

Crime Prediction and Analysis

Gayatri Sanjay Kolte¹, Nupur Rajesh Patel², Prof. Sonali Bodekar³

^{1,2} Students, Dept of Computer Science and Technology, Usha Mittal Institute of Technology, Mumbai, India

³ Professor, Dept of Computer Science and Technology, Usha Mittal Institute of Technology, Mumbai, India

Abstract— Crime Prediction is an approach for distinctive future crimes that are most seem to happen in an exceedingly particular location at a selected given timestamp. Daily there are a large variety of crimes committed frequently. By exploiting the knowledge set we'll understand the world of crime which can facilitate us in reducing the crime rate. The dataset is of Indore city. The dataset consists of crime data comparable to timestamp, style of crimes, latitude, and longitude. foremost we've done data preprocessing for removing the null values for obtaining high accuracy. We've done the testing and training using K-means clustering, Random forest, and Decision tree algorithms. We tend to have used these three algorithms for correct prediction and for obtaining the very best accuracy. The image of the dataset is completed in terms of graphical illustration comparable to feature selection, making the random forest tree of the complete dataset, employing a box plot for every style of crime. The only purpose of this project is to give a jest plan of how machine learning will be utilized by enforcement agencies to detect, predict and solve crimes at a way quicker rate and therefore reducing the crime rate.

Keywords— K-nearest neighbour, Random forest, Decision tree, feature selection, visualization

I. INTRODUCTION

Crimes are a vital threat to humankind maybe it's increasing and spreading at a speed in no time rate. Crime happens from little villages, cities to huge cities. Crimes are of various types-robbery, murder, rape, assault, kidnapping, etc. Since crimes are increasing so smartly there's a necessity to unravel the cases in an exceedingly abundant quicker way. Criminal activities are inflated at a faster rate and it is the responsibility of the local department to manage and scale back criminal activities. Crime prediction and criminal identification are making a nuisance to the police department as there's an amazing amount of crime data that exist. The aim of this project is to form crime predictions exploitation the options gift within the dataset. The dataset that's used has been retrieved from the official website. With the help of a machine learning algorithm, using python as core we foretold the crime which can transpire in an exceedingly explicit area. The target is to coach a model for prediction. foremost the information preprocessing is done. Then testing and coaching of the dataset is done. For creating the model we tend to use the K-nearest

neighbor, Random forest, decision tree algorithm program for getting the highest accuracy. Data visualization is completed for feature selection. Apart for this part, we've conjointly added one Independent section into our project which can offer the registered users with help such as to if any crime happens with our registered users they will get help in the type of helpline variety, nearest station address, automobile number. For this, we have used databases for storing the user details. The aim of making this project is to give a concept of how machine learning will be utilized by enforcement agencies to detect, predict and solve crimes at a way quicker rate and therefore reducing the crime rate. However even for local people in addition to grasp their area, the neighbourhood in terms of crimes and security.

II. PROPOSED SYSTEM

Predictive modeling is the way of building a model that's capable of constructing predictions. The method includes a machine learning algorithmic program that gains an understanding of sure properties from a coaching dataset so as to build those predictions. prognostic models make assumptions that support what's happening now. If arriving, new knowledge exhibit changes in what is happening now, the impact on the seemingly future outcome should be re-evaluated too. prognostic analytics depends heavily on untied access to spare volumes of accurate, clean, and relevant data. It reduces time, effort, and prices in prognostication business outcomes.

In our project, we've used the subsequent algorithmic programs k-nearest neighbor, random forest, call tree. we've used this algorithm for obtaining the very best accuracy for the prediction of crime and therefore the user can get applicable results for the same.

