Mathematical Statistician and Engineering Applications
ISSN: 2094-0343
2326-9865

Crime Prediction and Forecasting Using Deep Learning and Machine Learning Techniques

[1] Mohammed Abubakar, [2] Sohail Ali Khan, [3] Mir Raza Ali [4] Mrs. T. Anita

[1] B.E. Student, Dept. Of Computer Science and Engineering, ISL Engineering College

[2] B.E. Student, Dept. Of Computer Science and Engineering, ISL Engineering College

[3] B.E. Student, Dept. Of Computer Science and Engineering, ISL Engineering College

[4] Assistant Professor, Dept. Of Computer Science and Engineering, ISL Engineering College

Article Info

Page Number: 1601 - 1608

Publication Issue:

Vol 72 No. 1 (2023)

Article History

Article Received: *15 October 2022*

Revised: *24 November 2022*

Accepted: *18 December 2022*

Abstract

Understanding crime patterns is crucial for being better equipped to respond criminal behavior. We examine city-level crime statistics for our stuc Understanding crime patterns is crucial for being better equipped to respond criminal behavior. We examine city-level crime statistics for our stuc Predicting which type of crime is most likely to occur given the time a location in a city is the initial job. It is now possible to utilize AI and machi learning to identify crime using sound or video, and this technology is effectiv Although it is still mostly unknown, using AI/ML to predict crimes or person's propensity to commit one offers potential. The largest difficulty w likely be "proving" it to lawmakers. A mechanism that is intended to preve something from occurring It is challenging to demonstrate the opposi Businesses who are actively involved in providing governments with technology to monitor regions or predict crime will definitely benefit from positive feedback loop. As crime prevention technology develops, it projected that the overall cost of this equipment would increase. We also ma an effort to increase the significance of our categorization work by combini smaller groups into larger ones. Finally, we discuss potential directions f future research after reporting and reviewing our findings using seve classifiers.

Keywords: Crime Prediction, Random Forest, K-Nearest Neighbor (KNN) Algorithm, Decision Trees, Naïve Bayesian Model, K-means Clustering.

1. INTRODUCTION

1.1 Introduction

A deeper understanding of crime is helpful in several ways, including by enabling law enforcement to reduce crime through targeted and sensitive actions, as well as by encouraging greater cooperation between citizens and the state to foster safe neighborhoods. Understanding trends in crime from data is an active and expanding area of research because of the Big Data age and the accessibility of quick, effective algorithms for data analysis. Our algorithms' inputs include the location (latitude and longitude), kind of crime,

and time (hour, day, month, and year): Act 279 - Accident; Act 323 - Violence; Act 302 - Murder; Act 363 - Kidnapping; Act 379 - Robbery; Act 13 - Gambling; Act 279 - Accident; The class is the result. The type of crime that is most likely to have happened is the output. We test a variety of categorization techniques, including Decision Trees, Random Forests, and KNN (K-Nearest

Neighbors). Additionally, we carry out various categorization tasks. First, we attempt to determine which of the six categories of crimes are most likely to have occurred. Then, we attempt to distinguish between violent and non-violent crimes.

1.2 Objective

The goal of our effort is to:

- Predict crime as it occurs;
- Predict areas of greatest crime.
- An understanding of criminal trends.
- Arrange crime according to its location.
- A citywide crime analysis.

2. LITERATURE REVIEW

The issues surrounding crime control have been addressed by several scholars, who have also suggested various murder-prediction systems. The attributes used and the data set used as a reference both affect how accurate a prediction is. In [1], crime hotspots in London, UK, were predicted using human behavior data collected from wireless network activity paired with demographic data taken from actual crime data. In [6], WEKA, an open-source data mining program, and 10-fold cross-validation were used to compare Decision Tree and Naive Bayesian, two classification techniques. The 1990 US Census, 1990 US LEMAS survey, and 1995 FBI UCR were used to create the socioeconomic status, security, and crime databases for this study, respectively.[8] investigated the patterns of traffic accidents in Ethiopia while taking into account a variety of circumstances, including the driver, the environment, the vehicle, and the road conditions. On a dataset of 18,288 accidents, three different classification algorithms—KNN, Naive Bayesian, and Decision Tree—were applied. All three systems' prediction accuracy ranged from 79% to 81%. Accurate and effective analysis of huge crime datasets is a significant difficulty in crime prediction. Data mining is used to uncover hidden trends swiftly and effectively in huge crime datasets. The accuracy of crime prediction is raised by the improved efficiency and decreased mistakes in crime data-mining algorithms. Based on the Cop Link project's experience, carried out at the University of Arizona, a generic data-mining framework was created [7]. Most studies on crime forecasting are focused on locating crime hotspots, or places where crime rates are higher than the national norm. The authors' comparative examination of Kernel Density is presented in [9]. Modeling of the Risk Terrain and Estimation (KDE) employing limited data, techniques for developing hotspot maps, and proposed region-specific forecasting models. Using histogram-based statistical techniques and a spatial-temporal model, Linear KNN and discriminatory analysis (LDA) were used to forecast crime hotspots. To train an ANN (Artificial Neural Network) augmented by the Gamma test to forecast the crime hotspots in Bangladesh that a crime incidence-scanning technique was used in [11],[8] employed a data-driven machine-learning system to examine Taiwanese drug-related crime statistics and forecast new hotspots based on broken-window theory, geographical analysis, and visualization approaches. Statistical analysis was conducted to evaluate violent and non-violent crimes by using arrest data for the social crime network. Kang et al. used environmental context information to improve the prediction of models by proposing a feature-level data fusion method on deep neural networks [11]. Stec and Klabjan utilized the neural network idea

