International Journal on Future Revolution in Computer Science & Communication Engineering
Volume: 4 Issue: 11
ISSN: 2454-4248
32 – 36

Identifying Road Accidents Severity Problems Using Data Mining Approaches

Mr. P. Sampath*, Dr. Y. Jeevan Nagendra Kumar**

* Department of Information Technology, Gokaraju Rangaraju Institute of Engineering and Technology, Hyderabad

** Professor, Department of Information Technology, Gokaraju Rangaraju Institute of Engineering and Technology, Hyderabad

Abstract:- Roadway traffic safety is a major concernfor transportation governing agencies as well as ordinarycitizens. In order to give safe driving suggestions, carefulanalysis of roadway traffic data is critical to find outvariables that are closely related to fatal accidents. Inthis paper we apply statistics analysis and data miningalgorithms on the FARS Fatal Accident dataset as an attempt to address this problem. The relationship betweenfatal rate and other attributes including collision manner,weather, surface condition, light condition, and drunkdriver were investigated. Association rules were discoveredby Apriori algorithm, classification model was built byNaive Bayes classifier, and clusters were formed by simpleK-means clustering algorithm. Here we are also using one more classification technique for comparing with Naïve bayes classifier. Certain safety drivingsuggestions were made based on statistics, associationrules, classification model, and clusters obtained.

Keywords :- *Roadway fatal accidents, association, classification, clustering*

*****

I. INTRODUCTION

Road safety means development and management of roads, provision of safer vehicles, and a comprehensive response to accidents [1]. Modern traffic management systems, such as real-time adjustment of traffic flow, model predictive control (MPC) technique in traffic light control, tolling strategy, etc., can be used in design and maintenance of roads, and also for producing safer vehicles. BRT system of Ahmadabad city has achieved its objective of providing a safe mode of transportwith more than 50% decrease in road traffic [2].

According to the National Crime Records Bureau [3], there were 39,344 road accidents, which resulted in the death of 14,966 persons. Another point of concern is that, while 8.9% of all accidents in the country occur in the state, the percentage of all deaths is higher at 10.8% t. Statistics also reveal that most accidental deaths involve people traveling in three-wheelers. More than 25% of accident deaths involving passengers of auto-rickshaws throughout the country are in Andhra Pradesh [3]. About 1,734 persons died in road accidents involving auto-rickshaws, and the state has the highest number of such deaths in the country [3].

According to the report given in [4], road accidents are the ninth leading cause of death in 2004 and expected to be fifth leading cause of death by 2030 worldwide. This paper proposes a framework that is based on the cluster analysis using K-medoids and expectation maximization (EM) and association rule mining using a priori algorithm. Association rule mining is further applied on these clusters to generate association rules. Performance is analyzed using precision and recall measures.

The main objective of this research is to investigate the role of human-, vehicle-, and infrastructure-related factors in accident severity by applying machine learning techniques on road accident data. The overall architecture of the proposed system is shown in Fig. 1. The steps include data cleaning, data transformation, relevance analysis, clustering, association rules generation, and finally performance evaluation.

II. LITERATURE SURVEY

SAMI AYRAMO, PASI PIRTALA [5]: Results from the research study on applying large-scale data mining methods into analysis of traffic accidents on the Finnish roads are presented in . The main intension is to show that the selected data mining methods are able to produce understandable patterns from the data, finding more fertilized information could be enhanced with more detailed datasets.

S. SHANTHI, DR.R. GEETHA RAMANI [6]: The work of [emphasizes the importance of data mining classification algorithms in predicting the vehicle collision patterns occurred in training accident dataset. They followed a stepwise procedure which finally yields the required accident analysis results: data cleaning, data transformation, and relevance analysis. The feature selection algorithms have been explored to improve the classifier accuracy.

