

APPLICATION OF DATA MINING ALGORITHM TO INVESTIGATE THE EFFECT OF INTELLIGENT TRANSPORTATION SYSTEMS ON ROAD ACCIDENTS REDUCTION BY DECISION TREE

Mohammad Mehdi Khabiri *, Fatemeh Matin Ghahfarokhi, Sara Sarfaraz, Hasan Mohammadi Anaie

Geotechnic and Road Departement, Faculty of Civil Engineering, Yazd University, Yazd, Iran

*E-mail of corresponding author: mkhabiri@yazd.ac.ir

Resume

Due to the large amount of available data in this study, authors have utilized data mining algorithms, especially the decision tree, to process these data and obtain the information, which would result in increasing road safety, determining the causes affecting it and patterns leading to traffic accidents. The effective use of this tool and its role in controlling the number of driving accidents is the subject of this study with the help of data mining algorithms. The results show that the increase in the number of roadside assistances to more than 41; number of driving accidents (fatally injured) is not significantly different, hence one of the proposed strategies for intelligent relay stations and its organization with the intelligent transportation tool is available. The intelligent transportation system utilities comprise of monitoring, guidance and enforcement tools, plus service tools such as rescue, driver assistant and road improvement.

Article info

Received 1 March 2021

Accepted 23 November 2021

Online 7 March 2022

Keywords:

data mining algorithm
intelligent transportation systems (ITS)
driving accidents
road safety
decision tree

Available online: <https://doi.org/10.26552/com.C.2022.2.F36-F45>

ISSN 1335-4205 (print version)

ISSN 2585-7878 (online version)

1 Introduction

Annually more than twenty thousand people die on the roads of Iran because of car accidents. This issue is a social health problem not only for Iran but for other countries, as well. For example, more than 40000 people die on European roads annually. Every year, road accidents impose 200 million euros expenses in the European economy [1]. Expenses of road accidents in Iran are unprecedented in their kind and they account for about 7% of the country's GDP [2-6].

Presence and monitoring by police at mostly accident hotspots, modifying the geometric scheme of these places, using warning signs, increasing vehicle safety and using seat belts are some of the solutions that can reduce road casualties. Besides these solutions, applying intelligent transportation systems is another method that the majority of developed countries and many developing countries have followed that today and by applying this system they have been able to increase the road safety level [4-9].

Currently, the road conditions in Iran are such that, in spite of the implemented efforts by relevant organizations and bodies over the past few years, this country is still considered one of the most unsafe countries in terms of the traffic problems. Despite the new telecommunication and communication technologies, the intelligent transportation system

with fast informing and without human errors could reduce approximately 40% of the road casualties and accidents. The statistics and official information of the country have demonstrated that using the intelligent transportation system technologies, such as controlling cameras, speed and message boards and other facilities, such as the creation of databases and establishing telecommunication and computer communication networks, from 2009 to 2012 have increased 5 times on average [3-9].

A vast database with huge information is informing and developing by expanding the data collection tools and it can be used in many fields of knowledge and experiments. The main and widely used algorithms in knowledge exploration and data mining include categorization, anticipation, clustering, recognition of redundant information and obtaining association rules and analyzing the time series. One of the main advantages of the decision tree is the fact that knowledge in the form of if-then classification laws, can be extracted and displayed [2-10]

ITS are not only safety-related but primarily are applied for other purposes like traffic flow and mobility improvements. Therefore, by considering the importance of the intelligent transportation system in this context, for more perception and gathering this subject, it is tried to classify the performed researches in a concise way and it is asserted in a case study and analyzed. For



this aim, a summary of these researches is presented here and in the following they are discussed. Finally, the decision tree is described as an effective way in the road transportation accidents.

2 Reviewing the references

In one of these studies, the efforts are made to introduce some effective intelligent systems in road safety and it is tried to review some case studies in this issue. The results of this study show that these systems can decrease the number of accidents significantly or the time for assistance on roads. Video controlling systems are the main tools of traffic management and their advantage is to provide video information for deciding and law enforcement. Speed controlling cameras have a significant role in the speed reduction of passing vehicles and this speed decrease will reduce accidents and finally reduce road casualties [3-9]. In this regard, the locating issue of this tool is essential for making them more practical and optimized their performance. In this research, after identifying the effective layers in locating speed controlling cameras, first desired information and maps were collected and in following, by weighing them, these layers were put on each other in ARCGIS media by aids of the hierarchical analysis technique and maps of appropriate zones for installing speed controlling cameras were extracted [10-17].

