



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 12, Special Issue 2, April 2023

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

 9940 572 462

 6381 907 438

 ijirset@gmail.com

 www.ijirset.com

Crime Type Prediction and Analysis based on location using Machine Learning

Amrisha M¹, Dhivya R¹, Kalaimani G¹, S.Sridevi²

U. G Student, Department of Computer Science Engineering, Velammal Engineering College, Chennai,
Tamil Nadu, India

Assistant Professor, Department of Computer Engineering, Velammal Engineering College, Chennai,
Tamil Nadu, India

ABSTRACT: The prevention of crime is a crucial undertaking because it is one of our society's most significant and pervasive problems. The goal of this research is to evaluate a dataset that includes many crimes and make predictions about the kinds of crimes that could occur in the future depending on various factors. In this project, we will use data science and machine learning to predict crimes using a set of crime data of Indian States. It includes details on the offence, including the time, date, place, and kind of crime. Data preprocessing will be performed prior to training the model, and this will be followed by feature selection and scaling to ensure high accuracy. The activities are high. The soul purpose of this project is to give a just idea of how machine learning can be used by the law enforcement agencies to detect, predict and solve committing crimes at a significantly faster rate, which lowers the crime rate.

KEYWORDS: K-Nearest Neighbor Support, Vector Machine Autoregressive moving average, recurrent neuralnetwork, Recursive Feature Elimination, National Crime Records Bureau.

I. INTRODUCTION

Crimes are the significant trouble to the humankind. Some crimes occur often at predictable intervals. Maybe it's growing and dispersing quickly and widely. Crimes be from small vill, city to big metropolises. Crimes are of different type – thievery, murder, rape, assault, battery, false imprisonment, hijacking, homicide. Since crimes are adding there's a need to break the cases in a important faster way. The crime conditioning have been increased at a faster rate and it's the responsibility of police department to control and reduce the crime conditioning. Crime vaticination and felonious identification are the major problems to the police department as there are tremendous quantum of crime data that live. There's a need of technology through which the case working could be briskly. The purpose of this project is to forecast crime using the dataset's attributes. The official websites are where the dataset was taken. The type of crime that will occur in a specific region can be predicted with the aid of machine learning algorithms, which use Python as their basic language. To train a model for prediction would be the goal. The test dataset will be used to verify the training using the training dataset. Depending on the accuracy, a better algorithm will be used to build the model. For crime prediction, the K-Nearest Neighbor (KNN) classification and other algorithms will be employed. The dataset is visualised to examine potential crimes that may have occurred in the nation. This work aids the law enforcement agencies to predict and detect crimes in Indian States with improved accuracy and thus reduces the crime rate.

II. RELATED WORK

The issue of reducing crime has been the subject of numerous studies, and numerous methods for crime prediction have been put forth. The type of data used and the attributes used for prediction affect the prediction's accuracy. According to , the use of mobile network activities to gather information on human behaviour and anticipate crime hotspots When forecasting whether a particular neighbourhood in London's metropolis would be a hotspot for crime or not, the accuracy rate is around 70%. The Naive Bayes algorithm and decision trees were employed in to predict and categorise crime using data gathered from various websites and newsletters, and it was discovered that the former performed better. A thorough analysis of multiple crime prediction techniques, including Support Vector Machines (SVM) and Artificial Neural Networks (ANN), was conducted in , and it was found that no one technique can effectively address the issues associated with distinct crime datasets. In order to increase the predicted accuracy of crime, different

supervised learning techniques and unsupervised learning techniques on crime records have been carried out. These techniques address the relationships between crime and crime pattern. Several approaches to prediction, including data mining, deep learning, crime casting, and sentiment analysis, were examined in , and it was discovered that each method has both advantages and disadvantages. Every technique produces superior results in a specific situation. Criminal activity was detected using clustering techniques, and criminal activity was predicted using classification techniques . K-Means clustering was used, and the accuracy of their performance was assessed. By comparing the results of various clustering algorithms, DBSCAN produced results with the best degree of accuracy, and the KNN classification technique is used to predict crime. As a result, this approach aids in more precise and effective crime analysis for law enforcement authorities. With the help of the data mining programme WEKA, a comparison of the classification methods Naive Bayes and decision tree was carried out in . The USCensus 1990 provided the datasets for this investigation. examined the road accident pattern in Ethiopia while taking into account a number of variables, including the driver, vehicle, and road conditions. For a dataset with about 18000 data points, various classification techniques like K-Nearest Neighbor, Decision tree, and Naive Bayes were utilised. The three approaches have prediction accuracy ranging from 79% to 81%.

