ISSN: 2278-4632
Vol-12 Issue-11 No.02 November 2022

CRIME ANALYSIS AND PREDICTION USING MACHINE LEARNING

[#1]G.SOUMYA, M.Tech Student,
[#2]Dr.D.Ramesh, Professor of CSE & Principal CSI, ISTE,
Dept of Computer Science & Engineering, JNTUH University College of Engineering, Nachupally, Jagtial

ABSTRACT: Crime is a major and ubiquitous issue in our society, making crime prevention crucial. There are several crimes committed every day. Keeping tabs on criminal activity requires recording incidents and saving data for analysis down the line. The ability to keep a reliable database of criminal activity and use it to prevent or solve future crimes is currently unavailable. The purpose of this research is to examine a big dataset of criminal records in order to foretell the kind of crimes that might occur in the future based on a number of different variables. For this study, we will use crime data from Chicago and integrate data science with machine learning to make predictions about future crimes. The Chicago Police Department's official website is the source for the crime data. The specifics of the crime, such as its time, location, and type, are included. High accuracy in model training requires preprocessing the data and selecting and scaling features. There will be an evaluation of several different classification methods, including the K-Nearest Neighbor (KNN) classification, and the best algorithm will be used to train the others. Data sets will be represented graphically to show various scenarios, such as when or what month the crime rate is highest. The primary goal of this research is to lay the groundwork for how machine learning might be used by law enforcement to improve crime fighting efforts, including in the areas of crime prediction and detection. Depending on the availability of the dataset, this could be used in locations outside of Illinois.

Keywords: *K-Nearest Neighbor Support, Vector Machine Autoregressive moving average, recurrent neural network, Recursive Feature Elimination, National Crime Records Bureau.*

---

1. INTRODUCTION

Criminal activity poses a serious threat to society. There is a high rate of crime in the world. It could be moving rapidly and spreading. As with major cities, crime can occur in even relatively small communities. Theft, murder, rape, assault, battery, false imprisonment, and kidnapping are all instances of criminal acts. [1][2]. We need significantly quicker case resolution because crime is increasing. With crime on the rise, the onus is on law enforcement to intervene and make communities safer. [3][4]. Since there is a wealth of data on criminal activity, the two largest challenges facing law enforcement are crime prediction and perpetrator identification. Cases can be resolved more rapidly with the help of technological advancements. Learning to anticipate outcomes with a model would be the end objective. [5][6]. The test dataset will be used to validate the results of the training done with the first dataset. How well it works will determine whether or not a more advanced algorithm is utilized to construct the model. K-Nearest Neighbor (KNN) and other categorization algorithms will be used to make predictions about criminal activity. It is possible to ascertain whether or not there have been crimes in the country by utilizing this data set. [33][34][35].

The crime rate in Chicago can be reduced thanks to this project since the police have an easier time anticipating and solving crimes.

[7][8]. Algorithms based on machine learning have progressed to the point where they can anticipate criminal behavior based on historical data. The focus of this study is on predicting the future rate of crime across multiple states using machine learning models. [47] [48]. The primary objective is to develop a model for estimating the frequency of specific crimes in a given state. [9] [10]. Several different machine learning models, including K-NN and boosted decision trees, will be used to make accurate crime predictions in this study. [11][12]. Area A thorough analysis of the neighborhood can shed light on the causes of

Copyright @ 2022 Author

ISSN: 2278-4632
Vol-12 Issue-11 No.02 November 2022

criminal activity. When police use various plots and data visualization tools, they are better able to detect and anticipate criminal activity. By doing so, we can indirectly reduce crime rates while simultaneously making vital locations safer. Because of the criminals' hectic schedules and the fact that they like to operate in familiar environments, criminal activity will continue. To duplicate their initial success, they frequently return to the same locations. [13][14].

2. REVIEW OF LITERATURE

ML techniques used in crime prediction

Comparisons of trends in violent crime from the Communities and Crime Unnormalized Dataset were made using the open-source Waikato Environment for Knowledge Analysis data mining tool [42]. (WEKA). In this study, we use real crime data from neighborhoods and cities to compare three algorithms: linear regression, additive regression, and the decision stump. The exam takers were selected at random. The test data was somewhat random, which the linear regression method accounted for, hence it performed better than the other two. The study's objective was to demonstrate the efficacy of machine learning (ML) algorithms in predicting violent crime patterns and their utility in other areas, such as locating crime hotspots, creating criminal profiles, and identifying criminal trends.