by merging two techniques; convolutional neural network (CNN) and recurrent neural network (RNN), and achieved 75.6% accuracy [12].Catlett et al. proposed an approach that relies on Spatial-temporal to discover the crime in high-risk areas that are mostly urban and dependable trends for crimes forecast in every region while using clustering technique [13]. There have been multiple studies that are performed by using geographical locations, meta-association rules and specific detection system introduced to examine the crime rate in Chicago [13]. Almanie et al. used the dataset for the year 2014 to predict the potential crime type and applied the Apriori algorithm, Naïve Bayesian and Decision Tree [14]. Significant research effort has been made in different aspects, yet literature is still
pointing major concern towards better prediction accuracy, forecast and hotspot in large datasets

3. OVERVIEW OF THE SYSTEM

3.1 Proposed System

Here, we've implemented a variety of packages, including pandas, NumPy, sklearn, matplotlib, etc. Which supports the K-means clustering algorithm's elbow graph and data frame table plot? Datasets are imported into Spyder in CSV format as shown in Fig. 1 after being acquired from Kaggle datasets. Using the elbow approach, normalization is done to get the precise number of clusters (k). The elbow technique computes the SSE after performing k-means clustering on the acquired information for a range of values of k (2–15). For each value of k, the SSE is represented as a line chart.
In this study, machine learning methods are used to predict crime.
A prediction of the kind of crime is made based on crime data collected from a certain place. Before visiting a given area, this application enables users to learn about local crime statistics.

3.2 Methodology

In this paper work, I used five modules and each module has its functions, such as:

- Dataset collection
- Preprocessing
- Initialize algorithm
- Save Model
- Predict

3.2.1 Dataset Collection

The dataset is collected from the Kaggle website which has crime acts as features and location as labels. This dataset is used for the training model.

3.2.2 Preprocessing

In this step dataset features and labels are taken from a .csv file and divided into a testing set and a training set.

3.2.3 Splitting Data

In this step data sets are divided into 70 and 30 percent (x test, y test, x train, and y train) x test has features and the y test has labels (70 percent) whereas the x train has 30 percent features and labels

3.2.4 Model Training

In this step train data features and labels are fed as input machine learning algorithm after initializing and using fit function model is trained and model is saved to the system.

3.2.5 Prediction

A web application is designed using the Flask framework which is used for giving input and predicting crime for a given input.

4. ARCHITECTURE

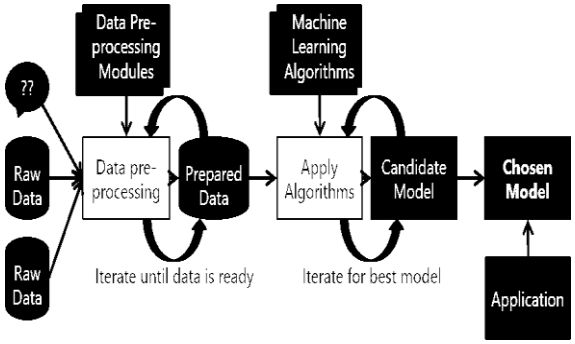

Fig 4.1: Framework of crime analysis

The above architecture diagram shows three stages of data flow from one module to another module. Data collection, preprocessing, and algorithm training.