III. METHODOLOGY

Predictive modelling is the process of creating a model that is capable of making predictions. A machine learning algorithm is used in the procedure to create those predictions by learning specific properties from a training dataset. The two subfields of predictive modelling are regression and pattern categorization.

Regression models are based on the investigation of relationships between variables and trends in order to make predictions about continuous variables. The goal of pattern classification, in contrast to regression models, is to assign discrete class labels to a specific data value as an output of a prediction. An illustration of a classification model is Predicting whether it will be sunny, wet, or snowy today could be a pattern classification job for weather forecasting. The duties of pattern classification can be split into two parts, Supervised and unsupervised learning. In supervised learning, the classification model's input dataset's class labels are predetermined. In a supervised learning problem, we would know which training dataset has the particular output through which we can predict unseen data.

Types of Predictive Models Algorithms

- a. **Classification and Decision Trees:** A decision tree is an algorithm that makes use of a graph or model of decisions in the shape of a tree, encompassing costs, utility, and outcomes of random events. That is one method of showing an algorithm. Naive Bayes classifiers are a family of straightforward probabilistic classifiers used in machine learning. They are based on the application of the Bayes theorem with independence assumptions between the features. The method creates classifier models that assign class labels to issue occurrences, represented as vectors of feature values, where the class labels are chosen at random from a small set.
- b. **Linear Regression:** The analysis is a statistical procedure for estimating the associations among variables. A method for modelling the connection between a scalar dependent variable (Y) and one or more explanatory variables (X) is known as linear regression. In cases where there is only one explanatory variable, simple linear regression is employed. Multivariate refers to multiple variables.
- c. **Logistic Regression:** A regression model known as logistic regression in statistics uses a categorical or binary dependant variable.

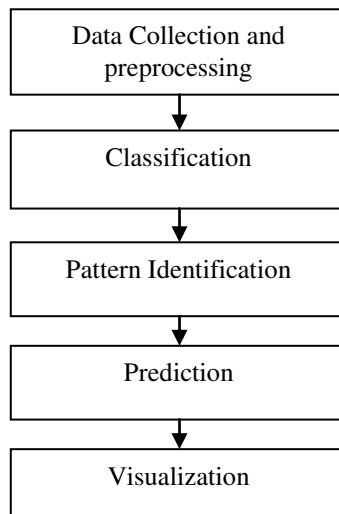


Figure 1 Architecture

A..Data Collection and Preprocessing

Data Collection is the process of gathering precise data from many sources and analysing it to find patterns, opportunities, and answers to study problems as well as to weigh potential repercussions. The CSV formatted crime dataset from kaggle is used.

Data pre-processing is the process includes measures to get rid of any infinite or null values that can harm the system's correctness.. Formatting, cleaning, and sampling are the key procedures. Data that is missing or needs to be fixed can be removed or fixed using the cleaning procedure. When adequate data are utilised in sampling, the algorithm's execution time may be shortened. The preprocessing is carried out in Python. The dataset has 10,000 entries. With `df = df.dropna()`, where `df` is the data frame, the null values are eliminated. Using Label Encoder, the categorical attributes (Location, Block, Crime Type, and Community Area) are transformed into numeric values. The date element has been divided into new properties like month and hour that can be used as model features.

B.Feature choice

The model-building features that can be employed are chosen. Block, Location, District, Community area, X, Y, Latitude, Longitude, Hour, and Month are the criteria used to pick features.