Another study is based on describing-analyzing method and it tried to determine the correct localization of intelligent road transportation equipment on the Shahrud-Sabzevar road in Semnan province by using analyzing and recognition of mostly accidental places and by means of the GIS localization method, localization of installing the intelligent transportation systems in Shahrud city was implemented. ARC GIS, Google Earth, RMG and MAP Source software were used for optimal localization in installing the mentioned systems. The results of this research can provide a scheme for installing intelligent systems in all of the provinces and it can provide approaches for optimal usage of intelligent systems [4].

The experimental Bayes method (EB) is commonly used by transportation safety analyzers for doing different safety analyses, such as before and after study and analyzing spots. Until today, most of implementation the EB methods were used by a negative two section method. Recent studies have shown that a limited mixture of the NB models with mixing GFMNB-K sections can be used in simulation of accident data in excessive closure and in overall it provides better statistical performance in comparison to the NB models [9-13]

A group of authors, in a case study in Semnan province, evaluated the conditions of the Shahrud-Sabzevar road before and after the installing the intelligent systems. The study on the share of different accidents on the roads shows that the highest number

of casualty statistics in the province in years 2008, 2009 and 2010, was because of overturning accidents and after that, vehicles accidents with each other. During these years, deaths due to vehicle overturning was approximately 59% of the total casualties in suburban road accidents in Semnan province and this statistics is more than twice of the national average. Semnan province had the first rank in country at 2011 for overturning vehicle casualties. It is worth mentioning that regarding to regulations from roads and transportation organization of the country, points or sections on road that has p number (quantity of measuring the accident rate of points) equal or more than 30, are considered as the high priority accident points and approaches should have provided for solving them in two-time length; short time and middle time. The changes after installing the intelligent transportation systems are in this way. Currently, 29 points of province roads have online controlling cameras, 40 critical and important axis have online traffic counting system, 6 sections of the province have variable message signs (VMS), 28 points have variant speed limit (VSL) and 34 sites have speed controlling cameras. Locating the installation of weigh in motion system (WIM) in two important sections of province (Semnan- Sarkhe and Shahrud- Sabzevar) were completed. The results of the mentioned research show that by considering the installation of intelligent transportation system and casualty statistics and traffic in Shahrud- Sabzevar road, the casualty numbers in the mentioned road during 2011 to 2013 (after installing intelligent systems) had a significant decreasing trend and by considering that the rate of traffic has increased 12% from 2011 to 2013, in spite of this increase in traffic, the number of casualties in Shahrud-Sabzevar has decreased [1].

Another case study, related to one of the intelligent transportation systems in Karaj city has been implemented. In this study, besides the introduction of effective parameters in selecting the location of variable message sign with the purpose of high efficiency of these tools in traffic distribution in city passages, ARC GIS software was used.

In this regard, with simulation of Karaj city passages in the mentioned software and defining the parameters involved in location selection, some points were proposed for locating that provided favorable visual success in passages by considering traffic volumes, as well as decision making by users in selecting other routs after passing the sign. In the following, after defining the favorable passages at the level of the urban passages network, in the next stage by using visual and geometrical parameters from the existing regulations, locating the variable message sign (VMS) along this passage with preparing logical points of view in the distances of vision and change of direction took place [5].

Application of data mining knowledge has been used in extensive safety researches in safety field, for example in studying the effects of variables and

Table 1 The comparison of effectiveness of intelligent transportation systems in roads safety [2]

Tool type	The comparison of safety performance of tools in intelligent transportation systems in different places		
Speedometer camera	Semnan: 23 to 43% decrease in accidents	Kerman: 86% decrease in fatal accidents	America: 9 to 51% decrease in accidents
Message board	Iran: 10 to 20 percent decrease in accidents	America: decrease in rate to 15 km/h or in another area: 8% decrease in travel time	
Ice and freezing prevention system	America: 16 to 80% decrease in urban accidents and 31 to 57 decrease in suburban accidents or in another area: 80% decrease in accidents due to bad weather		
Speed limit board	England: 55.7% decrease in casualties	America: 8 to 30% decrease in accidents	Spain: 11% decrease in air pollution and 16% decrease in accidents
Automatic accident recognition system	Japan: decrease in informing length to passengers from 8 minutes to 2 seconds and 50% decrease in secondary accidents	America: decreasing the recognition time and cleaning the accident scene from 1.5 hours to 31 minutes and 19 minutes decrease in relief to stopped cars	
Professional driver assistant system	Europe: the possibility of 10 to 82% decrease in accidents		
Short-range communication equipment	America: 20 to 50% increase in road capacity	Canada: 27% decrease in travel time and improving the safety index to 45%	
Controlling speed cameras	Decreasing the reaction time to accident less than 8 minutes and decreasing the time of cleaning the accident scene from 45 to 25 minutes		

