

CRIME DATA ANALYSIS USING MACHINE LEARNING

P.Kiran[1], B.Mounika[2], P.Naveen[3], N.Tejaswini[4], B.Karthik[5]

Assistant Professor (IT), Dept.of Information Technology, QIS College Of Engineering&Technology, Andhra Pradesh, India(1)

Final Year (B.Tech), Dept. Of Information Technology, QIS College of Engineering&Technology, Andhra Pradesh, India(2,3,4,5)

Corresponding author.

Correspondence P.Kiran

E-mail:

Article info

Received 8 th March 2022 Received in revised form 12 April 2022 Accepted 6 July 2022

Keywords

Crime analysis, prediction analysis, Machine Learning, decision trees, pattern detection.

Abstract

Crime is one of the biggest and dominating problems in our society and its prevention is an important task. Daily there are huge numbers of crimes committed frequently. This requires keeping track of all the crimes and maintaining a database for same which may be used for future reference. The current problem faced are maintaining of proper dataset of crime and analyzing this data to help in predicting and solving crimes in future. The objective of this project is to analyze dataset which consist of numerous crimes and predicting the type of crime which may happen in future depending upon various conditions. In this project, Machine Learning and data science techniques are used for crime prediction of Chicago crime data set. For this supervised classification Random Forest algorithm is used. This approach involves predicting crimes classifying, pattern detection and visualization with effective tools and technologies. Use of past crime data trends helps us to correlate factors which might help understanding the future scope of crimes. In this work, various visualizing techniques and machine learning algorithms are adopted for predicting the crime distribution over an area. In the first step, the raw datasets are processed and visualized based on the need.

I. INTRODUCTION

Crimes are the significant threat to the humankind. There are many crimes that happens regular interval of time. Perhaps it is increasing and spreading at a fast and vast rate. Crimes happen from small village, town to big cities. Crimes are of different type robbery, murder, rape, assault, battery, false imprisonment, kidnapping, homicide. Since crimes are increasing there is a need to solve the cases in a much faster way. The crime activities have been increased at a faster rate and it is the responsibility of police department to control and reduce the crime activities. Crime prediction and criminal identification are the major problems to the police department as there are tremendous amount of crime data that exist. There is a need of technology through which the case solving could be faster.

Latest technical developments in sophisticated tools of data analytics and visualization are helping the society in different ways to analyze the data of social relevance. One of such socially relevant activities is crime details of different demographic places. The analysis of the crime data will help decision making agencies to take precautionary steps to control the crime rate over demographic places.

Today, a high number of crimes are causing a lot of problems in many different countries. In fact, scientists are spending time studying crime and criminal behaviors in order to understand the characteristics of crime and to discover crime patterns. Dealing with crime data is very challenging as the size of crime data grows very fast, so it can cause storage and analysis problems. In particular,

issues arise as to how to choose accurate techniques for analyzing data due to the inconsistency and inadequacy of these kinds of data. These issues motivate scientists to conduct research on these kinds of data to enhance crime data analysis. Dealing with crime data is very challenging as the size of crime data grows very fast, so it can cause storage and analysis problems. In particular, issues arise as to how to choose accurate techniques for analyzing data due to the inconsistency and inadequacy of these kinds of data. These issues motivate scientists to conduct research on these kinds of data to enhance crime data analysis. Data mining and machine learning are inter-disciplinary fields involving computers and mathematics wherein the programming is done for the system to carry out the operation. Both are highly important in detection and prevention of crime. Crime analysis involves extraction of crime patterns, prediction and crime detection.

The objective would be to train a model for prediction. The training would be done using the training data set which will be validated using the test dataset. This work helps the law enforcement agencies to predict and detect crimes in Chicago with improved accuracy and thus reduces the crime rate. There has been tremendous increase in machine learning algorithms that have made crime prediction feasible based on past data. The aim of this project is to perform analysis and prediction of crimes in states using machine learning models. It focuses on creating a model that can help to detect the number of crimes by its type in a particular state.

The rest of the paper is organized as follows: Section II demonstrates the literature survey of different crime prediction models. The section III presents crime prediction and analysis using ML classifiers. The result analysis is demonstrated in section IV finally conclusion is presented in section V.

