

National Highway Road Accidents Analysis Using Data Mining Technique

¹K. Sridharan, ²Dr. T. Balasubramanian

¹Asst Prof/ Ph.D Research Scholar, ²Associate Prof/Head

Department of Computer Science & Applications, Sri Vidya Mandir Arts & Science College, Katteri, Uthangarai, Tamilnadu, India

ABSTRACT

Today, the government sets priority in enhancing the highway roads where the road safety is main focus. The increase in the vehicles, traffic lead to road accidents. Data Mining is frequently used technique in identifying the factors behind the road accident. The main aim of this research is to identify the factors behind the road accident in national highways in Tamilnadu, using Data Mining techniques. For this research the data is taken from National Highways Authority of India. This paper focuses the study of Clustering algorithms that have to be applied to the road accidents.

Keywords: Road Accident, Accident analysis, Data Mining, Clustering, Association Rule Mining

I. INTRODUCTION

Road accident is one of the major problems in our country. There are a lot of vehicles driving on the roadway every day, and traffic accidents could happen at anytime anywhere. Some accident results in fatal outcome. As per Ministry of Road Transport & Highways, a report was published in 2016 that more people died on roads accidents in India last year, as compared to the number of deaths in 2015. The data has further revealed that the states of Uttar Pradesh and Tamil Nadu have accounted for maximum number of deaths this year. Also as per the data cited in the report, the country recorded at least 4,80,652 accidents in 2016, leading to 1,50,785 deaths. The number suggests that at least 413 people died everyday in 1,317 road accidents.[1] Tamil Nadu also reported the highest number of persons injured in road accidents. Two-wheelers, cars and jeeps were involved in a majority of these accidents.

II. Review of Literature

Many works have been carried out by different researchers on the prediction of the causes of accidents,

their prone locations along various roads around the world using data mining algorithms.

Sachin Kumar and Durga Toshniwal [2] gathered road accident data from EMRI. then to analyze the accident data, first K-modes clustering algorithm has been applied and further association rule mining technique is applied to identify circumstances in which an accident may occur for each cluster.

Krishnaveni and Hemalatha [3] proposed with some classification models to predict the severity of injury that occurred during traffic accidents. Naive Bayes Bayesian classifier, AdaBoostM1 Meta classifier, PART Rule classifier, J48 Decision Tree classifier, and Random Forest Tree classifier are compared for classifying the type of injury severity of various traffic accidents. The final result shows that the Random Forest outperforms the other four algorithms.

Zhang and Fan [4] conducted data mining on Saskatchewan Highways traffic data to investigate major contributing factors to traffic collision. C4.5 algorithm is used to create decision tree model. In cited above, the factors affecting the road accidents have been used to create decision tree model using enhanced algorithm.

