

## CRIME ANALYSIS USING DECISION TREE CLASSIFICATION

V. Kirubha<sup>1</sup>, Dr. S. Manju Priya<sup>2</sup>

### ABSTRACT

Security is one of the major concerns and the issue is continuing to grow in intensity and complexity. It is an important aspect which is given higher priority by all political and government worldwide and aiming to reduce crime incidence. Law and enforcement is responsible for providing security to public. Nowadays, criminal incidences are increased and that are difficult to handle by police force. To reduce crimes, the crime incidences are to be analyzed. Law enforcement agencies like that of police today are faced with large volume of data that must be processed and transformed into useful information and hence data mining can greatly improve crime analysis and aid in reducing and preventing crime. Extractions of crime patterns are the essential need for the police force to identify the crime incidence in fast manner for further investigations. Since data mining is the tool for extracting hidden patterns, the data mining techniques can be applied for analyzing the crime dataset. Detecting crime incidence is not an easy task likewise preventing criminals from doing crime incidence is also not easy. But there is a chance to protect public from such crime attacks. To protect public, law and enforcement needs to analyze the

previously occurred crime patterns. Data mining is an exact field for analyzing a high dimensional crime data. The objective of this paper is to classify the crime incidence that was most occurred at each district. This paper helps the law and enforcement to provide controls over places based on the category of crime that was most happened in that particular district. It can also be used to detect the criminals easily based on this classification.

**Keywords :** Crime datasets, classification, decision tree.

### I. INTRODUCTION

Data mining is an iterative and interactive process of discovering novel, valid, potentially useful and ultimately understandable patterns and models in massive data sources. Here we did not know the patterns beforehand. Data mining has attracted a great deal of attention in the information industry and in society as a whole in recent years, due to the wide availability of huge amounts of data and the imminent need for turning such data into useful information and knowledge. The information and knowledge gained can be used for applications ranging from market analysis, fraud detection, and customer retention, to production control and science exploration. . It is also known as knowledge mining from data, knowledge extraction, data or pattern analysis, data archaeology, and data dredging.

---

<sup>1</sup>Research Scholar, Karpagam University, Coimbatore, Tamilnadu - India

<sup>2</sup>Associate Professor, Dept. of CS, Karpagam University, Coimbatore, Tamilnadu - India

Knowledge discovery as a process and consists of an iterative sequence of the steps like Data cleaning, Data integration, Data selection, Data transformation, Data mining, Pattern evaluation and knowledge presentation [7].

Classification analysis is the organization of data in given classes. It is also known as supervised classification, in which the classification uses given class labels to order the objects in the data collection. Classification approaches normally use a training set where all objects are already associated with known class labels. The classification algorithm learns from the training set and builds a model. The model is used to classify new objects. An algorithm that implements classification, especially in a concrete implementation, is known as a classifier. The term "classifier" sometimes also refers to the mathematical function, implemented by a classification algorithm that maps input data to a category.

Decision trees and decision rules are frequently the basis for data mining. They utilize symbolic and interpretable representations when developing methods for classification and regression. These methods have been developed in the fields of pattern recognition, statistics, and machine learning [8]. Symbolic solutions can provide a high degree of insight into the decision boundaries that exist in the data and the logic underlying them. This aspect makes these predictive mining techniques particularly attractive in commercial and industrial data mining applications.

## II. LITERATURE REVIEW

Kadhim B. Swadi Al-Janabi [1] presented a framework for the crime and criminal data analysis and detection using Decision tree Algorithms for data classification and Simple K Means algorithm for data clustering. It helps specialists in discovering patterns and trends, making forecasts, finding relationships and possible explanations, mapping criminal networks and identifying possible suspects. The classification is based mainly on grouping the crimes according to the type, location, time and other attributes.

Malathi. A and S. Santhosh Baboo [2] developed a crime analysis tool that focus on providing solutions that can enhance the process of crime analysis for identifying and reducing crime in India. The research work consist of developing analytical data mining methods that can systematically address the complex problem related to various form of crime. Vishnu Kumar Goyal [3] compared the five classification algorithms such as KNN, Naïve Bayes, Decision Tree, Decision Stump and Rule Induction using rapidminer studio and measured the performance parameters of the algorithms. Considered Weighting, Golf, Iris, Deals and Labor datasets for perform analysis. KNN algorithm gives best performance for these datasets.

Jyoti Agrwal, Reuka Nagpal and Rajni Sehgal [4] analyzed crimes using K means clustering to cluster the relevant crimes. They have used a crime dataset from offences recorded by the police in England and Wales by offence and police force area from 1990 to 2011-12. Crime analysis is based on crime homicide and plotted it with respect to year.

