Traffic accident severity analysis in Barcelona using a binary probit and CHAID tree

Ahmad Aiash, Francesc Robusté
Civil Engineering School, UPC – BarcelonaTech, Jordi Girona 1-3, 08034 Barcelona (Spain)

Abstract

Traffic accidents are still wide causation for fatalities around the globe. The set of alarm for this cause of deaths is still on, since the number of fatalities is still representing an enormous issue and a challenge for most governments. In Barcelona, similar to the rest of the world, traffic accidents are threatening lives and raising the need to lessen the number of both fatalities and severities. This study is conducted to grasp the correlations between different classification factors with accident severities and fatalities. A total of 47,153 traffic accident cases that occurred between 2016 and 2019 are utilized. Then, a binary probit model and Chi-square automatic interaction detector are exploited to grasp the impact of several risk factors. The results confirmed that males and 65 years and older injured persons are more exposed to severe or fatal injuries compared to other categories. Pedestrians and drivers are found to have higher probabilities compared to passengers in being involved in severe or fatal injuries. Weekends, afternoon, night timings all have higher odds of having severe or fatal traffic accidents. The findings of this study can help road authorities in targeting these risk factors to mitigate their impact to achieve Vision Zero.

Keywords: Barcelona, Accidents, Severities, Probit model, CHAID

1. Introduction

“Every fatal or serious crash on our roads is a tragedy. It is our moral obligation – our shared responsibility – to take road safety seriously.” Violeta Bulc. A traffic accident is still a major issue in Europe and the rest of the world. About 25 thousand lives were claimed, while more than a million persons were injured as a cause of traffic accidents in 2016 in Europe [1]. A report [2] was prepared to provide a performance index for 32 European countries regarding road safety. The report showed that there was a 3% decrease in the number of accidents in 2019 compared to 2018 for the total number of accidents that happened in the EU27. 7,023 road fatalities less in 2019 than in 2010 for the total number of EU27. Greece has the highest decreasing percentage with -44.4%, while Malta had the highest increasing percentage of 6.7%. For Spain, the percentage of road fatalities was 30.4% by comparing the two years 2010 and

2019. However, the medium-term target was halving the number of deaths, but as shown, the progress of lessening the number in some countries is stagnated.

In 2017, the EU transport ministers adopted the Valletta declaration to reduce the number of severe injuries, which can be considered as a consistent plan with the EU Commission [3]. The European Commission has also set a new target to lessen the number of serious injuries to half between 2020 and 2030. Most European countries, therefore, have set this time period for the new national road safety strategy that includes a target for death and severe injuries reduction. Key Performance Indicators (KPIs) were introduced to assess the performance of meeting the targets during the defined period [4]. Those indicators are speed compliance, utilizing the seat belt beside the child restraint system, protective equipment using, alcohol influence during driving, distractions while driving that include handheld devices, new cars safety, the safety of the infrastructure, and lastly, the care for post-crash. The reason for setting such indicators is acting in a proactive manner for identifying the problems due to their ability in providing a full picture of the status of the level of road safety [5]. Spain has set a strategy between 2011-2020 [6] to achieve safe and sustainable mobility. This strategy is mainly comprised of five main elements. The ecological aspect is one of those elements that is consisted of reducing air and noise pollution and improving energy efficiency besides other targets. The second aspect is the competitive aspect that aims to improve journeys quality and goods distribution alongside guaranteeing the regularity of journeys. For the safe aspect, this element is setting the plan to reduce the number of traffic accidents and severe injuries in parallel with enhancing the care that is given to road accident victims. Moreover, this aspect is aiming to decrease the rate of accidents among vulnerable road users. The healthy element is consisting of promoting cycling and walking and enhancing the mental and physical health of people. Lastly, the universal aspect that composed of securing an acceptable cost for all social sectors who use public transport. Additionally, the accessibility for all modes of transports should be improved for all people. The order for elements is negligible as all the aspects are important and part of the plan. Some aspects have more categories. However, the most relevant categories are mentioned. The vision for the Spanish strategy is “The citizens have the right to a Safe Mobility System in which everyone, citizens and agents involved, has a responsibility”. This vision is consisted of five different values shape this vision including shared rights and duties, sustainable mobility, safe users, roads, environments, and vehicle.

Catalonia is, indeed, on the same path of having a strategic plan related to traffic accidents. The first strategic plan was established in 1999, and now Catalonia has a total of six strategic plans including the last published plan [7]. The target of reducing the number of traffic accident severities and fatalities was met in the last three prepared plans, as all of them exceeded the planned reduction percentages. Based on the last report [7], the number of fatalities is one of the main challenges that still needs improvement in both urban and inter-urban areas. The priority for vulnerable road users should be maintained. Campaigns can also be conducted to target risky road users. Therefore, pillars were set to pledge the success of the implementation of the strategic plan. Law enforcement, search for efficiency and focus results, education and training, aspiring to leadership in Europe, shared responsibility in road safety, and lastly, the use of technologies associated with road safety are those main pillars. The overall target for the last strategic plan is to reach 50 % less in the number of fatalities in 2020 compared to 2010, while the overall target for 2050 is to reach zero fatalities and severe injuries with lifelong consequences under the Vision Zero.

