Available Online at www.ijcsmc.com/journal

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X
IMPACT FACTOR: 7.056

IJCSMC, Vol. 11, Issue. 3, March 2022, pg.95 – 101

Crime Prediction and Analysis Using Machine Learning

V Balu
Associate Professor, Department of Computer Science Engineering SCSVMV University, Kanchipuram, TamilNadu, India
Email- balukanchi@kanchiuniv.ac.in

T Navya Sri
Department of Computer Science Engineering SCSVMV University, Kanchipuram, TamilNadu, India
Email- 11189A159@kanchiuniv.ac.in

Medarametla Anand Bupathi
Department of Computer Science Engineering SCSVMV University, Kanchipuram, TamilNadu, India
Email- 11189A143@kanchiuniv.ac.in

DOI: https://doi.org/10.47760/ijcsmc.2022.v11i03.011

Abstract- It is critical to recognise crime patterns in order to be better prepared to respond to criminal behavior. We study crime data of states in india that was scrapped publicly available websites kaggle for our project. The goal is to estimate which type of crime is most likely to occur at a given time and location. The use of AI and machine learning to identify crime using sound or video is now in use, has been demonstrated to function, and is likely to grow. The use of AI/ML to forecast crimes or a person's chance of committing a crime has potential, but it is still a work in progress. The most difficult task will most likely be "proving" to legislators that it works. It's tough to establish the negative when a system is meant to prevent something from happening. A positive feedback loop would certainly benefit companies who are directly involved in providing governments with AI capabilities to monitor areas or predict crime. Improvements in crime prevention technology will almost certainly lead to an increase in overall spending on this technology. We also try to make our categorization work more relevant by grouping many classes together into larger groups. Finally, we present and discuss our findings using several classifiers, as well as future research directions.

Keywords– Machine Learning, Crime Prediction, Kaggle

I. INTRODUCTION

The rate of crime is rising on a daily basis as current technologies and high-tech ways assist criminals in carrying out their unlawful activities. According to the Crime Record Bureau, crimes such as burglary and arson have increased while crimes such as murder, sex abuse, and gang rap have climbed. Data about crime will be gathered from a variety of blogs, news outlet, and websites. The massive data is used to create a crime report database as a record. The police department can benefit from the information gathered through data mining tools. The police have discovered that identifying criminal "hotspots", or regions with a high concentration of crime, is an effective method. Data mining techniques can be used to extract useful information from crime report datasets. Crime analysis is the first phase in the investigation of a crime. Exploring, relating, and analyzing crime is what crime analysis is all about. Elements such as criminal intelligence, location security, and so on the effort followed the steps of data analysis, with data gathering data classification, pattern discovery, prediction, and visualization being the most essential parts. The proposed system employs a variety of visualization approaches to depict crime trends and several methods for predicting crime using a machine learning algorithm. Time(hour,day,month,and year) and location are the inputs to our algorithms(latitude and longitude) and class of crime: Act 379 - Robbery ,Act 13 – Gambling , Act 279 - Accident, Act 323 - Violence, Act 302 - Murder 3, Act 363 - Kidnapping .The output is the class of crime that is likely to have occurred. We try out multiple classification algorithms, such as KNN (K-Nearest Neighbors), Decision Trees, and Random Forests. We also perform multiple classification tasks – we first try to predict which of 6 classes of crimes are likely to have occurred, and later try to differentiate between violent and non-violent crimes.

II. PROPOSED ALGORITHM

2 KNN algorithm –

The K-nearest neighbors(KNN) algorithm is a simple, supervised machine learning algorithm that can be used to solve both classification and regression problems. It's easy to implement and understand, but has major drawback of becoming significantly slower as the size of that data in use grows.