Hsinchun Chen, Wingyan Chung, Jennifer Jie Xu, Gang Wang Yi Qin and Michael Chau [5] presented a general framework for crime data mining. In their paper, entity extraction, association, prediction, pattern and visualization are considered as the four major categories of crime data mining techniques. Each category represents a set of techniques for use in certain types of crime analysis.

S. Yamuna and N. Sudha Bhuvaneshwari [6] analyzed data mining techniques to predict crime. Discrimination discovery, link analysis, financial crime detection and crime reporting systems can be used to analyze the crime. The results of data mining could be used to lessen and even prevent crime in future.

### III. CRIME

In general, crime is defined as an action or an instance of negligence that is deemed injurious to the public welfare or morals or to the interests of the state and that is legally prohibited. The study of crime and law enforcement information in combination with socio-demographic and spatial factors to apprehend criminals prevents crime, reduce disorder, and evaluate organizational procedures.

Crime analysis uses both qualitative and quantitative data and analytical techniques. Qualitative data and analytical techniques refer to non-numerical data as well as the examination and interpretation of observations for the purpose of discovering

underlying meanings and patterns of relationships. Quantitative data are data primarily in numerical or categorical format. Quantitative analysis consists of manipulations of observations for the purpose of describing and explaining the phenomena that those observations reflect and is primarily statistical.

Crime analysis employs both types of data and techniques depending on the analytical and practical need. For example, crime data can be used in various ways, both quantitatively and qualitatively. The information such as date, time, location, and type of crime is quantitative in that statistics can be used to analyze these variables. On the other hand, narratives of crime reports are considered qualitative data in that a large number of narratives are nearly impossible to analyze statistically and are primarily examined to determine general themes and patterns.

In a law enforcement agency, the central focus is crime, both those reported to the police and those that are not. Thus, the central type of data analyzed is crime and the information surrounding it, such as arrests, offenders, victims, property, and evidence.

The location where crimes or activities occur is an important factor in the analysis of crime. The fourth goal of crime analysis is assisting with the evaluation of organizational procedures. Several examples include resource allocation, the assessment of crime prevention programs, realigning geographic boundaries, forecasting staffing needs, and developing performance measures for the police department.

#### IV. CRIME DATA MINING

Crime Analysis defined as analytical process which provides relevant information relative to crime patterns and trend correlations to assist personnel in planning the deployment of resources for the prevention and suppression of criminal activities.

It is important to analyze crime which helps to assist the police in reducing and preventing crime and disorder. Present cutting edge policing strategies, such as hotspots policing, problem-oriented policing, disorder policing, intelligence-led policing, and Comp Stat management strategies, are centered on directing crime prevention and crime reduction responses based on crime analysis results.

To perform crime analysis appropriate data mining approach need to be chosen and as classification is an approach of data mining which classifies objects based on the prior knowledge about the crime. In this research work decision tree classification is used and the most occurring crimes are found out in the district. It helps to reduce further occurrences of similar incidence and provide information to reduce the crime.

By improving communication within agencies, information and knowledge can be more effectively shared across jurisdictional boundaries. Since crime is now a global issue, with technological advances connecting all nations and places in an unprecedented manner, good crime analysis can help identify and solve regional, national, and international crime problems.

#### V. DATASET DESCRIPTION

The crime dataset for analysis taken from Open Government Platform of India titled with District-Wise Crimes under Various Sections of Indian Penal Code (IPC) Crimes [10].

The data refers to district wise details on crimes committed under Indian Penal Code (IPC) during the years. The nature of such crimes includes Murder, Attempt To Murder, Culpable Homicide Not Amounting To Murder, Kidnapping & Abduction, Kidnapping And Abduction Of Women And Girls, Kidnapping And Abduction Of Others, Dacoity, Preparation And Assembly For Dacoity, Robbery, Burglary, Theft, Auto Theft, Other Theft, Riots, Criminal Breach Of Trust, Cheating, Counterfeiting, Arson, Hurt/Greivous Hurt, Dowry Deaths, Assault On Women With Intent To Outrage Her Modesty, Insult To Modesty Of Women, Cruelty By Husband Or His Relatives, Importation Of Girls From Foreign Countries, Causing Death By Negligence and Other IPC Crimes.

#### VI. EXPERIMENTAL RESULTS

##### Rapidminer User Interface

RapidMiner [9] is a software platform developed by the company of the same name that provides an integrated environment for machine learning, data mining, text mining, predictive analytics and business analytics. It is used for business and industrial applications as well as for research, education, training, rapid prototyping, and

application development and supports all steps of the data mining process including results visualization, validation and optimization. RapidMiner supports all steps of the data mining process from data loading, pre-processing, visualization, interactive data mining process design and inspection, automated modeling, automated parameter and process optimization, automated feature construction and feature selection, evaluation, and deployment.

