International journal of basic and applied research

www.pragatipublication.com

ISSN 2249-3352 (P) 2278-0505 (E)

CRIME DATA ANALYSIS USING MACHINE LEARNING

[1]DR. T S GHOUSE BASHA,[2]V. CHITHRIKA,[3]S. NITYA SREE,[4]V. RAMYA,[5]S. VARSHA

[1]PROFESSOR, DEPARTMENT OF ECE, MALLA REDDY ENGINEERING COLLEGE FOR WOMEN, HYDERABAD

[2,3,4&5] UGSCHOLAR, DEPARTMENT OF ECE, MALLA REDDY ENGINEERING COLLEGE FOR WOMEN, HYDERABAD

ABSTRACT:

Data mining is the field containing procedures for finding designs or patterns in a huge dataset, it includes strategies at the convergence of machine learning and database framework. It can be applied to various fields like future healthcare, market basket analysis, education, manufacturing engineering, crime investigation etc. Among these, crime investigation is an interesting application to process crime characteristics to help the society for a better living. This paper survey various data mining techniques used in this domain. This study may be helpful in designing new strategies for crime prediction and analysis.

KEYWORDS: CRIME PREDICTION, DECISION TREES, LINEAR REGRESSION, K-MEANS

INTRODUCTION

Mining of data is a method of dealing with expansive data indexes to perceive outlines and set up an association to handle issues through information examination[10[22] The devices used, allow endeavors to accept future examples. Data mining is a procedure to analyze data from an informational collection to change it into a reasonable structure for additional utilization. It predicts future patterns and also enables the organization to make the learning driven decision. Generally utilized strategies for mining of data are artificial neural networks, decision tree, rule induction, nearest neighbor method and genetic algorithm[10][11][22][23]. They are applied in many fields. One such interesting application is crime investigation. A crime is an unlawful activity for which a man can be penalized by law. Crime against a person is called personal crime like murder, robbery, etc. Property crime means theft of property. Crime analysis is a law implementation task which includes an organized analysis that recognizes and determines the pattern of crime. Crime can be classified into different types but, in this, we focused on four types of crime i.e. Fraud detection, traffic violence,

ISSN 2249-3352 (P) 2278-0505 (E)

violent crime, web crime and sexual offense. The various techniques used for different crimes have been discussed with an introduction to the concerned crime.

2. TYPES OF CRIME

2.1. FRAUD DETECTION A fraud is misdirecting or taking unfair advantage of another. A fraud incorporates any act, exclusion, or concealment, including a breach of legal or equitable obligation or confide in, brings about the damage of other. Different types of frauds include check fraud, internet sale, insurance fraud and credit card fraud etc. Check fraud means issuance of a check when enough money is not present in account; internet sale means selling fake items; insurance fraud means fake insurance claimed for automobile damage, health care expenses and other; credit card fraudmeans obtaining credit card information from various means which is used for large amount of purchase without the permission of consumer.

2.2. VIOLENT CRIMEA violent crime is a crime in which a guilty party threatens to utilize compel upon a casualty. This entails the two crime of rough act called target, for example, killing or rape. Various sorts of this crime are as follows:

• Murdering of individual by other.

• Murder: Deliberate slaughtering of another individual.

• 1st degree murder: Used to allude to a deliberate slaughtering.

• 2nd degree murder: Used to allude to kill accidently in which the executioner shows, outrageous detachment to life of human.

2.3. TRAFFIC VIOLENCE Traffic violations happen when drivers damage laws that manage vehicle operation on roads and highways. The increasing number of cars in cities causes high volume of traffic, and implies that traffic violations become more critical which can cause severe destruction of property and more accidents that may endanger the lives of the people. To solve this problem and prevent such consequences, traffic violation detection systems are needed.

International journal of basic and applied research

www.pragatipublication.com

ISSN 2249-3352 (P) 2278-0505 (E)

2.4. SEXUAL ASSAULT Criminal attack is the risk or endeavor to physically strike a man, paying little respect to whether contact is really made, insofar as the casualty knows about the peril included. Level of Sexual assualts include:

• Simple Sexual Assault: It includes constraining a person to participate in any type of sexual action without unequivocal assent.

• Sexual Assault with a Weapon: It incorporates the utilization or danger of the utilization of a weapon or damage to an outsider.

• Aggravated Sexual Assault: It happen when the casualty is truly injured, mangled, fiercely beaten, or in threat of passing on because of a rape.

• Verbal assault: It is a sort of non-physical, oral ambush that outcomes in a passionate, mental, and additionally mental damage to the casualty, instead of physical substantial damage way.

