

Analysis of Road traffic fatal accidents using data mining.

Miss. Pooja Suhas Purwat, Miss. Sonali Anil Bhosale, Miss. Prajakta Ravijit Teke, Miss. Tejashri Bhalchandra Bagal

Guide: Dhotre V.A.

Computer Science & Engineering, Solapur University

Pooja.purwat@gmail.com

sonalibhosaleppr@gmail.com

prajaktateke15@gmail.com

tejubagal@gmail.com

Abstract: Data mining is a promising area for dealing with the increased, stored data that has been generated in our times. It is the extraction of implicit, previously unknown and useful data. In this paper we have analyzed some of the data mining techniques, tools, applications and search engines for accident investigation and traffic analysis. Most of the accident investigation methodologies are based on scenarios of the accident occurrence and simulation of accident situation. The costs of fatalities and injuries due to traffic accident have a great impact on society. Engineers and researchers in the automobile industry have tried to design and build safer automobiles, but traffic accidents are unavoidable. In recent years, researchers have been utilizing real-life data in studying various aspects of traffic accidents.

So measures have to be taken to reduce accidents. It is important that the measures should be based on scientific and objective surveys of the causes of accidents and severity of injuries. Our study highlights various tools, techniques and applications of data mining in accident analysis will eliminate deficiencies of other techniques but covers their advantages. Our main aim is to overcome the death rate and the increased rate of loss of lives by means of using some tools, techniques or various algorithms in the field of data mining using the traffic data bases.

Keywords: Road Traffic Accidents, Data Mining, Influential Factor, WEKA, Data Mining Techniques

I INTRODUCTION

Road safety experts and researchers deal with large volumes of quantitative information and collected statistics, in order to understand and estimate the social and economic cost of the accidents and to be able to introduce safety plans in order to prevent or reduce occurrences of accidents. The road traffic and accident statistics must be presented in such a way to make it easier to be both recognized and interpreted by a human operator. Previous works on accident analysis included statistical methods and formal techniques. Statistics tables and ordinary charting techniques are not sufficient for present day requirements and this causes difficulties in the effective visualization of results and patterns. Another disadvantage is that ordinary methods limit human involvement in the exploration tasks. In this

paper we have analyzed various studies done by various authors in accident databases and also measures proposed to prevent the accident and its death rate.

Simulation can help define deficiencies in the early design process with relatively low cost. A case is considered in this study is road traffic accident (RTA) reporting system in sample of Iran's road. Two main aspects of performance of the RTA system require improvement. Process execution need to speed up which provide a faster and more efficient service to vehicle drivers and the relatively high staffing cost associated with the process should reduce. This study simulates business process in current state and in presence of information technology systems to estimates potential cost savings.

II LITRETURE SURVEY

The cost of deaths and injuries due to traffic accidents has a great impact on society. In recent years, researchers have paid a great attention at determining the factor that significantly affects accident severity in traffic system. The author in presents a random forest & rough set theory to identify the factors significantly influencing single vehicle crash severity. The author in presents a decision tree which predicts cause of accident and accident prone locations. Papers predict traffic accident duration of incident and driver information system. The author in used various data mining techniques and tells the random forest outperforms than other classification algorithms. In paper, author talks about the significance of data mining classification algorithm in predicting the factor which influences road traffic accident.

The author in used to explore the possible application of data mining technology for developing a classification model and the result shows that developed model could classify accidents within a reasonable accuracy. It is important to analyse these datasets to extracts useful knowledge. Data mining is an effective tool for analysing data to extract useful knowledge. The severity of injuries measured for crash records has both continuous and categorical characteristics. Hence many previous studies have used models with ordered structure to analyse risk factor and their effect on severity of injuries sustained in traffic crashes.

III. METHODOLOGY

K-Mean Algorithm:

- K-Means clustering intends to partition n objects into k clusters in which each object belongs to the cluster with the nearest mean.
- This method produces exactly k different clusters of greatest possible distinction.
- The best number of clusters k leading to the greatest separation (distance) is not known as a priori and must be computed from the data.
- The objective of K-Means clustering is to minimize total intracluster variance, or, the squared error function.

Steps of algorithm:

1. Place K points into the space represented by the objects that are being clustered. These points represent initial group centroids.
2. Assign each object to the group that has the closest centroid.
3. When all objects have been assigned, recalculate the positions of the K centroids.
4. Repeat Steps 2 and 3 until the centroids no longer move. This produces a separation of the objects into groups from which the metric to be minimized can be calculated.

Mathematical Model of K-Mean Algorithm:

The diagram illustrates the mathematical formula for the K-Mean algorithm's objective function. The formula is $J = \sum_{j=1}^k \sum_{i=1}^n \|x_i^{(j)} - c_j\|^2$. Annotations include: 'number of clusters' pointing to k , 'number of cases' pointing to n , 'case i ' pointing to $x_i^{(j)}$, 'centroid for cluster j ' pointing to c_j , and 'Distance function' pointing to the norm $\|x_i^{(j)} - c_j\|^2$. The entire expression is labeled 'objective function' with an arrow pointing to J .

