



International Journal of Marketing Management

ISSN 2454 - 5007



www.ijmm.net

Email ID: editor@ijmm.net , ijmm.editor9@gmail.com

This article can be downloaded from <http://www.ijmm.net/currentissue.php>

CRIME DATA ANALYSIS AND PREDICTION USING DECISION TREE

CHILUKURI VAMSI CHENNA REDDY¹, KUKKALA RAJESH², SRIRAMOJU SHIVA PRASAD³, N.ROHITH⁴

ABSTRACT:

Using crime analysis and prediction is a methodical way to find the crime. This algorithm is able to identify areas with a high likelihood of crime and forecast where it will occur. We may take previously undiscovered, valuable information from an unstructured data set using the idea of data mining. Using the current datasets, the extraction of additional information is projected. Crime is a perilous societal issue that exists everywhere. Crimes have an impact on a country's reputation, economic prosperity, and quality of life. There is a need for cutting-edge methods and fresh ideas for enhancing crime analytics in order to safeguard communities from criminal activity. We provide a system that can analyze, find, and forecast different crime probabilities in a given area. Using different data mining approaches, this study describes several forms of criminal analysis and crime prediction.

Keywords: Data mining, Act379, Act302, and crime.

1.INTRODUCTION:

Criminal activity has a negative impact on a society's standard of living and ability to prosper economically [1]. It is regarded as a crucial aspect in determining whether or not individuals relocate to a new city and which locations they should avoid while traveling [2]. The repercussions of crime on society include unease that undermines the sense of community among the populace, the dissolution of social ties as a result of habitually avoiding specific locations, a reluctance to go out at night, and damage to property. the neighborhood's reputation. People may be discouraged from visiting a neighborhood and may decide to leave if they believe it to be one with a high crime rate. Damage is done to the economy as a result. Because of the increased

demand for police, courts, and correctional facilities as well as the intangible costs associated with psychological trauma and a lower standard of living for crime victims, crime has a negative impact on the economy by placing a financial burden on taxpayers and governments. In many nations nowadays, a large number of crimes are creating a lot of issues. In fact, researchers are examining criminal activity and criminal behavior in order to identify criminal trends and comprehend the nature of crime. Dealing with crime data is highly difficult since it develops in bulk quickly, which may lead to issues with storage and processing. Due to the irregularity and insufficiency

1,2, 3, 4 UG Student, Department of CSE, NARSIMHA REDDY ENGINEERING COLLEGE,
Maisammaguda, Kompally, Secunderabad, Telangana India. 500100

of various types of data, challenges emerge on how to choose precise procedures for data analysis. These problems drive scientists to study these types of data in order to improve crime data analysis. Dealing with crime data is highly difficult since it develops in bulk quickly, which may lead to issues with storage and processing. Due to the irregularity and insufficiency of various types of data, challenges emerge on how to choose precise procedures for data analysis. Various problems push researchers to study these types of data to improve crime data analysis [3]. The goal of this study is to use an appropriate machine learning algorithm on crime data to forecast if a county would have low, medium, or high violent crime rates.

2. BOOK SUMMARY:

A. Crime analysis and criminology Criminology is a field that focuses on the scientific investigation of criminal activity and law enforcement, and it is a method intended to pinpoint the features of crime [4]. It is one of the most crucial areas where the use of data mining methods may provide significant results. Crime analysis is a branch of criminology that entails investigating and identifying crimes as well as their connections to offenders. Criminology is a suitable topic for the use of data mining methods because of the large number of crime datasets and the intricate linkages between various types of data. The first step in doing additional analysis is identifying criminal characteristics. A particularly helpful instrument that may aid and assist police forces is the information obtained via data mining techniques [5]. [6] contends that investigating crimes is a difficult undertaking that calls for human intellect and expertise. Data mining is a method that may help Law Enforcement Agencies with their issues with crime detection. Here, data mining is being used to attempt to incorporate decades of human experience into computer models. B.