Barcelona has also set a plan for the period 2019-2022 regarding road safety [8]. A 20 % reduction in fatalities is set to be a goal to be accomplished in 2022 compared to the number of fatalities in 2018. For severe injuries, a 16 % reduction should be reached in 2022 compared also to 2018. Therefore, a number of aspects should be considered in order to achieve those goals. The accidents that can lead to severe injuries should have the priority to prevent the occurring of similar accidents. Vulnerable users should also have the priority in identifying certain measures to protect them. The reasons for causing a high number of accidents including accident types and specific road users should be resolved. The causes for fatalities related to traffic accidents are also considered so that the fatalities numbers can be decreased. Four main action domains are conducted to achieve the proposed goals. Education is one of the four spheres that composed of prevention, training, communication, and research action and projects related to road safety. Engineering is the second sphere that is considered to sustain the proposed targets by involving in the technical measures related to vehicles, roads, and other related aspects. The third sphere is enforcement that means the correction of conduct which includes the implication of different regulations and the operations planning. Lastly, support and patronage measurements for the road accidents victims are under the last sphere which is emergency. Identifying the reasons for traffic accident severities and deaths, as mentioned earlier, are one of the spheres that are laid to undermine their impacts and reducing the number of fatalities and severities. Several studies were conducted to comprehend the effect of different potential factors so that the risk factors can be targeted and neutralized.

A study [9] revealed the economic cost of traffic accidents in Barcelona in 2003. A 367 million is estimated for the total accident cost of accidents with 329 million for the direct cost. The gender differences in being involved in a risk injury due to traffic accidents by age, mode of transport, and severity were studied by research [10]. The data comprised of the time spent by people, who are residents and over three years old, in traveling in Catalonia in the period between 2004 and 2008. A Poisson regression model was exploited to delve into those differences. A significant impact in traffic accident injuries for both gender and age was distinguished. Males were found to have a higher risk of having severe and slight injuries amid the group of child pedestrians and young drivers. In contrast, females had a higher risk of injuries in older groups. Furthermore, the fatalities rate of males was higher than females.

Most of the pedestrian fatality who are hit by a motorized two-wheeled vehicle is found to be because of Traumatic Brain Injury [11]. Other predictors that can increase the level of severity when examining elderly pedestrians in rural traffic accidents were visibility conditions and the distribution of services besides other factors [12]. Another study [13] examined the gender, age, and type of user impact on injuries in Barcelona. The data represented the traffic injuries cumulative incident rate of disability in 1993. Males were more exposed to the incidence compared to females when considering motorcycle and car users. Female pedestrians were found to be more involved in incidents compared to males. Consistently, additional research [14] showed that males drivers were more engaged in a higher level of severities due to motorcycle/moped accidents compared to females.

For age categories, research [15] concluded that drivers who are older than 60 years were found to be more exposed to severe or fatal injuries compared to other age categories. Both driver age and accident type with also including lighting and the number of injuries resulting from the accident were confirmed as risk factors related to severe and fatal injuries [16]. Similarly, age, education level, time of the day, roadway condition and its type besides other factors were also confirmed to have an impact on the severity of traffic accidents by another study [17]. For traffic barrier crash severity, young drivers, driving during adverse weather, and having a turn before hitting the barrier were found to lessen the level of the injury severity [18]. Another study [19] examined several variables that may influence the level of injury that were resulted from truck crashes. The results revealed that weather condition was considered as a risk factor related to severity of injury. For non-fatal accidents, several factors that can affect the severity of injuries were found to be speeding, the age of the driver, and different seasons [20]. Consistently with factors that can increase the level of severity, a study [21] examined different type of vehicles and their consequential risk factors. Beginning with private vehicles, the age of the vehicle, driver gender, district board, temporal factor represented by the time when the accident happened, and lastly, the light condition were all found to be significant predictors in affecting the severity of the accident. Followingly, for goods vehicles, only seatbelt usage, and the temporal factor represented by the weekday were found to be the risk factors associated with severe injuries. Lastly, for the last type that was examined which was the motorcycle, three factors with respect to injury severity were detected including vehicle age and two temporal factors including weekday and the accident time. Passengers may also be exposed to a different level of severity based on their seats inside the vehicle. A rear left seat was found to have the highest risk of severe and moderate injuries followed by a rear right, rear middle, and front seat, respectively [22].

