IMPACT: International Journal of Information Engineering and Technology (IMPACT: IJIET)
ISSN (E): Applied; ISSN (P): Applied
Vol. 2, Issue 1, Jun 2016, 1-6

ROAD TRAFFIC ACCIDENTS WITH DATA MINING TECHNIQUES

BHAVNA KHATRI & HEMENDRA PATIDAR

Assistant Professor, SDITS, Khandwa, Madhya Pradesh, India

ABSTRACT

This paper emphasizes the importance of Data Mining classification algorithms in predicting the vehicle collision patterns oc curred in road accident data set. This paper is aimed at deriving decision tree which can be used fo r the prediction of manner of collision. A road traffic accident is defined as any vehicle accident occurring on a public highway. It includes collisions between vehicles an d animals, vehicles and pedestrians, or vehicles and fixed obstacles. Single vehicl e accidents, which involve a single vehicle, that means without other road user, are also included At all levels, whether at national or international level, road traffic accidents continue to be a growing problem. In connection with this, according to expected to grow from 28.1 million a year in 1990 to 49.7 million by 2020, which is an increase in absolute number of 76%. Traffic accidents are the main cause of this rise. Road traffic injures are expecte d to take higher place in the rank order of disease burden in the near future.

KEYWORDS: Decision Tree, Data Mining, WEKA

INTRODUCTION

Data Mining Software Tool

Notifications are normally reported by the drivers or any party being involved or having interest on it because the low requires doing so. On site investigation and recording is done with the aim of finding detailed and accurate information as to its cause, determine whether or not there has been violation of the law and ultimately to prevent the re-occurrence of further accidents. But sometimes as reported by the officers, due to time gap between the accident and the arrival of traffic officers, some details like the severity level and cause of an accident may not be identified effectively (yang liu et all [04]).

This accident record is basically used for various purposes in the office and for other stakeholder. National and regional transport offices use the data in directing their focus of attention in decision and policymaking's with regard to road safety. Different health offices and non-governmental organizations working in this area use the data in determining and managing health problem in society. Recent analysis proved that 81% of the accident all over the county is due to drivers fault and the other is due to vehicle, pedestrian and road faults. The main road safety problems are:

- drivers not respecting pedestrian priority
- over speeding
- unsafe utilization of freight vehicles for passenger transportation
- poor skill and undisciplined behavior of drivers
- less engineering effort in road design to consider safety

- poor vehicle conditions
- pedestrian not taking proper precautions
- week traffic law enforcement
- Lack of proper emergency medical services

The software package WEKA has number of ML tools for data analysis Decision Trees, Naïve Bias, Decision Table, Sequential Model Optimization, NN, Linear Regression and Voting Features. The learning methods are called classifiers. The performance of all classifiers is measured by a common evaluation module. The program outputs the mean absolute error and the root mean-squared error of the probability estimates. The root mean-squared error is the square root of the average quadratic loss. The mean absolute error is calculated in a similar way by using the absolute instead of the squared difference (alin dobra et all [8]). WEKA also implements cost-sensitive classification. When a cost matrix is provided, the dataset will be reweighted (or resampled, depending on the learning scheme). Because of its object oriented program code and good interface with several visual tools we prefer this program to the other three described above to conduct the experiments.

Types of Accident

Accidents and incidents (Reporting of Injuries, Disease and Dangerous Occurrences Regulations) can be classified into the following types (Witten H.I. et all [11]) :

- **Minor accidents** – are accidents which result in an injury, loss or damage but do not cause significant harm to a person,
- **Near miss incidents** – result in no apparent injury or damage and are not generally reportable under RIDDOR.
- **Lost time accidents -** are accidents which result in an employee being absent from work for more than ½ day.
- **Over three day Injury -** are accidents which result in an employee being absent from work for more than 3 days **.**
- **Major injury Accidents** - are accidents which result in a significant injury, loss or damage and are defined by RIDDOR.
- **Industrial Diseases** - are specific illnesses defined by RIDDOR which are linked to a work activity .
- **Dangerous Occurrences** - are specific incidents as defined by RIDDOR.
- **Fatality** - an accident or incident resulting in a fatality either immediately .
- **Public Reportable Accident -** their accident arose out of or in conjunction with Council work activity.

Accident / Incident Reporting Procedure

- All incidents, which result in an injury to an employee or people , must be recorded on Cheshire East Council's
- Electronic Accident Reporting System by the service or team responsible (Kweon, Y. J., et all [13]) .

METHODOLOGY

This section describes the process we followed to collect and analyze the academic performance. We discuss our selection of a data-mining tool, followed by the difficult task of preparing the data for analysis. We present our model of the academic performance prediction problem.

Source of Database and Description

Data base are extract from internet. Data base name is ACCIDENT2007-FullDataSet which is received from website. Data set description blow table (Martin, P. G. ET all [15].

