International Journal of Science, Technology and Management (IJSTM) ISSN (online): 2321-774X

Volume 9, Issue 1, 2022

Crime Prediction using Machine Learning

Sukhpender Panghal

sukhpenderpanghal123@gmail.com

M. Tech Scholar, Department of Computer Science & Engineering, BRCM CET, Bahal, Haryana

Mrs. Neha

neha@brcm.edu.in

Assistant Professor, Department of Computer Science & Engineering, BRCM CET, Bahal, Haryana

ABSTRACT

The two AI tools that are most frequently employed are data mining and machine learning. These tools are frequently used for data analysis and a wide range of problem-solving. In my research paper, "Crime Prediction Using Machine Learning," I compared various data sets theoretically and practically to determine the most suitable data mining and machine learning methods to analyze the information gathered from the verified government resources specializing in crime prevention. Additionally, I have elaborated on several study analyses in this research report. I have concluded that historical information about previous offences is quite helpful to predict the offences after analyzing the various sorts of data sets. In order to predict the hotspot where crime typically occurs, the research on this particular topic will enable early crime hotspot prediction. I'm hoping that this research will make it easier to anticipate crimes before they happen.

Keywords: Data mining, machine learning, artificial intelligence, and crime prediction.

INTRODUCTION

Crime has had a significant impact on people's lives across the nation. A crime or offence (sometimes known as a criminal offence) is an action that causes harm to a group of people, a society, or the government in general. The number of crimes is rising daily, as we can see in the newspaper. The Indian government's agencies are also thinking about various ways to minimise crime and are putting many different programmes into action. In order to prevent the general public, technical tools and artificial intelligence would be easy to use.

To identify and stop crimes from happening, many existing studies use machine learning and artificial intelligence to find patterns in crime. The majority of the existing efforts have a few drawbacks, such as a lack of ability to identify connections between various crime episodes and vulnerability analyses. In this essay, I outline several steps of crime detection that combine data collection, processing, and the application of various algorithms to data sets.

Indian major crimes include:

- Internet crimes
- Terrorism
- Crime directed at women
- Robbery and homicide
- Contraband trafficking (Drugs, Alcohol-related, Arms)
- Poaching
- Financial Crime
- Abduction

As shown in the image below, crime rates have increased recently, especially in major cities. It is concerning that major cities like Tamilnadu, Kerala, and New Delhi have the highest crime rates. I have combed through several data sources to gather information in order to analyse crime events and crime analyses in India. Implement Time Series Analysis for Hotspot Area Forecasting.

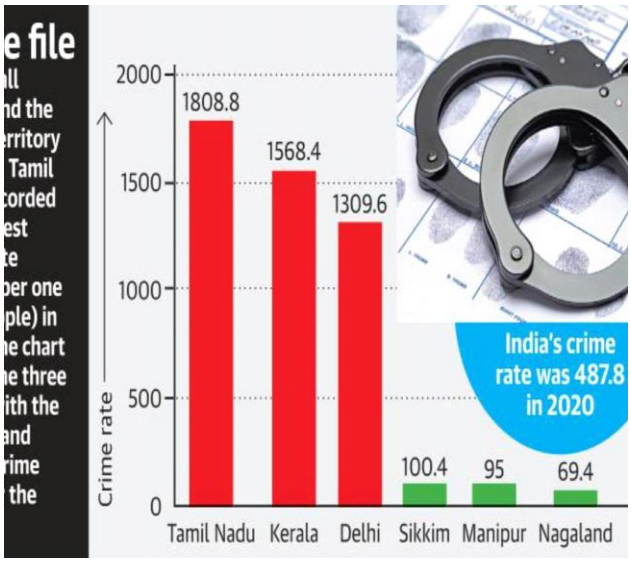

Figure 1: A countrywide crime status check

A drawback of the current system

Although the current approach of early criminal offence estimation helps us in some ways, the analysis procedure does not achieve the needed precision. An abnormality in the system could result from the incorrect forecast. Some cutting-edge approaches should be utilised to operate the system efficiently in order to make it practical. On the other hand, advanced techniques would address a range of crimes and provide the desired outcome depending on the nature of the offences.

RESEARCH REVIEW

Brown and Donald E., the authors of the research paper titled "Regional Crime Analysis Program (ReCAP): A framework for mining data to identify criminals," first introduced the concept of crime pattern analysis in 1998. Based on prior research, a limited number of data mining approaches and machine learning algorithms are used in crime prediction. Machine learning techniques like Random Forest, Neural Networks, K-Nearest Neighbors, and Logistic Regression will be utilised in a lot of research to anticipate events, and their performance will be compared depending on how the data was transformed and processed. [5,4]. However, a new study suggests that utilizing the SEMMA (Sample, Explore, Modify, Model, and Assess) model, following data processing, standardization, and cleaning, clustering is carried out using K-means, and subsequently the neural network is produced.

