International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2018 IJSRCSEIT | Volume 3 | Issue 1 | ISSN : 2456-3307

Case Study - Analysis the Reason for Road Accident In Dindugal District

C. Esther Rani[1], K. Rani[2]
[1]PG Scholar, Department of Computer Science, Jayaraj Annapackiam College for Women (Autonomous) Periyakulam-625601, Theni Dt, TamilNadu, India
[2]Assistant. Professor, Department of Computer Science, Jayaraj Annapackiam College for Women (Autonomous), Periyakulam,Theni, TamilNadu, India

ABSTRACT

Globalization has impacted many developing countries across the world. India is one such country, which benefited the most. Increased, economic activity raised the consumption levels of the people across the country. This created scope for increase in travel and transportation. The increase in the vehicles since last 10 years has lot of pressure on the existing roads and ultimately resulting in road accidents. It is estimated that Accidents involving two wheelers accounted for the largest number of deaths - 32,318 or 23.2 per cent - followed by Lorries - 26,678 (19.2 per cent). Bicycles accounted for 3069 deaths in the last year. Motor vehicles crashes are a common cause of death, disability and demand for emergency medical care. Globally, more than 1 million people die each year from traffic crashes and about 20-50 million are injured or permanently disabled. There has been increasing trend in road accidents in dindigul over a few years. So that analysis the reason for Road accidents in dindigul district.

Keywords : Casestudy, Datamining, Road Accident, Analysis The Reason

I. INTRODUCTION

Road and traffic accidents (RTA) are one of the important problems in India. MORTH [1] mentioned in its report that every year there are 0.4 million accidents reported in India, which makes India a country with large accident rate. This report shows that there is a negative trend of accidents from 2012 to 2013; however, as accidents are unpredictable and can occur in any type of situation, there is no guarantee that this trend will sustain in future also. Therefore, the identification of different geographical locations where most of the accidents have occurred and determining the various characteristics related to road accidents at these locations will help to understand the different circumstances of accident occurrence. Kannov and Janson [2] stated that systematic relationship between accident frequency and other variables such as geometry of road, road side features, traffic information and vehicle information can help to develop effective accident prevention measures.

Lee et al. [3] indicated that statistical models were a good choice in the past to analyze road accidents to identify the correlation between accident and other traffic and geometric factors. However, Chen and Jovanis [4] determined that analyzing large dimensional datasets using traditional statistical techniques may result in certain problems such as sparse data in large contingency tables. Also, statistical models have their own model specific assumptions and violation of these can lead to some erroneous results. Due to these limitations of

statistical techniques, data mining techniques are being used to analyze road accidents. Data mining is a set of techniques to extract novel, implicit and hidden information from large data. Barai [5] discussed that there are various applications of data mining in transportation engineering such as road roughness analysis, pavement analysis and road accident analysis. Various data mining techniques [6] such as association rule mining, classification and clustering are widely used for the analysis of road accidents.

Accident cases in India are usually recorded by police officer of the region in which the accident has occurred. Also, the area covered by a police station is limited and they keep record of accidents that have occurred in their regions only. Ponnaluri [7] discussed that the report prepared by police only contains the basic information that are not much useful for the research purpose. He suggests that data collection method used by police needs a lot of improvement. However, Indian researchers used these data and analyzed it for some highway portions using statistical methods [8,9]. Data mining can be described as a novel technique to extract hidden and previously unknown information from the large amount of data. Several data mining techniques such as clustering, classification and association rule mining are widely used in the road accident analysis by researchers of other countries. Geurts et al. [10] used association rule mining technique to understand the various circumstances that occur at high-frequency accident locations on Belgium road networks. Tesema et al. [11] used adaptive regression tree model to build a decision support system for the road accidents in Ethiopia. Abellan et al. [12] developed various decision trees to extract different decision rules for different trees to analyze two-lane rural highway data of Spain. They found that bad light conditions and safety barriers badly affect the crash severity. Depaire et al. [13] used clustering technique to analyze road accident data of Belgium and suggest that cluster-based analysis of road accident data can extract better information rather analyzing data without clustering. Kashani et al. [14] used classification and regression tree (CART) to analyze road accidents data of Iran and found that not using seat belt, improper overtaking and speeding badly affect the severity of accidents. Kwon et al. [15] used Naïve Bayes and decision tree classification algorithm to analyze factor dependencies related to road safety. Severity of accident is directly concerned with the victim involved in accidents, and its analysis only targets the type of severity and shows the circumstances that affects the injury severity of accidents. Sometime accidents are also concerned with certain locations characteristics, which makes them to occur frequently at these locations. Hence identification of these locations where accident frequencies are high and further analyzing them is very much beneficial to identify the factors that affect the accident frequency at these locations.

