Communications on Applied Electronics (CAE) – ISSN : 2394-4714
Foundation of Computer Science FCS, New York, USA
Volume 6 – No.4, December 2016 – www.caeaccess.org

A Survey on Extracting Hidden Patterns within Road Accident Data using Machine Learning Techniques

S. Vasavi
VR Siddhartha College of Engineering
Vijayawada
Andhra Pradesh, India

ABSTRACT
Road Accidents may not be stopped altogether, but can be reduced. Driver emotions such as sad, happy, anger etc can be one reason for accidents. At the same time environment conditions such as weather, traffic on the road, load in the vehicle, type of road, health condition of driver, speed etc can also be the reasons for accidents. Hidden patterns in accidents can be extracted so as to find the common features between accidents. This paper presents the literature survey and the results of the framework from the research study on road accident data of major national highways that pass through Krishna district (18 stations) for the year 2013 by applying machine learning techniques into analysis. Results showed that the selected machine learning techniques are able to extract hidden patterns from the data. Density histograms are used for accident data visualization.

General Terms
Machine Learning Techniques

Keywords
Machine learning techniques, Road Accident data analysis, Pre processing, Clustering, Association rule mining, Visualization

1. 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. In 2001, 28% of India's people lived in cities, but this is expected to rise to 40% by 2040[2]. BRT system of Ahmadabad city has achieved its objective of providing a safe mode of transport with more than 50% decrease in roads traffic [2]. During the last decade, 2001 to 2011, the number of registered motor vehicles recorded a CAGR of 9.9 per cent, while the road network increased at a CAGR of 3.4 per cent as shown in figure 1. Road traffic injuries are the eighth leading cause of death globally, and the leading cause of death for young people aged 15–29. More than a million people die each year on the world's roads, and the cost of dealing with the consequences of these road traffic crashes runs to billions of dollars. Current trends suggest that by 2030 road traffic deaths will become the fifth leading cause of death unless urgent action is taken. In road traffic, accident fatalities, deaths are mainly due to head injuries, especially in case of two wheelers.

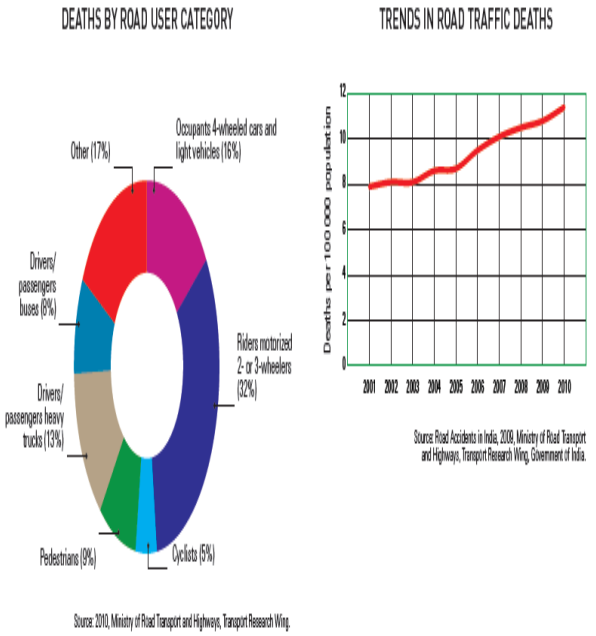

Figure 1: Deaths by road user category and trends in road traffic deaths (Source: Offices of State Transport Commissioners/UT administrations)

According to the National Crime Records Bureau[11], there were 39,344 road accidents in the state which resulted in the death of 14,966 persons. Another point of concern is that while 8.9 per cent of all accidents in the country occur in the state, the percentage of all deaths is higher at 10.8 per cent. Statistics also reveal that most accidental deaths involve people travelling in three-wheelers. More than 25 per cent of accidents deaths involving passengers of auto rickshaws throughout the country are in Andhra Pradesh. About 1,734 persons died in road accidents involving auto rickshaws and the state has the highest number of such deaths in the country [11]. According to the report given in [3] road accidents is the 9th leading cause of death in 2004 and expected to be 5th leading cause of death by 2030 worldwide. Figure 2 presents Statistics of Number of Road Accidents, Number of Persons Killed and Number of Persons Injured Per Lakh Population: 1970 – 2011.