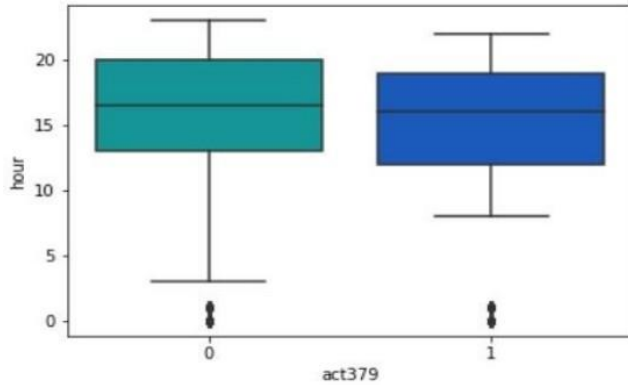
The Predictability of Crime The assumption that crime is predictable (in a statistical sense) is well supported by the facts, mostly because criminals like to operate in their comfort zones [7]. In other words, they often carry out the same or similar crimes that they have successfully accomplished in the past. Although not always the case, this happens often enough for these approaches to function rather effectively. Major theories of criminal behavior include the notions of routine activity, rational choice, and crime patterns. A blended theory is what results from the combination of several ideas.

C. Evaluation of Classification Methods techniques for classifying data that are often used in forecasts based on past information. Classification is a kind of supervised class prediction method. Under the condition that there are enough training instances, this approach can predict the label for classes. Support vector machines, k Nearest Neighbors, weighted voting, and artificial neural networks are just a few of the categorization techniques that are accessible. Each of these methods may be used on a dataset to find a group of models that predict an unidentified class label. The dataset is split into two sets for classification: the training set (dependent set) and the test set (independent set). The training set is used as the starting point for the machine learning method, while the test set serves as the final application for the predictive model. The categorization algorithms used in crime prediction are listed below.

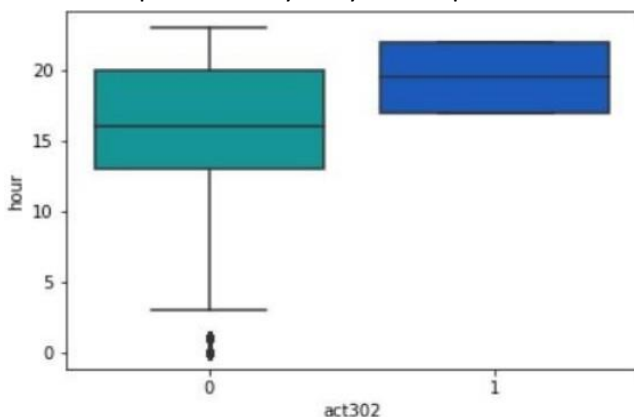
3. METHODOLOGY

Machine learning is the process of automatically finding significant patterns in data. It has developed into a standard tool in the last several decades for practically any activity requiring information extraction from massive data sets. A machine learning-based environment is all around us. Technology: Search engines learn to get us the best results (while putting up portable adverts), anti-spam software learns to screen our email communications, and credit card transactions

are safeguarded by a program that learns to



spot fraud. Intelligent personal assistance apps on smartphones learn to understand voice commands, much as digital cameras learn to identify faces. Systems for preventing accidents are incorporated into cars utilizing machine learning algorithms. Additionally, machine learning is often employed in fields of science including astronomy, bioinformatics, and medicine. All of these applications have the trait that, unlike more conventional uses of computers, a human programmer cannot offer an explicit, finely detailed specification of how such activities should be carried out in these circumstances owing to the intricacy of the patterns that need to be identified. Like intelligent creatures, many of our talents are developed or refined via experience-based learning (rather than following explicit instructions given to us). Giving programs the capacity to learn and adapt is the goal of machine learning technologies. A highly well-liked ensemble learning technique called Random Forests creates a lot of classifiers on the training data and then combines all of their results to provide the best predictions on the test data. Consequently, the Random Forests method is a variance-reducing technique that makes use of When making split decisions, unpredictability may assist prevent overfitting



on the training data. A random forests classifier is an ensemble classifier that combines the $h(x|x|x|x|k)$ family of classifiers. k is the number of trees selected from a model random vector, and each member of the family, $h(x|)$, is a classification tree. A parameter vector is also selected at random for each k . Each classification tree in the ensemble is generated using a distinct subset of $D_k(x,y)$ $D(x,y)$ of the training dataset if $D(x,y)$ signifies the training dataset..