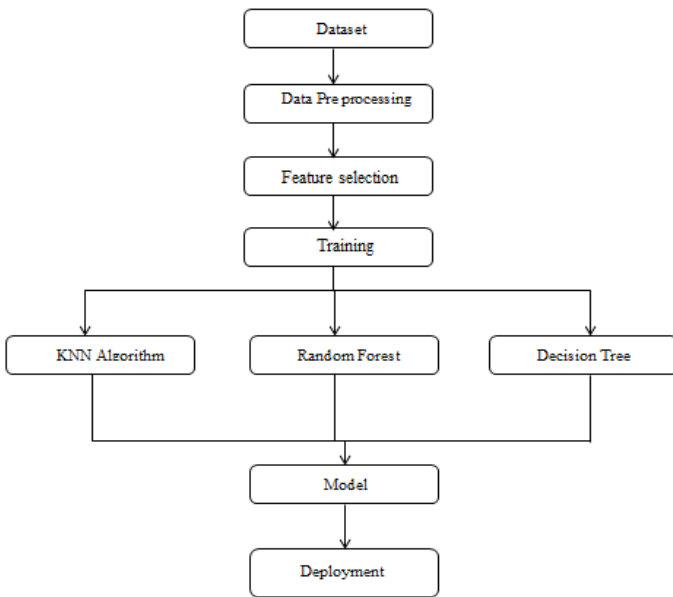


Fig.1 Proposed System

k-nearest neighbor- The k-nearest neighbors (KNN) algorithmic program could be a straightforward, supervised machine learning algorithm that may be wont to resolve each classification and regression problem. It's simple to implement and perceive however includes a major disadvantage of changing into remarkably slow because the size of that knowledge in use increases.

Random forest- Random forest is a supervised Machine Learning algorithmic program that's used extensively in Classification and Regression problems. It constructs call trees on totally different samples and takes their majority vote for classification and average just in case of regression.

Decision tree- call Tree is a supervised learning approach that may be used for each classification and Regression problem, however, in the main, it's favored for finding Classification problems. it's a tree-structured classifier, wherever internal nodes fill in for the options of a dataset, branches represent the choice rules and every leaf node represents the outcome.

III.RELATED WORK

B.Sivanagaleela, S.Rajesh, B.Sivanagaleela, S.Rajesh[1] planned crime analysis and prediction exploitation fuzzy c-means algorithm. They proposed a model within which however the prediction and analysis of the crimes are done through the improvement of the Fuzzy-c means that algorithm. exploitation they analyzed the crime patterns and prevented them. Also, they get the crime patterns to understand within which space the crime can occur frequently.

Sunil Yadav, Meet Timbadia, Ajit Yadav, Rohit Vishwakarma, Nikhilesh Yadav[2] introduced crime pattern detection, analysis, and prediction. This paper illustrates how social development would possibly lead to crime prevention. They used Association mining(Apriori),k-means, Naïve Bayes, correlation, and regression that were dead on the dataset with the help of the rail tool and R tool.

Varshishtha D N, Aishwarya P, K R Dhananjay Gupta, Vidyashree K P, Janya T S, Sahana R[3] introduced a unique approach for predicting like processing technique, deep learning technique, crime cast technique, sentimental analysis. Through this research, they discerned that every technique has its own professionals and cons. in order that they additional over that every technique provides higher results for a specific instance.

Jazeem Azeez, Dr. John Aravindhar [4] delivered a hybrid method to crime prediction the employment of deep learning. they need used deep learning, STI, sentiment analysis, LDA, RNN for the prediction of crimes specifically areas. they found that it' way possible to predict crime with the help of victimization specializing in designs and developments from various causative factors.

Rohit Mallula, Preetham Chowdary [5] introduced prognosticative policing in crime analysis using R programming. They built a model that offers the expected range of crimes that happen day by day terribly} given quantity in a specific city. They additional ended by adding a abstraction dimension thereto so it can predict the crime happening in a particular location at a particular time interval.

Wasim A.Ali, Husam Alalloush, Manasa K.N[6] introduced crime prediction and analysis victimization k-means cluster. They used the k means rule out NetBeans software, bunch the crime dataset and giving the result which will show the prime-prone areas with the speed that occurred in a very specific area. They more terminated that various different techniques or algorithms in the analysis are going to be accustomed predict outcomes and with higher efficiency.

Nikhil Dubey, Setu Kumar Chaturvedi[7] provide a comparative analysis of data mining techniques for the detection and prediction of future crime. The techniques introduced throughout this paper are Support vector machine, multi-variant time series, neural network

IV.IMPLEMENTATION

The dataset employed in this project is taken from the india.gov.in website. The dataset contains crime records of notably Indore city. It consists of nine rows and 2091 columns that includes timestamps, various crimes, latitude, and longitude.