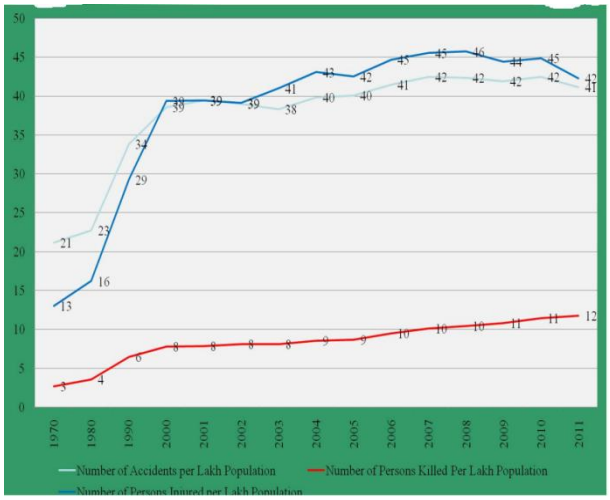

Figure 2: Number of Road Accidents, Number of Persons Killed and Number of Persons Injured Per Lakh Population: 1970 – 2011

As per the same report, figure 3 presents causes of road accidents in 2011.

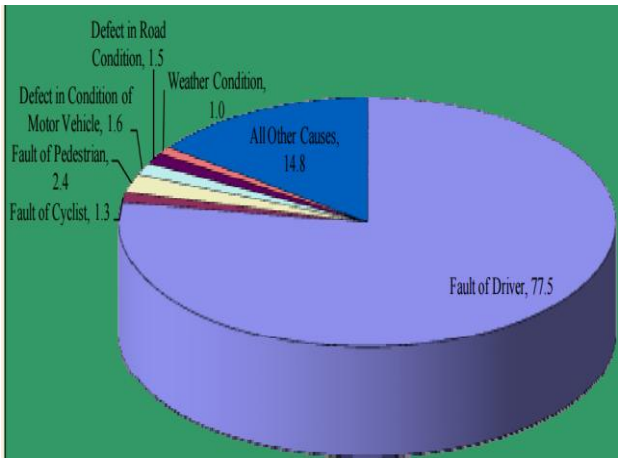

Figure 3: Causes of Accidents in 2011 year

The paper is organized as follows: Section 2 presents literature survey on various existing methods for accident data analysis. Methodology of proposed system is described in Section 3. Section 4 includes results obtained from our proposed system. Conclusions and future work are given in Section 5.

2. LITERATURE SURVEY

According to the work given in [4], presents the results from the research study on applying large scale data mining methods into analysis of traffic accidents on the Finnish roads. 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 data sets. 7 clusters such as Motor- and semi-motorway accidents, Alcohol-involved connecting road accidents, animal hits, Built-up area accidents, Low-speed multi-lane roadway accidents, Low traffic volume regional highways, Fatal main road accidents are considered. The work of [5], emphasizes the importance of Data Mining classification algorithms in predicting the vehicle collision patterns occurred in training accident data set. They followed a stepwise procedure which finally yields the required accident analysis results: Data Cleaning, Data Transformation and Relevance analysis. The feature selection algorithms including CFS, FCBF, Feature Ranking, MIFS and MODTree have been explored to improve the classifier accuracy. Classification algorithms like C4.5, ID3, C&RT, CS-MC4, Decision List, Naïve Bayes, and random Tree are compared and Random Tree got the best results. The important criteria under which the different feature selection or the classification algorithms can be done are explained well. The data preparation steps gave a clear vision of how the real data can be analyzed. The research work in [6] 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 classifier Arc-X4 to improve the accuracy of the classifier. The results have been evaluated using the accuracy measures such as Recall and Precision. Among the algorithms Random Tree classifier using Arc-X4 Meta gives high accuracy of 99.73% with 0.27% misclassification rate. In order to improve road safety, the authors of [7] 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. The basic hypothesis of research in [8] is that, accidents are not randomly scattered along the road network, and that drivers are not involved in accidents at random. This paper focus on the contribution of road related factors to accident severity in Ethiopia. This will help to identify the parts of a road that are risky, thus supporting traffic accident data analysis in decision making processes. Work presented in [9], is about discovering interesting rules from a set of generated rules using both association rules algorithms. Work reported in [10] is about investigation and analysis is to reduce the number of road accidents in main city of Tamilnadu by finding risks and circumstances which can be shown to be regular contributing factors to road accidents. Data mining WEKA tools and H-DTANN techniques are used to predict the road accident injury levels.

