International Journal of Control and Automation
Vol.10, No.8 (2017), pp.13-22
http://dx.doi.org/10.14257/ijca.2017.10.8.02

Intelligent Transport System for Road Safety Based Data Mining Approach

Addi Ait-Mlouk[1], Imane Kamsa[2], Fatima Gharnati[1] and Tarik Agouti[1]

[1]Laboratory of Intelligent Energy Management and Information Systems, Faculty of Sciences Semlalia, Cadi Ayyad University, Marrakech, Morocco
[2]Department of Mathematics and Computer Science, Faculty of Sciences, Moulay Ismail University, Meknes, Morocco
aitmlouk@gmail.com, i.kamsa@edu.umi.ac.ma, gharnati@ucam.ma, t.agouti@uca.ma

Abstract

Recently, road accidents considered a major public health problem worldwide, the aim of many studies is to identify the main factors that contribute to crash severity. To identify those factors this paper shows a large scale intelligent techniques, such as intelligent agents that can detect drivers' cognitive state and analyze the data in a central system, the intelligent agents use data mining techniques, especially association rules mining to identify future accident in advance and giving chance to drivers to avoid the dangers. However, the association rule technique produces a huge amount of decision rules, which does not allow the decision makers to make their own selection of the most relevant rules. In this context, we believe that the visualization techniques would be particularly useful for decision makers who are suffering from the redundancy and quantity of extracted rules. An analysis of accidents on highways in the province of Marrakech (Morocco) between 2004 and 2014 showed that the proposed approach serves our purpose and may provide meaningful information that can help to develop suitable prevention policies to improve road safety.

Keywords*: data mining, association rules, road accident, intelligent agents, visualization*

1. Introduction

Data mining defined as the non-trivial process of identifying valid, novel, potentially useful and ultimately understandable patterns in data [1], in fact, it is a vital part of many business analytics and the most important trends in information technology. It involves many common classes of tasks (clustering, classification, association rules [2], *etc.*) designed for Knowledge Discovery in Databases (KDD). Data mining techniques have been applied in many real-life fields; many researchers [3-7] studied the signification of data mining techniques in the road accidents field. This technique is based on statistical analysis and artificial intelligence for discovering correlation and association between the variable in the database. Association rules algorithms are used to find all frequent itemsets then generate the association rules by satisfying some parameters like the minimum support and the minimum confidence. In this paper, we proposed an intelligent approach based data mining to analyze road accident data and predict the future accident. This approach consists of two major steps; cognitive state analysis and associations rule generators. Furthermore, this paper presents a structured evaluation of intelligent transport systems. First, common causes of accidents are identified and specific systems that could remove those causes or reduce their effects are proposed. The objective of this study is to determine the possibilities that intelligent transport systems can provide for the improvement of road safety by targeting specific risk factors.

ISSN: 2005-4297 IJCA

2. Related Work

According to World Health Organization (WHO) [8], 1.24 million people die each year on the world roads and between 20 and 50 million injuries due to a road accident. In addition, the Centers for Disease Control and Prevention (CDCP), announced that road accidents cost 100 billion in medical care every year. In 2014, the Ministry of Equipment and Transport of Morocco [9] gives the statistics of road accidents between 2004 and 2014 see Table 1.

Table 1. Road Accident Statistics in Morocco between 2004 and 2014

	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
Death	3 894	3 617	3 754	3 838	4 162	4 042	3 778	4 222	1351	2632	2214
Injuries	80 150	77 264	82 651	89 264	98 907	102743	98 472	102 011	102011	61207	28150