C.Model Selection

We must choose one or a mix of modelling strategies based on the specified goal(s) (supervised or unsupervised) such as

- KNN Classification
- Logistic Regression
- Decision Trees
- Random Forest
- Support Vector Machine (SVM)
- Bayesian methods

D.Building and Training Model

- Verify the algorithm's underlying premises.
- Create/Train Model Using Available Data from Training Sample (Population)
- Check Model performance: Accuracy and Error

E.Prediction

Model predict is used to make predictions following model construction utilising the aforementioned procedure (`xtest`). Using metrics.accuracy score imported from metrics, the accuracy is determined (`ytest`, predicted).



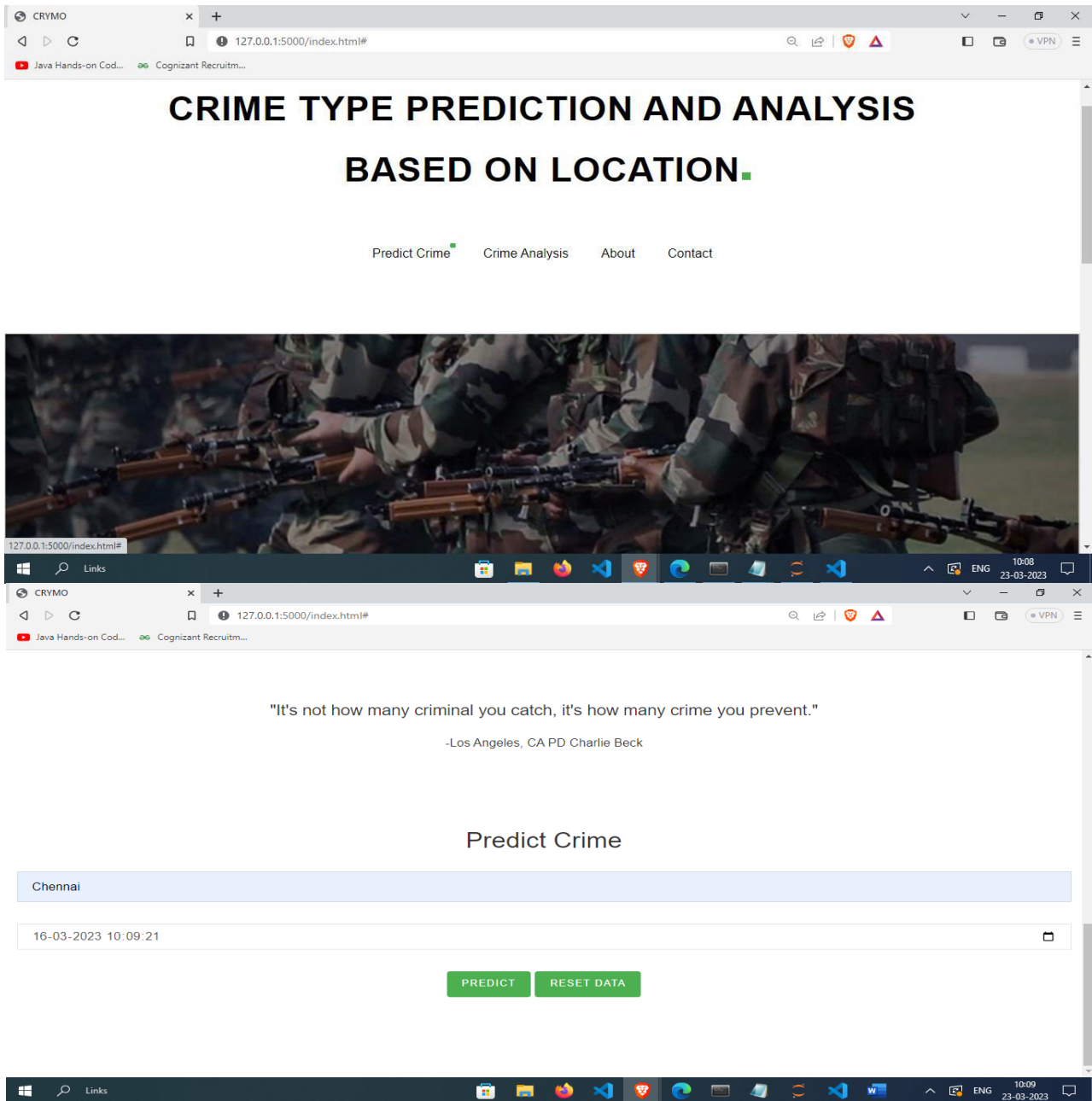
F.Validate/Test Model

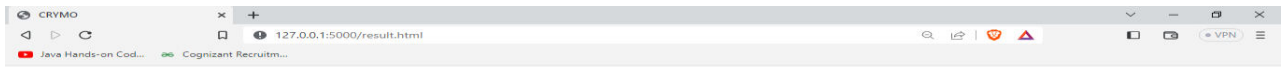
Using test sample and model performance, score and predict accuracy, etc.

G.Visualization

Use the Sklearn matplotlib library. The crime dataset is analysed using a variety of graphs.