Reference 34 investigates the viability of utilizing a predictive study of predicted criminal activity in a metropolitan area. A 200 m x 250 m grid was used to analyze burglaries, robberies, and battery in the past. Using three years' worth of crime data and the outcomes of logistic regression and neural network models, an ensemble model produced estimates for 2014. The predictions were evaluated based on their hit rate, accuracy, and prediction index. The predictive analysis strategy used to make the biweekly projections yielded correct predictions. They discovered that by distinguishing between day and night and comparing forecasts made every two weeks to forecasts made every month, they could significantly enhance the accuracy of their predictions.

Predicting criminal activity with ML is the subject of reference [44]. Vancouver, Canada's crime data from the past 15 years was analyzed to produce these estimates. Using machine learning, this investigation of criminal behavior gathers information, organizes it, searches for trends, makes forecasts, and displays the findings. To delve further into the crime dataset, we used the boosted decision tree and K-nearest neighbor (KNN) approaches. The researchers analyzed 560,000 crime statistics spanning 2003-2018. With an accuracy of 39%-44%, ML systems have the potential to foresee criminal activity. Researches discovered that by customizing the algorithms and crime data for each application, the prediction model's accuracy might be improved.

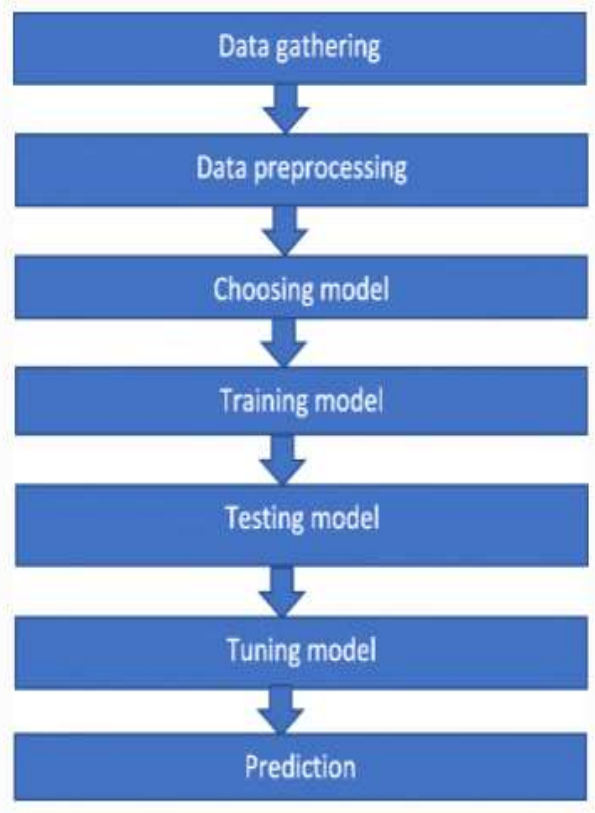
An ML technique for foreseeing Philly's crime rates is provided in [45]. The task might be broken down into three distinct parts: determining whether or not a crime was committed, identifying the nature of the crime, and selecting the most likely perpetrator. Algorithms like logistic regression, KNN, ordinal regression, and tree methods were used to train the datasets so that more precise quantitative crime estimates could be made. In addition, they predicted crime by creating a map that depicted the frequency of certain crimes in various Philadelphia neighborhoods during a given period of time. A separate hue represented each distinct category of criminal activity. To illustrate the evolution of criminal activity in Philadelphia through time, a variety of offenses were examined, from physical assaults to online fraud. An astounding 69% of the time, their system correctly predicted whether or not a crime would occur, and 47% of the time, between 0 and 32 times, it correctly predicted the number of crimes that would occur.

The estimated crime rates in reference [47] are determined using a machine learning (ML) technique that employs a graphical user interface (GUI). The primary purpose of this research was to evaluate and select the best effective machine

Copyright @ 2022 Author

ISSN: 2278-4632
Vol-12 Issue-11 No.02 November 2022

learning-based crime rate estimation techniques for this dataset. The dataset was analyzed using supervised machine learning methods for data validation, data cleaning, and data visualization. Many supervised machine learning systems' outputs were examined to make forecasts. The steps of the proposed system are laid out in Figure 1: data collection, preprocessing, model creation, dataset training/testing, algorithm comparison. Our research aims to demonstrate the feasibility and efficacy of using a machine learning (ML) system to forecast violent crimes.