Volume to capacity ratio (v/c) influence on accident rates was explored by a study [23] with including different other risk factors. The study exposed that there is an inverse relationship between the number of the v/c ratio and the number of traffic accidents. In the case of weekdays and weekends comparison related to traffic accidents and when the v/c ratio is high, weekdays had a higher number of traffic accidents. Contradictory, a recent study [24] revealed that crashes had higher odds of happening on weekends compared to weekdays when the impact of a new casino was examined in suburban Another study [25] presented the correlations between the spatiotemporal factors and injuries and fatalities related to traffic accidents in Columbia. The findings of this study showed that there was a higher risk of being involved in traffic accidents from Wednesday to Saturday compared to Sunday. Additionally, Dry and sunny weather was found to have an impact on decreasing the odds of having traffic accidents similar to the impact of a larger number of traffic lights. In Barcelona, a study [26] found that speed limits were not followed by drivers as the average measured speed limits were higher than legal speeds.

Therefore, the main objective of this study is to analyze different classification variables that are associated with traffic accidents' severities. The reason for this is to lessen the number of these types of traffic accidents by detecting the potential risk factors. Afterward, road authorities can legislate the required laws to undermine their influence on traffic accidents severities. Additionally, it can help the authorities to keep in mind these factors while constructing new roads to proactively act to prevent them. Eventually, this study can be considered as a path for the Vision Zero goal in Barcelona, Spain. Two approaches are utilized parametric and non-parametric models including the probit model and Chi-square automatic interaction detector model. This study exploited a large number of traffic accidents that occurred in Barcelona including 47,153 cases with focusing on five potential risk factors. Till now, no similar study is conducted in Barcelona with this amount of data and applying similar approaches. This study should highlight those potential independent variables alongside validating the models that are exploited for the data provided.

2. Methodology

A. Traffic accidents data in Barcelona

Barcelona province is providing a public access database for different data types which is called "Ajuntament de Barcelona's open data service". The dataset themes include administration, city and services, economy and business, population, and territory theme. Traffic accidents are classified under the city and services theme [27]. The data that is exploited in this study is comprised of 4 years duration from 2016-2019. The data is, also, consisting of different categories for the response variable that is the level of injury that consisted of different levels of slight injuries, severe injuries, and fatal injuries. In more detail, each category is also divided into other subcategories. However, slight injury cases are all merged together to facilitate the analysis procedure in this study. For different severe and fatal injuries, both are integrated together. Therefore, a total of only two categories are remained, so that that the response variable is, now, a binary variable. The reason for this is, both severe and fatal injuries are focused on to determine the factors that impact the occurrence for both classes and because of the number of fatal injuries cases to have a better statistic estimate. Moreover, the objective of this study is help road authorities in achieving the Vision Zero target for both fatalities and severe injuries resulting from traffic accidents. Therefore, both severe and fatal injuries are focused on in this study. Alongside combining response variable classes, blank and undefined data are eliminated from the final data set as a part of the data preparation. Moreover, some classification variables categories are combined, as well, based on the data and to have a better insight into the dataset.

Table 1, as shown, is presenting the classification variables frequencies with respect to injuries in the city of Barcelona during 2016-2019. The accidents reported by the Guàrdia Urbana in the city of Barcelona. A total of 47,153 report cases are utilized after the data preparation has been achieved. A total of five different classification variables are fetched for this study in order to apply a model that can exhibit the correlations and how each factor can impact the dependent variable. It is worth mentioning that the accidents occurred in the urban area as all the reported accidents happened inside the city region. The difference between rural and urban accidents can be noticed as higher number of pre-hospitals deaths were found at rural areas compared to urban area [28]. The study found that the lower use of seatbelt, helmets or other safety precautions were the reasons for the deaths in the rural region. Similarly, the accident injury level at a rural area was found to be higher compared to an accident at an urban area [29]. Weekends and weekdays are classified under the week which is an independent variable. Daytime is another independent variable that is also considered that includes morning, afternoon, and night timings. In addition, gender impact is considered. Four different age categories are considered in this study including injured who are less than 15 years old, between 16 and 24 years old, between 25 and 64 years, and lastly, injured who are 65 years and older. Three different user types are implicated including pedestrian, passenger, and drivers.

Table1. Classification variables frequencies.

		Injury level		
		Severe or fatal injury	Slight injury	Total
Week:				
	Weekend	246	9359	9,605
	Weekday	693	36,855	37,548
Daytime:				
	Afternoon	488	23274	23762
	Night	160	5215	5375
	Morning	291	17725	18016
Gender:				
	Female	305	18156	18461
	Male	634	28058	28692
Age:				
	0 - 15	25	1513	1538
	16-24	121	6893	7014
	25-64	655	34537	35192
	65 +	138	3271	3409
User:				
	Pedestrian	269	4574	4843
	Passenger	86	8989	9075
	Driver	584	32651	33235
Total		939	46214	47153

B- Binary probit model

In this study, a binary probit model is a traditional statistical method that lays under the parametric approaches that are employed to grasp the correlations between the classification variables and the response variable by following previous proceedings [30] [31]. SAS University Edition is exploited for employing the binary probit model [32]. The reason for choosing the probit model is having the level of injury as a dependent variable and establish, later on, a comparison between this traditional model and the data mining technique in the overall results and the concluded correlations. The employed data is comprised of a binary response variable which consists of the first category which is slight injury, and the second category is severe or fatal injury. Assume for each injured i, U_i is the net utility of injury level. The total number of exogenous variables (X_i) is I. U_i is related to X_i.

$$U_i = X_i\beta + \varepsilon_i \tag{1}$$

Where the parameter estimates and explanatory variables vectors are β and X_i, respectively. The error is ε_i and it is assumed to follow a standard normal distribution.