### Retrieving data from the Repository

Figure 6.1 shows the results of the retrieve operator. This result does not contain any class labels it just retrieve the data from the repository.

Row No.	DISTRICT	MURDER	ATTEMPT	KIDNAPING	ROBBERY	BURGLARY	CHEATING	CRIME
1	ANDHRA	18	32	47	8	41	11	Murder
2	CHENNAI	180	216	87	65	548	728	Chand
3	CHENNAI 1	5	2	18	1	8	8	Robbery
4	COMBONDUR	28	24	41	71	128	88	Chand
5	COMBONDUR URBAN	28	27	20	101	128	132	Chand
6	COIMBATUR	78	120	152	44	183	38	Robbery
7	DEWANPET	20	32	38	21	73	38	Kidnaping
8	CHENNAI 2	74	88	87	88	108	118	Chand
9	ERODE	48	31	38	44	54	118	Chand
10	CHENNAI 3	78	110	17	22	187	152	Robbery
11	VANARAJAN	53	118	23	88	203	122	Chand
12	KODAI	28	18	48	20	38	18	Kidnaping
13	SIRSAHANG	43	41	45	57	118	71	Robbery
14	MAHARAJAN	78	142	83	121	181	148	Chand
15	MAHARAJAN 1	74	85	78	88	284	284	Chand
16	MAHARAJAN 2	44	42	18	88	18	18	Robbery
17	MAHARAJAN 3	28	33	48	44	82	82	Chand
18	MAHARAJAN 4	14	17	8	8	85	28	Robbery
19	PERAMBALUR	11	15	21	13	37	21	Robbery
20	PULIANTHUR	71	38	18	18	34	42	Chand
21	PONDICHERRY	54	81	83	45	81	85	Atari Murd
22	SALIGRAM	48	48	78	28	88	82	Chand

Figure 6.1 Result of retrieve operator

### Dataset after generation of new attribute

Decision tree classifier can be applied on class labeled training tuples. Figure 6.2 presents the result of Generate Attribute operator. It has a new attribute as "CRIME".

Figure 6.2 Result after the generation of CRIME attribute

### Design of Decision Tree Classification

This involves analyzing the crime records and used to classify the crime category based on districts. Figure 6.3 displays the design of decision tree model in rapidminer tool.

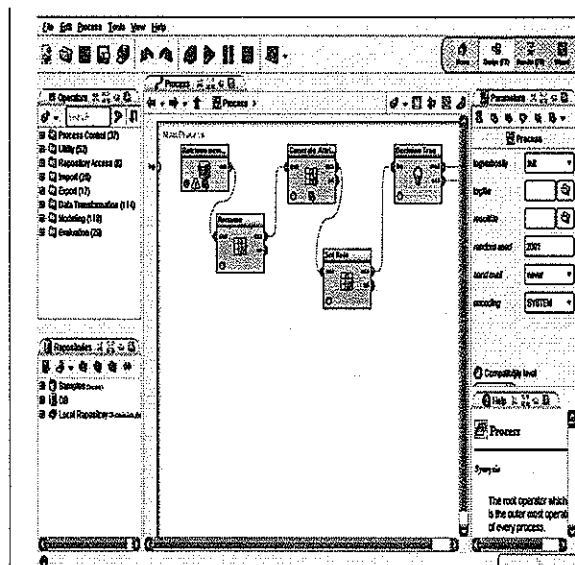


Fig.6.3 Design of decision tree classification

The Classified Tree Structure of Crime Data

Figure 6.4 shows the tree structure of crime data classification. District is a root node which is classified associate with crime category which is most occurred at that district. All leaf nodes denote the crimes and edges refer to the District names based on which a classification is performed.