Table 2 Overall summary from related studies with the current research [3-9]

	Subjects of study	Results
1	The severity of the driver injury in single car accidents	The safety of new cars influence the driver reaction.
2	Analyzing the effectiveness of human factors on predicting and categorization of accident severity in Iran	Safety belt, age and gender are human factor signs that have influence in road accident severities in Iran.
3	Defining the relation between injury severity of accidents with driver behavior, vehicle and environment	Type of variable vehicle is important that has the highest effect in flow.
4	Investigating the effect of weight and size of vehicle on the severity of accidents	Difference in weight and distance of vehicle axis in accidents increase the accidents.
5	Analyzing the effective factors on accident severity	The speed factor is very important in increasing the severity of the injury. So that, at high speeds the importance of age and gender is less than speed.

different parameters, the effects of sewage caps cover in the showoff of the motorcyclists were analyzed by decision tree, in the result of the mentioned research it is pointed out that more than 50% of motorcyclists make dangerous sudden change in location when they are faced with these cap covers [9-13]. Table 1 shows the comparisons between effectiveness of intelligent transportation system in country and different areas that are proposed by researchers, Sheikh Zeinodin and et al. [2].

Based on performed analyses by Niazi et al. (2018) [14], in reviewing the references it was clear that installing the speed controlling cameras have decreased the accidents 29%, or they say that 19% of all accidents and 44% of dangerous accidents in the UK have decreased after the installation of speed controlling cameras as a tool of intelligent transportation systems, with the practical implementation of intelligent transportation tools in China, the death statistics reached from 2.17%

at 2004 to 2.10% at 2007 [14]. In the studies before and after the installation of surveillance cameras in Jordan by Naqvi and et al. (2018), it was described that in the period between 2011 and 2014 the speed controlling cameras were installed in eight locations, an average of 59 to 63% of traffic accidents decreased, but in a location traffic accidents were increased to 35% [10-13]. A summary of related researches with the importance of subject are presented in Table 2.

3 Research method and data gathering

For analyzing data in this research by the SPSS software, a statistical pattern from the decision tree was plotted. There are different techniques for plotting a decision tree utilizing different algorithms, regarding the analyzed database, selecting an appropriate algorithm for applying on data is essential. The

advantages of the decision tree are:

- Using the knowledge of human
- Easy perception in training and test data
- Easy perception of categories
- Applicable for training data with low volume and easy problems [15-19].
- Facility in understanding the relationship between variables, but despite its easiness, it can work with complex data and make decisions from it.
- Ability to use a massive amount of data and a large number of variables and work with complex data are its advantages.
- It can be combined with other decision-making techniques to achieve better results.

For determining the appropriate branches in the categorization of data, a criteria for calculating the impurity amount for each group is defined that is obtained based on:

$$i_{(N)} = - \sum_j P\{(w_j) \times \log P(w_j)\}, \quad (1)$$

where:

$i_{(N)}$: entropy impurity criteria,

$P(w)$: the relation of samples that in N knot it is in J group.

When the amount of $i_{(N)}$ is maximum, all the samples with uniform distribution belong to all categories, this means that equal probability is created for all the groups or sub-branches [15].

One of the criteria for categorization in the decision tree is variance impurity, which is obtained for even groups based on:

$$i_{(N)} = P(w_1) \times P(w_2). \quad (2)$$

The Gini impurity $i_{j(N)}$ is the generalized impurity variance for the multi-group state that is obtained from:

$$i_{j(N)} = \sum (i \neq j) [P(w_i)P(w_j)] = \frac{1}{2} [1 - \sum (j=1) P_{(w-j)}^2]. \quad (3)$$

$i_{j(N)}$ Another impurity criterion is misclassification impurity that shows the minimum possibility of misclassification and in the highest state, in which all the groups are equal, is obtained from:

$$mi_{(N)} = 1 - \max P_{(w_j)}, \quad (4)$$

mi is the index of misclassification impurity.

The multi-branch state with a branching rate of B , can be used for the number of possible branches in classification and based on Equation (5) for normalization the number of impurity decrease is used.

$$\Delta_i B(S) = \frac{\Delta_i(S)}{(\sum_{k=1}^B (P_k \log 2) P_k)}, \quad (5)$$

where:

$i-B(S\Delta)$: appropriate index for roots

$i(S\Delta)$: difference or decrease in impurity variance [15].