In this paper, we are making use of data mining techniques to identify high-frequency accident locations and further analyzing them to identify various factors that affect road accidents at those locations. We use decision tree algorithm to structure the accident in the Dindigul District. Hence, our main emphasis will be the interpretation of the outcomes.

II. METHODS AND MATERIAL

2.1. Data Mining

Is defined as extracting information from huge sets of data. In other words, we can say that data mining is the procedure of mining knowledge from data. It can also be defined as an interactive process of discovering valid and novel, useful and understandable patterns or models in large database.
Data Mining is a process that uses a variety of data analysis tools to discover patterns and relationships in data that may be used to make valid predictions. Data mining uses advances in the field of Artificial Intelligence (AI) and Statistics.

2.2. Decision Tree

Decision tree is a "divide-and-conquer" approach to the problem of learning from a set of independent instances, which leads naturally to a tree-like style of representation called a decision tree.

A decision tree is a structure that includes a root node, branches, and leaf nodes. Each internal node denotes a test on an attribute, each branch denotes the outcome of a test, and each leaf node holds a class label. The topmost node in the node.

Decision Tree Algorithm

Method:

- create a node N;
- if tuples in D are all of the same class, C then
- return N as a leaf node labeled with the class C;
- If *attribute* list is empty then
- Return N as a leaf node labeled with the majority class in D;//majority voting
- Apply *Attribute selection _method* (D,*attribute_list*)to find the "best" *splitting criterion*;
- Label node N with *splitting criterion;*
- If *splitting attribute* is discrete-valued and multiday splits allowed then//not restricted to binary trees
- *Attribute list*<-*attribute _list_splitting_attribute*;//remove splitting attribute
- For each outcome j of *splitting criterion//*partition the tuples and grow sub trees for each partition
- Let Dj be a the set of data tuples in D satisfying the outcomes j;// a partition
- If Dj is empty then

Attach a leaf labeled with the majority class in D to node N;

- Else attach the node returned by generate_decision_tree(Dj, *attribute list*)to node N;
- Return N;

Decision tree Algorithms in R:

Decision tree algorithm is one of the most important classification measures in data mining. Decision tree classifier as one type of classifier is a flowchart like tree structure.

2.3. Data Modeling

The research considered the data of accident record between 32 km Dindigul to Batlagundu. The data were organized into a relation. The sample data used covered the period of 6 Months, from January 2016 to June, 2017. The output variable is the location and the locations can be divided into four distinct locations tagged location A, B, C and D. Location 1 – 11km is LocationA, above 11 km – 22km LocationB, and above 22km – 32km LocationC. Table 3.1 Showing variables given both continuous and categorical value.

Table 3.2

SN	Variables	Description	Value	Type
1	Type of vehicle	Small Car Heavy Car	A B	Categorical Categorica
2	Time of accident	Morning Afternoon Evening	A B C	Categorical Categorical
3	Causes of accident	OverSpeed LossOfControl WrongOvertaking TyreBlowouts PoorLights UncertainCause BrakeFailure	A B C D E F G	Categorical Categorical Categorical Cate

4	Location of Accident	LocationA LocationB LocationC LocationD	A B C D	Categorical Categorical Categorical

III. RESULTS AND DISCUSSION

R data mining Tool was used to mine the data using Id3 decision tree data mining algorithm. The algorithm is one of the most widely used and practical methods for inductive inference over supervised data. It represents a procedure for classifying and categorical data based on their attributes. It is also efficient for processing large amount of data, so is often used in data mining application.