The implementation of this project is divided into the sub-sequent steps-

A. Data Preprocessing

So, basically, there are 2091 entries into the dataset. The null values are removed using `df.dropna()`. Further, the timestamp attribute is split into new attributes resembling year, month, day of the month, hour, week, etcetera which may be used as a feature for the model

	timestamp	act379	act13	act279	act323	act363	act302	latitude	longitude
0	28-02-2018 21:00	1	0	0	0	0	0	22.737260	75.875987
1	28-02-2018 21:15	1	0	0	0	0	0	22.720992	75.876083
2	28-02-2018 10:15	0	0	1	0	0	0	22.736676	75.883168
3	28-02-2018 10:15	0	0	1	0	0	0	22.746527	75.887139
4	28-02-2018 10:30	0	0	1	0	0	0	22.769531	75.888772

In our Dataset the crime mentioned are

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

B. Building and Training Model

The training and testing of the dataset are done by first dividing the dataset into train, train, and test, test. All 3 of the algorithmic rule are foreign from sklearn package. Building the model is finished victimization the model. `Fit(train,train)`

Below is that the accuracy that we've got got when testing all the three models

ALGORITHM	ACCURACY
KNN	0.9323671497584541
Random Forest	0.9806763285024155
Decision Tree	0.9806763285024155

C. Prediction

When the model is totally engineered using the higher than process, we have done the prediction half using the model. `predict(test)`. The accuracy is calculated using `accuracy_score(xtest, ytest)`

D. Visualization

Knowledge visual image is done using matplotlib library from sklearn. Study of the crime dataset is done by plotting several graphs

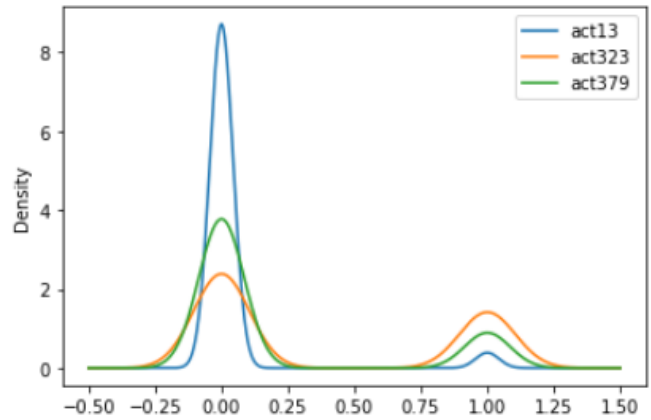


Fig.2 Crime Density

The graph shows the crime density between act 13(Gambling), act 323(Violence), act 379(Robbery). Through that we have a tendency to obtained that the best density we got is of crime Gambling

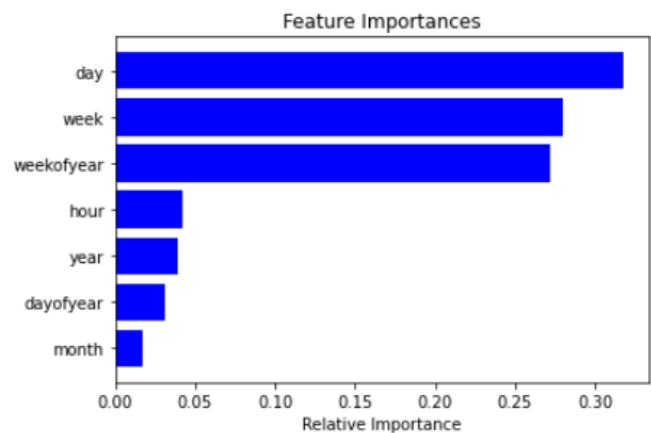


Fig.3 Feature Selection

Feature selection improves the machine learning technique and can increase the prophetic power of machine learning algorithms by selecting the foremost important variables and eliminating redundant and orthogonal features..