In the decision tree algorithm, the classification algorithm is considered with monitoring, thus, data are classified in learning and teaching media, in this kind of algorithm, the performance and accuracy of their classification in data mining is in lower entropy index [15-18]. The display of the resulting knowledge in the shape and structure of a tree makes it comprehensible. Each branch of a tree includes a combination of different variables that have close properties [15].

The optimum number of clusters in each model is calculated by using the Dun index based on Equation (6). The purpose of this index is to maximize the distances in clusters and minimize the distance between clusters [15-18].

$$D = \min \left\{ \min_{j=i+1} \frac{d(c_i, c_j)}{\max_k (diam(c_k))} \right\}, \quad (6)$$

where:

$d(c_p, c)$: the minimum distance between existing records in branch i that is calculated from Equation (7):

$diam_{(c_j)}$: the maximum distance between existing records in branch j that is calculated from Equation (8):

$$d_{(c_i, c_j)} = \min_{x,y} \{d_{(x,y)}\}, \quad (7)$$

$$diam_{(c_j)} = \max_{x,y,i,j} \{d_{(x,y)}\}, \quad (8)$$

where:

$d_{(x,y)}$: is the distance in each branch [15-18]. For analyzing the performance and the used algorithm, convergent validity is used. This index calculates the correct performance of the clustering algorithm, if a record is not placed correctly, it is recorded as "positive correct index" and if it is placed in an unrelated branch it is recorded as "negative correct index" and finally, the convergent validity is calculated from [15-18]:

$$\text{convergent validity} = \frac{\text{positive integer} + \text{negative integer}}{\text{positive data} + \text{negative data}}, \quad (9)$$

in that:

positive data: data with positive sign,

negative data: data with negative sign [15-18].

Equation (9) shows the degree of reliability of data segmentation and classification. The convergent validity parameter denotes the validity of the classified data to the total data. The larger this parameter is, the more reliable the decision tree algorithm is. Nevertheless, if a large amount of data is not properly segmented (negative data), it reduces the level of reliability. The decision tree is one of the non-parametric methods of data classification. Tree classification in audit analysis methods is based on logistic regression. In this research, the decision tree algorithm has been used to present the

Table 3 The summary of statistics from Road Transport Organization in the country from 2008 to 2017 [20]

Subjects	Unit	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
Freeways	km	1629	1770	1957	2053	2185	2102	2401	2201	2401	2401
Highways	km	7516	9061	10669	11652	12696	14155	14488	15463	16627	16810
Main roads	km	21202	20794	21276	22052	21222	21628	22886	23879	2528	25866
The number of on line traffic counter devices in the country	number	332	329	531	808	952	1197	1510	1786	1993	2199
Variable message systems (VMS) in the country	number	63	96	136	168	176	193	195	195	195	203
Video controlling cameras in the country	number	23	52	121	227	437	584	669	654	660	731
Controlling constant speed cameras in the country	number	137	199	186	200	225	382	-	229	982	1046
Weight in motion (WIM) system in the country	number	4	5	7	8	16	18	17	19	28	32
The number of suburb accidents in the country	Thousand items	137	144	153	117	116	112	102	101	102	121
The number of killed people in suburban traffic accidents	persons	13984	13556	14176	12232	11811	11568	10951	10860	10427	10729
Average number of deaths in a day	persons	38	37	39	32	32	32	30	30	29	29

pattern of traffic accidents.

In the decision tree, Chi-square Automatic Interaction Detector (CHAID) algorithm, this possibility exists that one knot is divided into more than two knots but it is used in the Cluster-Based Routing (CRT) algorithm for double trees (it means that each knot is divided into two other knots).

3.1. Introduction of research variables

In this study, variables of research include the number of accidents (death- injury) and independent variants and the intelligent control and surveillance of the road police or other intelligent transportation systems (ITS) in provinces from 2008 to 2017 (10 years' statistics). First, with using the available data from the summary of management statistics of roads and transportation in 2018 [20], the information statistics were inserted in Excel software and the next effort was modeling the decision tree for reaching an appropriate graphical display from the relationship of independent and dependent data, which in Table 3 input data for research are presented.

3.2. Introduction of statistical population

The statistical population includes a complete collection of recorded data or possible measurements of a qualitative characteristic or property, about the complete collection of units, which is considered, perceptions about it are performed. This research seeks to link safety variables and applications of intelligent transportation system tools, that the summary of statistical data from the statistical yearbook of the Road Transport Organization are extracted and are introduced in Table 4 in the form of descriptive statistics from statistical software [20].