IV. EXPERIMENTAL RESULTS

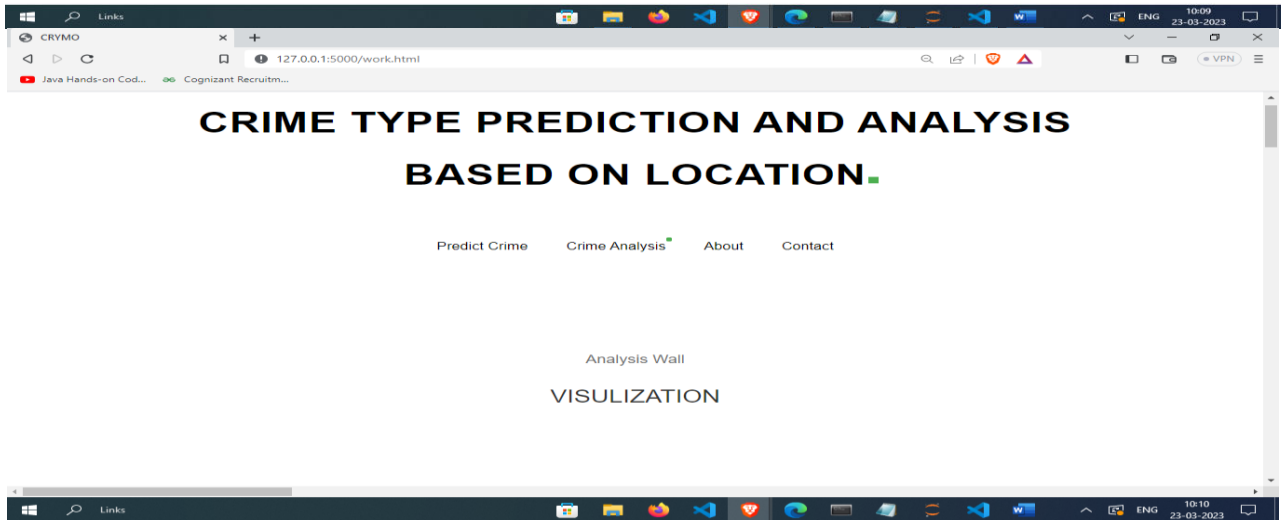