5. IMPLEMENTATION

The Python programming language is used to carry out the project. To be more precise, Anaconda is being utilised for machine learning. One of several Python distributions is Anaconda. Python now has a new distribution called Anaconda. Prior to then, it was known as Continuum Analytics. There are more than 100 new packages in Anaconda. Scientific computing, data science, statistical analysis, and machine learning all make use of Anaconda. In terms of Python technology, Anaconda is simpler, as we discovered. because it helps in following problems:

- Installing Python on multiple platforms.
- Separating out different environments.
- Dealing with not having correct privileges.
- Getting up and running with specific packages and libraries.

This information was taken from publicly accessible data on the Indore police website, which was created by officers at several police stations. In order to reduce the complexity of the prediction and limit the area it could cover, the notion was first put into practise in the city of Indore. The information was organised and changed into the new timestamp, longitude, and latitude format that would be used as the input for the computer to estimate the crime rate in a certain area or city.

The entries were made solely for the purpose of teaching the computer what to do with the data and what the actual output requirements are. After the machine learned algorithms and process, the accuracy of several algorithms was assessed, and the method with the highest accuracy—Random forest—was utilised for the prediction kernel.

Implementation Details

For the purpose of proper implementation and functioning several Algorithms and techniques were used. Following are the algorithms used:

5.1 KNN (K-Nearest neighbors)

A powerful classification algorithm used in pattern recognition K nearest neighbors stores all available cases and classifies new cases based on a similarity measure (e.g. distance function).One of the top data mining algorithms used today. A non-parametric lazy learning algorithm (An Instance based Learning method).

KNN: Classification Approach

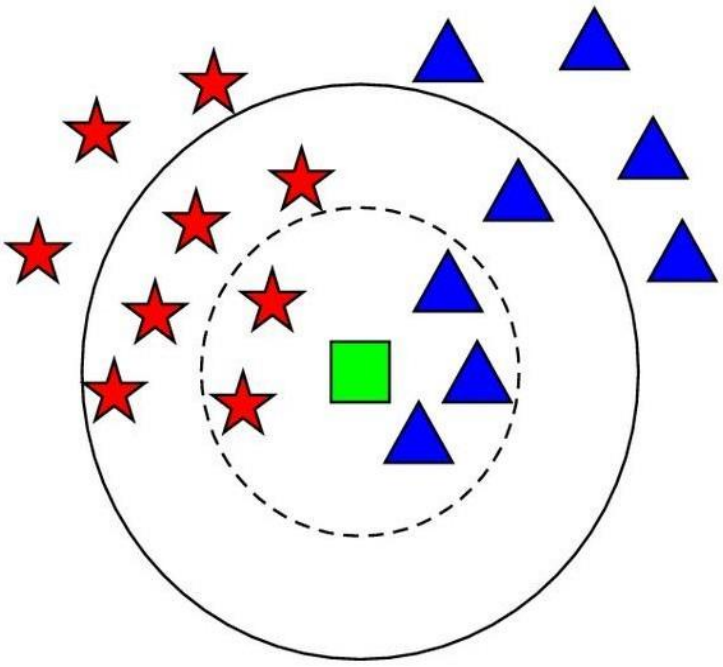

Fig: 5.1: Principle Diagram of KNN

A new instance of an object is categorised by a majority vote for its neighbour classes.

The item is allocated to the class with the highest frequency of occurrence among its K nearest neighbours.(as determined by the distance function)

SCREENSHOTS

Prediction Result:

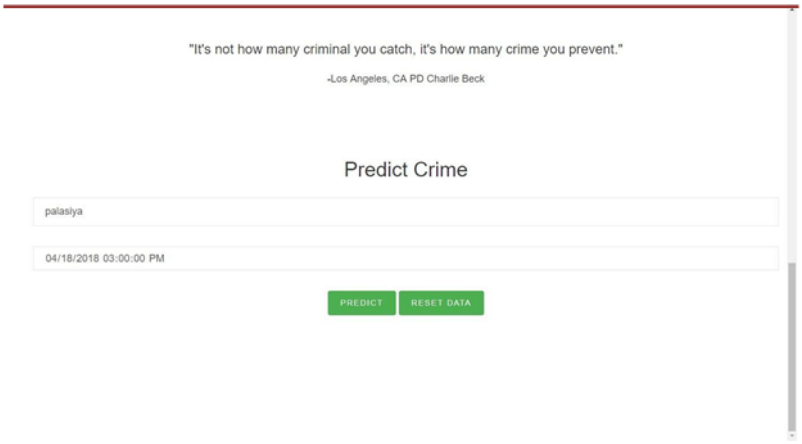

Fig 4.2 : Prediction Result

Analysis:

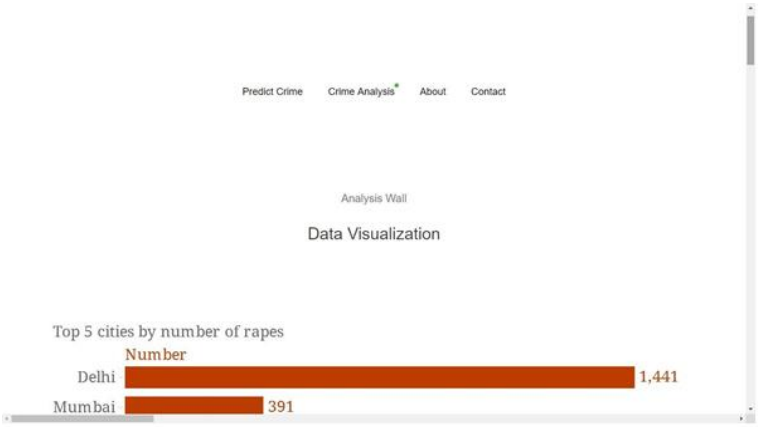

Fig 4.3 : Analysis

Feature analysis:

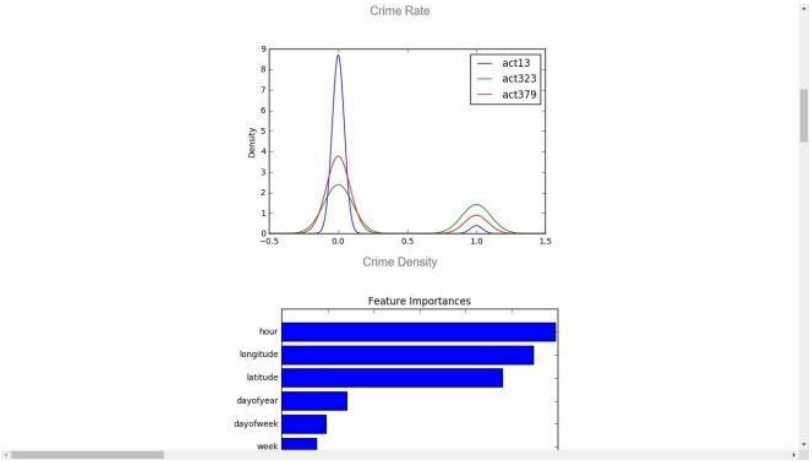

Fig 4.4: Feature Analysis.

Decision Tree:

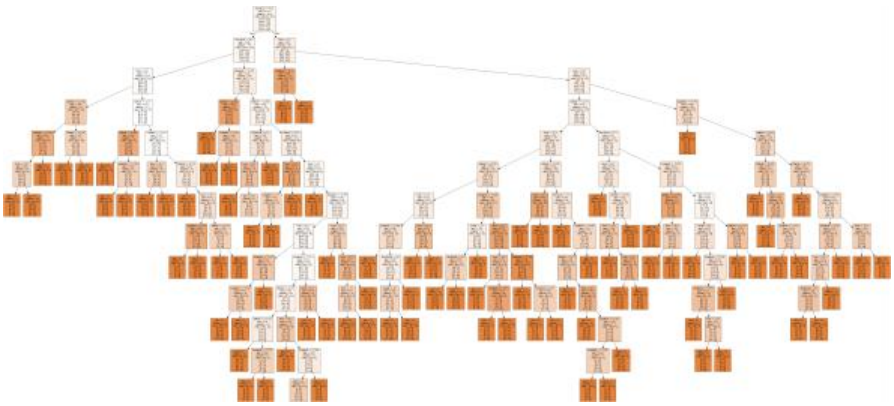

Fig 4.5: Decision Tree

5. CONCLUSION AND FUTURE ENHANCEMENT

This paper's premise is that crimes are largely foreseeable; all that is needed is the ability to sift through a vast amount of data to identify trends that are helpful to law enforcement. A few decades ago, this type of data analysis was technologically impractical, but current advancements in machine learning should make it viable again. AI and machine learning may now be used to detect crime from voice or video, and the system is reliable. The possibility of employing AI/ML to anticipate crimes or a person's tendency to commit one exists despite the fact that it is still mostly unknown. The biggest challenge will probably be "proving" its efficacy to legislators. It might be difficult to prove that a mechanism that is supposed to stop something from happening doesn't exist.Time-series modeling of the data to comprehend temporal relationships in it, which may subsequently be used to forecast surges in various kinds of crime, is one way to further this study. Additionally, it would be intriguing to investigate connections between surges in various crime categories. For instance, two or more crime categories may rise and fall concurrently, which would be an intriguing connection to discover. Implementing a more precise multi-class classifier and looking at better ways to visualize our results are other areas that need improvement.