2.5. CYBER CRIME Cyber-crime is the crime related to computer. It comprises of computer and a network for crime to occur .Offenses that are perpetrated against criminal process to hurt the victims by present day media transmission systems, for example, net and cell. Various types are web extortion, ATM misrepresentation, wire misrepresentation, document sharing and robbery, hacking, and so forth. Cyber-crime analysis is very important responsibility of law enforcement system in any country. It includes breakdown of protection, or harm to the PC framework properties, for example, documents, site pages or programming.

3.CRIME DATA ANALYSIS

Collection and analysis of crime related data are imperative to secure agencies. the use of a coherent methods to classify these data based on the rate and location of occurrences, detection of the hidden pattern among the committed crimes at different times, and prediction of their future relationship are the most important aspects that have to be addressed. One of the most popular approaches is hot spot analysis. Some of the most popular approaches used for this purpose of point pattern analysis and clustering/distances statistics. Another popular approach is the discovery of pattern or trends through various techniques from data mining, text mining and spatial analysis, and self-organizing maps.[1]An crime analysis tool should

International journal of basic and applied research

www.pragatipublication.com

ISSN 2249-3352 (P) 2278-0505 (E)

be able to identify crime patterns quickly and in an efficient manner for future crime pattern detection and action. The main purpose of crime analysis is:

- Extraction of crime pattern by crime analysis and based on available criminal information.
- Crime recognition [3].
- Problem of identifying techniques that can efficient and accurate.

3.1. CRIME ANALYSIS METHODOLOGY

The crime analysis methodologies are:-

- Data Collection
- Classification
- Pattern Identification
- Prediction
- Visualization

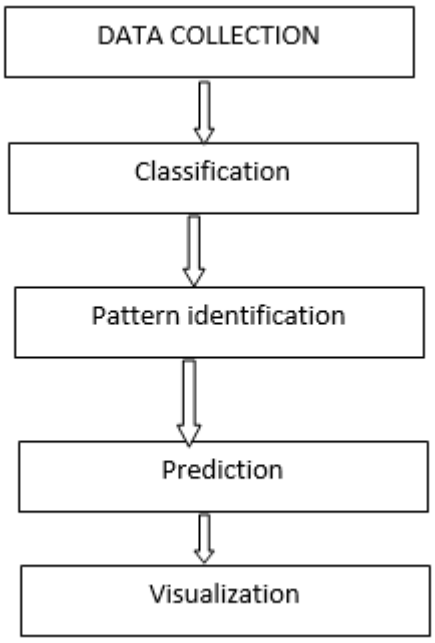

Fig:Crime analysis steps

International journal of basic and applied research

www.pragatipublication.com

ISSN 2249-3352 (P) 2278-0505 (E)

DATA COLLECTION The data collection is first methodology in crime analysis. Data's are collected from various different websites, news sites and blogs. The collected data is stored into database for further process. This is unstructured data and it is object oriented programming which is easy to use and flexible. Crime data is an unstructured data since no of field, content, and size of the document can differ from one document to another the better option is to have a schema less database. Also the absence of joins reduces the complexity. Other benefits of using an unstructured database are that:

- Large volume of structured, semi-structured, and unstructured data.
- Object-Oriented programming that is easy to use and flexible.

CLASSIFICATION In this step use Naive Bayes Algorithm which is supervised learning method. Naive Bayes classifier is a probabilistic classifier which when given an input gives a probability distribution of set of all classes rather than providing a single output. One of the main advantages of the Naïve bayes Classifier is simple, and coverage quicker than logistic regression [2].Compare to other algorithm like SVM (Support VectorUsing naïve Bays algorithm is create a model by training crime data related to vandalism, murder, robbery, burglary, sex abuse, gang rape, etc. Naive Bayes is that works well for small amount of training to calculate the classification parameter. Estimating probability sometimes while checking a probability P(A) * P(B/D) *P(C/D) * P(E/D) where P(C/D)=0[2].

Pattern Identification A third step is the pattern identification where we have identify trends and patterns in crime. For finding crime pattern that occurs frequently we are using apriori algorithm. Apriori can be used to determine association rule which highlight general trends in the database. By using pattern identification it will helps to the police officials in an effective manner and avoid the crime occurrences in particular place by providing security, CCTV, fixing alarms etc.