In the literature review, many techniques of data mining have been proposed to analyze road accident, Kuhnert *et al.* used CART and MARS to analyze of an epidemiological case-control study of injuries resulting from motor vehicle accidents and they identified potential areas of risk largely caused by the driver situation [10]. Ossenbruggen *et al.* [3] used logistic regression models to analyze the accident factors, and they found that the shopping sites are more dangerous than village sites. Sohn *et al.* [11] used three techniques of data mining such as decision tree, neural networks, and logistic regression for discovering significant factors for Korea Road traffic severity. Subsequently, Mio et al [12] used the decision tree to analyze the severity of traffic accident, and they found that fatal injury caused by many factors among them seat belts, alcohol, and light conditions. Chang and Wong [13] developed a CART model to analyze the relationship between drivers, injury severity and highway environment variable. Sze and Wong [14] used Binary Logistic Regression, Logistic Regression Diagnostics to controlling the influences of demographic and road environment. In addition, Abugessaisa [15] used clustering, classification trees to covers interactive explorations based on brushing and linking methods to detect and recognize interesting patterns. Moreover, Wong and Chang [16] used different methodologies to discover accident severity factors and they found that a dangerous accident caused by a combination of different factors. Anderson [17] studied the spatial patterns of road accident injury and results from the patterns in order to create a classification of road accident hotspots. Zelalem [18] studied the driver responsibility by using ID3, J48, and MLP algorithms to discover the related factors, and they found that many factors have a direct impact on severity accident such as license grades, driver age and experience. Pakgohar *et al.* [19] used CART and multinomial logistic regression (MLR) to study the role played by drivers' characteristics, and they found that the CART method provided results that are more precise. In other ways, Demirel *et al.* [20] used remote sensing for regional scale analysis and effective management of the environmental, and they concluded that this technology could be useful in the prevention of some type of accidents. Wu *et al.* [21] used the global positioning system (GPS) in the prevention of the collision accidents. Zhang *et al.* [22] concluded that the non-use of seat belts and inadequate training were also two important factors. Llus Sanmiquel [5] analyzes the main causes of those accidents by using Bayesian classifiers and decision tree.

3. Methodologies

In this section, we discuss the various steps constructing our proposed methodology, and we start with the association rules in big data as follows:

3.1. Association Rules Analysis

Association rules is a powerful technique for discovering the relationship between variables in large databases, it was initiated by Agrawal [2] for the first time to analyze transactional databases. An association rule is defined as an implication of the form: $A \rightarrow B$ Such as $A, B \subset I$ and $A \cap B = \phi$. Every rule is composed of two different sets of items A and B, where A is called antecedent and B called consequent. To improve the interestingness of association rules two measures are required, the minimum support and the minimum confidence. The support defined as the proportion of transaction in the database, which contains the items A, the formal definition is

$$Supp(A \rightarrow B) = Supp(A \cup B) = \frac{|t(A \cup B)|}{t(A)} \tag{1}$$

While the confidence determines how frequently items in B appear in the transaction that contains A, the formal definition is:

$$Confidence(A \rightarrow B) = \frac{Supp(A \cup B)}{Supp(A)} \tag{2}$$

3.2. The Proposed Approach

Road accident analysis can be conducted by three different categories of methods: analytical methods, statistical method and simulation. Each method has several good qualities as well as weakness. Generally, simulation methods required hardware and sophistical resources, this is leading to a time-consuming. In addition, analytical methods are fast to apply but cannot be used in complex problems. Due to the weakness of other methods, it seems statistical methods serve our mission to understand the complex road transportation system. Data mining is often used as an approach that integrates concepts from statistics and artificial intelligence; hence, it is a powerful tool that may discover complex and hidden relationships in large data.

The proposed approach is divided into two main steps. The first (Figure 1) is the detection of variables that contribute to road crashes, by analyzing the cognitive state of drivers using the detector agent and send back the data to the decision maker's agent. The second step (Figure 2) is the decision maker's agent, which is based on the association analysis. The following steps describe the details of proposed approach:

3.2.1. Detection of Driver Variables

Intelligent Transport Systems (ITS), based on advanced telecommunication and information technology, offer a great potential for improving the road safety situation for all types of road users. The systems are then assessed based on conclusions drawn from research studies and expert opinions. For each of the suggested systems the potential capabilities for improving driver behavior and road safety.