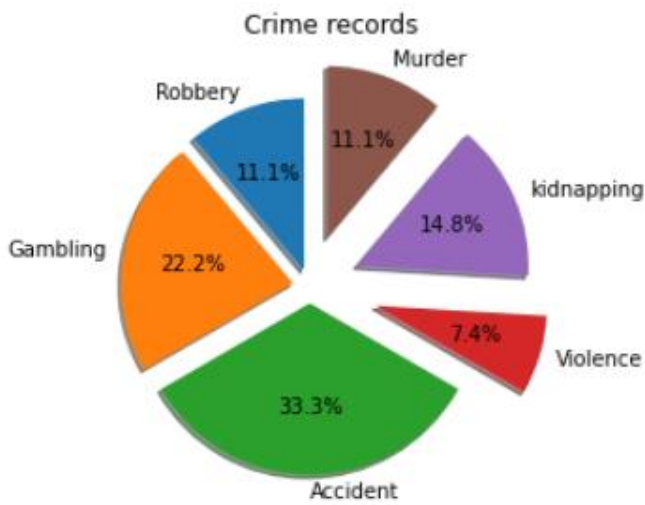


Fig.4 Percentage of crimes

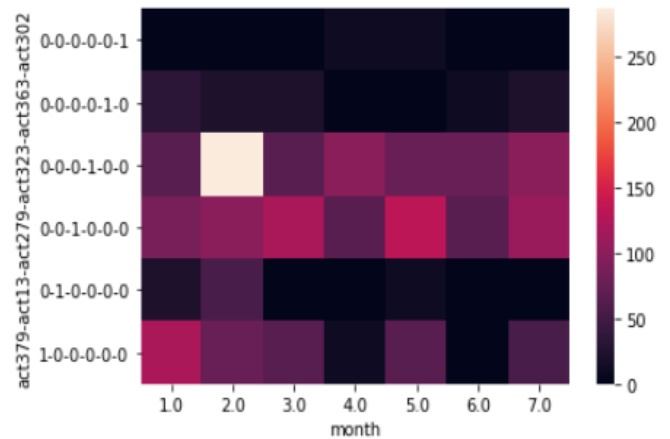


Fig.7 Heatmap of crimes by month

V. RESULT

Below shows the output images of our website which is named as ICA(Indian Crime Analysis). Hence we successfully developed a working model of the website which will help the users to view the crime prediction of a particular area which they prefer to check whether that locality is safe or not