Research Approaches

Various approaches, based on recent research, include:

- Crime Spatial-Temporal Analysis,
- Making use of Dynamic Features.
- Using old information

Crime Spatial-Temporal Analysis (STAC)

The next stage is to give the data a systematic view after collecting it from various sources. this could be accomplished by categorising the data by factors such as location, crime type, past crimes, etc. Due of this, hotspot analysis using homogeneous data is now widely used. Spatial analysis uses machine learning algorithms to assess crime data and forecast new crime hotspots that should receive more patrol attention or security measures. Predicting crime hotspots for the upcoming month, which are not in the same temporal environment, was the objective. The population of the city, the value of every house, and geographic

data about the city's roadways, interstate highways, malls, schools, and other congested areas are all measured in a census block group.

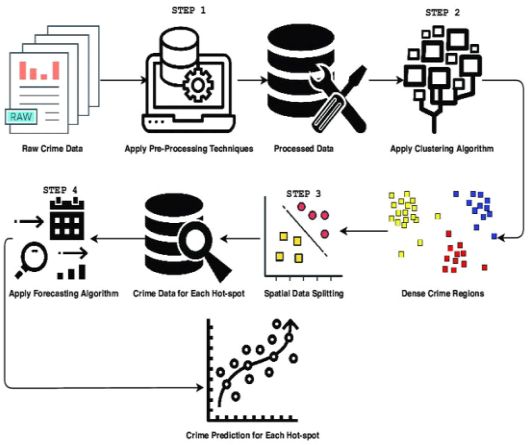

Figure 2: Spatial-Temporal Mechanism

Dynamic Features usage

We have dynamic approaches to reduce crime thanks to dynamic features. This function primarily focuses on tracking human movement inside a specific area. Social networking applications will be a great resource for learning about someone's current location, daily routine, likes and dislikes, occupation, and many other things. Dynamic features gather a wide variety of data that can assist us in comprehending regional data (such as human mobility) throughout a whole metropolis. Today's Location-Based Social Networks (LBSN) offer these chances to catch a variety of issues. Therefore, the dynamic approach would aid in our ability to anticipate offences.

Using old information

The best strategy to prevent crimes from happening further is to collect data from many sources to understand the types of offences that typically occur. Using information from past records to anticipate and prevent future occurrences is one strategy that can assist fight crime. The research primarily focuses on examining the trends of crime across various crime categories/types. Using crime statistics for various places, several studies have been conducted to forecast crime types, crime rates, and crime hotspots.

Data Collection & Pre-Processing | Methodology

Collecting the data from many authenticated sources is the first and most important phase in the data analysis process. The State Crime Record Bureau or law enforcement agencies, which serve as the repository for information on criminal offences, are where the data is gathered. For the best possible analysis, the data must be precise. A poor analysis might be the result of incorrect or inaccurate data.

	Attribute data						Spatial data
	Population-based data					*Crime type data*	*Location-based data*
Id	Sex	Offender age	Job	Victim age	Time	Type	Address
1	M	25	Nurse	18	17.02.12	Murder	.
2	F	47	Engineer	34	21.10.11	Injure	.
3	F	33	Driver	25	15.11.11	Rape	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
n	.	.	.	.	.	.	.

Figure 3: Data with preprocessing and characteristics

The data mentioned above is not organized. The provided data includes many characteristics, including sex, offender age, job, victim, time, and crime type. As a result, we preprocessed the data and created the report's tables in a systematic way. Data were combined from various sources while retaining consistency and validity. Using statistics on crime from several Indian cities, broken down into categories like theft offences, crimes against women, and crimes against children.

