

Crime Prediction and Analysis Using Machine Learning

Mrs. Karuna verma¹, M. Suchitha², D. Lakshmi Satwika³, G. Ashwini⁴

¹Associate Professor, Department of CSE, Sridevi Women's Engineering College, Hyderabad, Telangana, India.

Email Id: ¹karuna18882@gmail.com

^{2,3,4}UG Student, Department of CSE, Sridevi Women's Engineering College, Hyderabad, Telangana, India.

Email Id: ²suchithavarma123@gmail.com, ³satwika.datla1911@gmail.com,

⁴gollaashwini2002@gmail.com

ABSTRACT

Preventing criminal activity is crucial since crime is a major issue in modern society. Several violent crimes are perpetrated every day. For this reason, it is necessary to retain a record of any criminal activity for possible future reference. Current challenges include keeping accurate crime datasets and evaluating these datasets to aid in future crime prediction and resolution. This study aims to forecast future crime types based on an analysis of large datasets including information about past crimes. The goal of this research is to use data analysis and machine learning science to anticipate crimes committed in Chicago based on that city's extensive crime database. The statistics are taken directly from the Chicago Police Department's website. Details about the location, time, date, latitude, and longitude of a crime are all included. The data will be preprocessed before the model is trained, and then features will be selected and scaled for optimal accuracy. We will evaluate many different crime prediction algorithms, including K-Nearest Neighbor (KNN), and utilise the most effective one in our training procedures. Several scenarios, such as when crime rates or criminal activities tend to spike, will be graphically represented as part of the dataset's visualisation. This project's overarching goal is to provide a conceptual framework for how law enforcement organisations may make use of machine learning to improve their ability to identify, anticipate, and catch criminals at an accelerated pace, hence lowering the crime rate. This is not unique to Chicago; it may be used elsewhere, too, if a suitable dataset is made available.

1. INTRODUCTION

The greatest danger that humanity faces is from criminal activity. Many crimes occur at predictable intervals. Maybe it's becoming bigger and more widespread very quickly.

The rate of crime is the same in the country as it is in the city. There is a wide variety of criminal acts, including burglary, homicide, rapes, attack, battery, false imprisonment, abduction, and homicide. There has to be a significant increase in the rate at which crimes are solved. Crime has been on the rise, and it is now the job of the police to rein it in and bring it down to manageable levels.

Because there is so much data on criminal activity, authorities have a hard time both predicting crimes and identifying perpetrators. In order to solve cases more quickly, we need better technologies. Because of the aforementioned issue, I decided to look into possible means of simplifying the process of settling a criminal case. Reinforcement learning algorithms and data science were shown to simplify and expedite tasks via extensive research and case studies. The goal of this study is to use the properties of the dataset to

develop predictions about criminal behaviour. The data set is collected from authoritative sources. Crime patterns in a certain location may be anticipated with the use of a machine learning algorithm written in Python. The goal is to train the model to make predictions. The test dataset is used to verify the results of the training data set. Depending on the precision, a more suitable method will be used to construct the model.

Predictions of future criminal activity will be made using a classification system similar to the K-Nearest Neighbor (Knearest) technique. Data visualisation is used to examine potential criminal activity in the nation. The crime rate in Chicago may be reduced thanks to this effort since police can better anticipate and identify criminal activity.

2. RELATED WORK

Crime pattern detection, analysis & prediction. In Electronics, Communication and Aerospace Technology

Crime is a major drain on resources and a major source of societal discontent. Investing in research that speeds up the process of investigating crimes will be money well spent. Among all offenders, just around 10% are responsible for about 50% of all crime [9]. We provided the algorithm data from a reliable internet domain in India that records crimes from the previous year, including murder, kidnapping, dacoits, robbery, burglary, and rape. A regression model built with information from Indian statistics, which gives information on a range of crimes committed during the last 14 years (2001-2014), can be used to predict the crime rate in each state for the upcoming year [8]. To increase the crime prediction accuracy and make information discovery from the crime data easier [4]. The work you're doing will help local law enforcement organisations fight crime in your area.