S. SHANTHI, R. GEETHA RAMANI [7]: The research work in emphasizes the significance of data mining classification algorithms in predicting the factors which influence the road traffic accidents specific to injury severity. Further they applied feature selection methods to select the relevant road accident-related factors and Meta

 ISSN: 2454-4248
32 – 36

classifier Arc-X4 to improve the accuracy of the classifier. In order to improve road safety,

The Luis Martin[8]: Analyzed the Andalusia Complementary Road Network, by using advanced data mining techniques in order to discover hidden relationships between characteristic of the roads, ESM, and crashes.

Tibebe Beshah, Shawndra Hill [9]: The research work inis that accidents are not randomly scattered along the road network, and that drivers are not involved in accidents at random. Authors focused on the contribution of road-related factors to accident severity in Ethiopia.

Amira A. El Tayeb, Vikas Pareek, Abdelaziz Araar [10]: Work presented inis about discovering interesting rules from a set of generated rules using both association rule algorithms.

K. Geetha, C. Vaishnavi [11]: Work reported in is to reduce the number of road accidents in main cities of Tamil Nadu. They used WEKA tool and H-DTANN techniques in order to predict the road accident injury levels.

III. IMPLEMENTATION METHODOLOGIES

Dataset Creation

A total of 30 attributes that focus on various criteria, such as accident-specific attributes, driver specific attributes, FIR details, circumstance-specific attributes, and other attributes given in the FIR report, form the input dataset.

Data Preprocessing

Data preprocessing helps to remove noise, missing values, and inconsistencies. Missing values are replaced with NULL. Also each attribute data is discredited in order to make it appropriate for further analysis. Bellow information represents the attributes of dataset in different conditions.

Collision Conditions: Angle-front-to-side, Angle-front-to-side, Angle-front-to-side, Angle-front-to-side, Front-to-front, Front-to-rear, Not collision with vehicle in transport, Other, Rear-to-rear, Rear-to-side, Sideswipe-opposite direction, Sideswipe-same direction.
Light Condition: Dark, Dark but lighted, Dawn, Daylight, Dusk, Unknown.
Weather Condition: Blowing sand soil dirt, clear cloud no adverse condition, Frog smog smoke, Other, Rain mist, severe crosswinds, Snow or blowing snow, unknown.
Surface Condition: Dry, Ice frost, oil, Other, Sand dirt mud gravel, Snow or slush, unknown, Water, Wet.
Speed Conditions: 5, 10, 15, 20, 25, 35, 45, 55, 60, 65, 70, 80, 85, 95, 99.
And finally drunk and drive conditions also has taken.

AprioriAssociation

- Apriori is an algorithm for frequent item set mining and association rule learning over transactional the databases.
- It proceeds by identifying the frequent individual items in the database and extending them to larger and larger item sets as long as those item sets appear sufficiently often in the database.
- The frequent item sets determined by Apriori can be used to determine association rules.

Naive Bayes Classification

Step1: Convert the data set into a frequency table.
Step2: Create Likelihood table by finding the probabilities like Overcast probability = 0.29 and probability of playing is 0.64.

Step 3: Now, use Naive Bayesian equation to calculate the posterior probability for each class. The class with the highest posterior probability is the outcome of prediction.

$$\text{Precision} = TP/(TP + FP) \quad (1)$$

$$\text{Recall} = TP/(TP + FN) \quad (2)$$

$$\text{F-Measure} = (1+\alpha)/((1/\text{precision})+(\alpha/\text{Recall})) \text{ where } \alpha = 1 \quad (3)$$