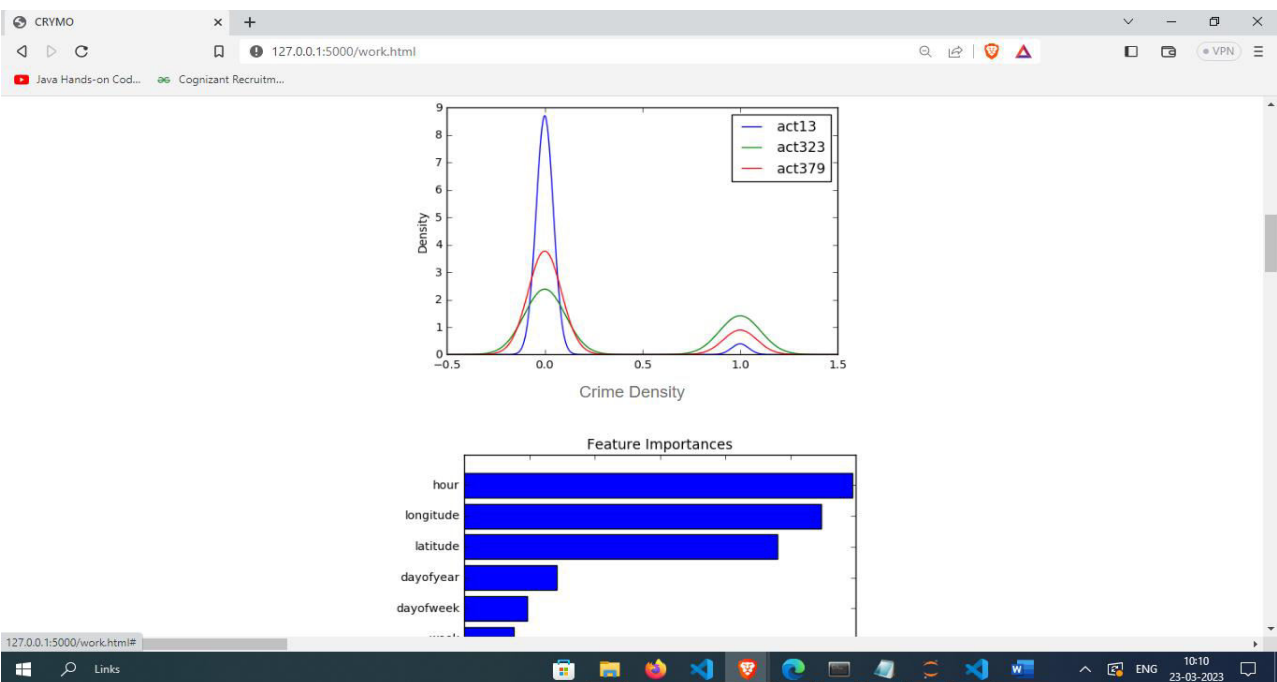
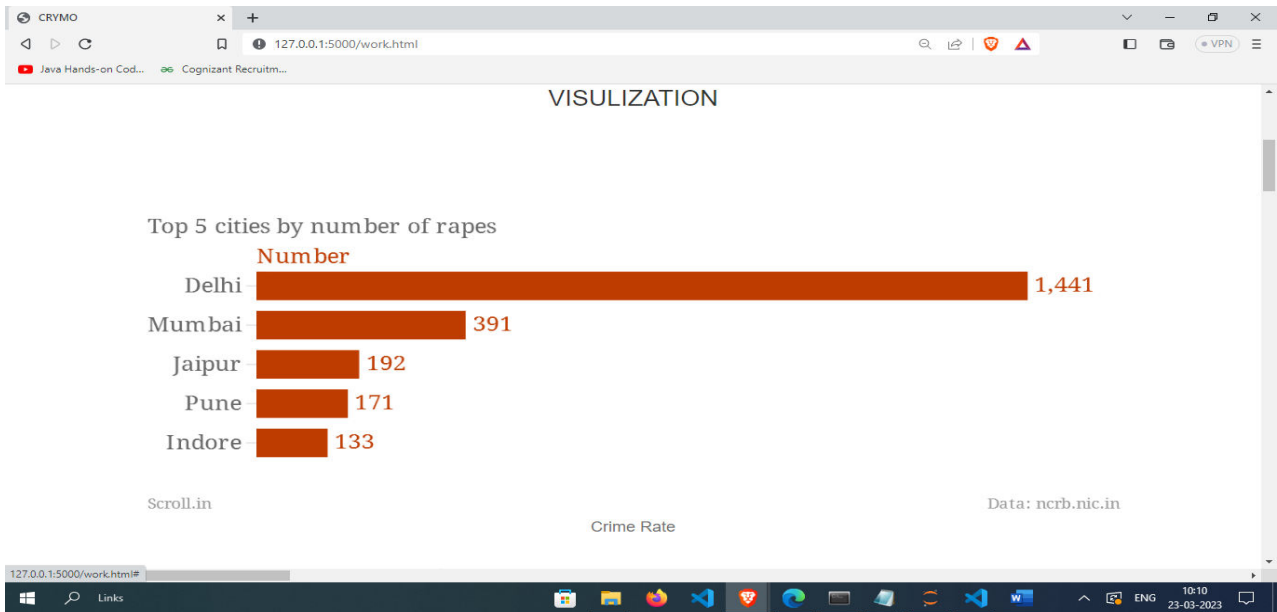