The latent model is

$$y_i = X_i\beta + \varepsilon_i \tag{2}$$

When the relationship between U_i and the observable injury level y_i satisfies:

$$y_i = \begin{cases} 1 & if \ U_i > 0 \\ 0 & otherwise \end{cases} \tag{3}$$

Pr $(y_t = 1|x_t)$ is the conditional probability. The change in this conditional probability is measured by j^{th} which is the coefficients vector element $\beta, \beta_j(j\epsilon\{1,2,\dots,I\})$. This change is measured when there is a unit change x_i^j. j^{th} is an element in vector X_i. The conditional probability is also assumed to have a normal distribution. Then, the standard normal cumulative distribution function is $\phi(.)$. The conditional probability function is:

$$\Pr(y_i = 1|x_i) = \phi(X_i\beta) \tag{4}$$

C- CHAID tree

Decision trees (DTs) models, in general, are considered as a classifier method that uses values of attributes to make a discrete prediction. DT is normally a non-parametric technique that does not have any dependency on any functional form and no prior probabilistic knowledge requirement [33]. Chi-square automatic interaction detector that is abbreviated as CHAID tree is part of DT family. This model is developed by [34] as an efficient approach for tree growing. [35]. To compare the results that are extracted from the binary probit model with a machine learning model, CHAID tree is chosen due to its ability in determining the correlations between the independent variables and the dependent variable [36]. Several reasons are behind choosing CHAID tree to analyze the data. Firstly, it can develop two levels and more for any tree level. This technique can also handle all types of variables and provide a good classification for the chosen variables by developing a simple chart that depicts the correlations for the interpretation process that follows the deployed model as the interpretation process can be a hideous process with other data mining techniques. It has to be mentioned that several other techniques were explored before choosing the CHAID tree as the data mining technique in this study. Neural network, XGBoost tree, and Linear Support Vector Machine have all provided similar accumulated accuracy with 98.059 accuracy similar to the employed CHAID tree. Additionally, CHAID exploits statistical significance test as criterion and evaluate all values of independent variables. Similar values are merged that are related to the dependent variable and other values that are not homogeneous are maintained. The first branch of the constructed tree is formed by the best independent variable. Then, each child node is developed from a group of the similar (homogeneous) values. Continually, this process is maintained till the tree is developed. Due to the dependent variable that is used in this paper is categorical variable, chi-square test is used as a measurement level of this variable.

The type of the dependent variable determines the process of calculating the p-values. In this study, the target variable is a categorical variable. Therefore, Pearson chi-square statistic and its corresponding p-values are calculated as shown in equation 5 and 6, respectively. Where a chi-square distribution $d = (J-1)(I-1)$ degrees of freedom is followed by X_d^2 . Where X is the predictor and Y is the target variable. The observed cell frequency is n_{ij}, while the expected cell frequency $(x_n = i\ ,\ y_n = j)$ is $\widehat{m}_{ij}$. Categories are compared during the merge process to consider only the records that belong to the comparison categories. All the current nodes are utilized during the split step as all the categories are considered in p-value calculations. Based on the observed and expected frequencies, the likelihood ratio chi-square and its p-value is calculated in equation 7 and 8, respectively. For the expected frequencies it is calculated based on two cases. The one with no case weights which is calculated as shown in equation 10 and 11. For the one with specified weights, the expected frequencies are calculated in equation 12 and 13. α_i and β_i are estimated parameters. This model is exploited using IBM Watson with using SPSS modeler set. More details and steps for this algorithm can be fround at guide [37].