3. PROPOSED SYSTEM

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 figure 4. The steps include data cleaning, data transformation, relevance analysis, clustering, Association rules generation and finally performance evaluation.

3.1 Visualization

Graphical representation techniques, such as histograms, pie charts, scatter Plots etc can be used for visualizing data. These visual representations will help in identifying the risk of the accident immediately by government officials. In this work we generated density histograms for visualizing region wise results basing on the following criteria using Matlab software.

1. Age and Fatal
2. Traffic and Fatal
3. Day, Time and fatal
4. Month and Fatal
5. Time and Fatal
6. Weather and Fatal

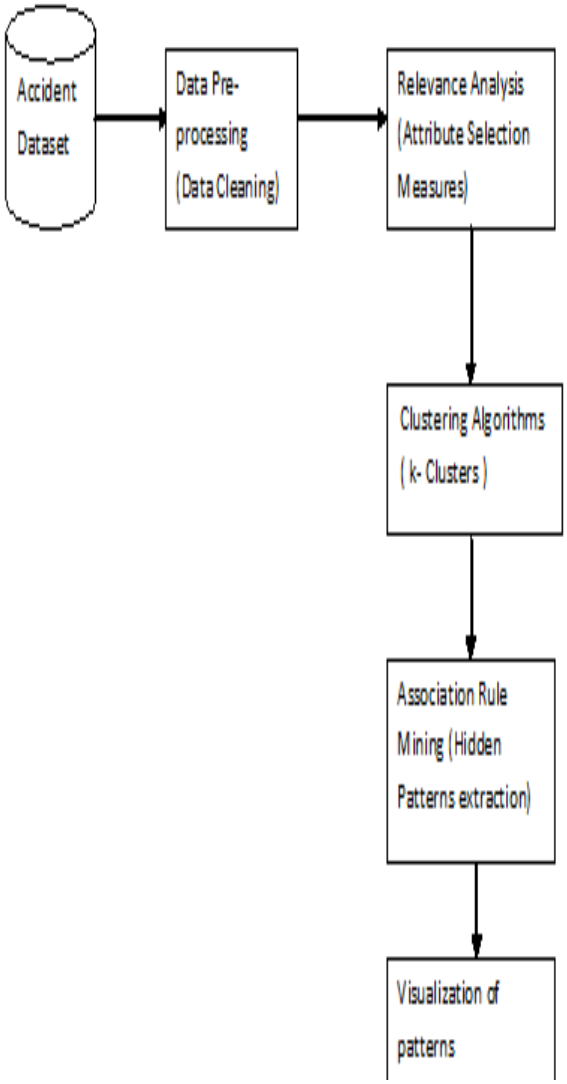

Figure 4: Proposed System Architecture

The following figure 5 to figure 10 present density histograms for sample dataset

