



## DEVELOPMENT OF HYBRID ALGORITHM FOR TAMILNADU CRIME ANALYSIS

C.Jayapratha

Assistant Professor,

Department of Computer Applications  
Karpaga Vinayaga College of Engineering  
Technology,  
Maduranthagam  
Research Scholar, Barathiyar University,  
Coimbatore

Dr.J.M.Gnanasekar

Professor & Head

Dept of Computer Science & Engineering,  
Karpaga Vinayaga College of Engineering &  
Technology, Maduranthagam

**Abstract-** It is evident from newspapers, TV, web and other sources of news, the occurrence of crime and terrorism in India is increasing year by year. They highly devastating the country's resources. The increase in crime rate and terrorism threat needs to be controlled in long run to be eradicated before it depletes the resources gradually. The crime occurrence and terrorist attacks have been recorded by the police department country wide. This huge volume of crime records needs to be thoroughly analyzed to reveal the frequency of crime occurrence crime type. Type of terrorist attack and other factors. The outcome of analysis should be interpreted and concluded. Integration of data mining techniques (clustering, association and classification) to enhance crime prevention through analysis of the relationships between type of crimes, locations, times and properties when crime happens. Being able to use the best datmining techniques to obtain the relationship between specific variables will help in coming up with a solution through crime prediction. Specific crime predictions will be of greater importance for the world at large. The First part of the paper is used to identify the most suitable algorithms from the six different clustering algorithms and reviews two types of Association algorithms are presented and compared. It is used to identify the most suitable algorithms from the two different Association algorithms. Such as Apriori and FP Tree Algorithms. And reviews Eight Classification algorithms C4.5, ID3, C&RT, CS-MC4, Decision List, Naive Bayes, Random tree and Rule Induction. It is used to identify the most suitable algorithms from the eight different Classification algorithms. My research focuses on to identify which algorithm give the best result. Finally I want to combine all the best three algorithms features into a single algorithm. The second part of this paper deals with an intelligent crime analysis and recording system designed to overcome problems that appear mainly in the

Tamilnadu police department. It is a GIS based system which comprises of data mining techniques such as Hotspot detection, Crime clock, Crime comparison, Crime pattern visualization, Outbreaks detection and the nearest police station detection. Salient features of the proposed system include a rich environment for crime data analysis and a simplified environment for location based data analysis. It facilitates the identification of various types of crimes in detail and assists the police personals to control and prevent such incident efficiently. The conclusion of the study will be recommended to the Tamilnadu police department as suggestions to reduce the crime level to a limit.

**Keywords** Datamining, Association, Clustering, Classification.

### 1. INTRODUCTION

Data mining or Knowledge Discovery in Databases (KDD) is the nontrivial extraction of implied, previously unknown, and potentially useful information from data. Knowledge discovery in databases (KDD) is the process of recognizing a valid, potentially, useful and Ultimately understandable structure in data. Data mining is a young and promising field, used to knowledge discovery from data. Data Mining is the process of automatic abstraction of novel, useful, and understandable patterns in very large databases. High-performance accessible and parallel computing is crucial for ensuring system scalability and interactivity as datasets grow inexorably in size and complexity [4]. It is the mechanized process of modeling large databases by means of discovering useful patterns. Consequently, data mining consists of more than Collecting and handling data, it includes analysis and prediction. Classification, Associations and clustering are three important techniques in data mining

Integration of data mining techniques (clustering, association and classification) to

enhance crime prevention through analysis of the relationships between type of crimes, locations, times and properties when crime happens. Being able to use the best datmining techniques to obtain the relationship between specific variables will help in coming up with a solution through crime prediction. Specific crime predictions will be of greater importance for the world at large.