3.3. Detector Agent (DA)

The driver can cause an accident if his physical or cognitive state is not well. The automatic and dynamic detection of this state is a complex task. For this purpose, we have implemented in our approach a detector agent replacing the man and carrying out cognitive actions. DA is a software entity that automatically and dynamically interacts with the driver to detect his current state [23]. This agent seeks to extract all the physical and moral information that shows that the driver is not in a good state. The extraction must be done automatically, either after a certain time of observation of the behavior of the driver (entertain in the road, unacceptable speed) or from his profile if he indicated at the beginning that he does not feel good (Figure 1).

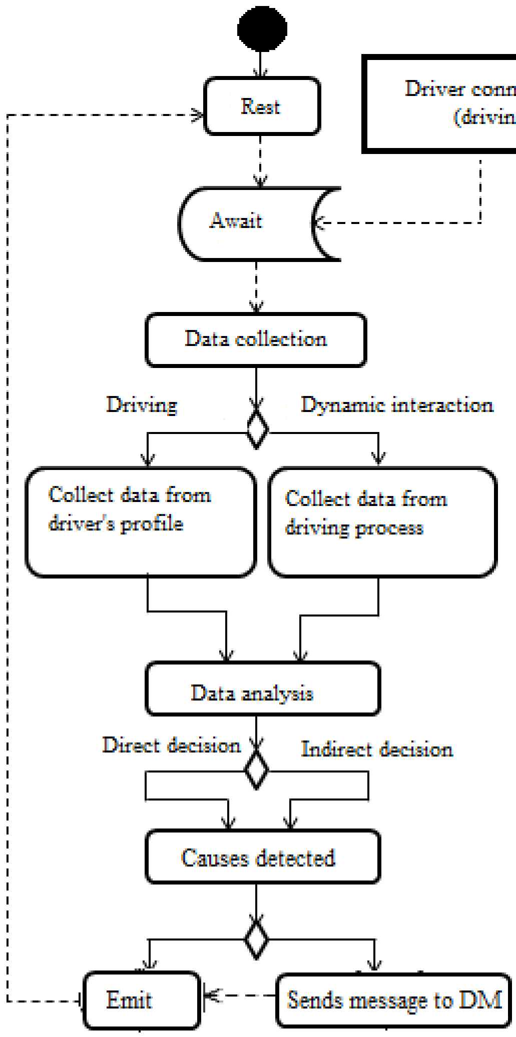

Figure 1. Activity Diagram of DA

The diagram given in Figure 1 shows the operation of DA agent to detect the physical and moral state of the driver. The DA objective is to collect the data necessary to detect the causes that can trigger an accident. The detected causes will be sent to the decision makers; if DA has identified causes otherwise it will go to rest.

3.4. Decision Makers Agent (DM)

The decision maker's agent DM is an intelligent and active entity, following a specific process in order to predict an accident [24]. The DM function is based on the message received from the DA, the data relating to the other variables indicated above and the rules of the association. Figure 3 describes in detail the DM features. As shown in Figure 2, the DM mission is to collect data and generate association rules in order to predict the future accident. The DM agent presents the result in a pictorial or graphical format that enables decision makers to see analytics presented visually, so they can grasp difficult concepts or identify new patterns. In addition, the DM works automatically and dynamically which allow us to realize a real-time assistance.

