

Crime Behaviour Analysis Using Machine Learning Algorithm

Packiya Lakshmi.P¹, Monisha alias selvi.S², Radha.E³, Priyanka.N⁴
^{1,2,3,4}CSE, PSRR College of Engineering, Sivakasi, Virudhunagar

Abstract— A criminal analysis is the method approach for analyzing and identifying patterns and trends in the crimes. With the increasing origin of the computerized system, crime data analysts can help Law enforcement officers to speed up the process of solving crimes. Using the concept of data mining, the system can analyze previously unknown useful information from unstructured data. Predictive policy means, using analytical and predictive techniques to identify the criminal and it is pretty and much effective in doing the same. Because of the increased crime rate over the years, the system will have to handle a huge amount of crime data stored in a warehouse which would be very difficult to analyze manually also nowadays criminals are becoming technologically advance, so there is a need to use the advance technologies to keep the police ahead of them. In our system, the crime dataset of India contains a record of serious fraud of property in all states and we will apply k-means clustering. we can implement machine learning algorithms. Finally, the experimental results show the accuracy, precision, recall, and f1-score.

I.INTRODUCTION

A crime rate has become a topic of major concern certainly limiting the development of good governance which will be increasing day by day. Crimes are either systematic or random otherwise crime cannot be analyzed. When the crimes like robbery, firebombing, etc. Have decreased, crimes like murder, sex abuse, gang rape, etc. We cannot analyze the victims of crime but can analyze the place, where the crime occurred or happened before. So, there is a need for effective analyzing tools which can be analyzed crime data efficiently and quickly, to give some useful crime patterns. Predictive policy means, that using analytical and predictive techniques to identify the criminal and it is pretty much effective. Because of the, increased crime rate over the years, we will have to handle a huge amount of crime data stored in warehouses which would be difficult to analyze manually, also nowadays criminals are becoming technologically advance, so there is a need to use the advance

technologies, to keep the police ahead of them. The crime rates are increasing continuously and crime patterns are constantly changing. As a result, the behaviors in crime patterns are difficult to explain. This paper illustrates, how social development may lead to crime prevention. The aim is to provide a comprehensive review of theory, and research concerning the prevention of crime in society and to implement different data analysis algorithms which address the connections between crimes and their patterns. Crime is one of the major issues, continuing to grow in intensity and complexity. In recent years, crime is one of the social problems influencing the nature of life and economic development in a community. Crime can be divided into a few types, such as crime against properties (theft, burglary, and robbery) and crime of aggression (homicides, assaults, and rape). The availability of information technologies has enabled law enforcement to collect detailed information on crime data. With the increasing numbers of crimes nowadays, crime analysis is needed which comprises measures and procedures that intend to reduce the risk of crime. Crime analysis can be done through both quantitative & qualitative methods. Qualitative approaches to predicting crime such as scenario writing or environmental scanning are valuable, in identifying the future of criminal activity. Meanwhile, the quantitative method is used to predict crime rates in the future specifically. Moreover, crime analysis is a practical approach to analyzing and identifying the pattern of crimes. Crime analysis is part of crime prevention, which has the task of discovering and detecting crimes and their relation with criminals.

A. Supervised Algorithm

The decision tree is the supervised learning technique that can be used for both classification and regression problems, but it is mostly preferred for solving the classification problem. It is a tree-structured classifier where the internal nodes represent the features of datasets and the branches represent the decision rules, each of the leaf nodes

represents the outcome. In a Decision tree, there are two nodes which are the Decision Nodes and Leaf Nodes. A decision node is used for making any decisions and also has multiple branches, whereas the leaf nodes are the outputs of those decisions and do not contain any further branches. It is a graphic representation of all the possible solutions to problems or decisions based on the given conditions. It is also called a decision tree because it is similar to a tree, it starts with the root nodes, which expand on the further branches and construct a tree-like structure.