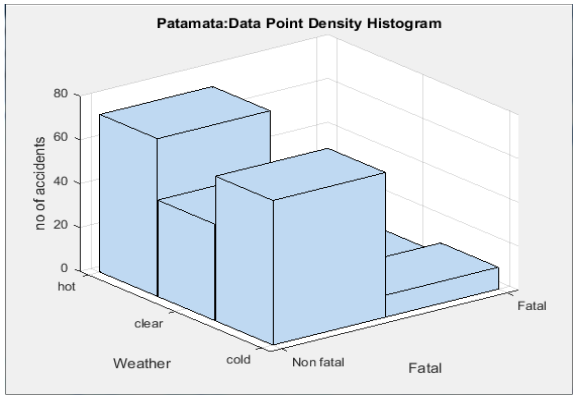

Figure 5: Fatal Vs Weather

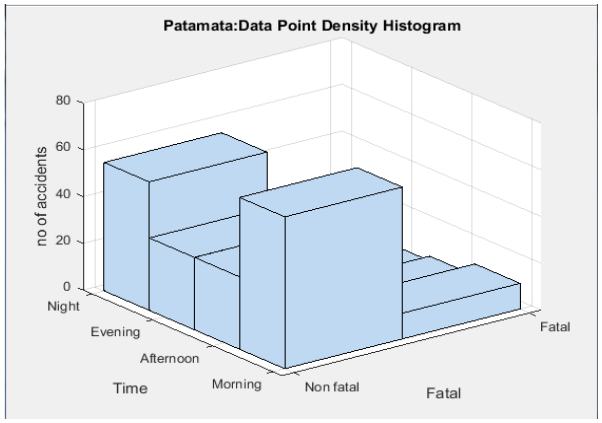

Figure 6: Fatal Vs time

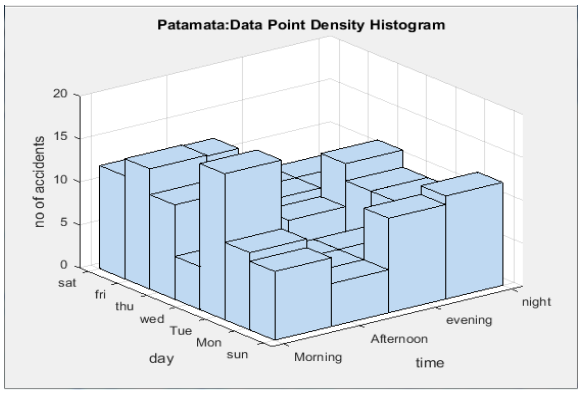

Figure 7: Time Vs Day

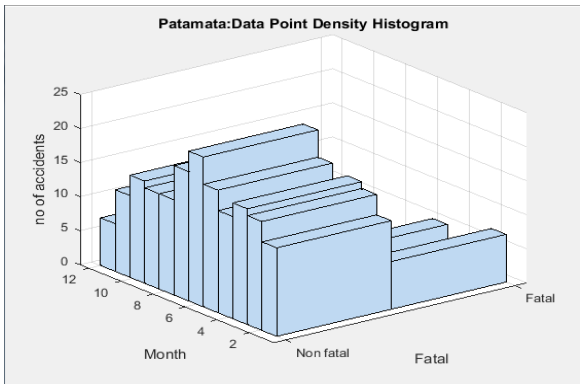

Figure 8: Fatal Vs Month

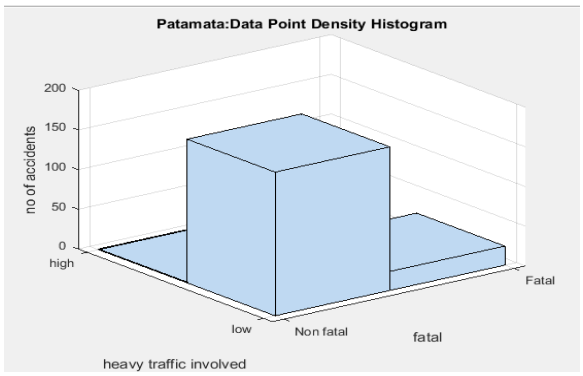

Figure 9: Fatal Vs Traffic

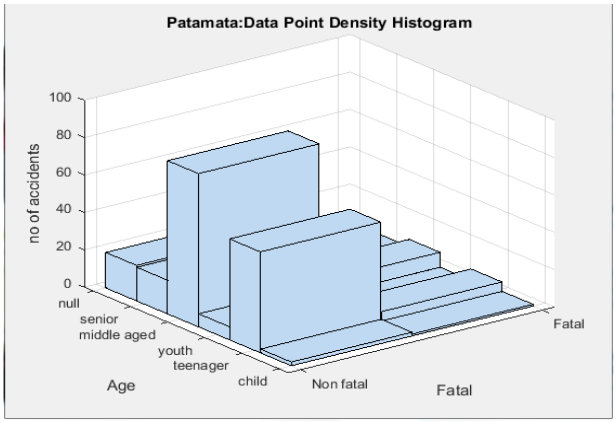

Figure 10: Fatal Vs Age

4. RESULTS AND ANALYSIS

Analysis such as, type of vehicle (2-wheeler, car, bus, lorry, jeep, truck etc) are not given in the FIR report and as such analysis is not done. Figure 11 to figure 24 presents Percentage distribution of accidents on various criteria such as 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.