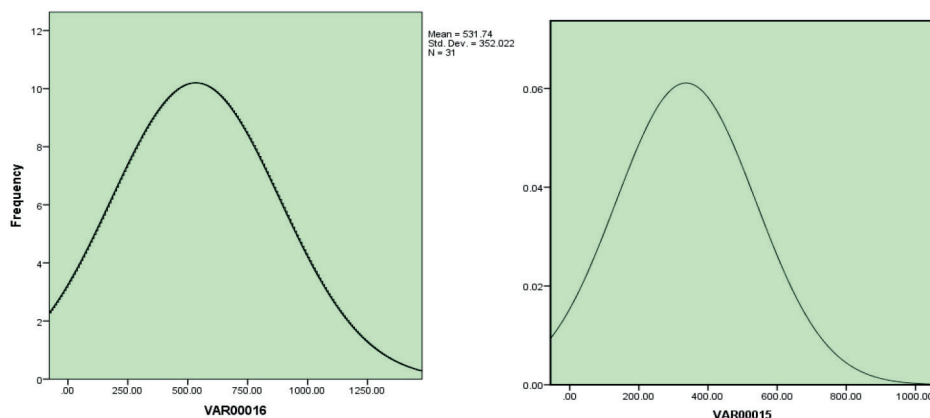
For analyzing the normal state of data, the tensile data are tested first; that its histogram is drawn in Figure 1. In Table 4 the tensile descriptive statistics or kurtosis show a normal distribution that its amounts are in (-2, +2) range.

4 Decision tree model

A model is a mental and non-physical display of the internal components and connections of a phenomenon

Table 4 Descriptive statistics data from intelligent transportation systems and variance experiments and average

Variable	N	Range	Minimum	Maximum	Mean	Std. Deviation	Variance	Skewness	Kurtosis	Std. Error	Statistic	Std. Error
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic			
RoLe	31	7617.00	393.00	8010.00	2779.8065	2011.30096	40431.561	1.316	0.421	0.788	0.821	
RuLn	31	2370.00	248.00	2618.00	1117.2581	673.93689	454190.931	1.107	0.421	0.079	0.821	
Orj	31	89.00	17.00	106.00	42.7419	25.79401	665.331	1.293	0.421	0.388	0.821	
ReCr	31	31.00	2.00	33.00	12.9677	6.86287	47.099	1.071	0.421	1.128	0.821	
Pol	31	12.00	4.00	16.00	7.6129	3.42225	11.712	0.983	0.421	0.437	0.821	
Loa	31	40747.00	499.00	41246.0	11612.000	10122.4305	1024600.66	1.526	0.421	1.881	0.821	
Pas	31	20114.00	1270.0	21384.0	5772.5806	5172.99881	267916.718	1.917	0.421	3.254	0.821	
Veh	31	2924.00	42.00	2966.00	830.6452	726.16736	527319.037	1.602	0.421	1.961	0.821	
CounPas	31	80.00	20.00	100.00	57.6129	18.52688	343.245	0.361	0.421	0.177	0.821	
VMS	31	16.00	1.00	17.00	6.2903	3.78764	14.346	1.172	0.421	1.352	0.821	
VSL	31	58.00	0.00	58.00	16.0000	15.90807	253.067	1.157	0.421	0.918	0.821	
Cam	31	40.00	9.00	49.00	21.0968	10.18284	103.690	1.068	0.421	0.934	0.821	
CamSp	31	56.00	0.00	56.00	14.4839	15.28152	233.525	1.366	0.421	1.089	0.821	
WIM	31	4.00	0.00	4.00	0.6129	1.02233	1.045	1.874	0.421	3.316	0.821	
Crash ₂₀₁₅	31	886.00	105.00	991.00	336.3548	202.42571	40976.170	1.301	0.421	2.064	0.821	
Crash ₂₀₁₆	31	1206.00	168.00	1374.00	531.7419	352.02159	123919.198	1.133	0.421	0.212	0.821	

**Figure 1** The normal distribution tensile statistical data (accidents) from the intelligent transportation system

that shows the relationships between data and different variables of that phenomenon. Models are used for drawing future phenomena and future situations. In the following, two types of statistical models, the multiple linear and the data mining model of the decision tree are used.

The graphical tool for displaying the classification and clustering data from decision support tools has the following advantages:

- Training samples are placed in the correct categorization as possible.