B. Unsupervised Algorithm

The algorithm, K-Means clustering is an unsupervised learning algorithm, and it is used to solve a clustering problem in machine learning and data science. In this topic, we will learn what is a K-means clustering algorithm, how this algorithm works along with the Python implementation of K-means clustering. K-Means Clustering is an Unsupervised Learning algorithm, which groups the unlabeled datasets into different clusters. Here K defines the number of pre-defined clusters, the clusters need to be created in the process, K=2 there will be two clusters, and K=3 there will be three clusters, and so on.

II. LITERATURE REVIEW

1) Crime Pattern Detection, Analysis & Prediction using Machine Learning, Rohit Patil¹, Muzamil Kacchi², Pranali Gavali³, Komal Pimparia,²⁰²⁰:

A criminal analysis is the method approach for analyzing and identifying patterns and trends in crimes. With the increasing origin of computerized systems, crime data analysts can help Law enforcement officers to speed up the process of solving crimes. Using the concept of data mining, the systems can be analyzed previously unknown, useful information from unstructured data. Predictive policing means, using analytical and predictive techniques, to identify the criminal and it is pretty much effective in doing the same. Because of the increased crime rate over the years, the system will have to handle a huge amount of crime data stored in a warehouse which would be very difficult to be analyzed manually also now a day's, criminals are becoming technologically advanced, so there is need to use the advanced technologies to keep police ahead of them. In this paper, the main focus is on the

reviews of algorithms and techniques used for identifying criminals.

2) Crime Pattern Detection, Analysis & Prediction, Sunil Yadav, Meet Timbadia, Ajit Yadav, Rohit Vishwakarma and Nikhilesh Yadav,²⁰²⁰

Crime is social irritation and costs our society deeply in several ways. Any research that can be helped in solving the crimes quickly will pay for itself. About 10 percent of the criminals committed about 50 percent of the crimes. The system is trained to feed the previous year's records of crimes taken from the legitimated online portal of India, which lists various crimes such as murder, abduction, dacoits, robbery, burglary, rape, kidnapping, and other such crimes. As per the data of Indian statistics, which gives the data of various crimes of the past 14 years (2001 to 2014) the regression model is created and the crime rate for the following years in various states can be predicted. The crime rates are accelerated continuously and the crime patterns are constantly changing. As a result, the behaviors in the crime patterns are difficult to explain. This paper illustrates, how social development may lead to crime prevention. We are using supervised, semi-supervised, and unsupervised learning techniques on the criminal records for knowledge discovery and to help in increasing the predictive accuracies of the crimes.

III. METHODS OF CRIMINAL BEHAVIOR

A. Data Selection

In our process, the crimes in India dataset is used. The data selection is the process of analyzing the crime. This dataset contains complete information about the various aspects of the crimes happened in India since 2001. Many factors can be analyzed from this dataset. In our process, we have to take the serious fraud dataset.

B. Data Filtering

Data pre-processing are the processes of removing unwanted data from the dataset. Pre-processing data transformation operations are used to transform the dataset into a structure suitable for machine learning. This step also includes cleaning the dataset by removing irrelevant or corrupted data that can affect the accuracy of the dataset, which makes it more efficient. Missing data removal Encoding Categorical data. Missing data removal: In this process, the null values such as missing values and

Nan values are replaced by 0. Missing and duplicate values were removed and data was cleaned of any abnormalities. Encoding Categorical data. Categorical data is defined as variables with a finite set of label values. Most of the machine learning algorithms require numerical input and output variables.

C. Clustering

K-means clustering is the method of vector quantization and originally it is from signal processing, that aims to partition the n observations into the k clusters in which, each observation belongs to the clusters with the nearest means, serving as a prototype of the clusters. K-means clustering is one of the types of unsupervised learning, which is used when you have unlabeled data (i.e., data without the defined categories or groups). The goal of this algorithm is to find the groups in the data, with the number of groups represented by the variable K. Data points are clustered by the base of feature similarity.