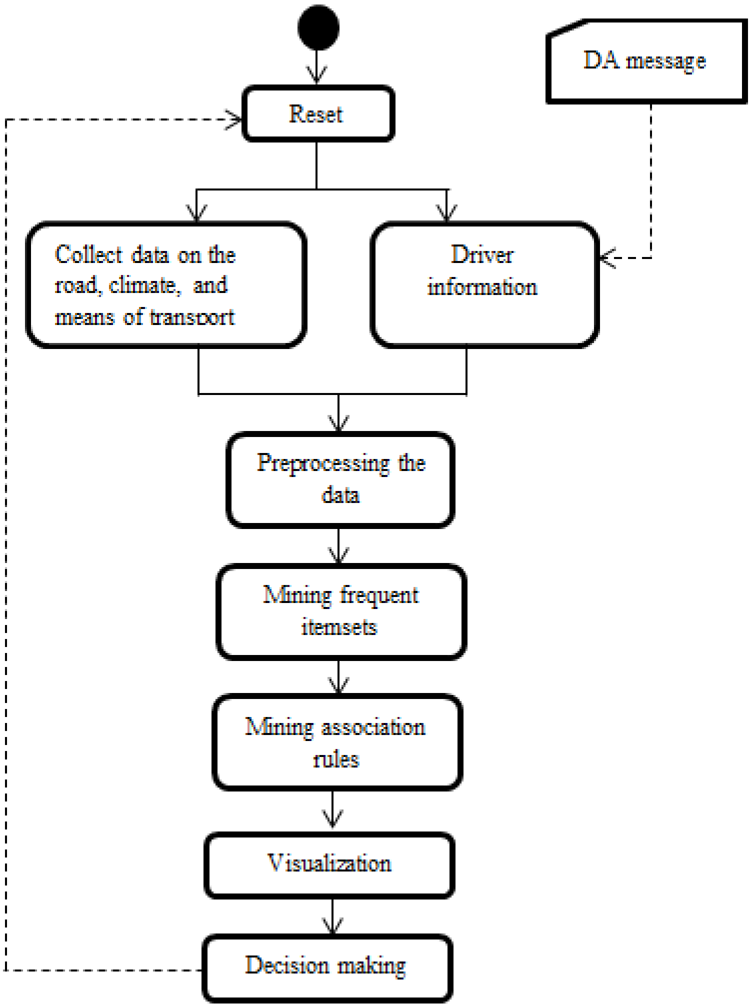

Figure 2. Activity Diagram of DM

Road accident data contains a rich source of information on different circumstances in which the accidents have occurred [25]. The variables describe characteristics related to the accident (type of collision, road users, injuries, *etc.*); traffic conditions (maximum speed, priority regulation, *etc.*); environmental conditions (weather, light condition, time of the accident, *etc.*); road conditions (road surface, obstacles, *etc.*); human conditions (fatigue, alcohol, *etc.*) and geographical condition (location, physical characteristics, *etc.*). To identify the main factors that affect transport and logistics, 21 variables were used, Table 2.

Table 2. Attributes and Factors of Road Accident

Attribute Name	Values	Description
Accident_ID	Integer	Identification of accident
Accident_Type	Fatal, Injury, Property damage	Accident type
Driver_Age	< 20, [21-27], [28-60] > 61	Driver age
Driver_Sex	M, F	Driver sex
Driver_Experience	<1, [2-4], >5	Driver experience
Vehicle_Age	[1-2], [3-4], [5-6] > 7	Service year of the vehicle
Vehicle_Type	Car, Trucks, Motorcycles, other	Type of the vehicle
Light_Condition	Daylight, Twilight, Public lighting, Night	Light condition

Weather_Condition	Normal weather, Rain, Fog, Wind, Snow	Weather conditions
Road_Condition	Highway, Ice Road, Collapse Road, Unpaved Road	Road conditions
Road_Geometry	Horizontal, Alignment, Bridge, Tunnel	Road geometry
Road_Age	[1-2], [3-5], [6-10], [11-20] > 20	The age of road
Time	[00-6], [6-12], [12-18],[18-00]	Accident time
City	Marrakesh, Casablanca, Rabat...	Name of city where accident occurred.
Particular_Area	School, Market, shops...	Where the accident occurred in school or Market areas.
Season	Autumn, Spring, Summer, Winter	Seasons of year
Day	Monday, Tuesday, Wednesday, Thursday, Friday, Saturday, Sunday	Days of week
Accident_Causes	Alcohol effects, Fatigue, Loss of Control, Speed, Pushed by another vehicle, Brake Failure	Causes of accident
Number_of_injuries	1, [2-5], [6-10],> 10	Number of injuries
Number_of_death	1, [2-5], [6-10],> 10	Number of death
Victim_Age	< 1, [1-2], [3-5] > 5	Victim Age