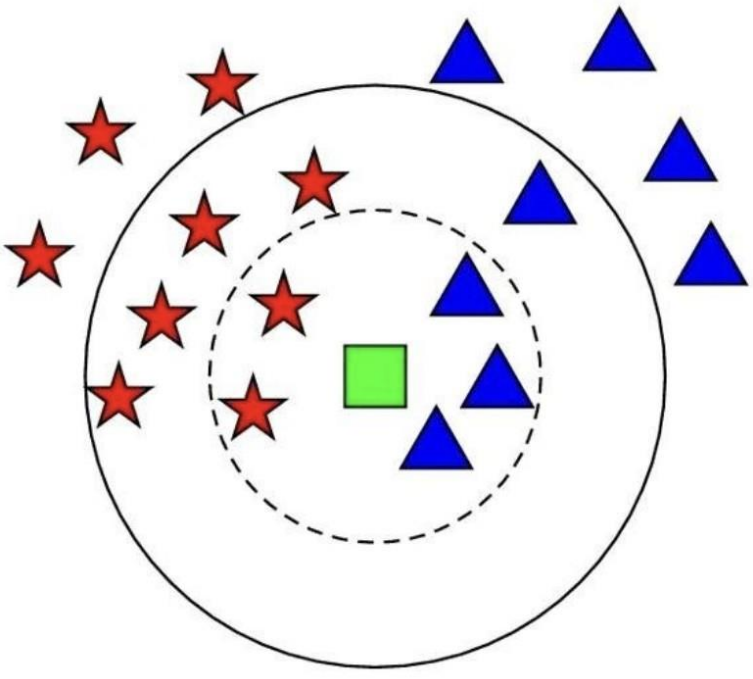

Figure 1. *Knn Principle Diagram*

III. IMPLEMENTATION

The project's dataset was obtained from kaggle.com. The dataset was downloaded from kaggle. This project's implementation is broken down into the following steps –

3.1. Data Collection

The Crime data set used in this project is in the CSV(Comma separated values) format.

3.2. Data Preprocessing

The dataset has 10000 entries. df=dropna( ),where df is the data frame,is used to remove null values. Label Encoder is used to translate category attributes(Location,Block,crime types,Community Area)into numeric values. The date attribute has been separated into additional characteristics such as month and hour, which can be utilized as model features.

3.3 Feature Selection

The features that will be utilized to build the model are chosen. Block,Location,District,Community area,X coordinate,Y coordinate,Latitude,Longitude,Hour and month are the attributes used to pick features.

3.4. Building and Training Model

The location and month attributes are used for training after feature selection. The dataset is split into xtrain ,ytrain and x test and y test pairs. Sklearn is used to import the algorithm model. Models are used to create models a good fit(xtrain,ytrain).

3.5. Prediction

Model predict is used to make predictions after the model has been developed using the above process(xtest).The accuracy is determined by importing accuracy score from metrics-metrics.accuracy score(y test,predicted).

Table1-Dataset before Preprocessing

	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

Table 2-Dataset after Preprocessing

	day	dayofweek	dayofyear	hour	month	quarter	week	weekday	weekofyear	year	act379	act13	act279	act323	act363	act302	latitude	longitude
0	28.0	2.0	59.0	21.0	2.0	1.0	9.0	2.0	9.0	2018.0	1	0	0	0	0	0	22.737260	75.875987
1	28.0	2.0	59.0	21.0	2.0	1.0	9.0	2.0	9.0	2018.0	1	0	0	0	0	0	22.720992	75.876083
2	28.0	2.0	59.0	10.0	2.0	1.0	9.0	2.0	9.0	2018.0	0	0	1	0	0	0	22.736676	75.883168
3	28.0	2.0	59.0	10.0	2.0	1.0	9.0	2.0	9.0	2018.0	0	0	1	0	0	0	22.746527	75.887139
4	28.0	2.0	59.0	10.0	2.0	1.0	9.0	2.0	9.0	2018.0	0	0	1	0	0	0	22.769531	75.888772

Data preprocessing entails removing rows that aren't needed for analyzing and converting any value that includes a value to Nan. Converting a textual variable to numerical variable so that it can be processed further.

IV. CRIME VISUALIZATION

This Section deals with the analysis performed on the datasets and plotting them using various graphs such as bar plot,box plot,pie chart….etc.