An overview on crime prediction methods

Recently, investigations of crime data have been essential for unravelling the dataset's intricacies. Involved parties in law enforcement will benefit from this procedure in terms of making arrests and guiding efforts to reduce crime. From a strategic or a tactical vantage point, they may benefit greatly from having the capacity to foresee future criminal activity based on location, pattern, and time. Yet, with the rising rates of crime in the current day, it is difficult to reliably anticipate future crimes with improved performance. As a result, a reliable strategy for predicting criminal activity is crucial for preventing more crime. Several academics have recently undertaken a study to foretell criminal activity using certain inputs. Several prediction approaches may be used to compare and contrast the efficacy of various models of prediction. Unfortunately, there are still certain gaps in their data that prevent an exact forecast of crime hotspots from being made.

There have been many earlier publications of academic papers on this subject. This publication provides a comprehensive overview of these studies and a summary of their findings. Our goal is to catalogue the many forms the crime prediction system has taken and the ways in which it may be improved in the future.

Crime prediction and analysis in kerala with the use of clustering approaches.

The fight against crime is an important one since it is so pervasive and disturbing in modern society. Analyzing crime data in a methodical manner allows for the discovery and exploration of trends and patterns. In this paper, we analyse crime statistics in Tamil Nadu using many data mining clustering methods. The statistics on crime in India come from the country's national criminal records bureau (NCRB). Information on criminal activity in the

six cities of Madras, Coimbatore, Salem, Madurai, Thirunelveli, and Thiruchirapalli is included, with a total of 1,760 separate incidents and 9 corresponding qualities spanning the years 2000 through 2014. Crime activities are clustered using many algorithms such as KMeans clustering, K medoids clustering, and Dense Dependent Spatial Clustering with Noise (DBSCAN) methods, and their results are compared to identify the most effective clustering methodology for crime detection. K-Means clustering results are shown on a Google Map for user interaction and better comprehension. Predicting criminal activity using K-Nearest Neighbor (KNN) analysis. Several measures, including accuracy, recall, and F-measure, are used to assess and compare the efficacy of various clustering methods. This study aids law enforcement in Tamilnadu in their efforts to minimise crime by increasing the precision with which they can anticipate and identify criminal activity.

3. METHODOLOGY

This process involves purging the system of any values that are infinite or unattainable. The three most crucial techniques are data editing, data sampling, and data formatting. A cleaning operation is carried out to close any gaps when the data is insufficient. Utilising pertinent data is referred to as sampling, and doing so could possibly hasten the algorithm's execution. Python is used for all of the initial research. Whether they are supervised or unsupervised in nature, modelling solutions must be adopted accordingly. As an illustration, consider the usage of methods like KNN classification, logistic regression, decision trees, random forests, support vector machines, and bayesian methods. It makes use of the CSV file known as the Kaggle crime dataset. Data The preprocessing dataset contains 10,000 raw records. To get rid of any nulls in the data frame, type `df = df.dropna()`.

Using Label Encoder, the category characteristics (Location, Blocks, Crime Type, Benefit Sharing) are encoded as numbers. New characteristics, such as month and hour, are created from the date attribute so that it may be utilized as a model feature. The feature selection process which will be used in model construction is performed. Attributes such as "block," "location," "district," "community area," "x" and "y" coordinates, "latitude" and "longitude," "hour" and "month," and "month" are used to choose features. System Design and Instructional Model After features have been selected, training may begin using the location and month attributes. The dataset is divided into four parts. they are xtrain, xtest, ytest, ytrain. The formklerin model is loaded by the algorithms. Model is used in the construction of models. Fit (xtrain, ytrain) (xtrain, ytrain).

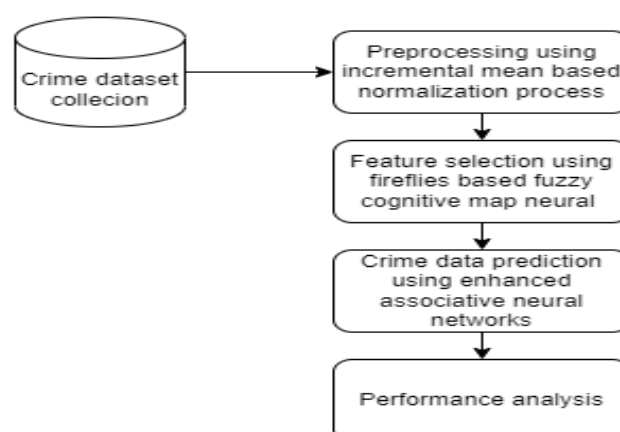


Fig 1 Model Diagram

4. RESULTS

First, we'll have to upload the dataset, and then we'll visualise it by separating the various sorts of assaults and showing them on a graph of annual rate and hourly rate. The annual rate graph is shown here.