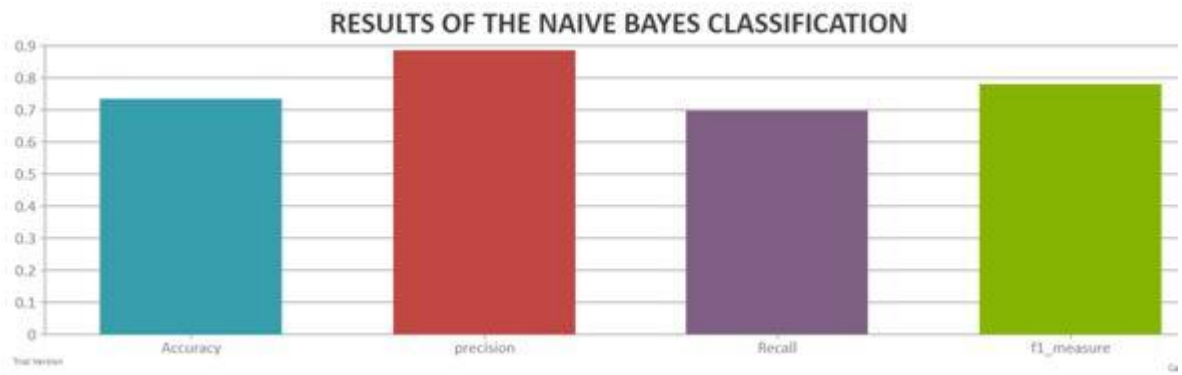


Fig: Performance Based analysis of Naïve Bayes Classification

 ISSN: 2454-4248
32 – 36

K-means Clustering algorithm

- K-means clustering is an iterative processing. The algorithm selects k initial cluster centers, and all data objects are allocated into the cluster whose center has the shortest distance to them.
- For each data object, it will be assigned to the cluster whose center has shortest distance to the object. Data objects and centers can be denoted as multi-dimensional vectors, and their distances can be measured using the square of Euclidean distance.
- After a round of clustering, centers of clusters are updated. Particularly, the new center of a cluster is generated by averaging each element over all data object vectors in the same cluster.

Algorithmatic Attribute Clustering Process

Weather Condition at the time of accident, in which 34.6% of accidents occurred when weather is cool, 33.5% of accidents occurred when weather is clear, and 31.9% of accidents occurred when weather is hot.

Lightening Condition at the time of accident, in which 33% of accidents occurred when lightening is dark, 25.5% of accidents occurred in dim light, and 41.5% of accidents occurred in bright light.

Type of accident, In which 69.5% of accidents occurred because of rash driving, 3.7% of accidents occurred because of single vehicle runoff, 0.33% of accidents occurred because of vehicle skidding, 0.33% of accidents occurred because of overlooking, 6.2% of accidents occurred because of overriding, 11.6% of accidents occurred because of hit by other vehicles, 8% of accidents occurred during lane change, and 0.34% of accidents are because of sudden turn back or animal hit, wrong direction.

Speed limit of vehicles at the time of accident, in which 32% of accidents occurred at normal speed limit, 44.5% of accidents occurred at high speed limit, and 33.5% of accidents speed limit value is missing.