II. LITERATURE SURVEY

Mohammed Boukabous, Mostafa Azizi et. al., [1] presented a “Crime prediction using a hybrid sentiment analysis approach based on the bidirectional encoder representations from transformers (BERT). In this paper, a hybrid approach is used that combines both lexicon-based and deep learning, with BERT as the DL model. The authors employed the lexicon-based approach to label our Twitter dataset with a set of normal and crime-related lexicons; then, we used the obtained labeled dataset to train this BERT model.

Ricardo Francisco Reier Forradellas et. al., [2] Applied Machine Learning in Social Sciences: Neural Networks and Crime Prediction. This study proposes a crime prediction model according to communes. For this, the Python programming language is used, due to its versatility and wide availability of libraries oriented to Machine Learning. For prediction, it is necessary to provide the model with the information corresponding to the predictive characteristics (predict); these characteristics being according to the developed neural network model: year, month, day, time zone, commune, and type of crime.

Neil Shah, Nandish Bhagat et. al., [3] presented a crime forecasting approach using ML and computer vision methods for the prediction and prevention of crime. In this paper, they described the results of certain cases where such approaches were used, and which are motivated to pursue further research in this field The sole purpose of this study is to determine how a combination of ML and computer vision can be used by law agencies or authorities to detect, prevent, and solve crimes at a much more accurate and faster rate.

Panagiotis Stalidis et. al., [4] Examining Deep Learning Architectures for Crime Classification and Prediction. In this paper, a detailed study on crime classification and prediction using deep learning architectures is presented. The author examine the effectiveness of deep learning algorithms in this

domain and provide recommendations for designing and training deep learning systems for predicting crime areas, using open data from police reports.

Paweł Cichosz, et. al., [5] has discussed about the Urban Crime Risk Prediction Using Point of Interest (POI) Data. This article demonstrates how they can be combined with ML algorithms to create crime prediction models for urban areas. Selected POI layers from Open Street Map are used to derive attributes describing micro-areas, which are -3assigned crime risk classes based on police crime records.

Shaobing Wu et. al., [6] presented Crime Prediction Using Data Mining and Machine Learning. The aim of the study is to show the pattern and rate of crime based on the data collected and to show the relationships that exist among the various crime types and crime Variable. By introducing formula and methods of Bayesian network, random tree and neural network in machine learning and big data, to analyze the crime rules from the collected data.

Sarah Brayne et. al., [7] presented Technologies of Crime Prediction: The Reception of Algorithms in Policing and Criminal Courts. We draw on ethnographic fieldwork conducted within a large urban police department and a midsized criminal court to assess the impact of predictive technologies at different stages of the criminal justice process. The author studied how predictive algorithms are used, documenting similar processes of professional resistance among law enforcement and legal professionals.

Shraddha Ramdas Bandekar et. al., [8] presented Design and analysis of Machine Learning Algorithms for the reduction of crime rates in India. This research work focuses on how machine learning algorithms can be designed and analyzed to reduce crime rates in India. By the means of machine learning techniques, determining the pattern relations among huge set of data has become easier. This research mainly depends on providing a prediction on crime type that might occur based on the location where it has already taken place. Machine learning has been used to develop a model by the use of training data set that have gone through the process of data cleaning and transformation.

Chao Huang, Junbo Zhang et. al., [9] presented DeepCrime: Attentive Hierarchical Recurrent Networks for Crime Prediction. In this paper, we develop a new crime prediction framework—*DeepCrime*, a deep neural network architecture that uncovers dynamic crime patterns and carefully explores the evolving inter-dependencies between crimes and other ubiquitous data in urban space. Extensive experiments on real-world datasets demonstrate the superiority of this framework over many competitive baselines across various settings.

Shubham Agarwa et. al., [10] presented Crime Prediction based on Statistical Models. Based on the previous year (s) crime details in Indian states, the author present statistical models *viz.* Weighted Moving Average, Functional Coefficient Regression and Arithmetic-Geometric Progression based prediction of the crime in coming years.

III. CRIME PREDICTION AND ANALYSIS USING ML MODELS

This data set includes criminal offenses in the City and County for the previous five calendar years plus the current year to date. The data is based on the National Incident Based Reporting System (NIBRS) which includes all victims of person crimes and all crimes within an incident. Machine learning agents work with data and employ different techniques to find patterns in data making it very useful for predictive analysis.