3.1 Causes of accident

Below are results obtained using Id3 decision tree for the causes of accident along Dindigul to Batlagundu Highway.

3.1.1. Id3 tree

AccidentLocation = LocationA: UncertainCause
AccidentLocation = LocationB
| VehicleType = SmallCar: PoorLights
| VehicleType = HeavyCar: LossOfControl
AccidentLocation = LocationC
| AccidentTime = Morning: OverSpeed
| AccidentTime = Afternoon: WrongOvertaking
| AccidentTime = Evening
| | VehicleType = SmallCar: WrongOvertaking
| | VehicleType = HeavyCar: BrakeFailure

3.1.2 Summary of the Result

Correctly Classified Instances	120	72.7273 %
Incorrectly Classified Instances	45	27.2727 %
Mean absolute error	0.0989	
Root mean squared error	0.2235	
Relative absolute error	49.0673 %	
Root relative squared error	70.7409 %	
Total Number of Instances	165	

3.1.3. Detailed Accuracy by Class

TP Rate	FP Rate	Precision	Recall	F-Measure	Class
0.636	0.098	0.618	0.636	0.627	OverSpeed
0.375	0	1	0.375	0.545	LossOfControl
0.84	0.381	0.68	0.84	0.751	WrongOvertaking
0	0	0	0	0	TyreBlowouts
1	0	1	1	1	PoorLights

1	0	1	1	1	UncertainCause
1	0	1	1	1	BrakeFailure

3.1.4. Confusion Matrix

a	b	c	d	e	f	g	<-- classified as
21	0	12	0	0	0	0	a = OverSpeed
0	6	10	0	0	0	0	b = LossOfControl
13	0	68	0	0	0	0	c = WrongOvertaking
0	0	10	0	0	0	0	d = TyreBlowouts
0	0	0	0 1	0	0	0	e = PoorLights
0	0	0	0	0	6	0	f = UncertainCause
0	0	0	0	0	0	9	g = BrakeFailure

3.2. Locations of Accident

Below are results obtained using Id3 decision tree for the locations of accident along Dindigul to Batlagundu Highway.

3.2.1. Id3 Tree

AccidentCause = OverSpeed
| AccidentTime = Morning: LocationC
| AccidentTime = Afternoon: null
| AccidentTime = Evening: LocationC
AccidentCause = LossOfControl
| VehicleType = SmallCar: LocationC
| VehicleType = HeavyCar: LocationB
AccidentCause = WrongOvertaking
| AccidentTime = Morning: LocationA
| AccidentTime = Afternoon: LocationB
| AccidentTime = Evening: LocationC AccidentCause =
TyreBlowouts: LocationB
AccidentCause = PoorLights: LocationB AccidentCause
= UncertainCause: LocationA AccidentCause =
BrakeFailure: LocationC

3.2.2. Summary of the Result

Correctly Classified Instances	133	80.6061 %
Incorrectly Classified Instances	32	19.3939 %
Kappa statistic	0.6838	
Mean absolute error	0.0951	
Root mean squared error	0.2231	
Relative absolute error	31.3524 %	
Root relative squared error	57.4578 %	
Total Number of Instances	Numbe	

3.2.3. Detailed Accuracy by Class

TP Rate	FP Rate	Precision	Recall	F-Measure	Class
1	0	1	1	1	LocationA
1	0	1	1	1	LocationB
0.831	0.208	0.69	0.831	0.754	LocationC

3.2.4 Confusion Matrix

a	b	c	d <-- classified as
6	0	0	0 | a = LocationA
0	16	0	0 | b = LocationB
0	0	49	10 | c = LocationC