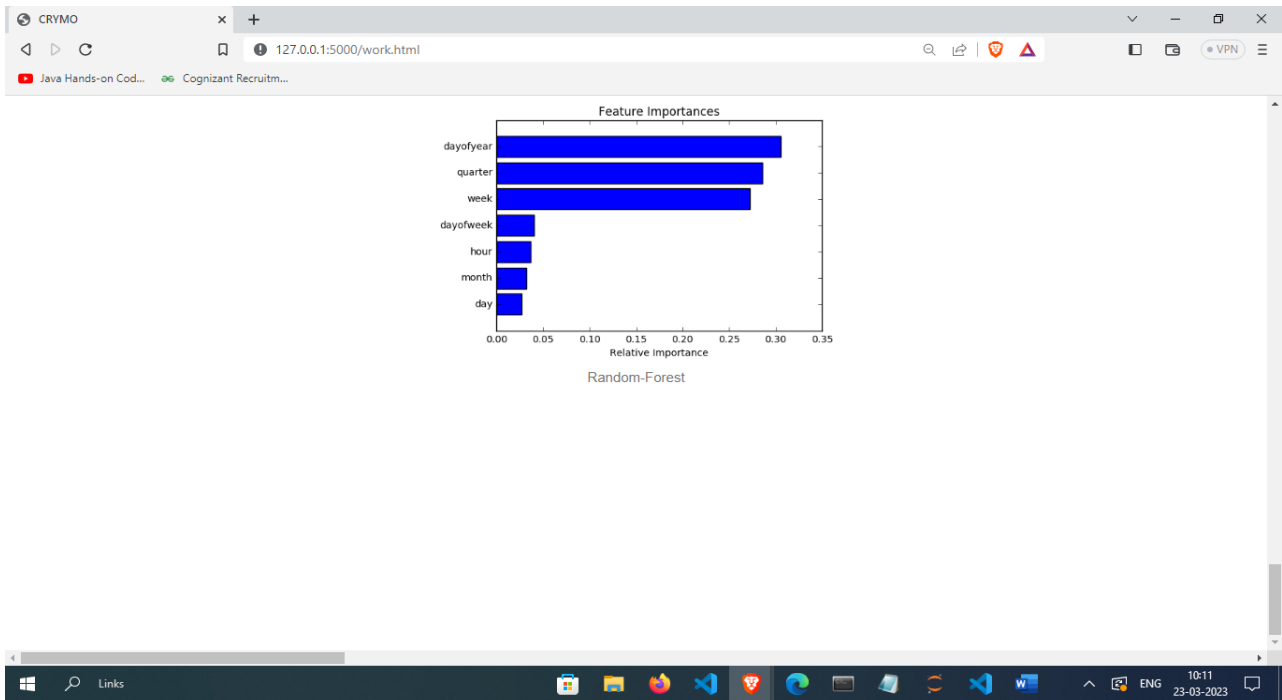
CRIME TYPE PREDICTION AND ANALYSIS BASED ON LOCATION.