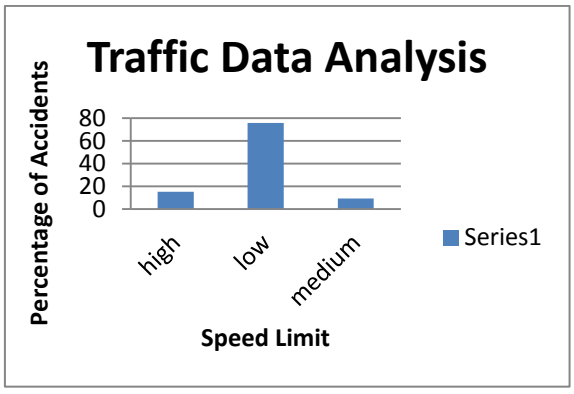

Figure 11: Percentage distribution of accidents by speed limit

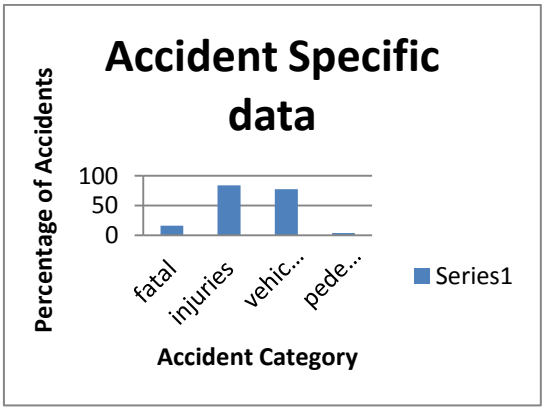

Figure 12: Percentage distribution of accidents by injury severity

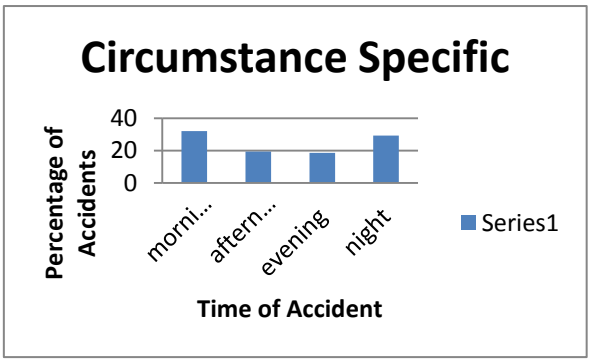

Figure 13: Percentage distribution of accidents by time of accidents and deceased age