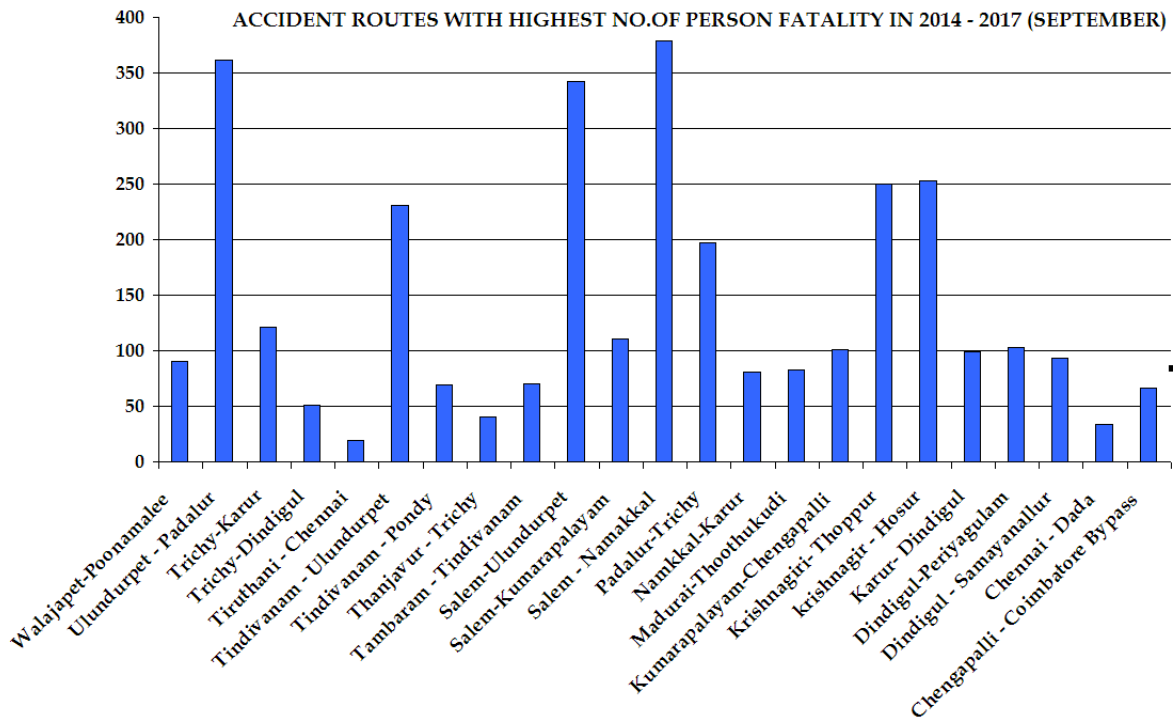
III. Accident Dataset

details of Tamilnadu with 19,705 records and 17 different attributes from 2014 – 2017 (September).[5]

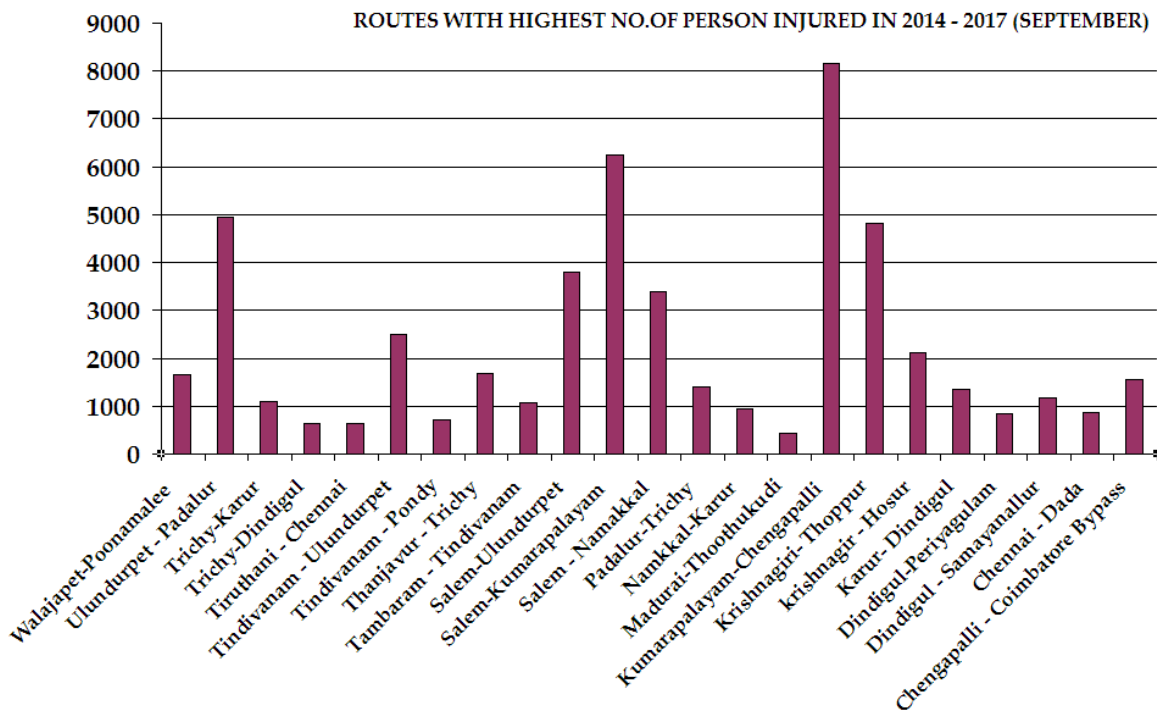
The dataset is collected from National Highways Authority of India. The dataset contains the accidental

