

# Development of Hybrid Algorithm for Road Accident Protection (Paper Id:195)

C. Jayapratha<sup>1</sup>

Assitant Professor, Dept of MCA

Karpaga Vinayaga College of Engg&Tech

Maduranthagam. Research Scholar

J. M. Gnanasekar<sup>2</sup>

Professor & Head, Dept of CSE

KarpagaVinayagaCollegeof engg&Tech

Maduranthagam. Barathiyar University,

Coimbatore

**Abstract:-** In this modern world the usage of automobiles by the people are increasing day by day. As such due to enhanced traffic in urbanized areas such as in highways and roads, Motor vehicle accident and rail accidents are increasing in our state as well as in our country. Though accidents are not wantonly being done, the causes of the accidents are many such as drunk and drive, violation of traffic rules, non application of protective appliances, defective roads, obstacles on the road, due to workload of continuous driving for hundreds of hours and due to defective mechanism in the motor vehicle. Only few accidents are due to actus Reu's. Consequent to the increasing number of accidents there are losses of precious human lives and limbs, loss of properties, Traffic Jam etc., and they are root cause for some social problems. So it is just and necessary to curtail the road accidents. By way of detecting the basic reasons for the occurring accidents it would be easier to prevent the accident in future. It will be useful to the police authorities as well as to the entire society for awareness. So the data analyzing of road accidents being done. My research aims to provide a review to extract useful information by means of Data Mining, in order to find new hybrid algorithm and predict accident trends for them using data mining techniques.

**Keywords:-** Datamining, Accident, Clustering, Classification.

## I. INTRODUCTION

Data mining refers loosely to finding relevant information or discovering knowledge, from a large volume of data. Datamining attempts to discover statistical rules and pattern automatically from stored data. However it differs from machine learning and deal with very large volume of data stored on disk [1]. Discovering knowledge from a large volume of data is not a simple process, but it is an iterative and interactive process. Data Mining should be "the non trivial process of identifying valid, novel, potentially, useful and ultimately comprehensible knowledge from databases such knowledge can be useful in making crucial decisions [6].

**Nontrivial** Means that rather than simple computations, complex processing is required to uncover the patterns that are built in the data. **Valid** The discovered patterns should be valid for all data including new data. **Novel** the discovered patterns should be

innovative. **Useful** The Organization Should be able to act upon these patterns to become more profitable and efficient. **Comprehensible** The new pattern should be understandable to the users and add to the knowledge.

Nowadays many industries are being used the electronic data repositories for storing the huge size of their data extract the knowledge from the huge size of the data source is to the analyst for better decision making process. The traditional techniques are insufficient to analyze these kinds of data. Today's world's data are collected and stored at enormous speeds. So it is essential to the industries to find a special tool for storing and accessing these databases. The Datamining tools are such type of tools. These tools are applied to both commercial and scientific data. The commercial data are mined to provide better service to customers and proactive their services. The tools help to extract, understand complex relationships predicting future status. Amazingly Data mining techniques applied on these databases discover relationships and patterns which are helpful in studying the road accident and analysis and prediction.

## II. CONTRIBUTION OF THE RESEARCH

### 1. Aim of the Research

The paper is organized as follows. First, background information on clustering, complex networks analysis techniques, and on the methods used to extract the relationship between crash involvement and risk factors, is provided. The study's methodology is then described including a description of: (1) To cluster the data; (2) the association rule data mining method; and the Characteristics of the dataset used. The clustering results are then presented, followed by a description of the discovered association rules for: (1) identifying hotspots and their Characteristics; and (2) understanding the factors affecting incident clearance time. A discussion of the difference between the association rules derived from the whole data set and those derived from each cluster is also included. Finally, the paper concludes by summarizing the major conclusions of the study.

- ⊕ Classifying the different types of Road accidents
- ⊕ To identify the most suitable algorithms from the different clustering algorithms.
- ⊕ To identify the most suitable algorithms from the different Classification algorithms.
- ⊕ To identify the most suitable algorithms from the different Classifier algorithms
- ⊕ Develop a hybrid algorithm for road accident protection
- ⊕ To represent graphically Hotspot detection And Nearest Hospital Location

2. *Need and Significance*

Many police departments all around the world lack good and efficient road accident recording and analysis systems. The vast geographical diversity and the complexity of accident patterns have made the analyzing and recording of accident data even difficult. According to the Tamilnadu police department, they face these problems for many years. They need good and efficient system to control and prevent various accident efficiently. Though earlier researchers have developed a lot of algorithms methodologies. It is imperative to find new algorithms. It is possibly unsuitable to apply the difficulty of mining in all item sets on very large datasets.

### 3. Scope of the Research

The Research addresses two problems related to Road accident analysis. The first part of this paper deals with data clustering. This paper reviews six types of clustering techniques are presented and compared. It is used to identify the new algorithms from the six different algorithms. 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. Salient features of the proposed system include a rich environment for accident data analysis and a simplified environment for location based data analysis. It facilitates the identification of various types of accidents 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 accident level to a limit. **Classification trees** are used to predict membership of cases or objects in the classes of a categorical dependent variable from their measurements on one or more predictor variables. Classification tree analysis is one of the main techniques used in Data Mining. Next subsections deals with the basic classification algorithms we used in our study [2],[3].