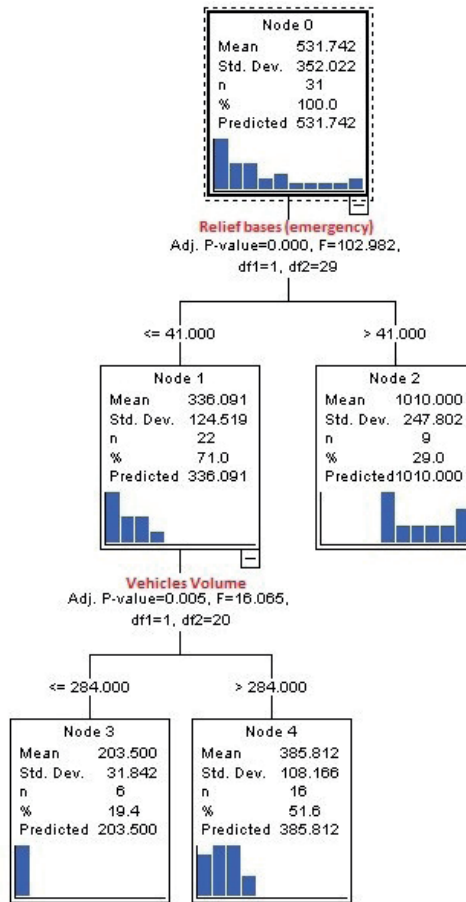


Figure 2 The model of constructed decision tree for accidents data at 2016 based on all 13 independent variables

- The emergence of training samples in a form that neglected samples are categorized with high accuracy.
- In the availability of new training samples, facile updating of trees is possible.
- It has as simple a structure as possible.

In the decision tree, the first leaf is divided into two leaves with various samples and different predictions. The used pattern in this modeling is the CHAID technique. Each category is divided into other categories with different numbers of specimens and different predictions, except for its own leaf. In addition, these steps continue to reach the final nodes called the leaf, the best branch and its final result is called the leaf, the category is found, the best branch and its final result that means leaf is selected according to the percentage of importance of the independent variables. Then, the constructed branches of the tree are removed or pruned to achieve the stopping criteria or reaching to the desired level of complexity.

The first sample of a constructed decision tree, from 13 variables in this study, is shown in Figure 2 based on the CHAID pattern; as it is seen, this decision tree is based on a double classification pattern, it has 5 main leaf or data groups, that describe variables

in these categories that are the roadside assistance (emergency). It is observed that in this node, with increasing the number of roadside assistance to more than 41 bases, the number of accidents (deaths) changes dramatically. As the number of vehicles increases, these are effective in accidents, as the number of crossings and the volume of vehicles increases by 50%, the number of accidents (death) increases. Based on this graph, the importance of existence of the relief stations and attention to the nature of their services is emphasized, especially when the possibility of making smart these Intelligent Equipment Relief Service in order to the in-time presence in traffic accidents. Another measure of relative reduction in traffic volume is the widening of road lanes, which in proportion to the passing traffic, reduces the intensity and number of accidents. In addition, it is recommended that the higher priority be given to locating intelligent traffic and transportation equipment on routes with higher passing traffic.

In Figure 3, the diagram of the extended decision tree model is considered as a dependent variable for the accidents in 2015, this pattern is based on the CHAID data clustering model. An important variable, affecting the data clustering in the first open node in these statistical data, is the number of roadside

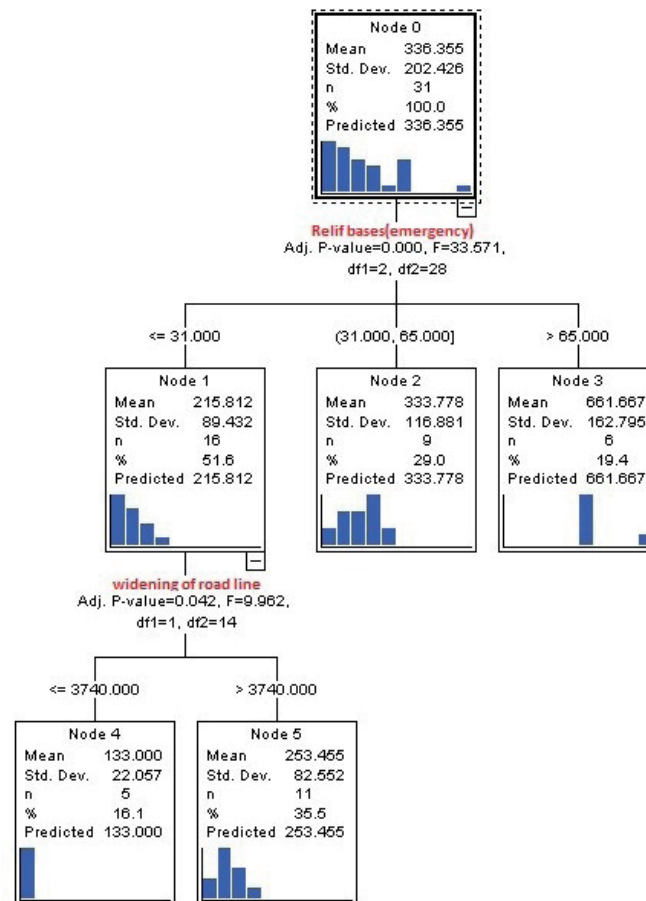


Figure 3 The decision tree model made based on all 13 independent variables for the accidents in 2015