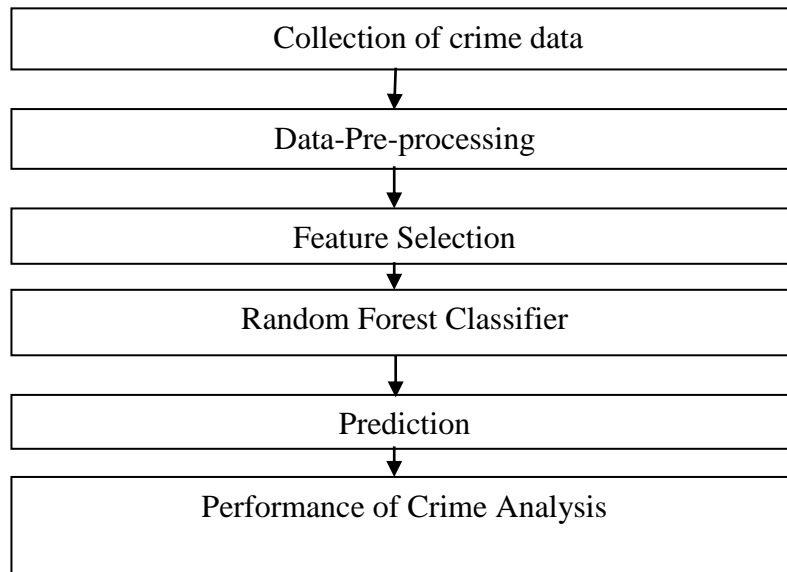


Fig 1: FLOW CHART OF PRESENTED CRIME PREDICTION AND ANALYSIS MODEL

This data reflects reported incidents of crime that have occurred in the City of Chicago during a specific time period. Data is extracted from the Chicago Police Department's CLEAR (Citizen Law Enforcement Analysis and Reporting) system. This data set contains 90 days of information and the most recent data available is seven days prior to today's date. Here a Data set is analyzed by using a object based approach with in the Python programming Language. Building the model will be done using better algorithm depending upon the accuracy.

Data processing is generally the collection and manipulation of items of data to produce meaningful information. In this sense it can be considered a subset of information processing, the change of information in any manner detectable by an observer. This simple observation led to the idea that it would be useful to use only some selected trees in classification. The selection of trees was based on their performance on similar instances, but without success. The step toward the analysis is preprocessing. If the data is dirty, it will generate incorrect visualizations, hence leading towards the incorrect conclusions. The crime data collected also has some level of dirtiness.

It contains some null values, inconsistent date formats, and some outliers.

Feature selection is also known as variable selection. It is the automatic selection of attributes in data that are most relevant to the predictive modeling problem. Random split selection does better than bagging; introduction of random noise into the outputs also does better; but none of these do as well as adaboost by adaptive reweighting (arcing) of the training set.

The importance of each feature variable in a training subset refers to the portion of the gain ratio of the variable compared with the total feature variables. The value of all feature variables are sorted in descending order and the top variable values are selected. Thus the number of dimensions of the dataset is reduced from feature variables in each sample to the number of the selected feature variables.

To make a prediction on a new instance, a random forest must aggregate the predictions from its set of decision trees. This aggregation is done differently for classification and regression. The data

gets split into many subsets and it compares the train and test data to and the best one. This process gets repeatedly on each subset and

find out the best prediction on each mapping. According to this process, each subset has its own predicted class. And comparing all the predicted class of its produce the final prediction based on training data.

Here in the proposed system we use the random forest algorithm in order to get good results and better accuracy when compared to the above or the existing algorithms. We use random forest for accuracy. Random forest is a most popular and powerful supervised machine learning algorithm capable of performing both classification, regression tasks, that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees.

Random decision forests correct for decision trees habit of over fitting to their training set .The data sets considered are rainfall, perception, production, temperature to construct random forest, a collection of decision trees by considering two-third of the records in the datasets. These decision trees are applied on the remaining records for accurate classification.

IV. RESULT ANALYSIS

The presented system analysis is analyzed by following steps. Python programming Language provides libraries which also focus on the implementation of data analysis and in representing the data in different visualizations. When you install Python directly from its official website, it does not include Jupyter Notebook in its standard library. In this case, you need to install Jupyter Notebook using the pip. The process is as follows: Open a new command prompt (Windows) or terminal (Mac/Linux). Execute the following command to install Jupyter Notebook.