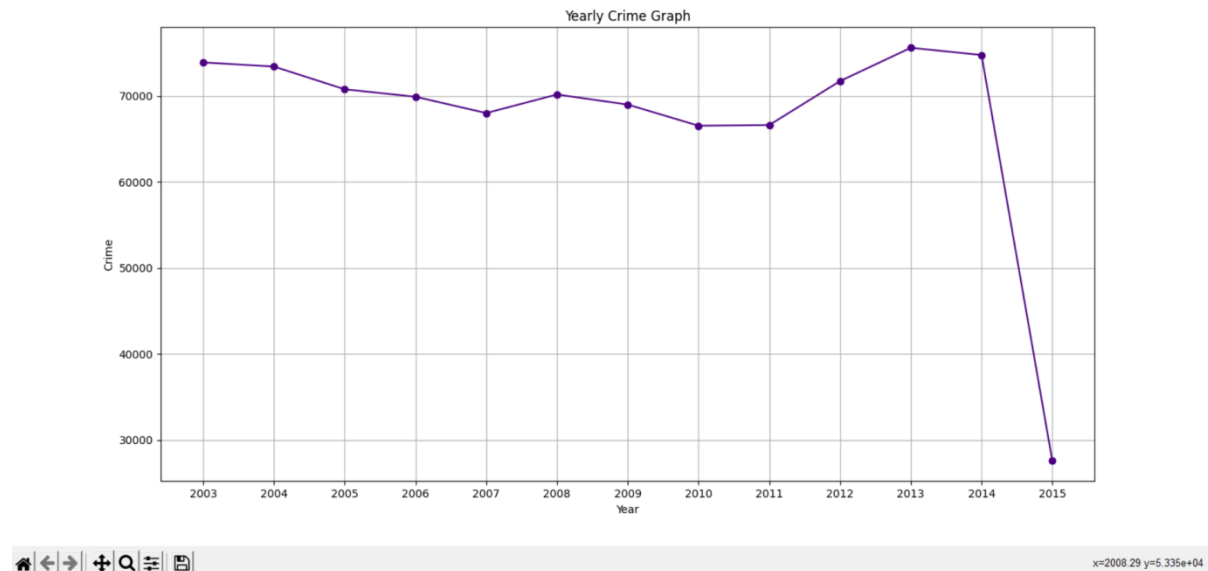


Fig 2 Yearly Crime Graph.

