
ISSN: 2321-2152
IJMECE
International Journal of modern
electronics and communication engineering
E-Mail
editor.ijmece@gmail.com
editor@ijmece.com
www.ijmece.com

ISSN2321-2152

www.ijmece.com

Vol 7, Issue 3, 2019

Crime Data Analysis and Prediction of Perpetrator Identity using Machine Learning Approach

Shahida Begum. K,Indira, Naveen Kumar. H

Asst. Professor, Asst. Professor, Asst. Professor

shahidahpt@gmail.com, indira.raj.06@gmail.com, navee2312@gmail.com

Department of CSE, Proudhadevaraya Institute of Technology, Abheraj Baldota Rd, Indiranagar, Hosapete, Karnataka-583225

ABSTRACT:

A methodical way to detect crimes is via crime analysis and prediction. With this technique, we can see where crimes are most likely to happen and identify areas with a high crime rate. Data mining is a technique for discovering new and valuable information in large amounts of unstructured data. Using the current datasets, additional information is anticipated to be extracted. All around the globe, people are dealing with the perilous issue of crime. The standard of living, economic development, and national prestige are all impacted by criminal activity. In order to safeguard society against criminal activity, people must find innovative ways to enhance crime analytics and use cutting-edge solutions. We provide a system that can examine a given area, identify potential crimes, and forecast the likelihood of specific crimes occurring there. Using a number of data mining approaches, this article describes several forms of criminal analysis and crime prediction.

Keywords: Crime, Act379, Act302, data mining.

1. INTRODUCTION:

Crimes are a common social problem affecting the quality of life and the economic growth of a society [1]. It is considered an essential factor that determines whether or not people move to a new city and what places should be avoided when they travel [2]. The effects of crime on society include feelings of fear that disrupt the population's sense of unity, the breakdown of social associations due to habitual avoidance of certain places, an unwillingness to go out at night and damage to the image of the community. The perception of a community as crime ridden can deter people from going there and induce residents to move away. This causes damage to the economy. Crime affects the economy by placing a financial burden on taxpayers and governments because of increased needs for police, courts and corrections facilities, as well as intangible costs including psychological trauma and reduced quality of life for crime victims. 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 [3]. The objective of this research is to apply suitable machine learning algorithm on crime data to predict the likelihood of a county having low, medium or high violent crimes.

2. LITERATURE SURVEY:

A. Criminology and Crime Analysis Criminology is an area that focuses on the scientific study of crime and criminal behavior and law enforcement and is a process that aims to identify crime characteristics [4]. It is one of the most important fields where the application of data mining techniques can produce important results. Crime analysis, a part of criminology, is a task that includes exploring and detecting crimes and their relationships with criminals. The high volume of crime datasets and also the complexity of relationships between these kinds of data have made criminology an appropriate field for applying data mining techniques. Identifying crime characteristics is the first step for developing further analysis. The knowledge that is gained from data mining approaches is a very useful tool which can help and support police forces [5]. According to [6], solving crimes is a complex task that requires human intelligence and experience and data mining is a technique that can assist Law Enforcement Agencies with crime detection problems. The idea here is to try to capture years of human experience into computer models via data mining. B. Why Crime Is Predictable There is a strong body of evidence to support the theory that crime is predictable (in the statistical sense) mainly because criminals tend to operate in their comfort zone [7]. That is, they tend to commit the type of crimes that they have committed successfully in the past, generally close to the same time and location. Although this is not universally true, it occurs with sufficient frequency to make these methods work reasonably well. There are major theories of criminal behavior, such as routine activity theory, rational choice theory, and crime pattern theory. These theories are consolidated into what is referred to as a blended theory.

C. Review of Classification Algorithms Classification algorithms that are mostly used in predictions basing on historical data. Classification is a class prediction technique, which is supervised in nature. This technique possesses the ability to predict the label for classes, provided that sufficient numbers of training examples are available. There is a variety of classification algorithms available, including Support vector machines, k Nearest Neighbors, weighted voting and Artificial Neural Networks. All these techniques can be applied to a dataset for discovering set of models to predict the unknown class label. In classification, the dataset is divided into two sets, namely the training set (dependent set) and a test set (independent set). The machine learning algorithm initially runs on the training set, than later the predicting model is applied on the test set. The following are classification algorithms that are used in crime predictions.