Figure.2: K-mean Algorithm Formula

Dataset Collection:

Sr.No	Variable	Description
1.	Place	In traffic accident, the place where people's accident got happened
2.	Year	In which year the traffic accident got happen.
3.	Fatal accident	In traffic accident, one or more persons dies within 30 days of the accidents
4.	Serious accident	In traffic accident one or more person injured and admitted in hospital for more than twelve hours

IV. IMPLEMENTATION DETAILS

The LOGIN FORM window contains two input fields: 'Username:' and 'Password:'. Below these fields are two buttons: 'SUBMIT' and 'CANCEL'.

The Select option window displays four options for user selection, each with a corresponding button: 'Do you want to insert:' with an 'INSERT' button, 'Do you want to find record:' with a 'find RECORD' button, 'Do you want to view:' with a 'VIEW RECORD' button, and 'Do you want to convert:' with a 'convert' button.

The Selection Form window features a list of input fields for data entry: 'Insert ID', 'Insert No of vehicle:', 'Insert type_of_vehicle:', 'Insert weather condition:', 'Insert gender:', 'Insert date:', 'Insert road surface:', and 'Insert casualty:'. At the bottom, there are three buttons: 'Submit', 'Cancel', and 'Back'.

The Searching Form window includes a 'Search record' input field and a 'Click here' label. Below these are two buttons: 'SEARCH' and 'CANCEL'. At the bottom, there is a 'BACK' button.

The 'View record Form' window has a title bar with a standard icon, a minus sign, a maximize button, and a close button. The main area is divided into two columns. The left column contains two text labels: 'If you want to view record click' and 'If you want to close click'. The right column contains three buttons: 'VIEW' (aligned with the first label), 'CANCEL' (aligned with the second label), and 'BACK' (at the bottom left of the main area).

The 'conversion Form' window has a title bar with a standard icon, a minus sign, a maximize button, and a close button. The main area is divided into two columns. The left column contains three text labels: 'convert', 'convert this file', and 'Redirect to weka tool'. The right column contains four buttons: 'convert' (aligned with the first label), 'browse' (aligned with the second label), 'Open' (aligned with the third label), and 'back' (at the bottom right of the main area).

V. CONCLUSION

Data mining in recent year with the database and artificial intelligence developed a new technology, its aim the large amount of data from the excavated useful knowledge, to achieve the effective consumption of data resources. A through literature review revealed a gap in published studies on the relationship between road characteristics and road traffic accident severity. The study on road traffic accident cause can identify the key factor rapidly and efficiently and provide instructional methods to the traffic accident prevention and road traffic accident reduction, which could greatly reduce personal casualty and property loss by road traffic accidents. Meanwhile, it would be helpful for improving the efficiency and security service level of the road transportation system.

REFERENCES

- [1]. Kim, K., Nitz, L., Richardson, J., & Li, L., Personal and Behavioral Predictors of Automobile Crash and Injury Severity. *Accident Analysis and Prevention*, Vol. 27, No. 4, 1995, pp. 469-481.
- [2]. Shankar, V., Mannering, F., & Barfield, W., Statistical Analysis of Accident Severity on Rural Freeways. *Accident Analysis and Prevention*, Vol. 28, No. 3, 1996, pp.391-401.
- [3]. Dia, H., & Rose, G., Development and Evaluation of Neural Network Freeway Incident Detection Models Using Field Data. *Transportation Research C*, Vol. 5, No. 5, 1997, pp. 313-331.
- [4]. Yang, W.T., Chen, H. C., & Brown, D. B., Detecting Safer Driving Patterns by A Neural Network Approach. ANNIE '99 for the Proceedings of Smart Engineering System Design Neural Network, Evolutionary Programming, Complex Systems and Data Mining, Vol. 9, pp. 839-844, Nov. 1999.
- [5]. Evanco, W. M., The Potential Impact of Rural Mayday Systems on Vehicular Crash Fatalities. *Accident Analysis and Prevention*, Vol. 31, 1999, pp. 455-462.
- [6]. Mussone, L., Ferrari, A., & Oneta, M., An analysis of urban collisions using an artificial intelligence model. *Accident Analysis and Prevention*, Vol. 31, 1999, pp. 705-718.
- [7]. Oliver, N. and Pentland, A. P., Graphical Models for Driver Behavior Recognition in a SmartCar, MIT IEEE intelligent vehicles symposium, 2000.
- [8]. Tavis, D. R., Kuhn, E. M, & Layde, P. M., Age and Gender Patterns In Motor Vehicle Crash injuries: Improtance of Type of Crash and Occupant Role. *Accident Analysis and Prevention*, Vol. 33, 2001, pp. 167-172.