$$n_{ij} = \sum\nolimits_n f_n I(x_n = i \wedge y_n = j) \tag{5}$$

$$X^2 = 2\sum_{j=1}^{J}\sum_{i=1}^{I}\frac{(n_{ij} - \widehat{m}_{ij})^2}{\widehat{m}_{ij}} \tag{6}$$

$$p = \Pr(X_d^2 > X^2) \tag{7}$$

$$G^2 = 2\sum_{j=1}^{J}\sum_{i=1}^{I} n_{ij}\ln(\frac{n_{ij}}{\widehat{m}_{ij}}) \tag{8}$$

$$p = \Pr(X_d^2 > G^2) \quad (9)$$

$$\widehat{m}_{ij} = \frac{n_{i.}n_{.j}}{n_{..}} \quad (10)$$

Where

$$n_{i.} = \sum_{j=1}^{J} n_{ij}, \quad n_{.j} = \sum_{i=1}^{I} n_{ij}, \; n_{..} = \sum_{j=1}^{J}\sum_{i=1}^{I} n_{ij}. \quad (11)$$

$$\widehat{m}_{ij} = \overline{w}_{ij}^{-1}\alpha_i\beta_i \quad (12)$$

Where

$$\overline{w}_{ij} = \frac{w_{ij}}{n_{ij}}, \quad w_{ij} = \sum_{n\in D} w_n f_n I\,(x_n = i \wedge y_n = j) \quad (13)$$

3. Results and discussions

Table 1 depicts how the model is significant. All likelihood ratio, score, and Wald are statistically significant with p-values less than 0.0001 for the applied probit model. Table 2 presents the significance of the classification variables that are used in the model. This table is showing the hypothesis tests for each variable individually. Weekday, the daytime, gender, age, and user are significantly improving the model-fit based on their Chi-square test and their p-values.

Table 1. Testing Global Null Hypothesis

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	419.0512	9	<.0001
Score	519.1789	9	<.0001
Wald	431.3261	9	<.0001

Table 2. Type 3 Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
Week	1	20.4102	<.0001
Daytime	2	48.5499	<.0001
Gender	1	21.2514	<.0001
Age	3	41.1812	<.0001
User	2	279.5212	<.0001

Table 3 portrays the coefficient estimates for all the classification variables alongside their standard errors, Chi-square test statistics, and their p-values. Accidents on weekends compared to accidents on weekdays increases the odds of having severe or fatal injuries by 0.1474. In Madrid, the probabilities of having traffic accidents, in general, on weekends were found to be higher compared to weekdays [38]. Weekends and public holidays were also found to have higher odds of severe injuries that are resulted from a traffic accident [39]. Accidents that happened in the afternoon and night period compared to accidents that happened in the morning period rise the log-odds of having severe or fatal injuries by 0.1263 and 0.3010, respectively. In contrast, a study found that the early morning period alongside other factors can lead to higher odds of having severe injuries [40]. Females injured decrease the odds by 0.1433 compared to males injured. This finding is consistent with a previously conducted study that found that males were more likely to have severe or fatal injuries compared to females [41]. For age categories, injured who was 15 years old and younger, between 16 and 24, and between 25 and 64, all decrease odds compared to injured who were older than 65 years old with 0.4948, 0.283, and 0.2282, respectively. To compare this age category, a study concluded that drivers who were between the age group of 46-55 were at higher risk compared to older drivers [42]. Pedestrians have an increased odd of having fatal or severe injuries with 0.5336 compared to drivers, while passengers have reduced the odds of having severe or fatal injuries by 0.2259 compared to drivers. A study that was conducted in France confirmed that pedestrians are at a higher risk compared to car occupants [43]. Besides the chosen and concluded risk factors in this study, there are other potential risk factors that can influence traffic accident severity such as the vehicle type. For instance, bus accidents were found to increase the odds of having fatal injuries when the bus crushed with non-motor vehicles [44]. Another conducted study [45] focused on the factors that impact bus-pedestrian accidents and found that intersections, darkness, pedestrians walking on carriageway with traffic, high speed zone alongside other factors can lead to a higher risk of pedestrian fatality.

Table 3. Analysis of the classification variables[1].

Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	-1.9796	0.0515	1477.5193	<.0001
Week	Weekend	1	0.1474	0.0326	20.4102	<.0001
	Weekday (Reference)	0	0	.	.	.
Daytime	Afternoon	1	0.1263	0.0308	16.7993	<.0001
	Night	1	0.301	0.0438	47.2094	<.0001
	Morning (Reference)	0	0	.	.	.
Gender	Female	1	-0.1433	0.0311	21.2514	<.0001
	Male (Reference)	0	0	.	.	.
Age	0 -15	1	-0.4948	0.0938	27.8364	<.0001
	16-24	1	-0.2833	0.058	23.8553	<.0001
	25-64	1	-0.2282	0.0475	23.0581	<.0001
	65+ (Reference)	0	0	.	.	.
User	Pedestrian	1	0.5336	0.0387	190.038	<.0001
	Passenger	1	-0.2259	0.0473	22.8035	<.0001
	Driver (Reference)	0	0	.	.	.

Figure 1 is depicting the developed tree for the applied model including all previously mentioned independent variables. The tree is presented for the 30 % testing set after the model 70 % training set. The thickness of the branch is presenting the number of accident injuries. The thicker branch means that this branch has a higher number of traffic accident injuries. Each node has the number of injuries that it represents including both slight injuries number and the severe or fatal injuries number. The categories are only mentioned that are branches of the node which represent a different independent variable. The most important variable is the one that the model started with which is the user that includes driver, passenger, and the pedestrian. In general, and through scanning the developed tree, males have a higher number of severe or fatal injuries compared to females during certain conditions for drivers. Older pedestrians are also at great risk compared to other age groups. Passengers have higher odds of having severe or fatal injuries during the evening or night period compared to the morning period during weekends. In contrast, drivers are more likely to have severe or fatal injuries resulting from traffic accidents during the morning period.