Crime Prediction The second Approach is predicting the crime type that might occur in a specific location within particular time. To predict an expected crime type is provide four related features of the crime. The features are: occurrence month, the occurrence day of the week, the occurrences time and the crime location. Prediction is stating probability of an event in future period time. A Classification approach is used crime prediction in data

ISSN 2249-3352 (P) 2278-0505 (E)

miningn[1]classify areas into hotspots and cold spots and to predictive an area will be a hotspot for residential burglary. Variety of classification techniques are used for predicting the crime:-[1]

- K-Nearest Neighbor (k-NN)
- Decision trees (J48)
- Support Vector Machine (SVM)
- Neural Networks
- Naïve Bayes and ensemble learning

Linear Regression methods are also used for predicting the crime prediction. Based on the crime probability. The formula for a regression line is Y=aX + b where, Y is the predicted score, b is the slope of the line, and A is the Y intercept. b = r sx\syAnd the intercept (A) can be calculated as A=MY –bMX.

Some Theories are used to predicting the crimes are:

- Integrated theory
- Biological theory
- Psychological theory
- Sociological theory
- Conflict theory
- Victimization theory
- Choice theory Visualization

The crime prone area can be graphically reoresented using a heat amp which indicates level of activity,dark colour indicates low activity and brighter colour indicates the high activity. Advantages of using heat map are [2]. • Numerical and category based color images •

International journal of basic and applied research

www.pragatipublication.com

ISSN 2249-3352 (P) 2278-0505 (E)

Gradient color range •Analyze only the data we want • Out of range data is automatically discarded.

4. ALGORITHMS Our experiment choose the algorithm are

- Instance based algorithm
- Decision tree
- Linear regression
- K-means algorithm

INSTANCE BASED ALGORITHM-The instance based algorithm is also called as tge machine based learning is a family of learning algorithm that, instead of performing explicit generalization, compares new problems instances with instance seen in training, which have been stored in memory. These stored their training set when predicting a value or class for a new instances, they compute distance training instances to make a decision. The algorithm in this category for numerical prediction can divided into two types: similarity- based, e.g., Euclidean or entropy based and regression-based e.g., LWL Since regression is one of the most popular methods for numerical prediction[14]. The advantages of the Instances based Algorithm is it over other methods of machine learning is its ability to adapt its model of machine learning is its ability to adapt its model to previously unseen data. Instance based learners may simply store a new instance or throw an old instance away. The Disadvantages of the instances based Algorithm are its need more storage and computational complexity.

LINEAR REGRESSION-It is simple form of regression. Linear regression attempts to model the relationship between the two variables by fitting a linear equation to observe the data. this is widely used in statistics. For this purpose ,linear functions are used for which the unknown parameter i.e., weight of the independent variables, are estimated from the training data[1].this can be used to predict the values One of the most common estimating method is least mean square. Linear regression algorithms for predicting include simple regression multiple regression and pace regression, which is suitable for data of high dimensionality and only accepts binary nominal attributes.[13]. The main advantages of the linear regressions is

International journal of basic and applied research

www.pragatipublication.com

ISSN 2249-3352 (P) 2278-0505 (E)

gain a far greater understanding of the variables that can impact its success in the coming weeks,months and years into the future.The disadvantages of the regression

DECISION TREE Decision tree is used for both the prediction and classification [15][16][17][18][19][20][21] for the classification purpose a function can be learned this is intervals defined by splits on the individuals attributes value

A Root node, that has incoming edges and zero or more outgoing edges.

- Internal nodes, each of which has one incoming edges and two or more outgoing edges.

- Leaf node or end node, each of which has exactly one incoming edge and no outgoing edges. For prediction purpose, the decision trees algorithm for classification have been adapted to output a numerical value the main difference

K-MEANS ALGORITHM:K –means is the simplest and most commonly used portioning algorithm among the clustering algorithm in scientific and industrial software[3].Acceptance of k means is mainly due to its being simple .This algorithm is also suitable for clustering of a large datasets since it has much less computational complexity grows linearly by increasing of the data points. Advantages of the k-means algorithm are relatively simple to implement, Scales to large dataset, Guarantees convergence, easily adapts to new examples. Disadvantages of the k-means algorithm are Choosing manually, Being dependent on initial values, clustering data of varying sizes and density.

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 be improved for enhanced prediction.

International journal of basic and applied research

www.pragatipublication.com

ISSN 2249-3352 (P) 2278-0505 (E)

REFERENCE

[1] Ginger Saltos and Mihaela Coacea, An Exploration of Crime prediction Using Data Mining on Open Data, International journal of Information technology & Decision Making,2017.

[2] Shiju Sathyadevan, Devan M.S, Surya Gangadharan.S, Crime Analysis and Prediction Using Data Mining, First International Conference on networks & soft computing (IEEE) 2014.