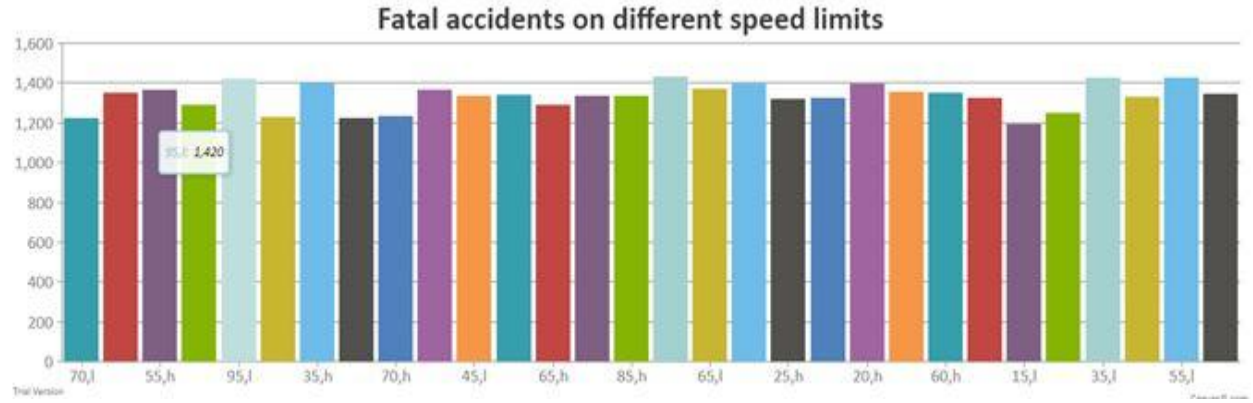
Results and Analysis

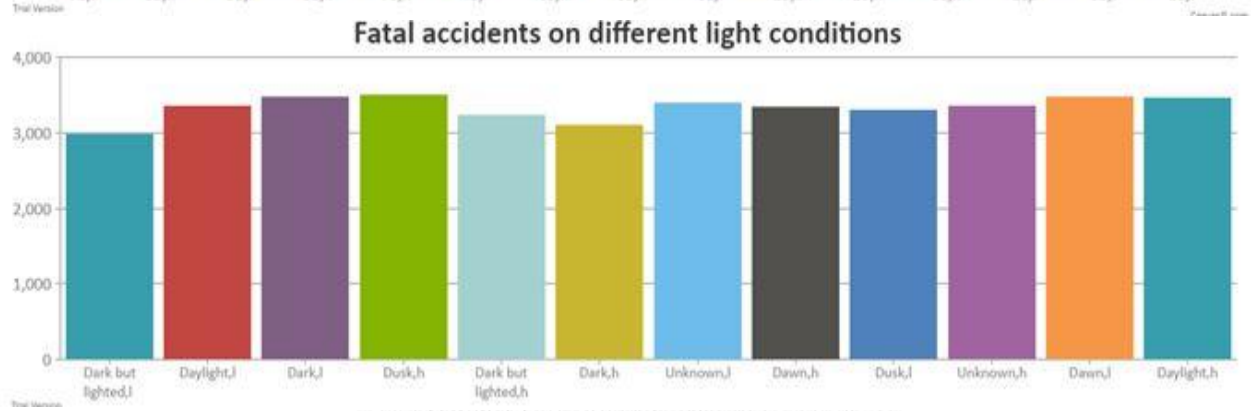
Road accidents and injuries occur because of human fault or vehicle fault or infrastructure fault or sometimes combinations of these factors. Each of these factors individually or in combination may cause accident. It was observed from the dataset that accidents mainly occurred because of combination of human fault and vehicle fault.

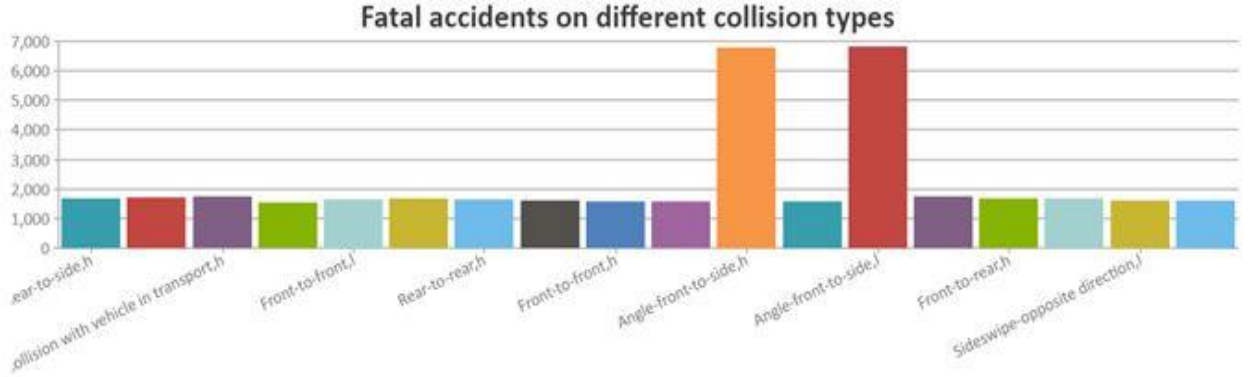
Human alone factors such as “helmet and seat belt not used” are not reported in the FIRs and as such are not known. Major contributing factors for accidents, highest being rash driving of the people.Analysis like type of vehicles (two-wheeler, car, bus, lorry, jeep, truck, etc.) is not given in the FIR report, and as such, analysis is not done. Above attributes presents in clustering process percentage distribution of accidents on various criteria, speed limit, weather and light conditions.