In reference [53], we see examples of the use of data mining and ML in criminological research. This paper contributes to the field by outlining the methods used in criminal data analytics. KNN, SVM, naive Bayes, and clustering were only some of the ML methods used to properly categorize, comprehend, and rank datasets in accordance with predetermined standards. You can learn more about the nature of the incident and the possible crime hotspot by reading and analyzing the details provided in the police report. Multiple operations, including feature selection, clustering, analysis, prediction, and evaluation, can be carried out on the provided datasets with the help of the suggested model. This research proves that machine learning algorithms can be helpful in analyzing and gauging future criminal activity.

In order to anticipate where crimes would occur in the future, the authors of reference [54] used a grid-based crime prediction model and categorized 84 different places around the city of Taipei. Each grid is analyzed using machine learning techniques to reveal trends and make crime forecasts for the following month. The DNN model performed the best when compared to the other ML methods used. The key contribution of this research is the use of state-of-the-art ML methods, particularly the idea of feature learning. In addition, testing of crime displacement showed that the proposed model method performed better than the baseline.

To build a model for foreseeing criminal conduct, the authors of reference [55] suggested using the decision tree (J48) technique. Using J48 in law enforcement and intelligence analysis has the potential to lower crime rates because it is the most successful ML algorithm for predicting crime data in the relevant literature. The WEKA toolbox was used to create the J48 classifier, which was then trained on a cleaned-up crime data set. When attempting to estimate the unknown crime data category, experimental results from the J48 algorithm showed an accuracy of 94.25287%. There is good reason to put faith in the system's ability to foresee criminal behavior, given the high level of accuracy it displays.

3. SYSTEM DESIGN AND RESULT

SYSTEM DESIGN

Machine learning system architecture refers to the process of tailoring a system's underlying hardware, algorithms, data, and software to achieve a set of predetermined goals. The software design documents the requirements for making the software used in web applications.

Copyright @ 2022 Author

ISSN: 2278-4632
Vol-12 Issue-11 No.02 November 2022

Fig: context diagram

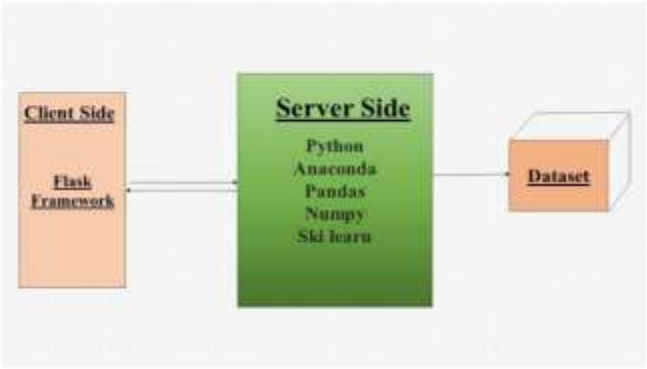


Fig: Architecture Design

IMPLEMENTATION

Predicting the future values of a dependent variable requires first establishing the link between independent and dependent variables, which can be done with a regression technique. By studying how different variables have interacted in the past, regression can provide predictions about how they will behave in the future. To differentiate between high-crime and low-crime areas, the researchers used a K-means clustering technique.

ALGORITHM IMPLEMENTATION

Regression Algorithm

Regression analysis is used to examine the connection between a set of independent variables and a set of dependent ones, with the ultimate goal of predicting the future values of the dependent ones. Predictions can be made using a model of the connection between the variables over time, which is what regression does. Linear regression analysis of the project data is used to draw findings for the methodology. Linear regression is one such statistical method that can be used to predict traffic crimes. The formula Y=b0+b1*x is used by the algorithm to determine the mean and variance of the dependent variables before making predictions about the future.

K-means Clustering

The CRIME hotspots and coldspots were evaluated using K-means clustering. Data is clustered into some fixed number of groups (let's call them k clusters) using a simple method. M All of these centers of gravity have different effects, so it's important to get them in the right places.

Decision Tree:

To identify and foretell information, a decision tree is employed. The production of intervals based on splits in individual attribute values is necessary for training a classification function.