[1] The reference for the response variable is severe or fatal injuries category.

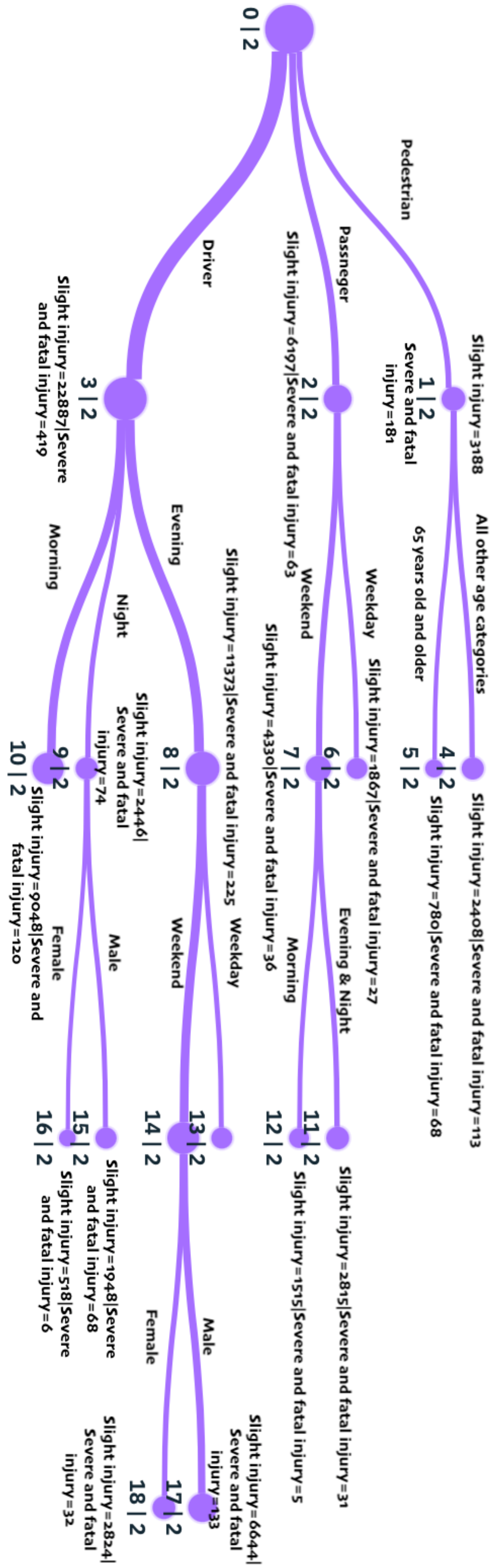

Figure 1. Predictors' classifications based on CHAID tree.

4. Conclusions

Traffic accidents are still wide causation for fatalities around the globe. The set of alarm for this cause of deaths is still on due to the fact that the number of fatalities is still representing an enormous issue and a challenge for most governments. In Barcelona, similar to the rest of the world, traffic accidents are threatening lives and raising the need to lessen the number of both fatalities and severities. Different strategies and plans are set to meet this target; however, the number of deaths is exemplifying a formidable issue as each life counts, alongside the impact of traffic accidents in causing severe injuries and disabilities. Eventually, this study is conducted to attempt to specify the correlations between certain chosen classification variables and locate the areas of high occurrence for traffic accident severities and fatalities. The results of this study can be considered as an initiation for future studies that deem traffic accident research in Barcelona. A binary probit model is exploited for this reason to determine the correlation between the classification variables and the response variable. In this study, five different classification variables are considered including temporal factors represented by two risk factors, age of the injured, gender, and lastly, user type including the driver, pedestrian, and passenger.

The impact of the selected factors is studied by the binary probit model. The results of the model show that females injured have a low probability of being involved in severe or fatal injuries compared to males. For the temporal factors, accidents that happen on weekends, afternoon, or night have higher odds of having severe or fatal injuries for the involved parties. Injured who are 65 years or older are more engaged in severe or fatal accidents compared to other age categories. Pedestrians are also more implicated in severe or fatal injuries. Following the results of the applied binary probit model, the results of the applied CHAID tree are consistent with the previous results. Driver, passenger, pedestrian are the most important predictors based on the applied CHAID. Older pedestrian, male drivers during the evening and night period are found to have a higher risk of fatal or severe injuries compared to other categories. These findings can help road planners and road safety auditors to detect these factors that can lead to a higher risk of having severe or fatal injuries. Then, legislate rules that can alleviate their impact or lessen the number of this type of accidents.

