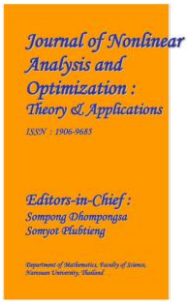
Journal of **N**onlinear **A**nalysis and **O**ptimization
Vol. 13, No. 1, (2022),
ISSN : **1906-9685**



ANALYSIS ON DETECTING AND PREDICTING VARIOUS CRIME PROBABILITY IN GIVEN REGION USING DECISION TREE ALGORITHM

[1]M Sampatha Kumar, [2]Gowry Sankar, [3] K Rajani, [4]Bandaru Bala Bhaskar
[1,2,3]Assistant Professor, [4]UG Student, [1,2,3,4]Department of Computer Science Engineering, Brilliant Grammer School Educational Society Group of Institutions-Integrated Campus, Hyderabad, India

ABSTRACT
Using crime analysis and prediction is a methodical way to find the crime. Our 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. With 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 tools and novel methods to enhance crime analytics in order to safeguard their communities from criminal activity. We propose a system which can analysis, detect, and predict various crime probability in given region. This paper explains various types of criminal analysis and crime prediction using several data mining techniques.
Keywords: Crime, Act379, Act302, data mining.

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 when travelling [2]. The repercussions of crime on society include unease that undermines a population's sense of community, the dissolution of social ties as a result of habitually avoiding specific locations, a reluctance to venture out at night, and harm to the town'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. 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 manydifferent countries. In fact, scientists are spendingtime 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 crimedata 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 crimedata 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.

LITERATURE SURVEY:

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 toidentify crime characteristics [4]. It is one of the most important fields where the application ofdata mining techniques can produce important results. Crime analysis, a part of criminology, is atask 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 thatis gained from data mining approaches is a very useful tool which can help and support policeforces [5]. According to [6], solving crimes is a complex task that requires human intelligence and experience and data mining is a techniquethat 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 ofcriminal 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 oftraining examples are available. There is a varietyof classification algorithms available, includingSupport 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.

METHODOLOGY

The term machine learning refers to the automated detection of meaningful patterns in data. In the past couple of decades it has becomea 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 ofthese 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 anexplicit, fine detailed specification of how such tasks should be executed. Taking example from intelligent beings, many of our skills are acquiredor reined through learning from our experience (rather than following explicit instructions given to us). Machine learning tools are concerned withendowing programs with the ability to learn and adapt.

Random Forests is a very popular ensemble learning method which builds a numberof classifiers on the training

data and combines all their outputs to make the best predictions on the test data. Thus, the Random Forests algorithmis a variance minimizing algorithm that uses randomness when making split decision to help 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 randomvector. 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.

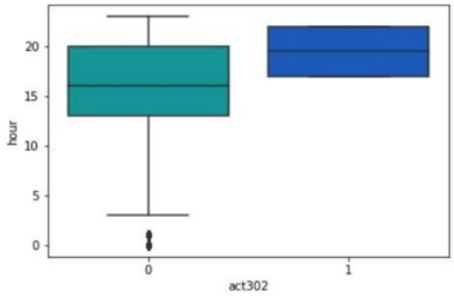


Fig.3.1. Act of 302 results.

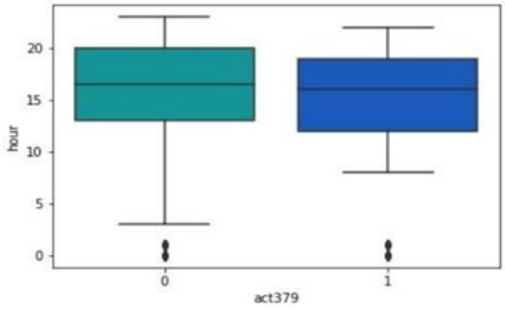


Fig.3.2.Act of 379 results.

CONCLUSION

In this paper focused on building predictive models for crime frequencies per crime type per month. The crime rates in India are increasing day by day due to many factors such as increase in poverty, implementation, corruption, etc. The proposed model is very useful for both the investigating agencies and the police official in taking necessary steps to reduce crime. The project helps the crime analysis to analysis these crime networks by means of various interactive visualization. Future enhancement of this research work on training bots to predict the crime prone areas by using machine learning techniques.Since, machine learning is similar to data mining advanced concept of machine learning can be used for better prediction. The data privacy, reliability, accuracy can beimproved for enhanced prediction.

REFERANCES

1. O. Karan, C. Bayraktar, H. Gümü¸skaya, and 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. 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.