The data model used to analyze road accident data is given in Figure 3.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Accident_Type	Drive_Age	Drive_Sex	Drive_Exp	Vehicle_Age	vehicle_Type	Light_Condition	Weather_Condition	Road_Condition	Road_Geometry	Time	Season	Day	Causes
2	Fatal	<20	M	<1	<2	Car	Day	Clear	Collapse road	Horizontal	[6-12]	Spring	Md	Loss of Control
3	Injury	[21-27]	F	>6	<5	Car	Day	Run	Highway	Crossing	[12-18]	Summer	S	Alcohol effects
4	Injury	[28-60]	F	>7	<10	Car	Night	Clear	Collapse road	Alignment	[18-00]	Autumn	W	Speed
5	Injury	>60	F	<1	<15	Car	Day	Run	Highway	Horizontal	[12-18]	Summer	Sa	Speed
6	Injury	<21	F	<2	<10	Truck	Day	Clear	Unpaved road	Alignment	[12-18]	Summer	T	Brake Failure
7	Injury	[21-27]	F	<3	<5	Car	Day	Wind	Highway	Alignment	[6-12]	Winter	Md	Speed
8	Property damage	[28-60]	M	[2-6]	<15	Car	Day	wind	Collapse road	Horizontal	[12-18]	Summer	T	Loss of Control
9	Injury	<21	F	[2-6]	<10	Truck	Day	wind	Unpaved road	Alignment	[12-18]	Autumn	S	Speed
10	Injury	[21-27]	F	[2-6]	<5	Truck	Day	Clear	Highway	Alignment	[12-18]	Summer	W	Pushed by another vehicle
11	Injury	[28-60]	F	[2-6]	<15	Pedestrian	Day	Clear	Collapse road	Crossing	[6-12]	Autumn	Md	Alcohol effects
12	Injury	>61	F	>6	<5	Truck	Day	Clear	Unpaved road	Alignment	[6-12]	Summer	S	Speed

Figure 3. Data Model

3.5. Results and Discussion

In this paper, accident data were obtained from METM [9] in the period of 2004 and 2014. After data preprocessing, we selected a set of significant records, which identifies the factors related to the road accident, then we applied the proposed approach through two steps. The first step is the extraction of association rules from datasets by using Apriori algorithm [2] with the minimum support = 0.33 to count frequent itemsets, see Figure 4. This figure illustrates the itemsets by frequency, the result is sensitive to the minimum support introduced in the first step of Apriori algorithm.