REFERENCES

[1]H. Adel, M. Salheen, and R. Mahmoud, "Crime in relation to urban design. Case study: the greater Cairo region," Ain Shams Eng. J., vol. 7, no. 3, pp. 925-938, 2016.

[2] "Overall crime rate in Vancouver went down in 2017, VPD says," CBC News, Feb. 15, 2018. [Online] Available: https://www.cbc.ca/news/canada/british-columbia/crime-rate-vancouver2017-1.4537831 . [Accessed: 09- Aug- 2018].

[3] J. Kerr, "Vancouver police go high tech to predict and prevent crime before it happens," Vancouver Courier, July 23, 2017. [Online] Available:https://www.vancourier.com/news/vancouver-police-go-high-tech-topredict-and-prevent-crime-before-it-happens-1.21295288 . [Accessed: 09- Aug- 2018]

[4] J. Han, Data mining: concepts and techniques, Morgan Kaufmann, 2012.

[5] R. Iqbal, M. A. A. Murad, A. Mustapha, P. H. Shariat Panahy, and N. Khanahmadliravi, "An experimental study of classification algorithms for crime prediction," Indian J. of Sci. and Technol., vol. 6, no. 3, pp. 4219- 4225, Mar. 2013.

[6] H. Chen, W. Chung, J. J. Xu, G. Wang, Y. Qin, and M. Chau, "Crime data mining: a general framework and some examples," IEEE Computer, vol. 37, no. 4, pp. 50-56, Apr. 2004.

[7] T. Beshah and S. Hill, "Mining Road traffic accident data to improve safety: role of road-related factors on accident severity in Ethiopia," Proc. of Artificial Intell. for Develop. (AID 2010), pp. 14-19, 2010.

[8] M. Al Boni and M. S. Gerber, "Area-specific crime prediction models," 15th IEEE Intl. Conf. on Mach. Learn. and Appl., Anaheim, CA, USA, Dec. 2016.

[9] Q. Zhang, P. Yuan, Q. Zhou, and Z. Yang, "Mixed spatial-temporal characteristics based crime hot spots prediction," IEEE 20th Intl. Conf. on Comput. Supported Cooperative Work in Des. (CSCWD), Nanchang, China, May 2016.

[10]Khan, P. A., Ali, M. R., Assistant Professor, & Assistant Professor. (n.d.). Methodological implementation of academic organizational operated data with clustering mechanism using an improved k- means algorithm. Ijmec.com. Retrieved May 10, 2023, from https://doi-ds.org/doilink/10.2022-22255487.

[11]H.-W. Kang and H.-B. Kang, ‘‘Prediction of crime occurrence from multi-modal data using deep learning,’’ PLoS ONE, vol. 12, no. 4, Apr. 2017, Art. no. e0176244.

[12]A. Stec and D. Klabjan, ‘‘Forecasting crime with deep learning,’’ 2018, arXiv:1806.01486. [Online]. Available: http://arxiv.org/abs/1806.01486

[13] C. Catlett, E. Cesario, D. Talia, and A. Vinci, ‘‘A data-driven approach for spatio-temporal crime predictions in smart cities,’’ in Proc. IEEE Int. Conf. Smart Comput. (SMARTCOMP), Taormina, Italy, Jun. 2018, pp. 17–24.

[14]T. Almanie, R. Mirza, and E. Lor, ‘‘Crime prediction based on crime types and using spatial and temporal criminal hotspots,’’ Int. J. Data Mining Knowl. Manage. Process, vol. 5, pp. 1–19, Aug. 2015.

[15].Mohammed Abdul Bari, Shahanawaj Ahamad, Mohammed Rahmat Ali,” Smartphone Security and Protection Practices”, International Journal of Engineering and Applied Computer Science (IJEACS) ; ISBN: 9798799755577 Volume: 03, Issue: 01, December 2021 (International Journal,U K)

[16] Dhabliya, D., & Sharma, R. (2019). Cloud computing based mobile devices for distributed computing. International Journal of Control and Automation, 12(6 Special Issue), 1-4. doi:10.33832/ijca.2019.12.6.01

[17] Dhabliya, D., Soundararajan, R., Selvarasu, P., Balasubramaniam, M. S., Rajawat, A. S., Goyal, S. B., . . . Suciu, G. (2022). Energy-efficient network protocols and resilient data transmission schemes for wireless sensor Networks—An experimental survey. Energies, 15(23) doi:10.3390/en15238883