assistance stations The effectiveness of the roadside assistance parameter in this collection is such that in the first node three separate leaves are made for the data categorization. The second factor in the separation and categorization of data is the road networks. Of course, as the length of the route network increases, the number of accidents increases obviously, but this result emphasizes the correct construction of this graphic model.

5 Results and propositions

5.1. Summary of results

Analysis of accidents occurring on suburban roads with the aim of identifying the parameters affecting the increase in the frequency of accidents, can be effective in the decision of those involved in improving a road safety in order to reduce the road accidents. The purpose of this study was holistic, in other words, minor indicators such as speed were not included in the study. In this study, according to the published general data related to the last ten years on the status of the road network and intelligent transportation system tools and statistical modeling techniques, clustering models are used that in summary, the main results of this study in a divided form include:

- Dependent variables, such as network length and volume of passing traffic, were of a great importance in constructing the accident prediction model, which, due to the specificity of this issue, is proof of the accuracy of the constructed models.
- Compared to more important variables, such as passing volume and route length, variables related to intelligent transportation did not have a significant impact on the model construction, hence it is essential to pay attention to develop the appropriate primary network infrastructures.
- The influential factor in the number of accident models in the drawing techniques was the relief situation, that equipping the relief stations and coordinating them with other equipment of the intelligent transportation system can be effective in reducing the severity of accidents.

5.2. Recommendations for increasing the traffic safety

According to the research results, in general, approaches for increasing the effectiveness of the application of road intelligence tools and road transport are provided.

- Despite the widespread use of smart devices in the road transport, there are still many shortcomings, for example, in-time informing in accidents and the golden time of rescue to rescue forces can decrease a significant proportion of the number of accidents and casualties and the severity of injuries.
 - Decentralization and proportional distribution of intelligent transportation system infrastructures between the provinces and accident-prone places with high transit volume and providing its related tools are among the essentialities of increasing the efficiency of this system.
 - Completing the primary infrastructures and then equipping them with intelligent transportation is an obligation, but along with it, the development of telecommunication and internal electronic systems of used cars in the country such as car-car systems and car infrastructure needs more attention.
- Finally, in order to complete the present study, several study horizons are suggested as follows.
- Providing an appropriate data mining framework to implement appropriate models in determining the influence of effective parameters on the performance of intelligent transportation tools.
 - Using other data mining and neural network models to calculate the effectiveness of intelligent transportation tools in the severity and the number of events such as CART and K-MEANS methods and phase neural network.

Nomenclature

Variables	The meaning of the variables
RoLe	The main way
RuLn	The length of important roads
Orj	Emergency station
ReCr	Red crescent station
Pol	Population
Loa	The length of all directions (km)
Pas	Number of vehicles passing
Veh	The average of all vehicles in a certain time
CounPas	Online traffic count
VMS	Variable message systems
VSL	Variant speed limit
Cam	Video surveillance camera
CamSp	Camp and police station
WIM	Weigh-in-Motion
Crash2015	Number of crashes in 2015
Crash2016	Number of crashes in 2016

References

- [1] AHMADZADEH, T., KARKEABADI, Z. Investigating the effect of installing intelligent transportation systems on reducing road losses case study of Shahroud - Sabzevar. *Quarterly Journal of Semnan Police Science*. 2008, **8**(27), p. 91-29. ISSN 2322-1771.
- [2] SHEIKHOZEDDIN, H., SHARIFIAN, V., QURESHI, S. F. The need to use intelligent transportation systems to improve road safety. In: *The 1st National Symposium to Improve Road Safety: proceedings*. 2015.
- [3] ZIARI, H., KHABIRI, M. M. Analysis characteristics and provide a prediction model of public bus accident in Tehran. *Journal of Applied Sciences* [online]. 2006, **6**(2), p. 247-250. ISSN 1812-5654, eISSN 1812-5662. Available from: <https://doi.org/10.3923/jas.2006.247.250>
- [4] ASHRAFZADEH, A., GHORBANI, A., DOLAMA, H. Application of intelligent transportation systems in road traffic management. In: *International Conference on Civil Engineering, Architecture and Urban Infrastructure: proceedings*. 2015.
- [5] AKBARI NIA, F., JAFARI, M. H. Locating variable message sign (VMS) in intra urban networks using ARCGIS. *Journal of Civil Engineering and Structures*. 2017, **1**(1), p. 44-54. ISSN 2588-3275.
- [6] TAHERI, V. Evaluation of car accident in Iran and the rule of intelligent transportation systems in decreasing of collisions in the highways. In: *3rd National Conference on Road Accidents, Rail and Air Accidents: proceedings*. 2014. p. 1-13.