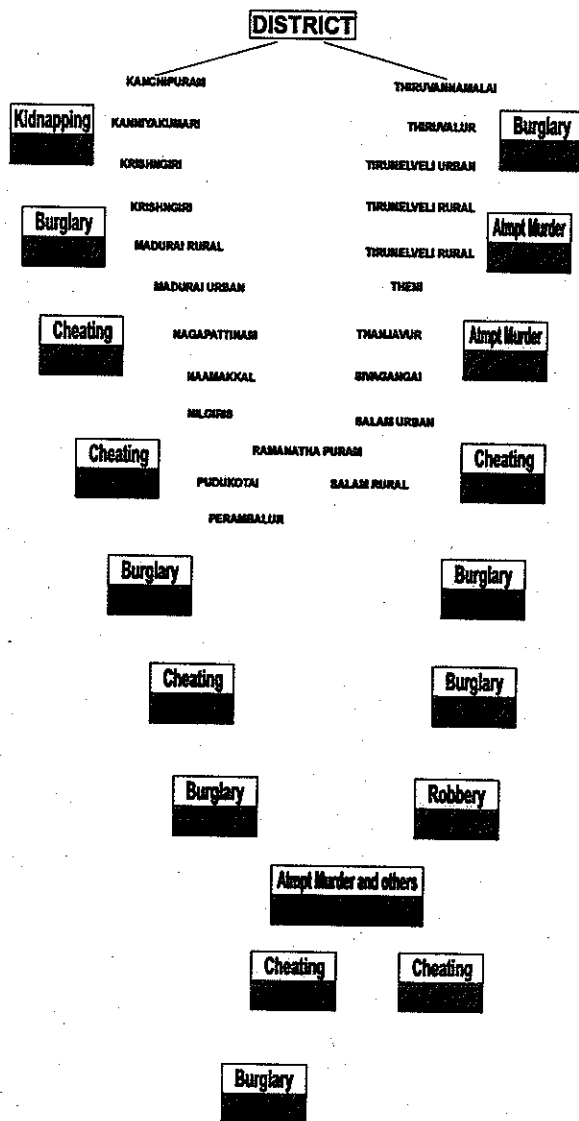


Figure 6.4 Resultant Tree Structure

From the results, the most occurring crime at each district is classified. It is useful to improve the protection based on the crime category.

VII. CONCLUSION

To get a satisfactory model for data mining and to get excellent results of analyzing crime data set, it requires huge historical data that can be used for both creating and testing the model. This research work focuses on crime analysis by implementing classification algorithm on crime dataset using rapidminer tool and here crime analysis is done by classifying crime categories based on places. From the classified results it is easy to identify crime over specific areas and it can be used by the law and enforcement to prevent such criminal incidence in future. Further, it can be enhanced in to more states.

REFERENCES

- [1] Kadhim B.Swadi al-Janabi, "A Proposed Framework for Analyzing Crime Dataset using Decision Tree and Simple K-means Mining Algorithms", Journal of Kufa for Mathematics and Computer, Vol.1, No.3, May 2011.
- [2] Malathi.A, Dr. S. Santhosh Baboo, "Algorithmic Crime Prediction Model Based on the Analysis of Crime Clusters", Global Journal of Computer Science and Technology, Volume 11 Issue 11 Version 1.0 July 2011.
- [3] Vishnu Kumar Goyal, "A Comparative Study of Classification Mehods in Data Mining

using RapidMiner Studio”, International Journal of Innovative Research in science and Engineering, International Journal of Innovative Research in Science & Engineering, ISSN(online)2347-3207.

[4] Jyoti Agarwal, Renuka Nagpal, Rajni Sengal, “Crime Analysis using K-Means Clustering”, International Journal of Computer Applications(0975-8887), Volume 83-No 4, December 2013.

[5] Hsinchun Chen, Wingyan Chung, Jennifer Jie Xu, Gang Wang Yi Qin, Michael Chau, “Crime Data Mining: A general framework and some examples”, The IEEE Computer Society (0018-9162), April 2004.

[6] S. Yamuna N. Sudha Bhuvaneswari, “Data Mining Techniques to Analyze and Predict Crimes”, The International Journal of Engineering And Science (IJES), Volume 1, Issue 2, ISSN: 2319 – 1813 ISBN: 2319 – 1805, December 2012.

[7] Jiawei Han and Micheline Kamber, “Data Mining: Concepts and Techniques”, Morgan Kaufmann Publishers, second Edition, 2006.

[8] K.P Soman, Shyam Diwakar, V. Ajay, “Insight into Data Mining Theory and Practice”, Prentice-Hall of India Private Limited, 2008.

[9] Rapidminer is an open source-learning environment for data mining and machine learning. <https://rapidminer.com>

[10] Open Government Data India, District-Wise Crimes Under Various Sections Of Indian Penal Code (IPC) Crimes [http://data.gov.in/catalogs#path=group\\_name/crime-statistics4188/keywords/ipc-crimes-15550](http://data.gov.in/catalogs#path=group_name/crime-statistics4188/keywords/ipc-crimes-15550)

#### AUTHOR’S BIOGRAPHY



**Kirubha. V** received her Master Degree (M.Sc(CS)) in 2015 and the Bachelor Degree (B.Sc(CS)) in 2013 from Karpagam University, Coimbatore. She is currently pursuing as M. Phil Research Scholar in the Department of Computer Science at Karpagam University. Her areas of interest are Data Mining and Medical Data Mining.



**Dr. S. Manju Priya**, working as Associate Professor, Dept. of Computer Science, Karpagam University. She has completed her Ph.D in Karpagam University in 2014.

She has published more than 15 papers in national and international journals. Her area of interest is sensor networks, cloud storage, communications.