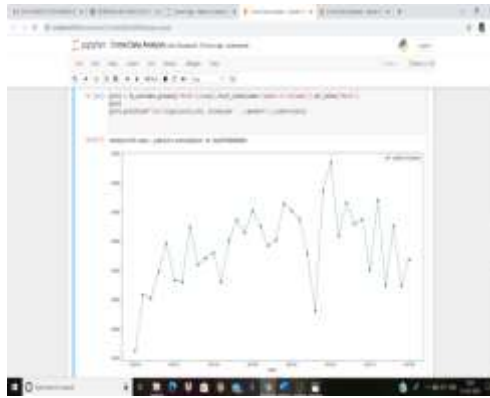


Fig. 2: Python3 -m Pip Install Jupyter Window

After you have installed the Jupyter Notebook on your computer, you are ready to run the notebook server.Keep the terminal open as it is. It will then open the default web browser with the URL mentioned in the command prompt or terminal. When the notebook opens in your browser, you will see the Notebook Homepage as shown. This will list the notebook files and subdirectories in the directory where the notebook server was started. After executing the code, we get the resultant output. The following screenshots shows the resultant output.



Fig. 3: SCREENSHOTS OF RESULTANT OUTPUTS



The Fig. 4 represents the crimes per month.

The Fig. 5 represents the bar graph of crime wise arrests.

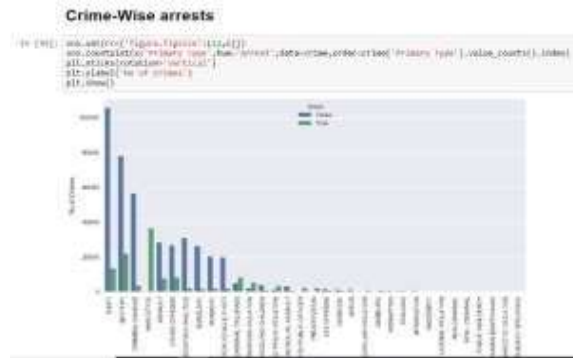


Fig. 5: CRIME WISE ARRESTS BAR GRAPH

The Fig. 6 represents the number of crimes occurred per an hour i.e. HEAT MAT of crime.

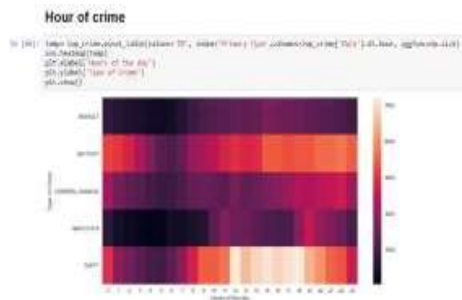


Fig. 6: CRIMES HEAT MAP (HOUR OF CRIME OCCURRED)

V. CONCLUSION

With the help of machine learning technology, it has become easy to find out relation and patterns among various data's. The work in this project mainly revolves around predicting the type of crime which may happen if we know the location of where it has occurred. Using the concept of machine learning we have built a model using training data set that have undergone data cleaning and data transformation. The model predicts the type of crime with accuracy of 0.789. Data visualization helps in analysis of data set. The graphs include bar, pie, line and scatter graphs each having its own characteristics. We generated many graphs and found interesting statistics that

helped in understanding Chicago crimes datasets that can help in capturing the factors that can help in keeping society safe. The tool we have developed provides a framework for visualizing the crime networks and analyzing them by various machine learning algorithms using the Google Maps. The project helps the crime analysts to analyze these crime networks by means of various interactive visualizations. The interactive and visual feature applications will be helpful in reporting and discovering the crime patterns. Many classification models can be considered and compared in the analysis. It is evident that law enforcing agencies can take a great advantage of using machine learning algorithms to fight against the crimes and saving humanity. For better results, we need to update data as early as possible by using current trends such as web and Apps.