- [7] TOLOUEI, M., ZOGHI, H., SIAMARDI, K. Application of intelligent driver assistance system in order to increase traffic safety on Iranian roads (case study of highways in the south of Tehran. In: 10th International Conference on Transportation and Traffic Engineering: proceedings. 2016.
- [8] AMJADI TALEBI, M. H., KIAMEHR, R. Location of freeway intelligent transportation systems for traffic safety using GIS (study of speed control cameras on the Qazvin-Rasht freeway. In: The 1st Conference on New Ideas and Technologies in Geographical Sciences: proceedings. 2017.
- [9] KUEHN, M., HUMMEL, T., BENDE, J. Benefit estimation of advanced driver assistance systems for cars derived from real-life accidents. In: 21st International Technical Conference on the Enhanced Safety of Vehicles ESV 2009: proceedings. Vol. 15. 2009.
- [10] SONG, B. Potential safety benefit analysis of cooperative driver assistance systems via vehicle-to-vehicle communications. *The Journal of the Korea Institute of Intelligent Transport Systems* [online]. 2018, 17(2), p. 128-141. ISSN 1738-0774, eISSN 2384-1729. Available from: <https://doi.org/10.12815/kits.2018.17.2.128>
- [11] ZOU, Y., ASH, J. E., PARK, B. J., LORD, D., WU, L. Empirical Bayes estimates of finite mixture of negative binomial regression models and its application to highway safety. *Journal of Applied Statistics* [online]. 2018, 45(9), p. 1652-1669. ISSN 0266-4763. Available from: <https://doi.org/10.1080/02664763.2017.1389863>
- [12] NAGHAWI, H., AL QATAWNEH, B., AL LOUZI, R. Evaluation of Automated Enforcement Program in Amman. *Periodica Polytechnica Transportation Engineering*, [online]. 2018, 46(4), p. 201-206. ISSN 0303-7800, eISSN 1587-3811. Available from: <https://doi.org/10.3311/PPtr.10939>
- [13] CHANG, L.-Y. Analysis of effects of manhole covers on motorcycle driver maneuvers: a nonparametric classification tree approach. *Traffic Injury Prevention* [online]. 2014, 15(2), p. 212-206. ISSN 1538-9588, eISSN 1538-957X. Available from: <https://doi.org/10.1080/15389588.2013.792110>
- [14] NIAZI, M., ASGARI, A., NOURANI, E. The role of police surveillance cameras in *reducing driving violations*. *Traffic Management Studies Quarterly*. 2018, 47(4), p. 67-82. ISSN 2008-4005.
- [15] VEISI, H. Machine learning, decision tree. Tehran: Faculty of Modern Sciences and Technologies, University of Tehran, 2014.
- [16] SOLEIMANPOUR, S. M., MESBAH, S. H., HEDAYATI, B. Application of CART decision tree data mining to determine the most effective drinking water quality factors (case study: Kazeroon plain, Fars province). *Iranian Journal of Health and Environment* [online]. 2018, 11(2), p. 1-14. ISSN 2008-2029, eISSN 2008-3718. Available from: <http://ijhe.tums.ac.ir/article-1-5881-en.html>
- [17] KHALILI, S., PASHAZADEH, S. Classification of current bank account data using the decision tree. In: 2nd Conference of Computer and Information Technology Scholars: proceedings. 2014. p. 1-8.
- [18] ASHOORI, M., NAJIMOGHADAM, V., ALIZADEH, S., SAFI, M. Classification and clustering algorithm application for prediction of tablet numbers: case study diabetes disease. *journal of the Health Information Management Association of Australia*. 2013, 10(5), p. 739-749. ISSN 1833-3575.
- [19] SHARM, H., KUMAR, S. A survey on decision tree algorithm of classification in data mining. *International Journal of Science and Research*. 2016, 5(4), p. 2024 -2097. eISSN 2319-7064.
- [20] Statistical yearbook of road transportation. Summary of highways and road transportation statistics of the country in management view. Roads and Road Transportation Organization, Data and Communication Management and Public Relations Management, 2016.