## 2.CONTRIBUTION OF THE RESEARCH

### i. Aim of the Research

The Main objective of this analysis is to come up with the best model for crime analysis. The aim of the Project is

1. Classifying the different types of crimes
2. To identify the most suitable algorithms from the different clustering algorithms
3. To identify the most suitable algorithms from the different Association algorithms
4. To identify the most suitable algorithms from the different classification algorithms
5. Develop a hybrid algorithm for Tamilnadu crime analysis
6. To represent graphically Hotspot detection and Nearest Police Station Detection.

### ii. Need and Significance

Many police departments all around the world lack good and efficient crime recoding and analysis systems. The vast geographically diversity and the complexity of crime patterns have made the analyzing and recording of crime data even difficult. According to the police department, they face these problems for many years. They need good and efficient systems to control and prevent various crimes efficiently.

### iii. Scope of the Research

The scope of the project is to develop an analysis system that is able to take different variables (Type of Crime, Location and Times) and come up with a model that is able to predict crime. The system shall have different breakdown parts in aims to build the sequence of the model and system. The system should be able to take in data from different sources pre-process it for analysis. After the pre-processing then the analysis will produce the relationship between the variables and aim of all this is to produce a model that can predict the outcome of the interaction between the variables. For example if an individual is given a location, time, type of offence and properties one should predict the class of crime. The Main focus is prediction through analysis of past data.

### iv. Data Analysis

- a. Year 2000-2014 crime report collected from State crime records bureau, Mylapore, Chennai, Tamilnadu-28.
- b. SPSS16.0 Software used for finding the statistical report
- c. The technology used for this project is WEKA and RAPIDMINER. RapidMiner is a powerful programming statistical tool for graphic and analysis and will be used or the analysis. Information about the crimes usually is obtained using the quantitative methods since most of the data is non-numeric hence will use the datamining technique (WEKA SYSTEM AND Rapid Miner (to analyze the data and this comprises training, building and validation of models.

## 3. ALGORITHM DESCRIPTION

### 1.Clustering Algorithm

The basic clustering algorithms we used in our study. K-means Clustering, Hierarchical clustering, DB Scan Clustering .OPTICS, EM Algorithm .

### 2. Classification Algorithms

The basic classification algorithms we used in our study.

C4.5, ID3, C&RT, CS-MC4, Decision List Naïve Bayes, Random Tree, Rule Induction Decision Trees

### 3. Association Rules

The basic classification algorithms we used in our study Apriori Algorithm and FP Tree.

## 4. DESIGN AND ARCHITECTURE

For this analysis I want to reviews two types of Association algorithms are presented and compared. It is used to identify the most suitable algorithms from the two different algorithms. Such as Apriori and FP Tree Algorithms. Next I want to reviews Eight Classification algorithms C4.5, ID3, C&RT, CS-MC4, Decision List, Naive Bayes, Random tree and Rule Induction. It is used to identify the most suitable algorithms from the eight different algorithms. Next I want to reviews six types of clustering algorithms are presented and compared. It is used to identify the most suitable algorithms from the six different algorithms. Such as k-means clustering, Hierarchical clustering, DbSCAN Clustering, Density based Clustering, Optics and EM Algorithms. I want to apply association algorithms to find more frequent set of attributes to a particular class. I want to apply classification algorithms to classify set of attributes to a particular class. I want to used cluster analysis so

as to identify how the various variables can be grouped together in relation to how more similar to each other than to those in other groups. My research focuses on to identify which algorithm give the best result. Finally I want to combine all the best three algorithms features into a single algorithm. The Architecture used for the structure of the system is the integrated approach. The architecture is designed for a specific purpose or workload. It is used when there is need for fit for purpose. The first step is to upload the dataset and then pre-process and clean the data so that it is ready for analysis. According to Rouse (2010) noisy data can adversely affect the results of any data mining analysis therefore it is of essence to clean up the data. Clean data can be analyzed using different data mining techniques and visualization of data is very important for descriptive analysis and also final output. For my analysis I want to used RapidMiner programming language and Weka machine learning tool. Data modeling is an important aspect of analyzing data and getting the best model for an analysis is also crucial hence model validation is also an imperative step. The system followed for analysis is shown in Figure 1.