3. METHODOLOGY

The term machine learning refers to the automated detection of meaningful patterns in data. In the past couple of decades it has become a common tool in almost any task that requires information extraction from large data sets. We are surrounded by a machine learning based technology: search engines learn how to bring us the best results (while placing portable ads), anti-spam software learns to filter our email messages, and credit card transactions are secured by a software that learns how to detect frauds. Digital cameras learn to detect faces and intelligent personal assistance applications on smart-phones learn to recognize voice commands. Cars are equipped with accident prevention systems that are built using machine learning algorithms. Machine learning is also widely used in scientific applications such as bioinformatics, medicine, and astronomy. One common feature of all of these applications is that, in contrast to more traditional uses of computers, in these cases, due to the complexity of the patterns that need to be detected, a human programmer cannot provide an explicit, fine detailed specification of how such tasks should be executed. Taking example from intelligent beings, many of our skills are acquired or reined through learning from our experience (rather than following explicit instructions given to us). Machine learning tools are concerned with endowing programs with the ability to learn and adapt.

Random Forests is a very popular ensemble learning method which builds a number of classifiers on the training data and combines all their outputs to make the best predictions on the test data. Thus, the Random Forests algorithm is a variance minimizing algorithm that uses randomness when making split decision to help

ISSN2321-2152

www.ijmece.com

Vol 7, Issue 3, 2019

avoid overfitting on the training data. A random forests classifier is an ensemble classifier, which aggregates a family of classifiers h(x|θ1),h(x|θ2),..h(x|θk). Each member of the family, h(x|θ), is a classification tree and k is the number of trees chosen from a model random vector. Also, each θk is a randomly chosen parameter vector. If D(x,y) denotes the training dataset, each classification tree in the ensemble is built using a different subset Dθk(x,y) ⊂ D(x,y) of the training dataset.

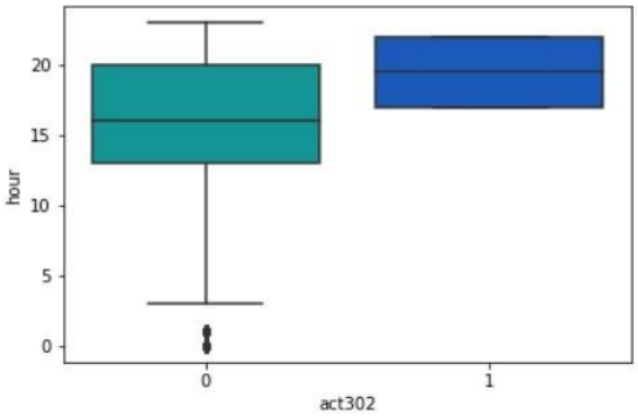

Fig.3.1. Act of 302 results.

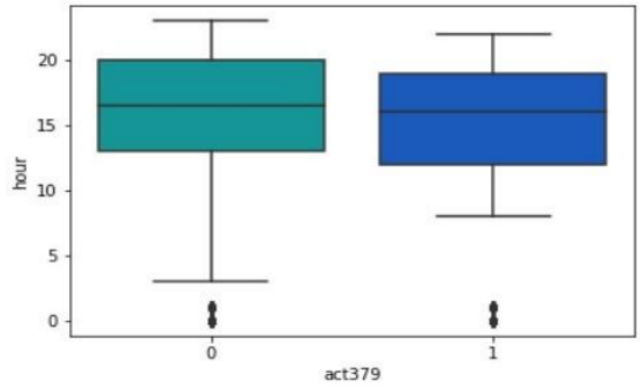

Fig.3.2.Act of 379 results.

CONCLUSION

Predictive models for monthly crime rates by crime category were the primary topic of this article. Many causes, including rising poverty, implementation, corruption, etc., are contributing to India's steadily rising crime rates. Law enforcement officials and investigative agencies alike may benefit greatly from the suggested approach as they work to decrease crime. The project's interactive visualisations aid in the understanding of these criminal networks. Improving this study in the future will include teaching bots to use machine learning to identify high-crime regions. Machine learning's similarities to data mining make it possible to use its more sophisticated concepts for more accurate prediction. For better prediction, there is room to increase data privacy, dependability, and accuracy.

REFERANCES

[1] O. Karan, C. Bayraktar, H. Gümü¸skaya, and B. Karlık, “Diagnosing diabetes using neural networks on small mobile devices,” Expert Syst. Appl., vol. 39, no. 1, pp. 54–60, 2012.

[2] F. Ma, R. Chitta, J. Zhou, Q. You, T. Sun, and J. Gao, “Dipole: Diagnosis prediction in healthcare via attention-based bidirectional recurrent neural networks,” in Proc. 23rd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2017, pp. 1903–1911.

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

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

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

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

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

[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," Academic Emergency Med., vol. 19, no. 9, pp. 993–1003, 2012.

[9] O. Hasan et al., "Hospital readmission in general medicine patients: A prediction model," J. General Internal Med., vol. 25, no. 3, pp. 211–219, 2010.

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

[11] J. F. Fernandez, O. Sibila, and M. I. Restrepo, "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., "Developing and evaluating a machine learning based algorithm to predict the need of pediatric intensive care unit transfer for newly hospitalized children," Resuscitation, vol. 85, no. 8, pp. 1065–1071, 2014.