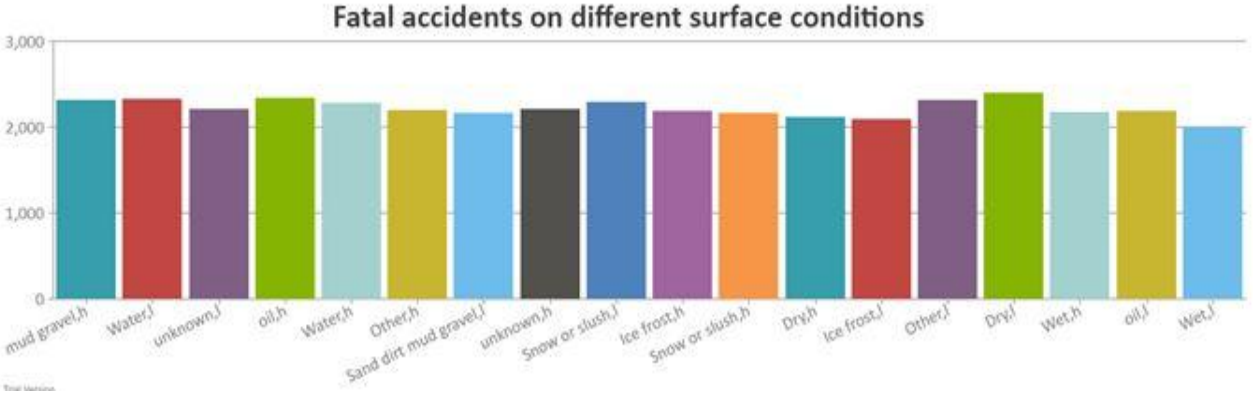
Similar analysis is done on other criteria such as distribution of accidents by time of accidents and deceased age, distribution of accidents by month and weather during the accident, distribution of accidents by lightness and speed limit, distribution of accidents by accident type (human factors), distribution of accidents by day of accident and deceased age, distribution of accidents by deceased emotions, distribution of accidents by hospital reported and ambulance used. Because of space limit, all graphs are not listed here.