References

[1] European Road Safety Observatory, «Annual Accident Report 2018,» European Commission, 2018.

[2] J. Carson, D. Adminaité-Fodor y G. Jost, «14th Road Safety Performance Index Report,» ETSC, Brussels, 2020.

[3] Council of the European Union, «Council conclusions on road safety - endorsing the Valletta Declaration of March 2017,» Brussels, 2017.

[4] European Transport Safety Council, «Briefing | EU Strategic Action Plan on Road Safety,» ETSC, 2019.

[5] European Transport Safety Council, «Briefing: 5th EU Road Safety Action Programme 2020-2030,» ETSC, 2018.

[6] «Spanish Road Safety Strategy 2011-2020. Executive Summary,» Traffic General Directorate, 2011.

[7] Servei Català de Trànsit, «Strategic Road Safety Plan for Catalonia 2014-2020,» Generalitat de Catalunya, 2014.

[8] Servei catala de Transit, «Pla Local de Seguretat Viària 2019-2022,» Ajuntament de Barcelona, Barcelona, 2019.

[9] A. García-Altés y K. Pérez, «The economic cost of road traffic crashes in an urban setting,» Injury Prevention, vol. 13, pp. 65-68, 2007.

[10] E. Santamariña-Rubio, K. Pérez, M. Olabarria y A. M. Novoa, «Gender differences in road traffic injury rate using time travelled as ameasure of exposure,» Accident Analysis and Prevention, vol. 65, pp. 1-7, 2014.

[11] M. C. Rebollo-Soria, C. Arregui-Dalmases , D. Sánchez-Molina , J. Velázquez-Ameijide y I. Galtés, «Injury pattern in lethal motorbikes-pedestrian collisions, in the area of Barcelona, Spain,» Journal of Forensic and Legal Medicine, vol. 43, pp. 80-84, 2016.

[12] N. Casado-Sanz, B. Guirao y D. Gálvez-Pérez, «Population ageing and rural road accidents: Analysis of accident severity in traffic crashes with older pedestrians on Spanish crosstown roads,» Research in Transportation Business & Management, vol. 30, p. 100377, 2019.

[13] J. Ferrando, A. Plasència, E. MacKenzie, M. Orós, P. Arribas y C. Borrell, «Disabilities resulting from traffic injuries in barcelona, spain: 1-year incidence by age, gender and type of user,» Accident Analysis & Prevention, vol. 30, nº 6, p. 723–730, 1998.

[14] S. Hidalgo-Fuentes y M. J. Sospedra-Baeza, «Factores asociados a los accidentes de motocicleta en Barcelona, España Contributing factors of motorcycle crashes in Barcelona, Spain,» Ciencias Psicológicas, vol. 13, nº 2, pp. 265-274, 2019.

[15] S. AlKheder, F. AlRukaibi y A. Aiash, «Risk analysis of traffic accidents' severities: An application of three data mining models,» ISA Transactions, vol. 106, pp. 213-220 , 2020.

[16] J. de Oña, R. O. Mujalli y F. J. Calvo, «Analysis of traffic accident injury severity on Spanish rural highways using Bayesian networks,» Accident Analysis & Prevention, vol. 43, nº 1, pp. 402-411, 2011.

[17] G. O. Kadilar, «Effect of driver, roadway, collision, and vehicle characteristics on crash severity: a conditional logistic regression approach,» International Journal of Injury Control and Safety Promotion, vol. 23, nº 2, pp. 135-144, 2016.

[18] M. Rezapour, S. S. Wulff y K. Ksaibati, «Bayesian hierarchical modelling of traffic barrier crash severity,» International Journal of Injury Control and Safety Promotion, vol. 28, nº 1, pp. 94-102, 2021.

[19] C. Dong, S. H. Richards, B. Huang y X. Jiang, «Identifying the factors contributing to the severity of truck-involved crashes,» International Journal of Injury Control and Safety Promotion, vol. 22, nº 2, pp. 116-126, 2015.

[20] D. Potoglou, F. Carlucci, A. Cirà y M. Restaino, «Factors associated with urban non-fatal road-accident severity,» International Journal of Injury Control and Safety Promotion, vol. 25, nº 3, pp. 303-310, 2018.

[21] K. K. W. Yau, «Risk factors affecting the severity of single vehicle traffic accidents in Hong Kong,» Accident Analysis & Prevention, vol. 36, nº 3, pp. 333-340, 2004.

[22] G. Zhai, H. Yang y J. Liu, «Is the front passenger seat always the "death seat"? An application of a hierarchical ordered probit model for occupant injury severity,» International Journal of Injury Control and Safety Promotion, vol. 27, nº 4, pp. 438-446, 2020.

[23] M. Zhou y V. P. Sisiopiku, «Relationship Between Volume-to-Capacity Ratios and Accident Rates,» Journal of the Transportation Research Board, vol. 1581, nº 1, pp. 47-52, 1997.