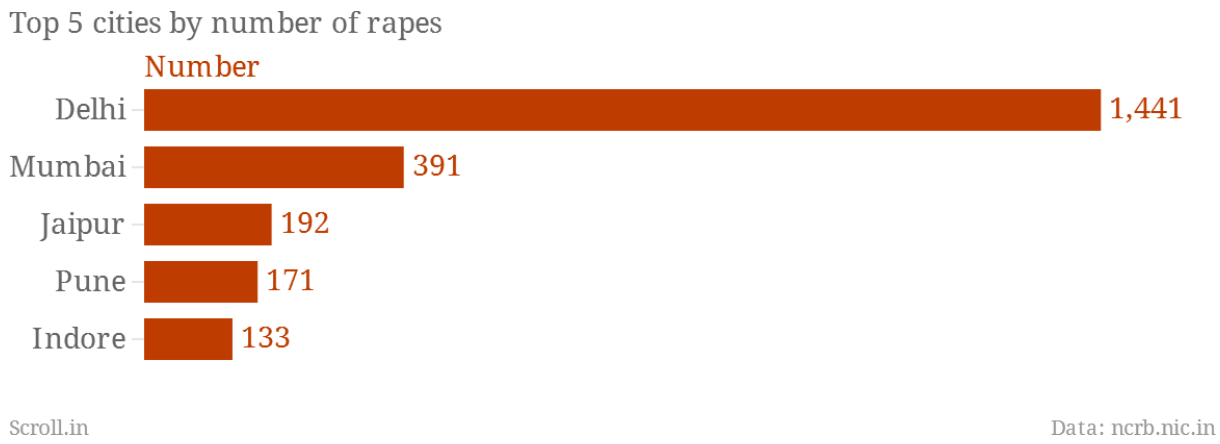

Figure 1-Most Number of rapes occurred vs Cities

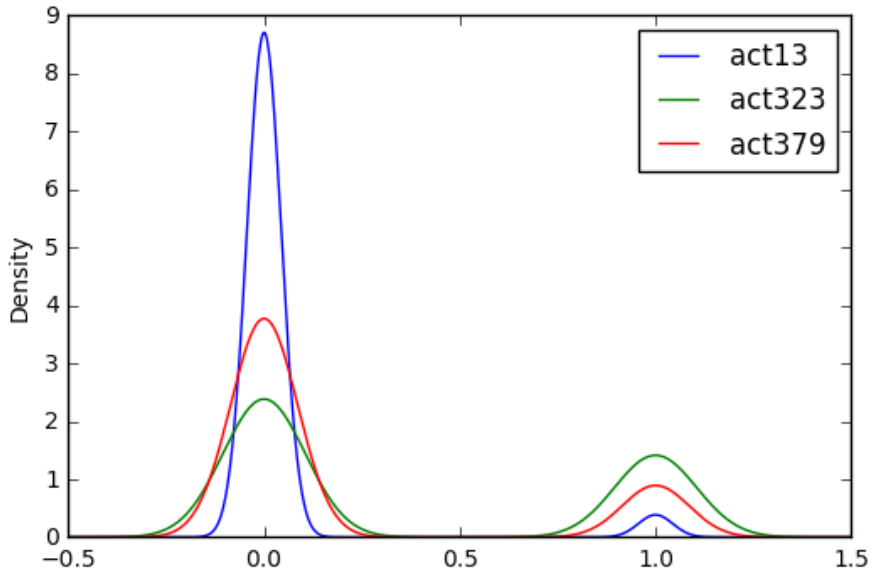

Figure 2- Types of crimes vs Density

The above graph shows the density of the crimes and it is evident that most of the crimes occurring are of the kind Gambling, violence and Robbery.

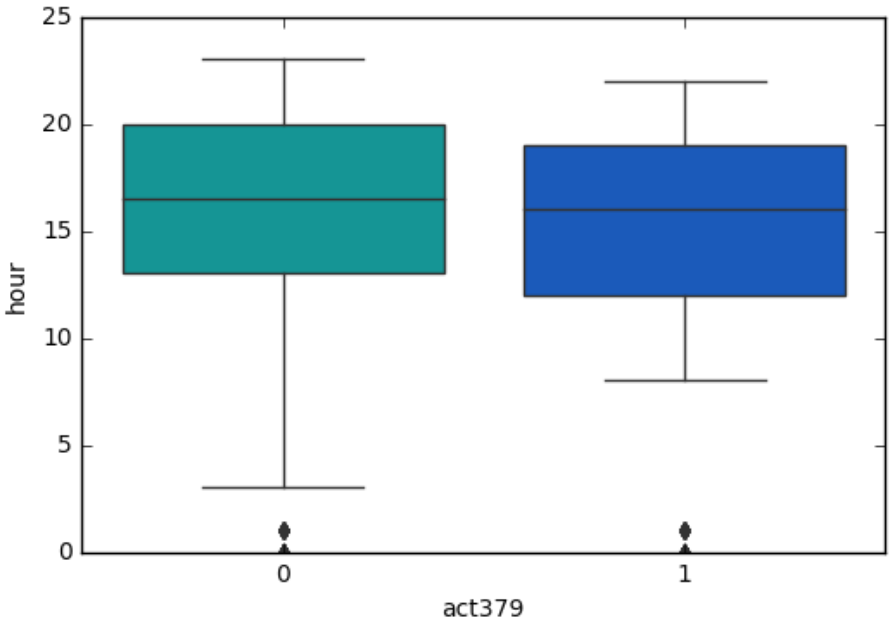

Figure 3-Act379(Robbery) vs hour

The above is the box plot showing the hours during which there is a high probability of occurrence of robbery.