X (year)	Y(value)	A1=(x-mean of x)	B1=(y—mean of y)	A1 *B1	(A1)2	(b1)2
2008	3496	-4	-329	1316	16	108241
2010	3500	-3	-325	975	9	105625
2010	3987	-2	162	-324	4	26244
2011	2987	-1	-838	838	1	702244
2012	3019	0	-806	0	0	649636
2013	3999	1	174	174	1	30276
2014	4015	2	190	380	4	36100
2015	4786	3	961	2883	9	923521
2016	4018	4	193	772	16	37249
2021	4445	5	620	3100	25	384400

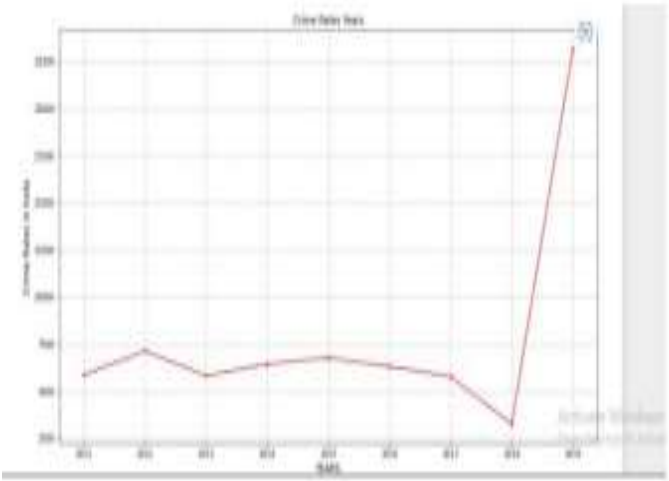


Fig. To display the crime rates year in India

4. CONCLUSION

Copyright @ 2022 Author

ISSN: 2278-4632
Vol-12 Issue-11 No.02 November 2022

Since the advent of machine learning, it has been much simpler to spot connections and patterns in large amounts of data. For the most part, our efforts in this study have focused on creating tools to determine the nature of future criminal activity based on the specifics of past cases. We used machine learning techniques to clean and process training data in order to construct a model. The algorithm has a 0.789% success rate when attempting to predict the category of crime. The analysis of data is made easier with the help of data visualization. Several different types of graphs with their own unique characteristics are presented here. Through the creation of various graphs and the discovery of intriguing statistics, we were able to improve our understanding of Chicago's crime databases and their capacity to collect factors that contribute to public safety.

REFERENCES

- 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.
- 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.
- K. Zakir Hussain, M. Durairaj and G. R. J. Farzana, "Criminal behavior analysis by using data mining techniques," IEEE-International Conference On Advances In Engineering, Science And Management (ICAESM -2012), Nagapattinam, Tamil Nadu, 2012, pp. 656-658.
- Keyvanpour, Mohammad & Javideh, Mostafa & Ebrahimi, Mohammadreza. (2011). Detecting and investigating crime by means of data mining: A general crime matching framework. Procedia CS. 3. 872-880. 10.1016/j.procs.2010.12.143.
- Ioannis Kavakiotis OlgaTsave Athanasios Salifoglou Nicos Maglaveras Ioannis Vlahavas Ioanna Chouvarda, Machine Learning and Data Mining Methods in Diabetes Research, Computational and Structural Biotechnology Journal Volume 15, 2017, Pages 104-116
- Frank, Eibe & Hall, Mark & Holmes, Geoffrey & Kirkby, Richard & Pfahringer, Bernhard & Witten, Ian & Trigg, Len. (2010). Weka-A Machine Learning Workbench for Data Mining. 10.1007/978-0-387- 09823-4_66.
- Ginger Saltos and Mihaela Coacea, 2017 International Journal of Information Technology and Decision Making, An Exploration of Crime Prediction Using Data Mining on Open Data.
- Shiju Sathyadevan, Devan M.S, Surya Gangadharan.S, Crime Analysis and Prediction Using First International Conference on networks & soft computing (IEEE) 2014.
- Lakshman Narayana Vejendla and A Peda Gopi, (2019),” Avoiding Interoperability and Delay in Healthcare Monitoring System Using Block Chain Technology”, Revue d'Intelligence Artificielle , Vol. 33, No. 1, 2019,pp.45-48.
- Gopi, A.P., Jyothi, R.N.S., Narayana, V.L. et al. (2020), “Classification of tweets data based on polarity using improved RBF kernel of SVM” . Int. j. inf. tecnol. (2020). https://doi.org/10.1007/s41870-019-00409- 4.
- A Peda Gopi and Lakshman Narayana Vejendla, (2019),” Certified Node Frequency in Social Network Using Parallel Diffusion Methods”, Ingénierie des Systèmes d' Information, Vol. 24, No. 1, 2019,pp.113-117.. DOI: 10.18280/isi.240117

Copyright @ 2022 Author