[24] J. A. Kufera, A. Al-Hadidi, D. G. Knopp, Z. D. Dezman, T. J. Kerns, O. E. Okedele, G. L. Rosenthal y J. K. Tracy, «The impact of a new casino on the motor vehicle crash patterns in suburban Maryland,» Accident Analysis and Prevention, vol. 142, 2020.

[25] A. F. Ramírez y C. Valencia, «Spatiotemporal correlation study of traffic accidents with fatalities and injuries in Bogota (Colombia),» Accident Analysis & Prevention, vol. 149, 2021.

[26] F. Robuste, A. Lopez-Pita y M. Velez, «Optimal social speed limits in a highway: Suitability of the 120 km/h speed limit in Barcelona, Spain,» de Actes de la European Transport Conference, Estrasburg, 2003.

[27] Barcelona's City Hall Open Data Service, «People involved in accidents managed by the Police in the city of Barcelona,» 2020. [En línea]. Available: https://opendata-ajuntament.barcelona.cat/data/en/dataset/accidents-persones-gu-bcn.

[28] M. Li , J. Doong , K. Chang , T. Lu y M. Jeng, «Differences in urban and rural accident characteristics and medical service utilization for traffic fatalities in less-motorized societies,» Journal of Safety Research, vol. 39, nº 6, pp. 623-630, 2008.

[29] W. Hao y C. Kamga, «Difference in rural and urban driver-injury severities in highway–rail grade crossing accidents,» International Journal of Injury Control and Safety Promotion, vol. 24, nº 2, pp. 174-182, 2017.

[30] R. Yu y M. Abdel-Aty, «Using hierarchical Bayesian binary probit models to analyze crashinjury severity on high speed facilities with real-time traffic data,» Accident Analysis & Prevention, vol. 62, p. 161–167, 2014.

[31] R. Garrido, A. Bastos , A. de Almeida y J. P. Elvas , «Prediction of road accident severity using the ordered probit model,» Transportation Research Procedia, vol. 3, p. 214 – 223, 2014.

[32] SAS, «Chapter 79 The PROBIT Procedure,» de SAS/STAT® 13.1 User's Guide, North Carolina, SAS, 2013.

[33] J. Abellán, G. López y J. de Oña, «Analysis of traffic accident severity using Decision Rules via Decision Trees,» Expert Systems with Applications, vol. 40, nº 15, pp. 6047-6054, 2013.

[34] G. V. Kass, «An Exploratory Technique for Investigating Large Quantities of Categorical Data,» Journal of the Royal Statistical Society: Series C (Applied Statistics), vol. 29, nº 2, pp. 119-127, 1980.

[35] A. Dobra, «Decision Tree Classification,» de Encyclopedia of Database Systems, Springer, New York, NY, 2018.

[36] M. J. A. Berry y G. S. Linoff, Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management, Wiley, 2004.

[37] IBM, «IBM SPSS Modeler 18.0 Algorithms Guide,» IBM Corporation, 2016.

[38] A. Kanaan, P. Huertas, A. Santiago, J. A. Sánchez y P. Martínez, «Incidence of different health factors and their influence on traffic accidents in the province of Madrid, Spain,» Legal Medicine, vol. 11, nº 1, pp. S333-S336, 2009.

[39] G. Zhang, K. K. Yau y G. Chen, «Risk factors associated with traffic violations and accident severity in China,» Accident Analysis & Prevention, vol. 59, pp. 18-25, 2013.

[40] M. A. Quddus, R. B. Noland y H. C. Chin , «An analysis of motorcycle injury and vehicle damage severity using ordered probit models,» Journal of Safety Research, vol. 33, nº 4, pp. 445-462 , 2002.

[41] A. Vorko-Jović, J. Kern y Z. Biloglav, «Risk factors in urban road traffic accidents,» Journal of Safety Research, vol. 37, nº 1, pp. 93-98, 2006.

[42] G. O. Kadilar, «Effect of driver, roadway, collision, and vehicle characteristics on crash severity: a conditional logistic regression approach,» International Journal of Injury Control and Safety Promotion, vol. 23, nº 2, pp. 135-144, 2016.

[43] L. Bouaoun, M. M. Haddak y E. Amoros, «Road crash fatality rates in France: A comparison of road user types, taking account of travel practices,» Accident Analysis & Prevention, vol. 75, pp. 217-225, 2015.

[44] S. A. Samerei, K. Aghabayk, A. Mohammadi y N. Shiwakoti, «Data mining approach to model bus crash severity in Australia,» Journal of Safety Research, vol. 76, pp. 73-82, 2021.

[45] S. A. Samerei, K. Aghabayk, N. Shiwakoti y S. Karimi, «Modelling bus-pedestrian crash severity in the state of Victoria, Australia,» International Journal of Injury Control and Safety Promotion, vol. 28, nº 2, pp. 233-242, 2021.