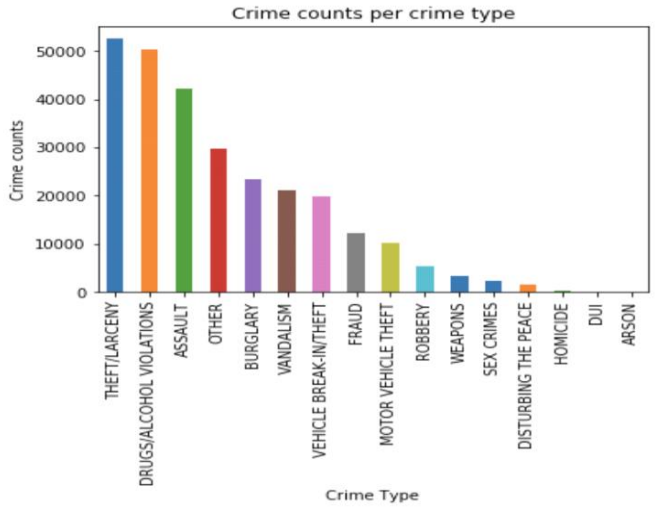

Figure 4: The structured data

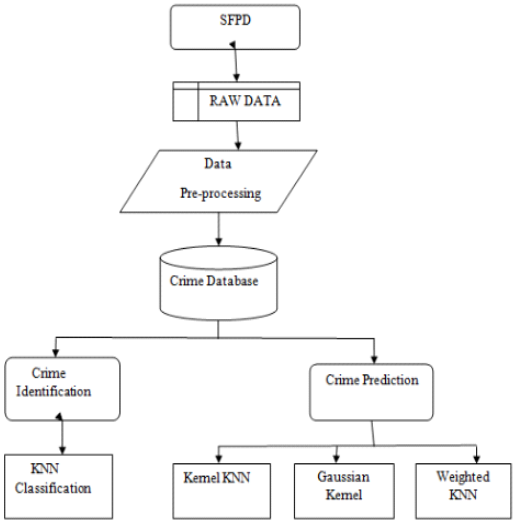

Figure 5: KNN classification

The Pattern is Recognized Using Machine Learning Algorithms

KNN (K-nearest neighbors)

Another technique used in categorization and regression is K-Nearest Neighbors. It is the most straightforward yet efficient algorithm. A new data point is classified into the target class using the K Nearest Neighbor (KNN) algorithm, which chooses the best data point from the available data sets based on the attributes of the neighboring data points. The non-parametric method that employs a visual representation of the data is known as KNN. This method is quite trustworthy and capable of producing exact results.

Support Vector Machine

This classification model is used to predict crime. For the purpose of predicting dynamic crime, SVM mechanism is applied. This strategy can be quite helpful in predicting where crime hotspots will be. By first segregating the data and then evaluating it, this approach might produce the desired outcome. The mechanical function utilized for the separation is known as the Kernel function. This classifies the data by adding curb in between the two different sets of data.

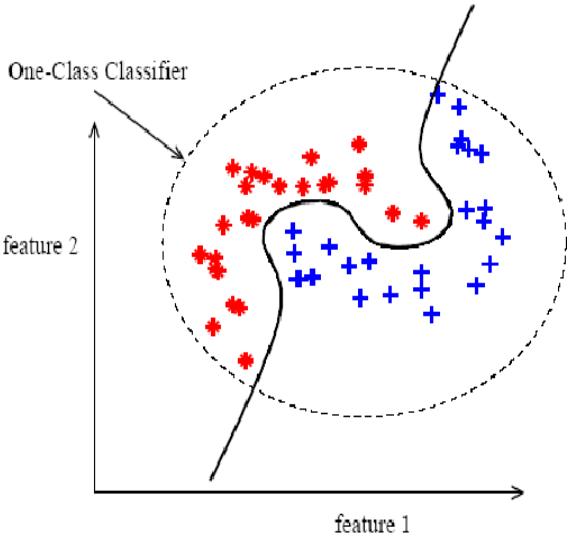

Figure 6: SVM-based forecast of crime hot spots

Classifier using a decision tree (DT)

It is a predictive modeling technique used in data mining, statistics, and machine learning. Using a decision tree as a predictive model, decision tree learning converts observations about an object into inferences about the goal value of the item. The observations regarding the specific processed data are recorded in the tree's branches, and the tree's leaves will reveal the value of the observations. A limited range of values that are regarded as variables that lead to the intended outcome could be included in the predictive model.

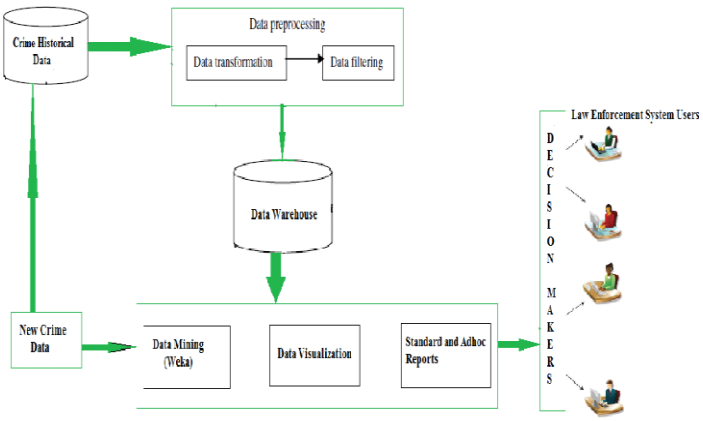

Figure 7: Decision Tree Classifier

CONCLUSION

Machine learning algorithms are essential in examining criminal charges in order to obtain a more accurate outcome. This paper's research focuses on identifying crimes that have already happened. Machine learning techniques can be used to identify patterns by looking at various data sets and normalizing the data.