D. Identification of crime behavior

Machine learning is one of the methods of data analysis, and that automates the analytical model buildings. It is a branch of artificial intelligence and it is based on the idea, of a system that can learn from data, identify patterns and make decisions with minimal human intervention. The decision Tree is simple to understand, and interpret. Trees can be visualized. it requires little data preparation. In other techniques that often required data normalization, dummy variables need to be created, and blank values are removed. Note that, however, this module does not support missing values.

E. Result Generation

The Final Result will be get generated based on the overall classification and prediction. The performance of this proposed approach is evaluated by using some measures like,

Accuracy: The accuracy of the classifier refers to the ability of the classifier. It predicts the class label correctly and the accuracy of the predictor refers to how well a given predictor can guess the value of a predicted attribute for new data.

$$AC = (TP+TN) / (TP+TN+FP+FN)$$

Precision: Precision is defined, as the number of true positives are divided by the number of true positives plus the number of false positives.

$$Precision = TP / (TP+FP)$$

Recall: Recall is the number of correct results divided by the number of results that should have been returned. In binary classification, the recall is called sensitivity. It can be viewed as the probability that the relevant document is retrieved by the query.

$$Recall = TP / (TP+FN)$$

IV. CONCLUSION

With the advancement in technologies that are coming recently in data science and especially in machine learning, it becomes easy and efficient to discover patterns and information which might get useful for future prediction in crime analysis and behavior segmentation. Clustering is the process of grouping similarities in a dataset so that it can get useful for analysis, discovering patterns, and prediction.

APPENDIX

The machine learning algorithm implemented in this study(Logistic Regression, SVM, Naive Bayes, KNN, Decision tree, MLP, Random Forest, XGBoost)K-means and Decision tree.

REFERENCE

- [1] Suhong Kim; Param Joshi; Parminder Singh Kalsi; Pooya Taheri. Crime Analysis Through Machine Learning, 2018 IEEE 9th Annual Information Electronics and Mobile Communication Conference (IEMCON)
- [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] 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 ieeewic/acm international conference on (pp. 41-44). IEEE
- [4] 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.
- [5] A.Singh, A. Rana, J. Ranjan, "Proposed analytical customer-centric model for an automobile industry", in *International Journal of Data Mining, Modelling, and Management*, Vol. 7, Issue 4, pp 314-330 (2015).
- [6] N. Tyagi, A. Rana, "Fuel your growth with integration: Hybrid cloud computing", in *International Journal of Applied Engineering Research*, Vol. 10, Issue 13, pp 32761-32762 (2015).
- [7] A.Singh, A. Rana, J. Ranjan, "Data mining techniques and its effect in customer relationship management", in *International Journal of Data Analysis Techniques and Strategies*, Vol. 7, Issue 4, pp 406-427 (2015).
- [8] G. Dubey, A. Rana, N. K. Shukla, "User reviews data analysis using opinion mining on the web", in 2015 1st International Conference on Futuristic Trends in Computational Analysis and Knowledge Management, ABLAZE 2015, pp 603- 612 (2015).
- [9] Alkesh Bharati, Dr. Sarvanaguru RA. K, Crime Prediction and Analysis Using Machine Learning, *International Research Journal of Engineering and Technology (IRJET)*, Volume: 05 Issue: 09 | Sep 2018.
- [10] Arpita Nagpal, Aman Jatain, Deepti Gaur. Review based on Data Clustering Algorithms, *Proceedings of 2013 IEEE International Conference on Information and Communication Technologies (ICT 2013)*.
- [11] Rasoul Kiani, Siamak Mahdavi, Amin Keshavarzi. Analysis and Prediction of Crimes by Clustering and Classification, (IJARAI) *International Journal of Advanced Research in Artificial Intelligence*, vol 4, No.8, 2015
- [12] 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.