data. In addition, it uses a Viterbi algorithm to find the double layers hidden states sequences corresponding to the just driven trajectory. Finally, it proposes a new algorithm for vehicle trajectory prediction based on the hidden Markov model of double layers hidden states, and predicts the nearest neighbour unit of location information of the next k stages.

# DISADVANTAGES OF EXISTING SYSTEM:-

- By this methodology they had less accuracy in prediction.
- By this methodology results are not perfect.

# PROPOSED SYSTEM:-

The proposed system is made on the basis of the research work that is done by going through various such documentations. Nearly all of the crimes are predicting based on the location and the types of crimes that are occurring in those areas. On surveying previous works, Linear Regression, Decision Tree and Random Forest tend to give good accuracy so these models are used in this paper to predict crimes. The dataset used in this paper is fromdata.world.com. The data set contains different types of crimes that being committed in India according to the

state and year respectively. This paper takes types of crimes as input and gives the area in which crimes are committed as output. The data pre-processing involves data cleaning, feature selection, dropping null values, data scaling by normalizing and standardizing. After data pre-processing the data is free of null values which m ay alter the accuracy of the model significantly and feature selection is used to select only the required features that won't affect the accuracy of model. After data pre-processing the models chosen i.e., Logistic Regression, Decision Tree and Random Forest are trained by splitting the data into as train and test data. As the output required is a categorical value classification models are used here. Python language is used for the data prediction.

# ADVANTAGES OF PROPOSED SYSTEM:

- We had a high accuracy in this model prediction methodology. In this algorithm for a data mining approach to help predict the crimes patterns and fast up the process of solving crime.
- The results are perfect and accurate using this technology.

---

# FUNCTIONAL REQUIREMENTS:-

1.Data Collection

2.Data Pre-processing

3.Training and Testing

4.Modiling

5.Predicting

---

# NON FUNCTIONAL REQUIREMENTS:-

NON-FUNCTIONAL REQUIREMENT (NFR) specifies the quality attribute of a software system. They judge the software system based on Responsiveness, Usability, Security, Portability and other non-functional standards that are critical to the success of the software system. Example of non-functional requirement, "how fast does the

website load?" Failing to meet non-functional requirements can result in systems that fail to satisfy user needs. Non- functional Requirements allows you to impose constraints or restrictions on the design of the system across the various agile backlogs. Example, the site should load in 3 seconds when the number of simultaneous users are > 10000. Description of non-functional requirements is just as critical as a functional requirement.

- Usability requirement
- Serviceability requirement
- Manageability requirement
- Recoverability requirement
- Security requirement
- Data Integrity requirement
- Capacity requirement

# HARDWARE REQUIREMENTS

Minimum hardware requirements are very dependent on the particular software being developed by a given in thought Python / Canopy / VS Code user. Applications that need to store large arrays/objects in memory will require more RAM, whereas applications that need to perform numerous calculations or tasks more quickly will require a faster processor.

| | |
|---|---|
| • Operating system | • windows, Linux |
| • Processor | • minimum intel i3 |
| • Ram | • minimum4gb |
| • Hard disk | • minimum 250gb |

# SOFTWARE REQUIREMENTS

The functional requirements or the overall description documents include the product perspective and features, operating system and operating environment, graphics requirements, design constraints and user documentation.
The appropriation of requirements and implementation constraints gives the general overview of the project in regards to what the areas of strength and deficit are and how to tackle them.

- **Python idel 3.7 version (or)**
- **Anaconda 3.7 (or)**
- **Jupiter (or)**
- **Google colab**

# ARCHITECTURE DIAGRAM:-

# SYSTEM COMPONENTS(MODULES):

- **Import libraries-**In this module different libraries are imported, which are useful for data processing, visualization and prediction etc.
- **Data Processing-**In this module data undergoes pre-processing, like cleaning of dataset and removing null and unwanted values etc.
- **Visualization of crime against rape-**We have different times of crime data but initially we are taking rape cases. In this module data undergoes Visualization of crime against rape.
- **Model Generation-**In this module model is build.
- **Build Naive Bayes Gaussian Classifier-**In this module Naive Bayes Gaussian Classifier is used to data analysis.
- **Build Decision Tree Classifier-**Decision Tree Classifier is used to data analysis.
- **Build KNN Classifier-**In this module KNN Classifier is used to data analysis
- **Build Boosted Decision Tree Classifier-**In this module Boosted Decision Tree is used to data analysis
- **Accuracy Comparison-**In this module Accuracy comparison is shown. Accuracy is compared among the different kinds of classifiers used.