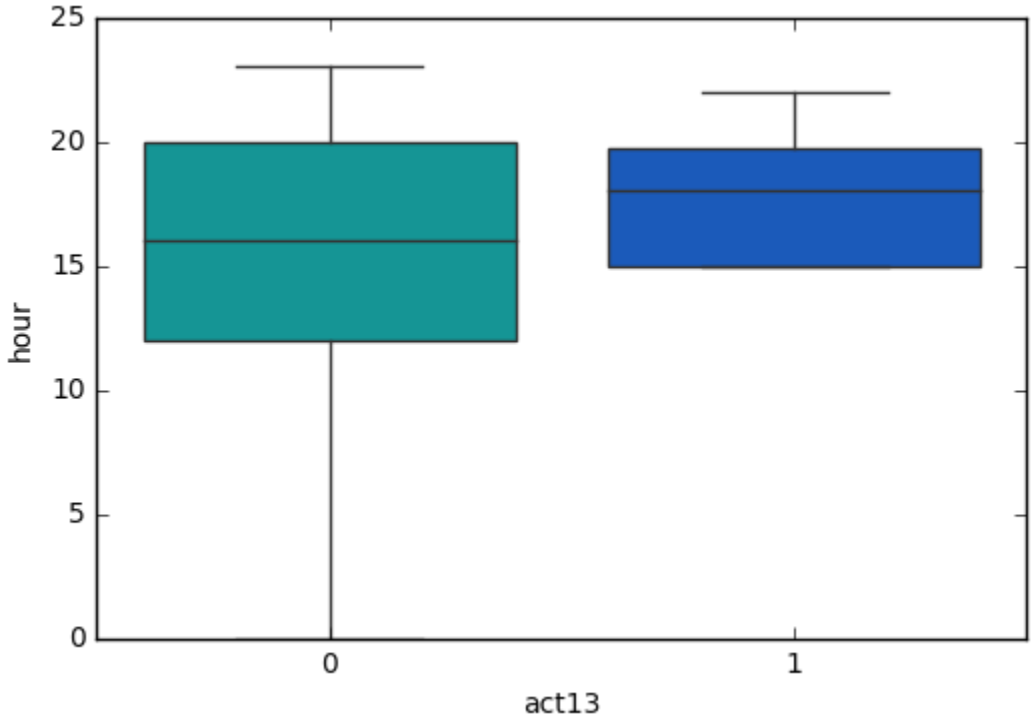

Figure 4-Act13(Gambling) vs hour

The above box plot is showing the hours during which there is a higher occurrence of Gambling.

V. CONCLUSION

The initial task of identifying six different crime categories was a difficult multiclass classification problem, and our initial data set lacked enough predictability to achieve high accuracy. In order to identify organization in the data, we discovered that collapsing the criminal categories into fewer, larger groups was a more meaningful method. On prediction, we had a high level of accuracy and precision. However, using the same classifiers, the Violent/Non Violent crime categorization did not produce impressive results-this was much more difficult classification task. As a result, collapsing crime categories is a difficult classification task. As a result, collapsing crime categories is a difficult operation that necessitates careful selection and study. Time-series modeling of the data to identify temporal relationships in it, which may subsequently be used to forecast sures in different types of crime, is one way to extend this work. It would be also be interesting to investigate links between surges in other types of crimes–for example, it's possible that two or more classes of crimes surge and sink at the same time, which would be a fascinating relationship to discover. Implementation a more accurate multi-class classifier and investigating better ways to show and results are two other areas to work on.

REFERENCES

[1]. Pratibha kumari,Akanksha Gahalot,Suraina,Lokesh Chouhan(2020 ,february).An overview on 'crime prediction and analysis' .In Conference: 2020 2nd International Conference on Data, Engineering and Applications (IDEA)

[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]. 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. 225- 230). IEEE.

[4]. Sivaranjani, S., Sivakumari, S., & Aasha, M. (2016, October). Crime prediction and forecasting in Tamilnadu using clustering approaches. In Emerging Technological Trends (ICETT),International Conference on (pp. 1-6). IEEE.

[5]. 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.

[6]. 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