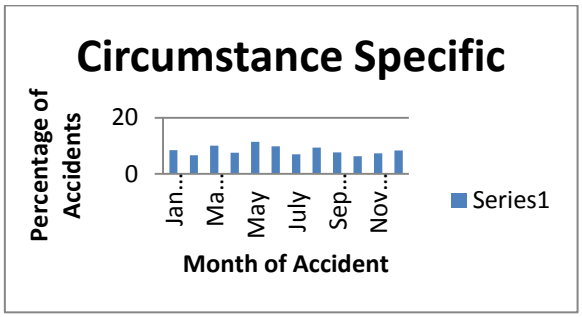

Figure 14: Percentage distribution of accidents by month

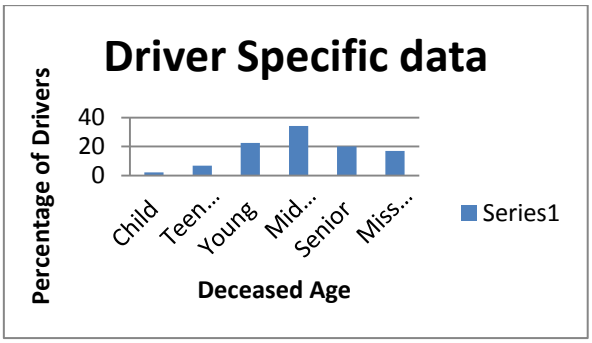

Figure 15: Percentage distribution of accidents by deceased age

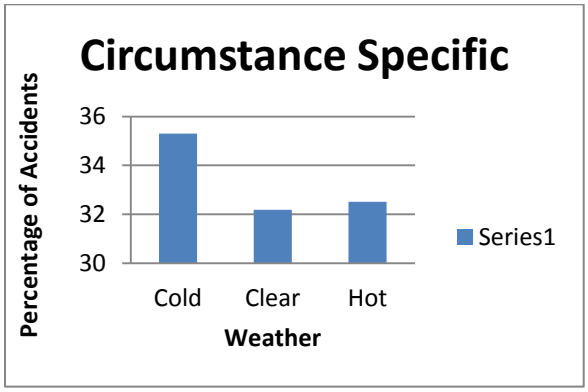

Figure 16: Percentage distribution of accidents by weather during the accident

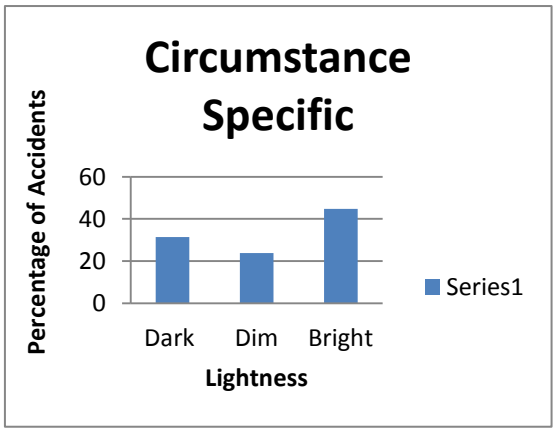

Figure 17: Percentage distribution of accidents by lightness

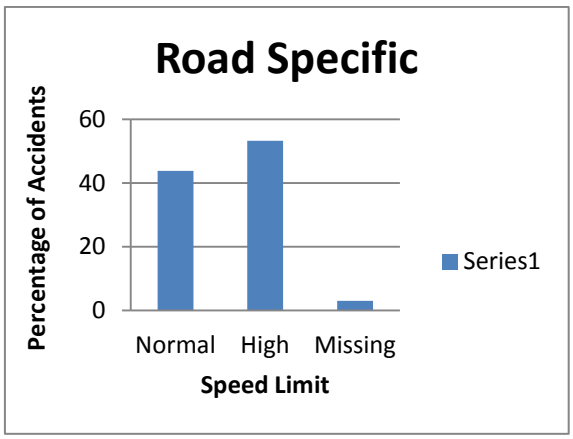

Figure 18: Percentage distribution of accidents by speed limit

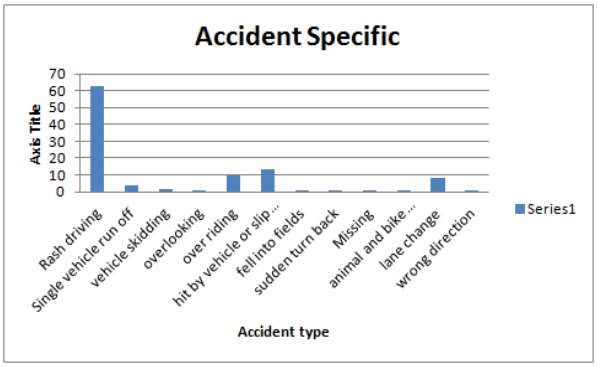

Figure 19: Percentage distribution of accidents by accident type (Human factors)