[Predict Crime](#) [Crime Analysis](#) [About](#) [Contact](#)

Predicted crime : Act 279-Accident





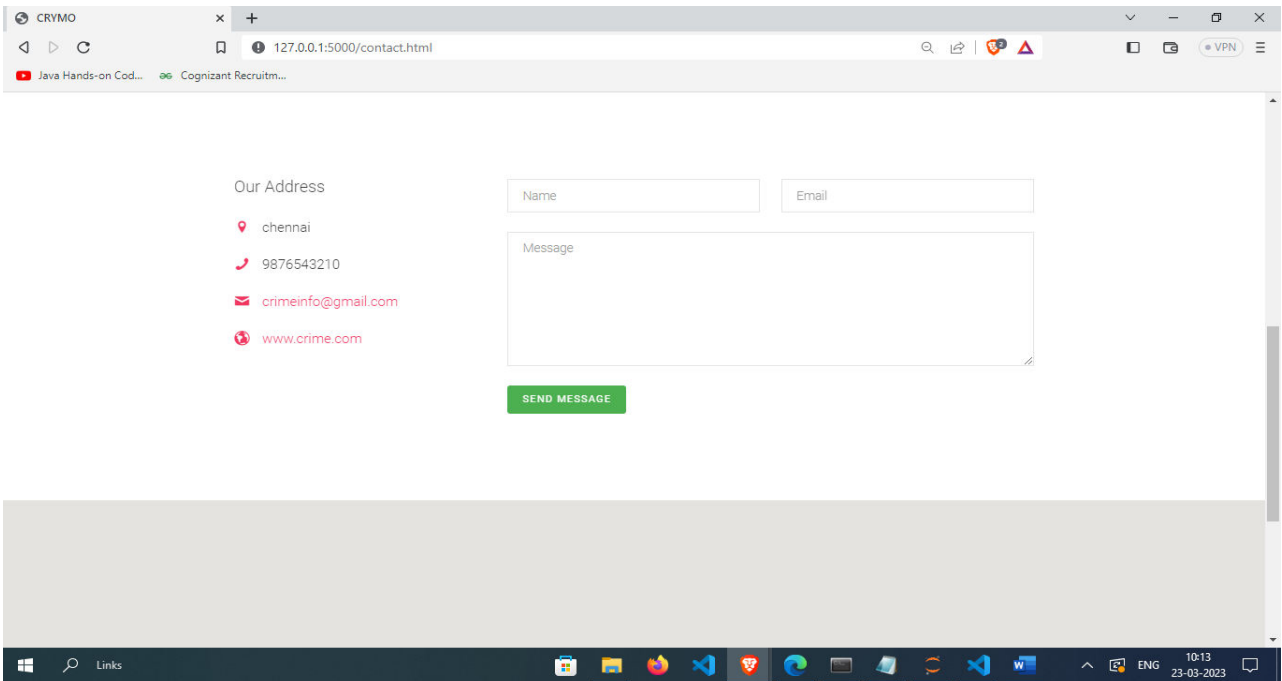
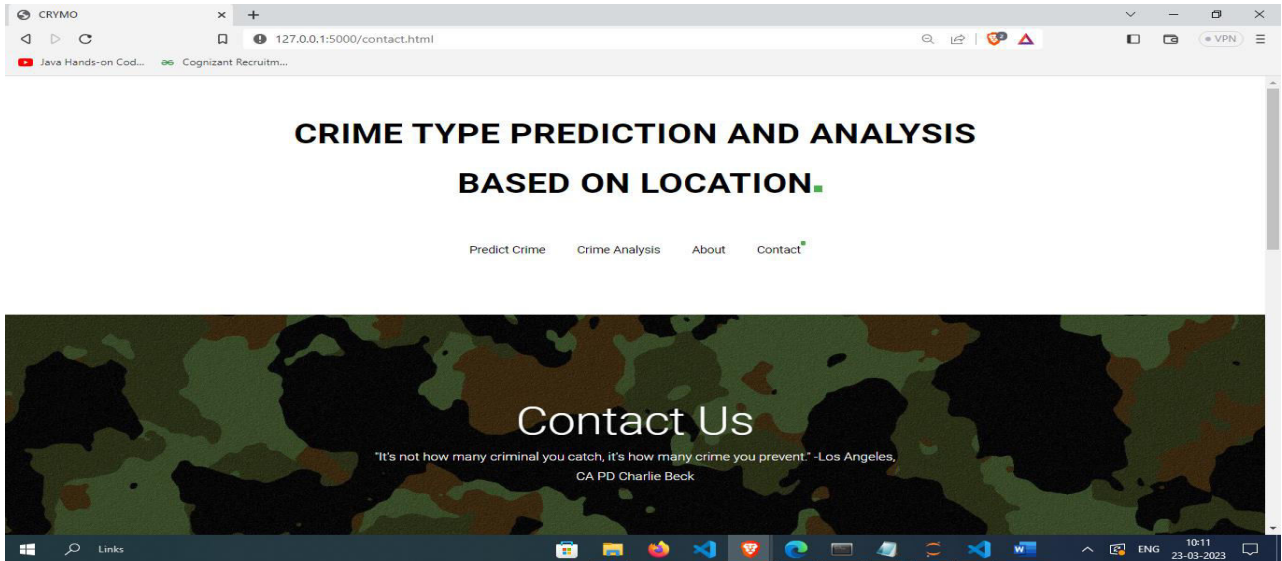