IV. REFERNCES

1. Mohammed Boukabous, Mostafa Azizi, "Crime prediction using a hybrid sentiment analysis approach based on the bidirectional encoder representations from transformers", Indonesian Journal of Electrical Engineering and Computer Science Vol. 25, No. 2, February 2022, pp. 1131-1139 ISSN: 2502-4752, DOI: 10.11591/ijeecs.v25.i2.pp1131-1139
2. Ricardo Francisco Reier Forradellas, Sergio Luis Nández Alonso, Javier Jorge-Vazquez and Marcela Laura Rodriguez, "Applied Machine Learning in Social Sciences: Neural Networks and Crime Prediction", Soc. Sci. 2021, 10, 4. <https://doi.org/10.3390/socsci10010004>
3. Neil Shah, 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 (2021) 4:9, <https://doi.org/10.1186/s42492-021-00075-z>
4. Panagiotis Stalidis, Theodoros Semertzidis and Petros Daras, "Examining Deep Learning Architectures for Crime Classification and Prediction", Forecasting 2021, 3, 741-762. <https://doi.org/10.3390/forecast3040046>
5. Paweł Cichosz, "Urban Crime Risk Prediction Using Point of Interest Data", ISPRS Int. J. Geo-Inf. 2020, 9, 459; doi:10.3390/ijgi9070459
6. Shaobing Wu, Changmei Wang, Haoshun Cao, and Xueming Jia, "Crime Prediction Using Data Mining and Machine Learning", Springer Nature Switzerland AG 2020 AISC 905, pp. 360-375, 2020, https://doi.org/10.1007/978-3-030-14680-1_40
7. Sarah Brayne and Ange`le Christin, "Technologies of Crime Prediction: The Reception of Algorithms in Policing and Criminal Courts", Advance Access Publication Date: 5 March 2020 Article, doi: 10.1093/socpro/spaa004
8. Shraddha Ramdas Bandekar, C. Vijayalakshmi, " Design and analysis of Machine Learning Algorithms for the reduction of crime rates in India", 2020, Elsevier B.V., doi: 10.1016/j.procs.2020.05.018
9. Chao Huang, Junbo Zhang, Yu Zheng, Nitesh V. Chawla, "DeepCrime: Attentive Hierarchical Recurrent Networks for Crime Prediction", 2018 Association for Computing Machinery, doi.org/10.1145/3269206.3271793
10. Shubham Agarwa, Lavish Yadav, and Manish K Thakur, "Crime Prediction based on Statistical Models," Proceedings of 2018 Eleventh International Conference on Contemporary Computing (IC3), 2-4 August, 2018, Noida, India.
11. Alkesh Bharati, Dr Sarvanaguru RA.K, " Crime Prediction and Analysis Using Machine Learning", International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 05 Issue: 09 | Sep 2018 www.irjet
12. Nelson Baloian1, Enrique Bassaletti, Mario Fernández, Oscar Figueroa, Pablo Fuentes, "Crime Prediction using Patterns and Context", Proceedings of the 2017 IEEE 21st International Conference on Computer Supported Cooperative Work in Design
13. Emmanuel Ahishakiye, Elisha Opiyo Omulo, Danison Taremwa, Ivan Niyonzima, " Crime Prediction Using Decision Tree (J48) Classification Algorithm", International Journal of Computer and Information Technology (ISSN: 2279 – 0764) Volume 06 – Issue 03, May 2017

14. Lawrence McClendon and Natarajan Meghanathan, "Using Machine Learning Algorithms To Analyze Crime Data", *Machine Learning and Applications: An International Journal (MLAIJ)* Vol.2, No.1, March 2015
15. Shamsuddin, N. H. M., Ali, N. A., & Alwee, R, " An overview on crime prediction methods", In *Student Project Conference (ICT-ISPC), 2017 6th ICT International* (pp. 1-5). IEEE.
16. Jeyaselvi, M., M. Sathya, S. Suchitra, S. Jafar Ali Ibrahim, and N. S. Kalyan Chakravarthy. "SVM-Based Cloning and Jamming Attack Detection in IoT Sensor Networks." *Advances in Information Communication Technology and Computing*, pp. 461-471. Springer, Singapore, 2022. https://link.springer.com/chapter/10.1007/978-981-19-0619-0_41
17. Thangamani, M., and S. Jafar Ali Ibrahim. "Ensemble Based Fuzzy with Particle Swarm Optimization Based Weighted Clustering (Efpso-Wc) and Gene Ontology for Microarray Gene Expression." In *Proceedings of the 2018 International Conference on Digital Medicine and Image Processing*, pp. 48-55. 2018. <https://dl.acm.org/doi/abs/10.1145/3299852.3299866>