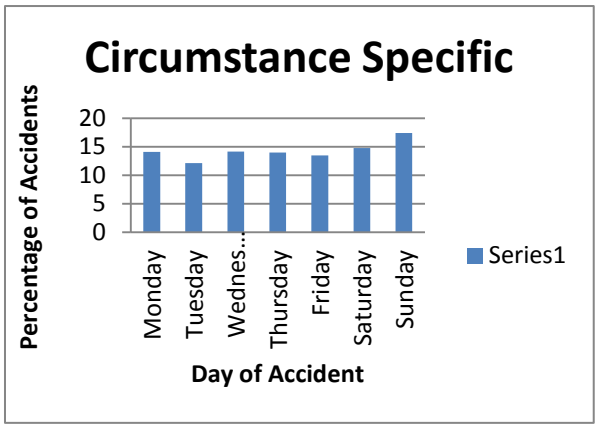

Figure 20: Percentage distribution of accidents by day of accident

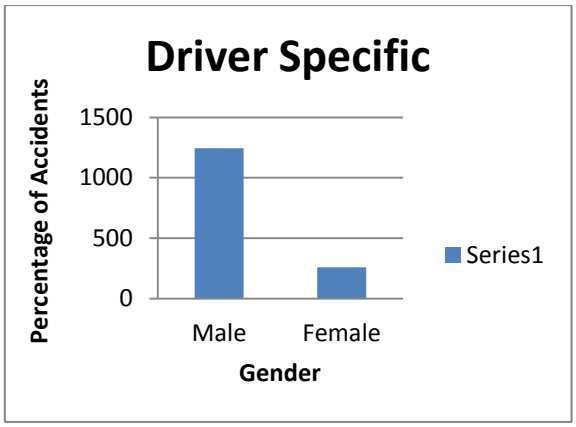

Figure 21: Percentage distribution of accidents by gender

Figure 22: Percentage distribution of accidents by deceased emotions

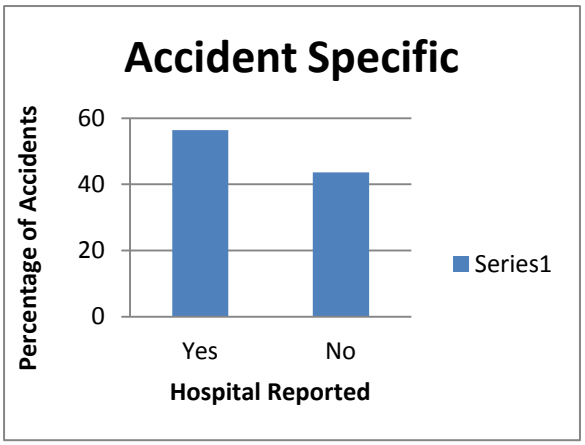

Figure 23: Percentage distribution of accidents by hospital reported

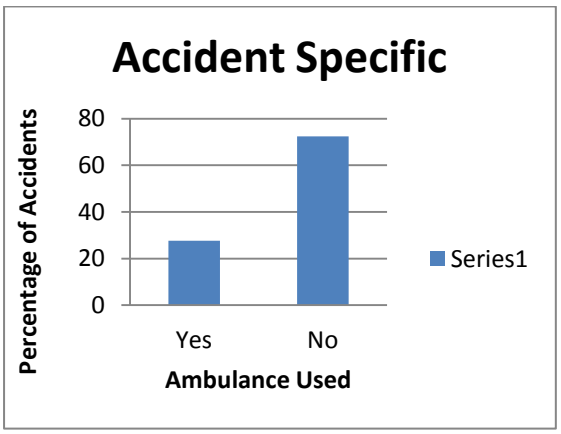

Figure 24: Percentage distribution of accidents by ambulance used

From the data analysis, accident distribution is even in normal days and it is observed to be higher in weekends. Accidents occurrence is high at cold nights compared to hot and clear conditions. Most accident prone area is observed to be Kesarapalli village road and Venkateswara theater in Gannavaram. It is observed to be fatal accidents are high among the old aged group and non-fatal in young aged and middle aged people. Accidents are high in the month of August and low in the month of June. Female involved in accidents are observed to be 20.16% of overall accidents to 73.45% of male.

5. CONCLUSIONS AND FUTURE WORK

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 Vs Weather, Fatal Vs time, Time Vs Day, Fatal Vs Month, Fatal Vs Traffic, Fatal Vs 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 are 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.

6. ACKNOWLEDGMENTS

I thank University Grants Commission (UGC), for funding this project. I also thank Police authorities, Andhra Pradesh for providing the required information. I am also thankful to the management of Siddhartha Academy for providing me resources and environment for successfully completing this project. Finally I thank my students who helped me during implementing this project.

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

[4] 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)

[5] 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)

[6] 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)

[7] 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)

[8] 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)

[9] 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)

[10] 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)

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

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