ISSN: 2454-4248
32 – 36

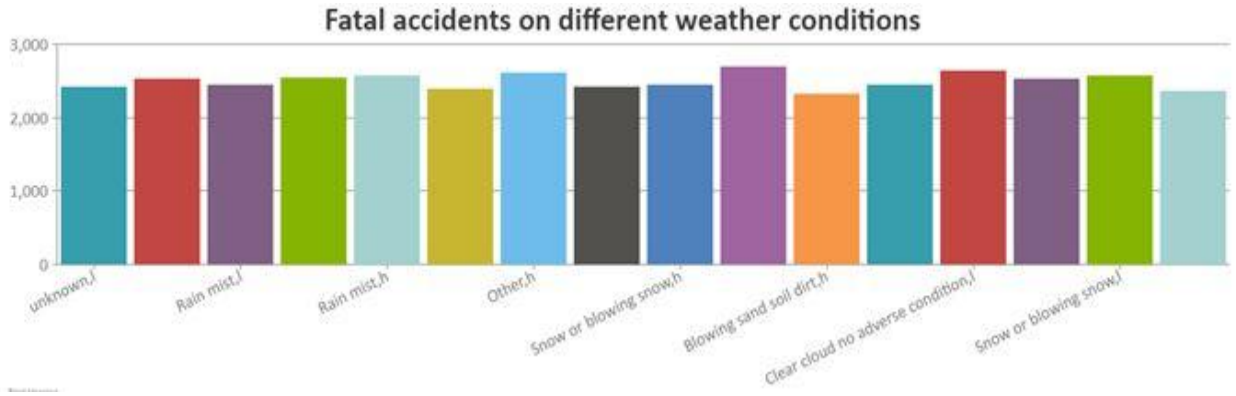








 ISSN: 2454-4248
32 – 36



IV. CONCLUSIONS

The aim of this paper is to generate association rules that will analyze how to discover hidden patterns that are the root causes for accidents among different combinations of attributes of a larger dataset. Density histograms for visualizing region wise such as fatal versus weather, fatal versus time, time versus day, fatal versus month, fatal versus traffic, and fatal versus age are performed. Percentage distribution of accidents on various criteria, speed limit and injury severity, distribution of accidents by time of accidents and deceased age, distribution of accidents by month and weather during the accident, distribution of accidents by lightness and speed limit, distribution of accidents by accident type (human factors), distribution of accidents by day of accident and deceased age, distribution of accidents by deceased emotions, distribution of accidents by hospital reported and ambulance used is also made. Future work is to make analysis on road accidents' dataset by considering more features and clusters and also to use deep learning techniques so as to better cluster the records.

References

[1] Road safety and traffic management : Report of the committee Planning Commission, Government of India in February 2007 (2007)

[2] Rayle L, Pai M. Scenarios for future urbanization: carbon dioxide emissions from passenger travel in three Indian cities. Transportation Research Record: Journal of the Transportation Research Board, 2193:124–131 (2010)

[3] http://www.deccanchronicle.com/130629/news-current-affairs/article/andhra-pradesh-ranked- 3rd-road-accidents last accessed June 29th 2013

[4] Road Accidents in India Issues & Dimensions, Ministry of Road Transport & Highways Government of India (2014)

[5] SAMI AYRAMO, PASI PIRTALA, Mining road traffic accidents, Reports of the Department of Mathematical Information Technology Series C. Software and Computational Engineering No. C. 2/2009 (2009)

[6] S. SHANTHI, DR.R. GEETHA RAMANI, Classification of Vehicle Collision Patterns in Road Accidents using Data Mining Algorithms, International Journal of Computer Applications (0975–8887) Volume 35– No.12, December 2011 (2011)

[7] S. SHANTHI, R. GEETHA RAMANI, Feature Relevance Analysis and Classification of Road Traffic Accident Data through Data Mining Techniques, Proceedings of the World Congress on Engineering and Computer Science 2012 Vol I WCECS (2012)

[8] Luis Martín, Leticia Baena, Laura Garach, Griselda López, Juan de Oña Using Data Mining Techniques to Road Safety Improvement in Spanish Roads, Volume 160, Pages 607–614, XI Congreso de Ingenieria del Transporte (CIT 2014)

[9] Tibebe Beshah, Shawndra Hill, Mining Road Traffic Accident Data to Improve Safety: Role of Road- related Factors on Accident Severity in Ethiopia, AAAI Spring Symposium Series (2010)

[10] Amira A. El Tayeb, Vikas Pareek, Abdelaziz Araar Applying Association Rules Mining Algorithms for Traffic Accidents in Dubai, International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-5 Issue-4, (2015)

[11] K. Geetha, C. Vaishnavi Analysis on Traffic Accident Injury Level Using Classification, International Journal of Advanced Research in Computer Science and Software Engineering, Volume 5, Issue 2, (2015)

[12] Jiawei Han and Micheline Kamber, Data Mining Concepts and Techniques, 2 ed, Elsevier publishers.