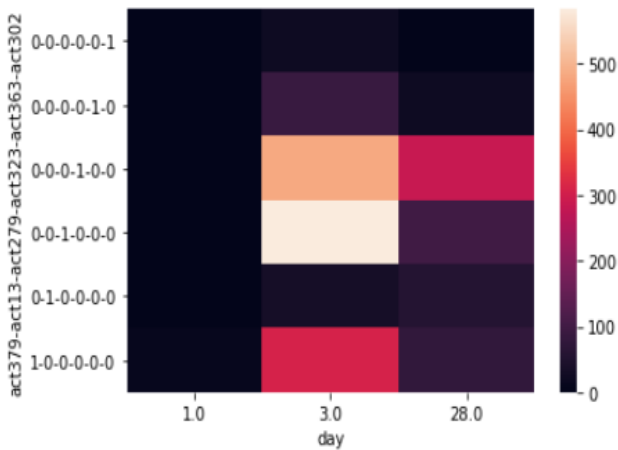


Fig.5 Heatmap of crimes by day

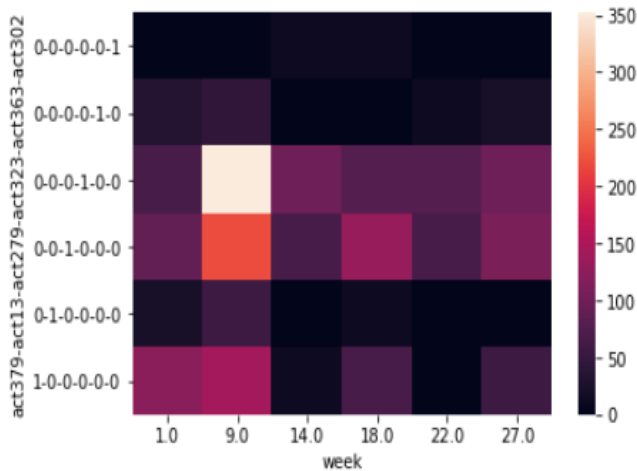


Fig.6 Heatmap of crimes by week



Fig.8 Home Page

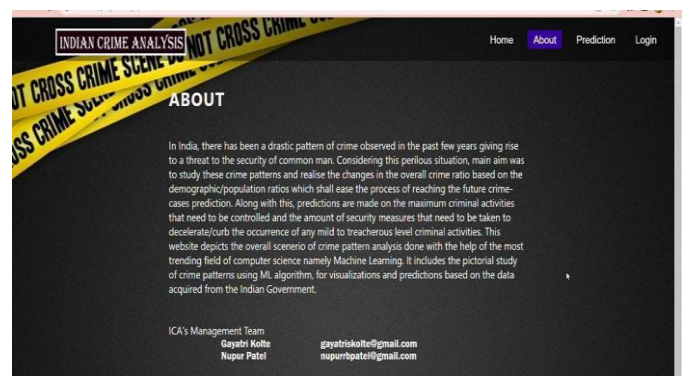


Fig.9 About Page

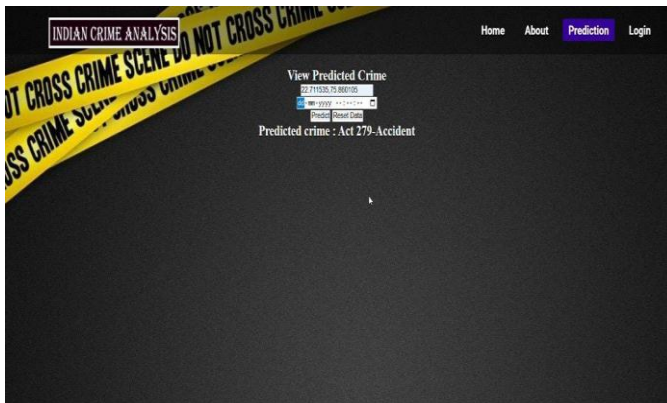


Fig.10 Result

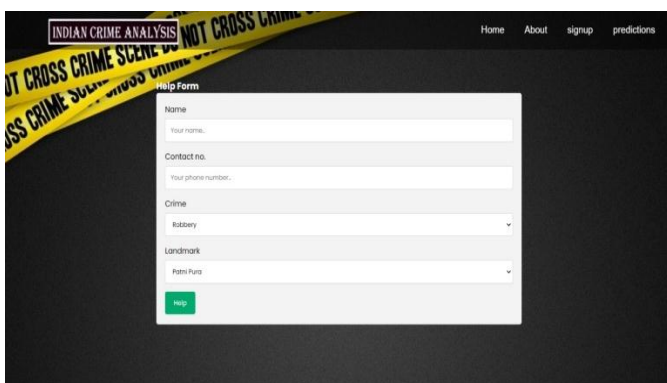


Fig.11 Help Form

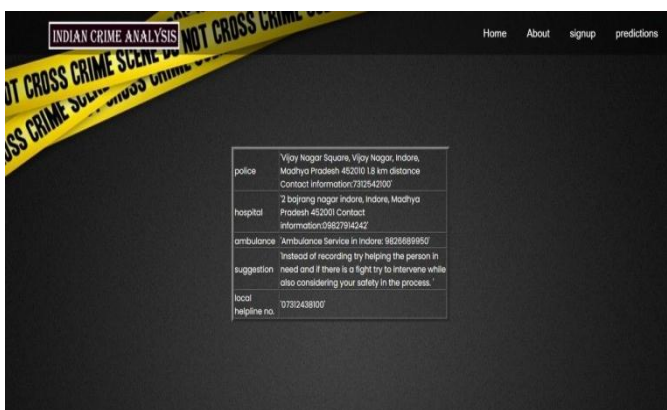


Fig.12 Help output

VI. CONCLUSION

With the help of machine learning technology, it's become simple to search out out relations and patterns among varied knowledge. The add this project chiefly concentrates on predicting the kind of crime that's presumably to happen at a specific location at the given time interval. victimization the conception of machine learning we've got engineered a model employing a coaching dataset that has gone underneath data preprocessing and data cleaning. The model

provides high accuracy. knowledge visual image helps in the analysis of the dataset. we've got included graphs, pie charts, and heatmaps. victimization this analysis we have a tendency to found fascinating characteristics that facilitated in understanding the dataset alright and might help in capturing the factors which will help to keep society safe.

VII. REFERENCES

- [1]. B.Sivanagaleela,S.Rajesh,Crime analysis and prediction using fuzzy c-means algorithm.In 3rd International conference on trends in Electronics and Informatics(ICOEI),April 2019,IEEE
- [2]. Sunil Yadav,Meet Timbadia ,Ajit yadav, Rohit Viswakarma,Nikhilesh Yadav, Crime pattern detection ,analysis and prediction. In international conference of Electronics, Communication and Aerospace Technology(ICECA),April2020,IEEE
- [3]. Jazeem Azeez, Dr.John Aravindhar Hybrid approach to crime prediction using deep learning .In International Conference on advances in Computing, Communications and Informations (ICACCI),August 2019,IEEE
- [4].Rohith Mallula,Preetham Chowdary.Predictive Policing in crime analysis using R.In 5th International Research Journal of Engineering and Technology(IRJET).July 2018,IEEE
- [5].Wasim A.ali,Husam Alalloush,Manasa K.N.Crime Analysis and Prediction using K-means clustering technique.In 5th International Journal of Research and Development(IJRD).July 2020.IEEE
- [6].Nikhil Dubey,Setu Kumar Chaturvedi.A survey paper on cime prediction technique using data mining.In 4th International Journal of Engineering Research and Applications(IJERA).March 2021.IEEE
- [7].Alkesh Bharati,Sarvanaguru RA.K.Crime prediction and analysis using machine learning. In 5th International Research journal of engineering and technology(IRJET).SEP 2018.IEEE