IV. DISCUSSION

There are 7 identified causes of accidents along the Dindigul to Batlagundu Highway which include; over speed, loss of control, wrong overtaking, tyre blowouts, poor lights, uncertain causes and brake failure. The result showed that out of the 165 instances of the accident, between January, 2016 to June, 2017, 81 instances of the accident occurred as a result of wrong overtaking, 33 as a result of over speed, 16 instances as a result of loss of control, 10 instances as a result of tyre blowout, 10 instance also as a result of poor light, 9 instances as a result of brake failure and 6 instances of the accident was uncertain. The best decision tree result was obtained with Id3 with 165 instances, 120 instances were correctly classified and 45 instances were incorrectly classified, which represent 72.7273%, 27.2727% respectively. The mean absolute error is 0.0989, root mean squared error is 0.2235, relative absolute error is 49.0673 % and root relative squared error is 70.7409 %.

V. CONCLUSION

The historical data collected for the accidents occurred between January, 2016 to June, 2017 along Dindigul to Batlagundu Highway was analyzed using R data mining Tool using Id3 decision tree and

predicted the causes of the accident, its prone location and time. The result showed that mostly the cause of the accident is wrong overtaking, followed by loss of control, then tyre blowout, poor lights, uncertain causes and brake failure. The result indicated that, accident mostly occurred in location C, followed by L

VI. REFERENCES

[1]. Afukaar FK, Antwi P, Ofosu-Amaah S. Pattern of road traffic injuries in Ghana: Implications for control. Inj Control Saf Promot 2003;10:69-76.

[2]. Centers for Disease Control. CDC Surveillance Update. Atlanta:Centers for Disease Control and Prevention; 1988.

[3]. Government of India. Code of Criminal Procedure-1973. New Delhi:Ministry of Home Affairs, Government of India; 1973.

[4]. Holder Y, Peden M, Krug E, Lund J, Gururaj G, Kobusingye O (eds). Injury surveillance guidelines. Geneva:World Health Organization; 2001.

[5]. Jacobs G, Aaron-Thomas A, Astrop A. Estimating global road fatalities. TRL Report 445. London:Transport Research Laboratory; 2000.

[6]. Joint Transport Commissioner and Secretary. Vehicles registered in twin cities. Hyderabad:Government of Andhra Pradesh; 2003. http://www.aptransport.org (accessed October 2003). 7Kapp C. WHO acts on road safety to reverse accident trends. Lancet 2003;362:1125.

[7]. Krug E (ed). Injury: A leading cause of the global burden of disease. Geneva: World Health Organization; 1999.

[8]. Krug EG, Sharma GK, Lozano R. The g l o b a l burden o f injuries. Am J Public Health 2000; 90:523-6.

[9]. Mohan D. Road traffic injuries-a neglected pandemic. Bull World Health Organ 2003;81:684-5.

[10]. Murray CJL, Lopez AD, Mathers CD, Stein C. The Global Burden of Disease 2000 Project: Aims, methods, and data sources (revised). Global Program on Evidence for Health Policy Discussion Paper No. 36. Geneva:World Health Organization; 2001.

[11]. Nantulya VM, Reich MR. The neglected epidemic: Road traffic injuries in developing countries. BMJ 2002;324:1139-41.

[12]. Nantulya VM, Sleet DA, Reich MR, Rosenberg M, Peden M, Waxweiler R. Introduction: The global challenge of road traffic injuries: Can we achieve equity in safety? Inj Control Saf Promot 2003;10:3-7.

[13]. National Crimes Records Bureau. Accidental deaths and suicides in India- 1999. New Delhi:Ministry of Home Affairs, Government of India; 2001.

[14]. Odero W, Garner P, Zwi A. Road traffic injuries in developing countries: A comprehensive review of epidemiological studies. Trop Med Int Health 1997;2: 445- 60.

[15]. Odero W, Khayesi M, Heda PM. Road traffic injuries in Kenya: Magnitude, cases and status of intervention. Inj Control Saf Promot 2003;10:53-61.

[16]. Registrar General of India. Population totals: India, Census of India 2001. New Delhi: Ministry of Home Affairs, Government of India; 2001. http://www. censusindia.net (accessed October 2003).

[17]. World Health Organization. The World Health Report 1999: Making a difference. Geneva:World Health Organization; 1999.