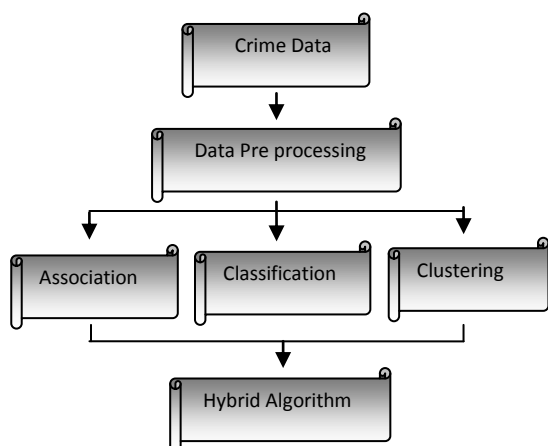


Figure1. Data Mining Techniques used in Crime Data

## 5. GRAPHICAL REPRESENTATION

### 1. Hotspot Detection

To identify hotspots with high accident density, cluster analysis is used for identifying the clusters of accident spots.

### 2. Accident Clock

The Accident clock is a representation of the number of accident scenes that has been taken place within the 24 hours of a day. A Crime clock is represented as a bar chart. The 24 hour clock is represented using 24 bars on the graph and the height of each bar represents the number of crime scenes per hour. Three extra bars are used to represent the crime scenes without an exact time of incident. The “day bar” represents the crime scenes which were taken place in the day time, the

“night bar” represents the Crime scenes which were taken place in the night time and the “unknown bar” represents the crime scenes which cannot be assigned to any time duration.

### 3. Crime Comparison

Comparing different types of accident is very important to get an idea about the growth of a particular crime over the other types of accidents. A pie-graph is used to satisfy this requirement by allowing the analyst the maximum freedom to compare the different types of accidents s in an optimal way. It shows the percentage comparison between different accident types.

### 4. Accident Pattern Visualization

Time series analysis comprises methods for analyzing time series data in order to extract meaningful statistics and other characteristics of the data. A time series plot is used to represent the changes in frequency of crime occurrence. The Y-axis represents the frequency of accidents and the X-axis represents the time.

### 5. Nearest Police Station Detection

The J48 decision tree is a predictive machine-learning model that decides the target value (dependent variable) of a new sample based on various attribute values of the available data.

## 6. CONCLUSION

The project is a good starting point for implementation of data mining for real world examples. This project has brought us insight into various techniques not only in the field of data mining but also in database utilization, visualization, etc. Few points of consideration are for the project itself are Data quality is an extremely important aspect, and we have realized during the course of implementing the project that more time should have been spent in checking how *sane* the data we had was. This, however, would have had no effect at all on the work done, but it would definitely result in much more useful information about the data. Although the problem of parsing crime reports wasn't tackled in this work, we realize how important it is, and how challenging it can be. From the variation we've seen among the different datasets, we believe that some sort of standardization should be enforced among the different police departments in order to make automatic parsing of crime reports more reliable. One more issue that could be considered is the use of open-source data mining tools, even though WEKA and RapidMiner is a very useful alternative many other tools exist that are more robust and feature rich. Utilization of such tools would prove for more open and feature rich application.

## REFERENCES

[1]Agrawal R, Srikant R 1994 Fast algorithms for mining association rules in 20<sup>th</sup> International Conference on very large Databases (Newyork; Morgan Kaufmann)

[2]Abernethy, M. (2010) 'Data mining with WEKA, Part 2: Classification and clustering' [Online].

[3] Chen, H., W.Chung, et al.(2004). Crime data mining: a general framework and some examples. Computer 37 (4):50-56.

[4] Fayyad, U, M, Piatetsky-Shapiro.G and Smith P 1996."Knowledge discovery and Data Mining: Towards A unifying framework" Proceedings of the 2<sup>nd</sup> International Conference on Knowledge Discover and Data Mining Portland.Oregan, August2-4PP(82-88)