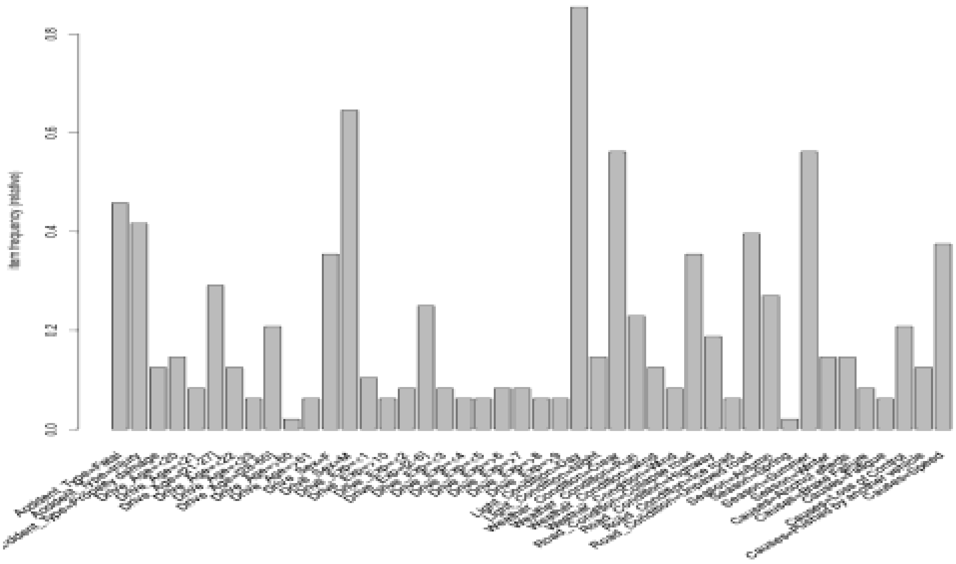

Figure 4. Frequent Itemsets

Figure 5. Grouped Matrix-Based Visualization

The second step is to generate the association rules from frequent itemsets previously extracted [26-28]. To visualize the extracted rules we used arulesViz [29] as an R-package extension, which implements several known and novel visualization techniques such as the matrix, group and graphs based visualization. The Matrix-based visualization technique present the antecedent and consequent items on the X and Y axes, this technique is enhanced by grouped matrix by grouping extracted rules via clustering, the grouped matrix based visualization is given in Figure 5. In addition, Figure 6 shows a parallel coordinates plot for 312 rules. The width of the arrows represents support and the intensity of the color represents confidence.

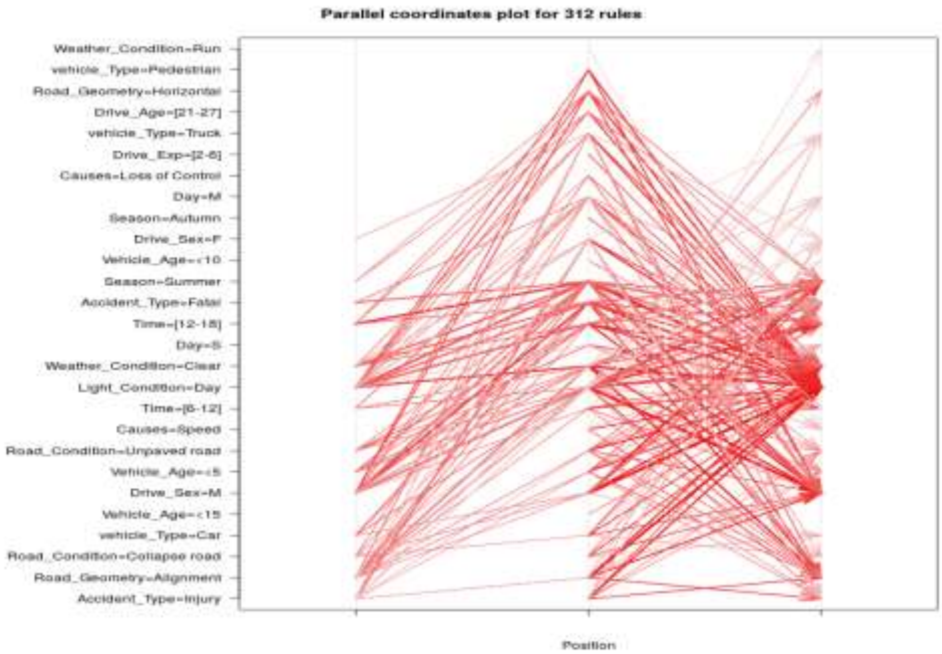

Figure 6. Parallel Coordinate Plot

Previous studies as [12] have found an association between drivers' behaviors weather condition, light conditions and accident severity. However, the size of the database leads to a very large number of results, which not explored further. The result of our study not only confirms an association between different variables but also shows that the intelligent system based agent allows the decision makers to make the right decision.

In summary, what we observe in results is that intelligent transport system favors with higher signification. Finally, the proposed approach has the flowing major strengths.