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

**Association rules** is a data mining method for investigating the associative property of different events, which can be used in traffic accident data mining to mine the importance of attributes, that is, the associative

relationship of events with certain types of accident. Its basic idea is to treat each characteristic as an item. Accident site, number of death, and so on can all be called an item. The higher the association, the more likely one event is directly linked to the cause of a certain type of accident. To decide how related two items are, we need to identify how many times some characteristics appear at the same time in a large number of similar events [3][4]. **Classifier analysis** is one of the main techniques used in Data Mining. Next subsections deals with the basic classification algorithms we used in our study. [4][5]

- a. Naive Bayesian Classifier
- b. J48 Decision Tree Classifier
- c. AdaBoostM1 Classifier
- d. PART (Partial Decision Trees)
- e. Random Forest Tree Classifier

### 4. Data Analysis

1. 2000-2014 Road accident record report collected from State Crime Records Bureau, Tamil Nadu, Chennai – 600 028. 2. SPSS16.0 software used for finding the statistical report. 3. The clustering techniques are implemented and analyzed using a clustering tool WEKA. Performance of the 6 techniques are presented and compared 4. IRTAD, GLOBESAFE ACCIDENT RESOURCE.

### 5. Limitations

Research in respect of whole India is a tedious task. So the research focused here is to do research only about Tamilnadu.

### 6. Process of Data Mining

#### a. Data Modeling

The sample data used covered the period of 24 Months, that is, January 2000 to December 2014.

b. Deployment		
Sno	Variable	Description
1	Vehicle Type	Small Cars Heavy Vehicle vehicle body type, vehicle age, vehicle role Vehicle Condition
2	Time of the Day	Morning Afternoon Evening Night/Midnight
3	Season	Wet/Dry/muddy
4	Causes	Wrong Overtaking,Careless Driving,Loss of control TyreBust,Over Speeding,Obstruction,Pushed by another Vehicle,Broken Shaft,Broken Spring,Break Failure Road Problem,Unknown Causes,Robbery Attack,Alcohol/drug
5	Person,	Age,Gender,Driver/ passenger,Race/Ethnicity
6	Injury	No injury, Possible injury,Non- incapacitating injury, Incapacitating injury, Fatal injury.
7	Initial Point of Impact	nodamage/non-collision, front, right side, left side, back, front right corner, front
8	Accident Information	month, Region, primary sampling unit, the number of the police jurisdiction,case number, person number, vehicle number, vehicle make andModel,RoadSeparation ,RoadOrientation,RoadSurfaceType,Road adSurfCondition WeatherCondition , LightCondition

Table1.Deployment data for Road accident Protection

*C. Graphical Representation**a. Hotspot Detection*

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

*b. 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 accident 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 accident scenes per hour. Three extra bars are used to represent the accident scenes without an exact time of incident. The “day bar” represents the accident scenes which were taken place in the day time, the “night bar” represents the accident scenes which were taken place in the night time and the “unknown bar” represents the accident scenes which cannot be assigned to any time duration.

*c. Accident 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 shows the percentage comparison between different accident types.

*d. 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.

*e. Nearest Hospital 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.

**III. SYSTEM ARCHITECTURE**

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.

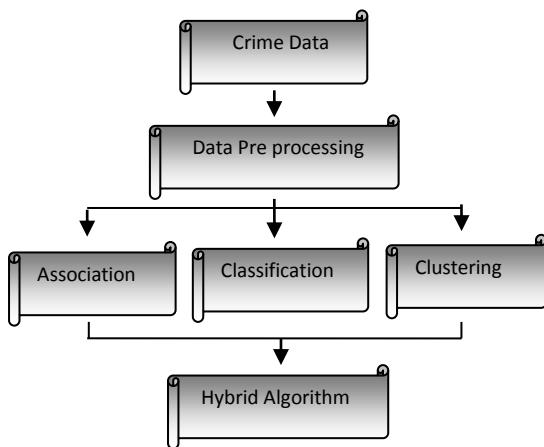


Figure1.Data Mining Techniques used in Road Accident

#### IV.CONCLUSION

The aim of this study was to show the applications of data mining techniques in the field of accident investigation. It was done by reviewing various papers. We are currently enhancing it by considering several issues; variation in crash occurrence may have some consequence for traffic safety measures in some places in Tamilnadu. The modeling will be to combine road-related factors with driver information for better predictions, and to find interactions between the different attributes. 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 accident reports more reliable. It

will be useful to the police authorities as well as to the entire society for awareness. So the data analyzing of road accidents being done.

#### V.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] Abdel-Aty, M., & Abdelwahab, H., Analysis and Prediction of Traffic Fatalities Resulting From Angle Collisions Including the Effect of Vehicles' Configuration and Compatibility. *Accident Analysis and Prevention*, 2003
- [3] Abdelwahab, H. T. & Abdel-Aty, M. A., Development of Artificial Neural Network Models to Predict Driver Injury Severity in Traffic Accidents at Signalized Intersections. *Transportation Research Record* 1746, Paper No. 01-2234.
- [4] Bedard, M., Guyatt, G. H., Stones, M. J., & Hireds, J. P., The Independent Contribution of Driver, Crash, and Vehicle Characteristics to Driver Fatalities. *Accident analysis and Prevention*, Vol. 34, 2002, pp.
- [5] Beshah, T. and S. Hill. Mining road traffic accident data to improve safety: Role of road-related factors on accident severity in Ethiopia. 2010.
- [6] Gorrilla, J. and V. Lobo, Improvements on the visualization of clusters in geo-referenced data using Self-Organizing Maps. *Computers and Geosciences*, 2012. 43: p. 177-186.
- [7] S.Shanthi, Dr.R.Geetha Ramani, "Classification of Vehicle Collision Patterns in Road Accidents using Data Mining Algorithms", International Journal of Computer Applications, Vol.35, No.12, Pp.30-37, 2011.
- [8] S.Shanthi, Dr.R.Geetha Ramani, "Classification of Seating Position Specific Patterns in Road Traffic Accident Data through Data Mining Techniques", Proceedings of Second International Conference on Computer Applications, ICCA 2012, Vol.5, pp. 98-104, January, 2012.
- [9] 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.Oregon, August 2-4PP(82-88)