CRIME TYPE PREDICTION AND ANALYSIS BASED ON LOCATION.

[Predict Crime](#) [Crime Analysis](#) [About](#) [Contact](#)

About Us

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.



V. CONCLUSION

With the help of machine literacy technology, it has come easy to find out relation and patterns among colorful data's. The work in this design substantially revolves around prognosticating the type of crime which may be if we know the position of where it has passed. Using the conception of machine literacy we've erected a model using training data set that have experienced data drawing and data metamorphosis. The model predicts the type of crime with delicacy of 0.789 Analyzing a data set is made easier by data visualisation. The graphs include bar, pie, line and smatter graphs each having its own characteristics. We generated numerous graphs and set up intriguing statistics that helped in understanding Indian States crimes datasets that can help in landing the factors that can help in keeping society safe.



REFERENCES

- 1] Yadav, S., Timbadia, M., Yadav, A., Vishwakarma, R., & Yadav, N. (2017, April). Crime pattern detection, analysis & prediction. In Electronics, Communication and Aerospace Technology (ICECA), 2017 International conference of (Vol. 1, pp. 225230). IEEE.
- [2] Shamsuddin, N. H. M., Ali, N. A., & Alwee, R. (2017, May). An overview on crime prediction methods. In Student Project Conference (ICT-ISPC), 2017 6th ICT International (pp. 1-5). IEEE.
- [3] Sathyadevan, S., & Gangadharan, S. (2014, August). Crime analysis and prediction using data mining. In Networks Soft Computing (ICNSC), 2014 First International Conference on (pp. 406-412). IEEE.
- [4] Nath, S. V. (2006, December). Crime pattern detection using data mining. In Web intelligence and intelligent agent technology workshops, 2006. wi-iat 2006 workshops. 2006 IEEE/WIC/ACM International Conference on (pp. 41-44). IEEE.
- [5] Zhao, X., & Tang, J. (2017, November). Exploring Transfer Learning for Crime Prediction. In Data Mining Workshops (ICDMW), 2017 IEEE International Conference on (pp. 1158-1159). IEEE.
- [6] Al Boni, M., & Gerber, M. S. (2016, December). AreaSpecific Crime Prediction Models. In Machine Learning and Applications (ICMLA), 2016 15th IEEE International Conference on (pp. 671-676). IEEE.
- [7] Tayebi, M. A., Gla, U., & Brantingham, P. L. (2015, May). Learning where to inspect: location learning for crime prediction. In Intelligence and Security Informatics (ISI), 2015 IEEE International Conference on (pp. 25-30). IEEE.



SJIF: 8.423



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INTERNATIONAL CENTRE



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 **9940 572 462**  **6381 907 438**  **ijirset@gmail.com**



www.ijirset.com

Scan to save the contact details