Table 1 : Record of accident along the national high way in Tamilnadu between, 2014 to June, 2017 (September).

Accident Routes	2014	2015	2016	2017	Total
Walajapet-Poonamalee	377	315	98	19	809
Ulundurpet – Padalur	114	719	759	474	2066
Trichy-Karur	141	184	170	117	612
Trichy-Dindigul	181	120	41	6	348
Tiruthani – Chennai	38	158	55	63	314
Tindivanam – Ulundurpet	128	282	366	225	1001
Tindivanam – Pondicherry	168	152	162	5	487
Thanjavur – Trichy	171	149	0	2	322
Tambaram – Tindivanam	97	174	201	152	624
Salem-Ulundurpet	242	663	648	341	1894
Salem-Kumarapalayam	108	337	311	162	918
Salem – Namakkal	172	364	368	231	1135
Padalur-Trichy	99	220	173	103	595
Namkkal-Karur	111	215	150	118	594
Madurai-Thoothukudi	84	198	41	6	328
Kumarapalayam-Chengapalli	148	425	390	224	1187
Krishnagiri- Thoppur	260	781	804	493	2338
Krishnagir – Hosur	350	325	22	143	840
Karur- Dindigul	100	208	234	147	689
Dindigul-Periyagulam	52	265	190	9	516
Dindigul – Samayanallur	101	272	171	0	544
Chennai - Dada	179	228	237	85	729
Chengapalli - Coimbatore Bypass	45	96	476	198	815



This comparison denotes that the National highway **Salem to Namakkal** results in highest death rates through road accidents when compared to other national highways from the year 2014 2017.



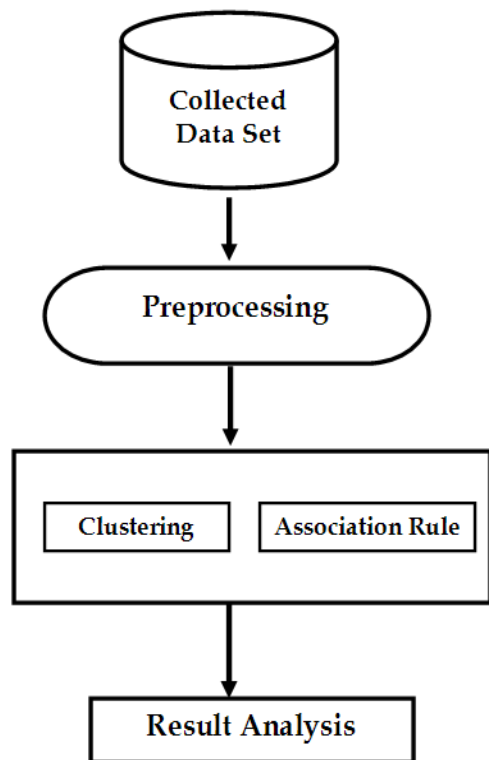
This comparison figure denotes that the National highway Kumarapalayam to Chengapalli results in highest injuries through road accidents when compared to other national highways from the year 2014 2017. Based on the above comparisons, to improve the

accuracy of the result and improvement, this study has to go further with the clustering techniques.