- Manage the complex decision situations by taking into account all the objective and subjective factors.
- Predict road accident based on historical data
- Visualization of association rules
- Allows the logistics managers to make their own choice of relevant rules.
- Allow transport managers to make the right decisions and help them to improve road safety, road network planning, transport and shipping services.

4. Conclusion

In many countries, road transport suffers from accidents, which affects the transport and shipping services. Understanding the road traffic is extremely important in improving the road security. This paper shows a large scale techniques such as associations rule analysis to optimize transport security by identifying hotspots in advance and giving chance to drivers to avoid the dangers. The analysis showed that by generating association rules the identification of accident circumstances that frequently occur together is facilitated. This leads to a strong contribution towards a better understanding of the occurrence of traffic accidents. Furthermore, the results indicate that although human and behavioral characteristics play an important role in the occurrence of all traffic accidents.

Finally, this analysis shows that a special traffic policy towards black spots and black zones should be considered since these high- frequency accident locations are characterized by specific accident circumstances, which require different measures to improve the road safety and transport services.

This study has raised some ideas for further work, first it would be useful to analyze the whole road accident covering recent years, and it would be good to integrate operational research techniques to optimize transportation and shipping services. Another study subject would be to focus on the internet of things and big data analytics in transportation and shipping safety.

References

[1] U. M. Fayyad, G. Piatetsky-Shapiro and P. Smyth, “From data mining to knowledge discovery: an overview”, In Advances in Knowledge Discovery and Data Mining. AAAI Press/The MIT Press, **(1996)**, pp. 134.

[2] R Agrawal, T. Imielinski and A. Swami, “Mining association rules between sets of items in large databases”, In: Proceedings of ACM SIGMOD Conference on Management of Data, **(1993)**, pp. 207-216.

[3] P. J. Ossenbruggen, J. Pendharkar and J. Ivan, “Roadway safety in rural and small urbanized areas." Accidents Analysis and Prevention, vol. 33, no. 4, **(2001)**, pp. 485-498.

[4] J. Oa, G. Lpez and J. Abelln, “Extracting decision rules from police accident reports through decision trees, Accident Analysis and Prevention”, vol. 50, **(2013)**, pp. 1151-1160.

[5] L. Sanmiquel, J. M. Rossell and C. Vintr, “Study of Spanish mining accidents using data mining techniques”, Safety Science, vol. 75, **(2015)**, pp. 49-55.

[6] A. Mirabadi and S. Sharifian, “Application of association rules in Iranian Railways (RAI) accident data analysis”, Safety Science, vol. 48, no. 10, **(2010)**, pp. 1427-1435.

[7] B. T. Eur. “Transp. Res. Rev, Common before-after accident study on a road site: a low-informative Bayesian method”, vol. 1, no. 3, **(2009)**, pp. 125-134.

[8] The World Health Organization. http://www.who.int/gho/road safety/en/ , accessed, **(2016)**.

[9] The Ministry of Equipment and Transport of Morocco, http://www.equipement.gov.ma/en/Pages/home.aspx, accessed, **(2016)**.

[10] P. M. Kuhnert and D. R. McClure, “Combining non-parametric models with logistic regression: an application to motor vehicle injury data”, Computational Statistics and Data Analysis, vol. 34, no. 3, **(2000)**, pp. 371-386.

[11] S. Sohn and S. Hyungwon, “Pattern recognition for a road traffic accident severity in Korea”, Ergonomics, vol. 44, no. 1, **(2001)**, pp.101-117.

[12] M. Chong, A. Abraham and M. Paprzycki, “Traffic Accident Analysis Using Decision Trees and Neural Networks”, IADIS International Conference on Applied Computing, Portugal, IADIS Press, Pedro Isaias (Eds.), vol. 2, **(2004)**, pp. 39-42.