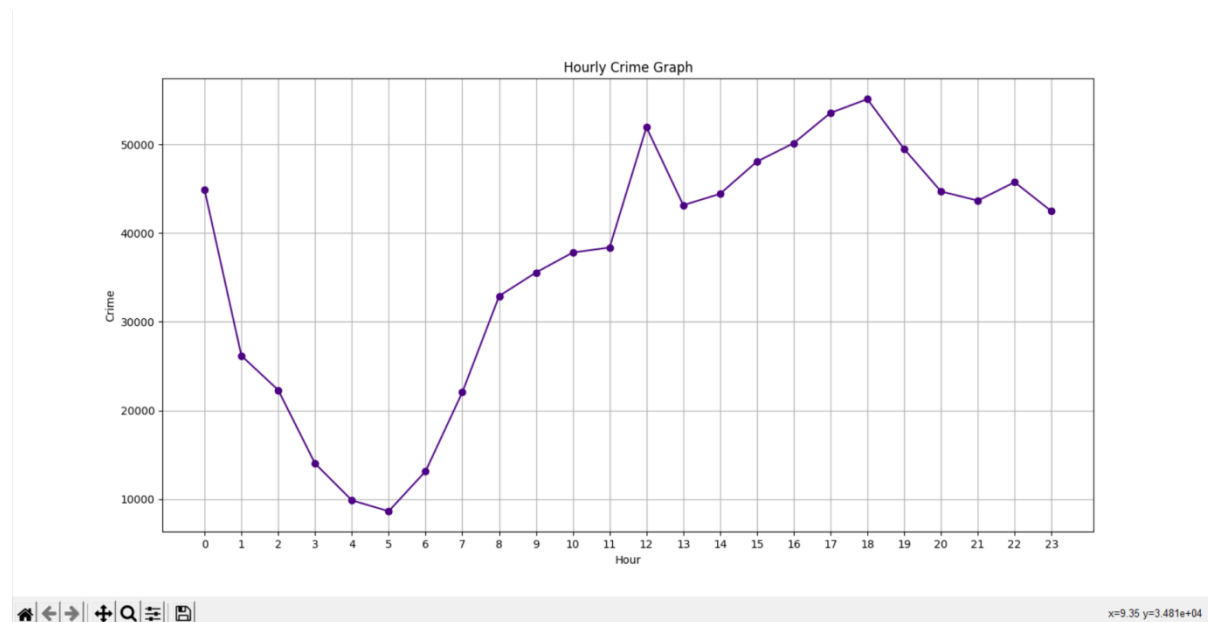


Fig 3 Hourly Crime Graph

And we use two algorithms—neural networks and long short-term memory networks—to foresee future criminal activity. Initially, you have to train our algorithms, and then we'll utilise those trained algorithms to make predictions and see which ones result in the lowest loss rates so we may apply those methods in the future.

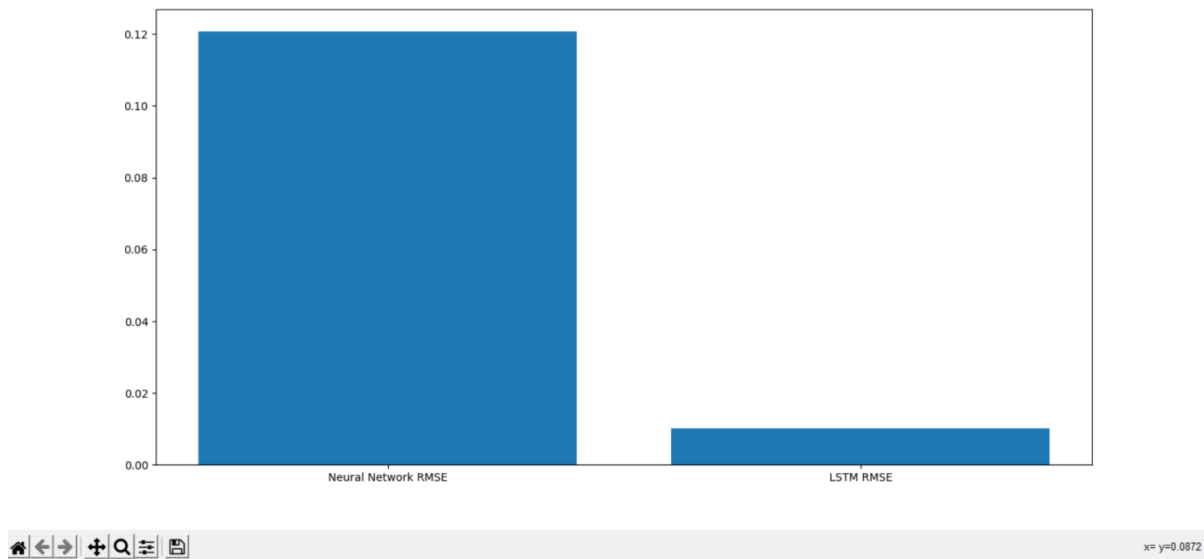


Fig 4 RMSE Comparison Graph

The above graph shows the comparison between the algorithms and also both axis for representing the algorithms, the comparison is RMSE. According to the above graph, LSTM has superior performance since it has a lower RMSE. The RMSE error of a prediction model is said to be minimised if it has a high degree of accuracy.

5. CONCLUSION

Machine learning tools have made it easy to identify links and patterns in datasets that were previously unrelated. Our study is primarily focused on developing methods for determining the kind of crime that was committed at a certain location. Using machine learning principles, we changed and cleaned up training data to create a model. With a 0.789% accuracy rate, the model can classify the type of crime. In order to analyse information, data visualisation is helpful. Here, a variety of graphs, including bar, pie, line, and scatter graphs, are exhibited. Each has unique characteristics. An array of graphs were made, and intriguing data was discovered, all of which provided insight into Chicago crime databases and their capacity to record factors that improve.

6. REFERENCES

- [1] Yadav, S., Timbadia, M., Yadav, A., Vishwakarma, R., & Yadav, N. (2017, April). Crime pattern detection, analysis & prediction. In Electronics, Communication and AerospaceTechnology (ICECA), 2017
- [2] International conference of (Vol. 1, pp. 225- 230). 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] 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.

- [4] 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.
- [5] 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. 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.