Table 1: Data Set Description

No	Attribute Name	Type	Description	Distinct Values
1	Girth	Numerical	Tree diameter in inches	0.1
2	Height	Numerical	Tree Height in ft	0.012
3	Volume	Numerical	Volume of timber in cubic ft	

Figure 1: Attribute Value View

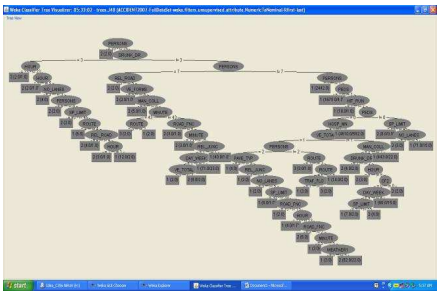

Figure 2: Generated Decision Tree from WEKA Tool

Preparing the Data and Selecting the Relevant Attribute

Table 2: Data and Selecting the Relevant Attribute

Classification Name	Number of Instance	Classified In %
Correctly Classified	2995	90.9505 %
Instances		
Incorrectly Classified	298	9.0495 %
Instances		

CONCLUSIONS

In this paper we have applying data mining tools. There are various types of data mining software are available but we have used WEKA software because it is freeware and it has many classifiers. We have used J48 classifier which is the extension of ID3 algorithm. We have observed that training data set are classified 89.076 % correctly and 8.02540 % are incorrectly. Can be extracting decision tree from Figure 1 shows the decision tree which is generated from accident data. Table 1 shows the result of experiment.

REFERENCES

1. PROF. BRIAN D. RIPLEY, “Study of the pure interaction dataset with CART algorithm”, Professor of Applied Statistics
2. xindongwu, vipin kumar, ross quinlan, joydeep ghosh, qiang yang, hiroshi motoda, geoffrey j. Mclachlan, angus ng, bing liu, philip s. Yu, zhi-hua zhou, michael steinbach, david j. Hand, dan steinberg, “top 10 algorithms in data mining”
3. NATHAN ROUNTREE, “Further Data Mining: Building Decision Trees”, first presented 28 July 1999
4. YANG LIU, “Introduction to Rough Set Theory and Its Application in Decision Suppot System”
5. WEI-YIN LOH, “Regression trees with unbiased variable selection and interaction detection”, University of Wisconsin–Madison.
6. S. RASOUL SAFAVIAN AND DAVID LANDGREBE, “A Survey of Decision Tree Classifier Methodology”, School of Electrical Engineering ,Purdue University, West Lafayette, IN 47907.
7. DAVID S. VOGEL, OGNIAN ASPAROUHOV AND TOBIAS SCHEFFER, “Scalable Look-Ahead Linear Regression Trees” 2000.
8. ALIN DOBRA, “Classification and Regression Tree Construction”, Thesis Proposal, Department of Computer Science, Cornell university, Ithaca NY, November 25, 2002
9. YINMEI HUANG, “Classification and regression tree (CART) analysis: methodological review and its application”, Ph.D. Student,The Department of Sociology, The University of Akron Olin Hall 247, Akron, OH 44325-1905,
10. Tang Z., MacLennan J., Data Mining with SQL Server 2005, USA, Wiley Publishing Inc., 2005.
11. Witten H.I., Frank E., Data Mining: Practical Machine Learning Tools and Techniques,Second edition, Morgan Kaufmann Publishers, 2005
12. Data Mining: Bagging and boosting available at: http://www.icaen.uiowa.edu/~comp/Public/Bagging.pdf
13. Kweon, Y. J., & Kockelman, D. M., Overall Injury Risk to Different Drivers: Combining Exposure, Frequency, and Severity Models. Accident Analysis and Prevention, Vol. 35, 2003, pp. 441-450.
14. Miaou, S.P. and Harry, L. 1993, “Modeling vehicle accidents and highway geometric design relationships”.

Accidents Analysis and Prevention, (6), pp. 689–709.27. Desktop Reference for Crash Reduction Factors Report No. FHWA-SA-07-015, Federal HighwayAdministrationSeptember,2007http://www.ite.org/safety/issuebriefs/Desktop%20Reference%20Complete.pdf

15. Martin, P. G., Crandall, J. R., & Pilkey, W. D., Injury Trends of Passenger Car Drivers In the USA. Accident Analysis and Prevention, Vol. 32, 2000, pp. 541-557.
16. National Highway Traffic Safety Administration, Traffic Safety Facts 2005, 2007, P. 54. http://wwwnrd.nhtsa.dot.gov/Pubs/TSF2006.PF
17. Ossenbruggen, P.J., Pendharkar, J. and Ivan, J. 2001, "Roadway safety in rural and small urbanized areas". Accidents Analysis and Prevention, 33 (4), pp. 485–498.