REFERENCES

[1] Y. Lee, T. Song, H. Kim, D. K. Hant, and H. Ko, "Hostile intent and behaviour detection in elevators," in *4th International Conference on Imaging for Crime Detection and Prevention 2011 (ICDP 2011)*, pp. 1–6, London, 2011

[2] Oren Anava, Kr Y. Levy "k_-Nearest Neighbors: From Global to Local",arXiv:1701.07266v1 [stat.ML] (2017).

[3] Ping Li, Jianping Gou, Hebiao Yang, "The Distance-Weighted K-nearest Centroid Neighbor Classification", Volume 8, Number 3, May 2017 © 2017 ISSN 2073-4212.

[4] J. Borges, D. Ziehr, M. Beigl et al., "Feature engineering for crime hotspot detection," in *2017 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computed, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation(SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*, pp. 1–8, San Francisco, CA, USA, 2017.

[5] S. Yadav, M. Timbadia, A. Yadav, R. Vishwakarma, and N. Yadav, "Crime pattern detection, analysis & prediction," in *2017 International conference of Electronics, Communication and Aerospace Technology (ICECA)*.

[6] Ranjan, Shruti, et al. "Framework for image forgery detection and classification using machine learning." *2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI)*. IEEE, 2018.

[7] U. V. Navalgund and K. Priyadharshini, "Crime intention detection system using deep learning," in *2018 International*

Conference on Circuits and Systems in Digital Enterprise Technology (ICCSDET), pp. 1–6, Kottayam, India, 2018.

[8] M. Nakib, R. T. Khan, M. S. Hasan, and J. Uddin, "Crime scene prediction by detecting threatening objects using convolutional neural network," in *2018 International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2)*, pp. 1–4, Rajshahi, Bangladesh, 2018.

[9] B. Sivanagaleela and S. Rajesh, "Crime analysis and prediction using fuzzy c-means algorithm," in *2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)*, pp. 595–599, Tirunelveli, India, 2019.

[10] S. Ranjan, P. Garhwal, A. Bhan, M. Arora, and A. Mehra, "Framework for image forgery detection and classification using machine learning," in *2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS)*, pp. 1872–1877, Tirunelveli, India, 2018. View at:

[11] E. E. Eryilmaz, D. O. Ahin, and E. Kl, "Machine learning based spam e-mail detection system for Turkish," in *2020 5th International Conference on Computer Science and Engineering (UBMK)*, pp. 7–12, Diyarbakir, Turkey, 2020.

[12] B. S. Aldossari, F. M. Alqahtani, N. S. Alshahrani et al., "A comparative study of decision tree and naive bayes machine learning model for crime category prediction in Chicago," in *Proceedings of 2020 the 6th International Conference on Computing and Data Engineering, ser. ICCDE 2020*, p. 3438, New York, NY, USA, 2020.

[13] S. A. Dudani, The distance-weighted k-nearest neighbor rule, IEEE Transactions on System, Man, and Cybernetics, 6 (1976), 325-327.

[14] Krysovatyy, Andriy, et al. "Economic Crime Detection Using Support Vector Machine Classification." *MoMLeT+ DS*. 2021.

[15] Shah, Neil, Nandish Bhagat, and Manan Shah. "Crime forecasting: a machine learning and computer vision approach to crime prediction and prevention." *Visual Computing for Industry, Biomedicine, and Art* 4.1 (2021): 1-14.

[16] Mahor, Vinod, et al. "Machine Learning based Detection of Cyber Crime Hub Analysis using Twitter Data." *2021 IEEE 4th International Conference on Computing, Power and Communication Technologies (GUCON)*. IEEE, 2021.

[17] Kanimozhi, N., et al. "CRIME type and occurrence prediction using machine learning algorithm." *2021 International conference on artificial intelligence and smart systems (ICAIS)*. IEEE, 2021.

[18] Pastaltzidis, Ioannis, et al. "Data augmentation for fairness-aware machine learning: Preventing algorithmic bias in law enforcement systems." *2022 ACM Conference on Fairness, Accountability, and Transparency*. 2022.

[19] Saraiva, Miguel, et al. "Crime Prediction and Monitoring in Porto, Portugal, Using Machine Learning, Spatial and Text Analytics." *ISPRS International Journal of Geo-Information* 11.7 (2022): 400.

[20] Mukherjee, Anupam, and Anupam Ghosh. "Predictive Geospatial Crime Data Analysis and Their Association with Demographic Features Through Machine Learning Approaches." *International Conference on Computational Intelligence in Pattern Recognition*. Springer, Singapore, 2022.