[13] L. Chang, and H. Wang, “Analysis of traffic injury severity: An application of non-parametric classification tree techniques Accident analysis and prevention”, Accident analysis and prevention vol. 38, no. 5, **(2006)**, pp. 1019-1027.

[14] N. N. Sze and S. C. Wong, “Diagnostic analysis of the logistic model for pedestrian injury severity in traffic crashes”, Accident Analysis and Prevention, vol. 39, **(2007)**, pp.1267-1278.

[15] I. Abugessaisa, “Knowledge discovery in road accidents database integration of visual and automatic data mining methods”, International Journal of Public Information Systems, vol. 1, **(2008)**, pp. 59-85.

[16] J. Wong and Y. Chung, “Comparison of Methodology Approach to Identify Causal Factors of Accident Severity”, Transportation Research Record 2083, **(2008)**, pp. 190-198.

[17] T. K. Anderson, “Kernel density estimation and K-means clustering to profile road accident hotspots”, Accident Analysis and Prevention, vol. 41, no. 3, **(2009)**, pp. 359-364.

[18] R. Zelalem, “Determining the degree of drivers responsibility for car accident: the case of Addis Ababa traffic office”, Addis Ababa, Addis Ababa University, **(2009)**.

[19] A. Pakgohar, R. S. Tabrizi, M. Khalilli and A. Esmaeili, “The role of human factor in incidence and severity of road crashes based on the CART and LR regression: a data mining approach”, Procedia Computer Science, vol. 3, **(2010)**, pp.764-769.

[20] N. Demirel, M. K. Emil and H.S. Duzgun, “Surface coalmine area monitoring using multi-temporal high-resolution satellite imagery”, Int. J. Coal Geol., vol. 86, no. 311, **(2011)**.

[21] H. Wu, J. Tao, X. Li, X. Chi, H. Li, X. Hua, R. Yang, S. Wang and N. Chen, “A location based service approach for collision warning systems in concrete dam construction”, Saf. Sci., vol. 51, no. 338346, **(2013)**.

[22] M. Zhang, V. Kecojevic and D. Komljenovic, “Investigation of haul truck-related fatal accidents in surface mining using fault tree analysis”, Saf. Sci., vol. 65, **(2014)**, pp.106-117.

[23] I. Kamsa, R. Elouahbi and F. E. Khoukhi, "Intelligent Agents for Dynamic Optimization of Learner Performances in an Online System", Journal of Information Technology Education: Research, vol. 16, no. 1, **(2016)**, pp. 31-45.
[24] I. Kamsa, R. Elouahbi and F. El Khoukhi, "Interaction in Online System is a Favor key for Learners' success", International Journal on Advanced Science, Engineering and Information Technology, vol. 7, no. 2, **(2017)**, pp. 519-525.
[25] A. M. Addi, A. Tarik and G. Fatima, "An approach based on association rules mining to improve road safety in Morocco", International Conference on Information Technology for Organizations Development (IT4OD), Fez, **(2016)**, pp. 1-6.
[26] A. Ait-Mlouk, T. Agouti F. Gharnati and B. Derbali, "A choice of relevant association rules based on multi-criteria analysis approach", 5th International Conference on Information and Communication Technology and Accessibility Marrakech, **(2015)**, pp. 1-6.
[27] A. Ait-Mlouk, F. Gharnati and T. Agouti, "Multi-agent-based modeling for extracting relevant association rules using a multi-criteria analysis approach", Vietnam Journal of Computer Science, vol. 3, no. 4, **(2016)**, pp. 235-245.
[28] A. M. Addi, A. Tarik and G. Fatima, "Comparative survey of association rule mining algorithms based on multiple-criteria decision analysis approach", 3rd International Conference on Control, Engineering & Information Technology, Tlemcen, **(2015)**, pp. 1-6.
[29] M. Hahsler and S. Chelluboina, "Visualizing association rules: Introduction to the R-extension package arulesViz", R project module, **(2011)**, pp. 223-238.