Fig.3.1. Act of 302 results.

Fig.3.2. Act of 379 results.

CONCLUSION

This study focused on developing forecasting models for monthly crime rates by crime category. Due to a variety of variables, including a growth in poverty, unemployment, corruption, etc., crime rates in India are rising daily. In order to take the required actions to minimize crime, the suggested model is highly helpful for both the investigative agency and the police official. Through numerous interactive visualizations, the project aids in the understanding of these criminal networks. Future improvements to this study will include teaching bots how to identify crime hotspots using machine learning methods. Advanced machine learning concepts may be employed for improved prediction since they are comparable to data mining. For better prediction, the data privacy, dependability, and accuracy may be increased.

REFERANCES

- [1] Karan,. Bayraktar,. Gümü skaya, and Diabetes detection using neural networks on portable devices by A. Karlk, Expert Syst. Appl., vol. 39, no. 1, pp. 54–60, 2012.
- [2] R. Chitta, J. Zhou, Q. You, T. Sun, and F. Ma [2] Dipole: Healthcare diagnosis prediction using attention-based bidirectional recurrent neural networks, by J. Gao, in Proceedings of the 23rd ACM SIGKDD International Conference on

Knowledge Discovery and Data Mining, 2017, pp. 1903–1911.

[3] "Deep learning for healthcare decision making using emrs," in Proc. IEEE Int. Conf. Bioinformatics Biomed., 2014, pp. 556–559; Z. Liang, G. Zhang, J. X. Huang, and Q. V. Hu.

[4] Pattern Recognition and Machine Learning by C. M. Bishop. Springer, 2006; Berlin, Germany.

[5] "Curriculum learning," in Proc. 26th Annu. Int. Conf. Mach. Learn, 2009, pp. 41–48. Y. Bengio, J. Louradour, R. Collobert, and J. Weston.

[6] V. A. Convertino et al., "Utilization of sophisticated machine-learning methods for noninvasive hemorrhage monitoring," J. Trauma Acute Care Surgery, vol. 71, no. 1, pp. S25-S32, 2011.

[7] P. Gueth et al., "Patient-specific prompt-gamma dose monitoring in proton therapy: a machine learning-based approach," Phys. Med. Biol., vol. 58, no. 13, p. 4563, 2013.

Academic Emergency Medicine, vol. 19, no. 9, pp. 993–1003, 2012. [8] J. Labarère, P. Schuetz, B. Renaud, Y.-E. Claessens, W. Albrich, and B. Mueller, "Validation of a clinical prediction model for early admission to the intensive care unit of patients with pneumonia."

\Hospital readmission in general medicine patients: A prediction model, O. Hasan et al. J. General Internal Med., 2010, Volume 25, Number 3, Pages 211–219

[10] "Use of calendar and weather data to estimate walk-in attendance," by A. K. Diehl, M. D. Morris, and S. A. Mannis Southern Medical Journal, vol. 74, no. 6, 1981, pp. 709–712.

[11] "Predicting ICU admission in community-acquired pneumonia: Clinical scores and biomarkers," Expert Rev. Clin. Pharmacology, vol. 5, no. 4, pp. 445-458, 2012.

[12] H. Zhai et al., "Development and assessment of a machine learning based algorithm to predict the requirement of pediatric intensive care unit transfer for newly

admitted children," Resuscitation, vol. 85, no. 8, pp. 1065-1071, 2014.