[3] KhushabuA.Bokde, TiskshaP.Kakade, Dnyaneshwari S. Tumasare, Chetan G.Wadhai B.E Student, Crime Detection Techniques Using Data Mining and K-Means, International Journal of Engineering Research & technology (IJERT) ,2018

[4] H.Benjamin Fredrick David and A.Suruliandi,Survey on crime analysis and prediction using data mining techniques, ICTACT Journal on Soft computing, 2017.

[5] Tushar Sonawanev, Shirin Shaikh, rahul Shinde, Asif Sayyad, Crime Pattern Analysis, Visualization And prediction Using Data Mining, Indian Journal of Computer Science and Engineering (IJCSE), 2015.

[6] RajKumar.S, SakkaraiPandi.M, Crime Analysis and prediction using data mining techniques, International Journal of recent trends in engineering & research,2019.

[7] Sarpreetkaur, Dr.Williamjeet Singh, Systematic review of crime data mining, International Journal of Advanced Research in computer science , 2015.

[8] AyisheshimAlmaw, Kalyani Kadam, Survey Paper on Crime Prediction using Ensemble Approach, International journal of Pure and Applied Mathematics,2018.

[9] Dr .M.Sreedevi, A.Harha Vardhan Reddy, ch.Venkata Sai Krishna Reddy, Review on crime Analysis and prediction Using Data Mining Techniques, International Journal of Innovative Research in Science Engineering and technology ,2018.

[10] K.S.N .Murthy, A.V.S.Pavankumar, Gangu Dharmaraju, international journal of engineering, Science and mathematics, 2017.

[11] Deepiikak.K, Smitha Vinod, Crime analysis in india using data minigtechniques , International journal of Enginnering and technology, 2018.

[12] Hitesh Kumar Reddy ToppyiReddy, Bhavana Saini, Ginika mahajan, Crime Prediction &Monitoring Framework Based on Spatial Analysis, International Conference on Computational Intelligence Data Science (ICCIDS 2018).

[13]P. Kshirsagar and S. Akojwar, "Classification & Detection of Neurological Disorders using ICA & AR as Feature Extractor", *Int. J. Ser. Eng. Sci. IJSES*, vol. 1, no. 1, Jan. 2015.

[14]Pravin Kshirsagar, Dr.SudhirAkojwar, "Classification and Prediction of Epilepsy using FFBPNN with PSO", IEEE International Conference on Communication Networks, 2015.

[15]P. Kshirsagar, S. Akojwar, Nidhi D. Bajaj , "A hybridised neural network and optimisation algorithms for prediction and classification of neurological disorders" International Journal of Biomedical Engineering and Technology ,vol. 28,Issue 4,2018.

[16]P. Kshirsagar and S. Akojwar, "Novel approach for classification and prediction of non linear chaotic databases," *2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT)*, 2016, pp. 514-518, doi: 10.1109/ICEEOT.2016.7755667,2016

[17]Kshirsagar, P.R., Akojwar, S.G., Dhanoriya, R, " Classification of ECG-signals using artificial neural networks", In: Proceedings of International Conference on Intelligent Technologies and Engineering Systems, Lecture Notes in Electrical Engineering, vol. 345. Springer, Cham (2014).

[18]P. Kshirsagar and S. Akojwar, "Optimization of BPNN parameters using PSO for EEG signals," ICCASP/ICMMD-2016. Advances in Intelligent Systems Research. Vol. 137, Pp. 385-394,2016

[19]Pravin Kshirsagar, Nagaraj Balakrishnan & Arpit Deepak Yadav "Modelling of optimised neural network for classification and prediction of benchmark datasets" , Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, 8:4, 426-435, DOI: 10.1080/21681163.2019.1711457,2020

ISSN 2249-3352 (P) 2278-0505 (E)

[20]Dr. Sudhir Akojwar, Pravin Kshirsagar, Vijetalaxmi Pai "Feature Extraction of EEG Signals using Wavelet and Principal Component analysis", National Conference on Research Trends In Electronics, Computer Science & Information Technology and Doctoral Research Meet, Feb 21st & 22nd ,2014.

[21]S. Akojwar and P. Kshirsagar, "A Novel Probabilistic-PSO Based Learning Algorithm for Optimization of Neural Networks for Benchmark Problems", Wseas Transactions on Electronics, Vol. 7, pp. 79-84, 2016.

[22]Sudhir G. Akojwar, Pravin R. Kshirsagar, " Performance Evolution of Optimization Techniques for Mathematical Benchmark Functions". *International Journal of Computers*, **1**, 231-236,2016.

[23]Pravin Kshirsagar And Sudhir Akojwar "Hybrid Heuristic Optimization for Benchmark Datasets", International Journal of Computer Applications (0975 – 8887) Volume 146 – No.7, July 2016