IV. Survey of Algorithms proposed

1 Preprocessing

Data preprocessing is one of the important task to be done prior to analyzing data. Data preprocessing leads to data cleaning such as removing noise, handling missing values and applying various data transformations in order to get the data ready for the analysis



4.2 Clustering algorithm

The objective of clustering algorithm is to divide the data into different clusters or groups such that the objects within a group are similar to each other whereas objects in other clusters are different from each other.[6]

Partitioning based Clustering algorithms

All objects are considered initially as a single cluster. The objects are divided into no of partitions by iteratively locating the points between the partitions. The partitioning algorithms like K-means, K-medoids and K-modes.[7]

Hierarchical Clustering algorithms:

There are two approaches to perform Hierarchical clustering techniques Agglomerative (top-bottom) and Divisive (bottom- top). In Agglomerative approach, initially one object is selected and successively merges the neighbor objects based on the distance as minimum,

maximum and average. The process is continuous until a desired cluster is formed. The Divisive approach deals with set of objects as single cluster and divides the cluster into further clusters until desired no of clusters are formed.

Density based Clustering algorithms:

Data objects are categorized into core points, border points and noise points. All the core points are connected together based on the densities to form cluster.

Grid based Clustering algorithms:

Grid based algorithm partitions the data set into no number of cells to form a grid structure. Clusters are formed based on the grid structure. To form clusters Grid algorithm uses subspace and hierarchical clustering techniques. Compare to all Clustering algorithms Grid algorithms are very fast processing algorithms.

Model based Clustering algorithms:

Set of data points are connected together based on various strategies like statistical methods, conceptual methods, and robust clustering methods. There are two approaches for model based algorithms one is neural network approach and another one is statistical approach.

4.3 Association rules

Association rules are if/then statements that help to uncover relationships between unrelated data in a database, relational database or other information repository. Association rules are used to find the relationships between the objects which are frequently used together. Applications of association rules are basket data analysis, classification, cross-marketing, clustering, catalog design, and loss-leader analysis etc.[8]

There are two basic criteria that association rules uses, support and .confidence. It identifies the relationships and rules generated by analyzing data for frequently used if/then patterns.

$$\begin{aligned} \text{support}(A \Rightarrow B) &= P(A \cup B) \\ \text{confidence}(A \Rightarrow B) &= P(B|A) \\ &= \frac{P(A \cup B)}{P(A)} \end{aligned}$$

V. CONCLUSION

Data mining has proved as a consistent technique in analyzing the road accident data. Using data mining techniques, many researches has been carried out to analyze road accident data of different countries. Certain data mining techniques such as clustering, classification and association rule mining are widely used in the literature to identify the factors, reasons behind the severity of road accidents. In this paper, study of clustering algorithms has been highlighted. Hope this algorithm will give much accuracy of results and would reduce the road accidents.

VI. REFERENCES

- [1]. <http://indianexpress.com/article/india/road-accidents-in-india-2016-17-deaths-on-roads-every-hour-chennai-and-delhi-most-dangerous-4837832/>
- [2]. Sachin Kumar , Durga Toshniwal ,“Analyzing Road Accident Data Using Association Rule Mining”, IEEE 2015 International Conference on Computing, Communication and Security (ICCCS).
- [3]. S. Krishnaveni and M. Hemalatha. “A perspective analysis of traffic accident using data mining techniques”. International Journal of Computer Applications,
- [4]. X-F Zhang, and L. Fan. “A decision tree approach for traffic accident analysis of Saskatchewan highways”, 26th Annual Canadian Conference on IEEE, 5-8 May, 2013
- [5]. www.nhai.org.in/home.aspx
- [6]. Fraley C, Raftery AE. “Model-based clustering, discriminant analysis, and density estimation”. J Am Stat Assoc. 2002;97(458)
- [7]. T.Sajana, C.M. Sheela Rani and K.V.Narayana .”A Survey on Clustering Techniques for Big Data Mining”
- [8]. Trupti A. Kumbhare Prof. Santosh V. Chobe “An Overview of Association Rule Mining “Algorithms” International Journal of Computer Science and Information Technologies, (IJCSIT)