ROAD SAFETY PREVENTION BY ACCIDENT RISK ASSESSMENT

Hugo S. Oliveira

University of Porto

ABSTRACT

To achieve sustainable long-term transport infrastructure development, there is a growing need for fast, reliable, and effective methods to evaluate and predict the impact of traffic safety measures. Several cities require this need and need for an active traffic safety approach. Road traffic accidents account for a large percentage of several casualties and fatalities in cities every year. It is estimated that road traffic cost 3% of the country's domestic products, with more than half of the fatalities belonging to the most vulnerable road users. This work aims to perform a detailed analysis of the genre of accidents, the most common places, and correlate with the existing infrastructures to enable city authorities to develop efforts to minimize this problem. For the task, we employ detailed exploratory analysis to find correlations between the several intervenients and infrastructures present in cities, identifying the most critical points and establish association rules to determine the leading cause for the severe accidents (fatalities). Several essential aspects of traffic accidents are identified, with strong relations of causality leading to the occurrence of fatal victims.

Index Terms**—** Road Prevention, Fatalities identification, Fatalities prediction

1. INTRODUCTION

Approximately 1.35 million people die each year as a result of road traffic crashes. Several counties and organizations have established an ambitious target of halving the global number of deaths and injuries from road traffic crash [1]. It's estimated that road traffic crashes account for more than 3% of their gross domestic product [2]. Half of all road traffic deaths are among vulnerable road users such as pedestrians, cyclists, and motorcyclists. To mitigate this problematic, this work presents a careful analysis of the leading causes of road accidents, finding tricky sections and areas where a vast number of fatalities occur, and establishing a relation of causality for the most severe cases.

Modern cities nowadays need to accommodate automobiles, pedestrians, and other smooth locomotion genres in harmony, grating all intervenients' safety, protecting the more fragile.

The study in [3] shows that 93% of the world's fatalities occur on the roads of low- and middle- income countries, even though these countries have a 60% of the world's vehicles. Road traffic injuries have a considerable economic loss to the individuals, their relatives, and a nation as a whole, leading to an increase in the cost of treatments and the lost productivity for those killed and disabled by their injuries, creating a time burden for close relatives.

Several groups are particularly prone to be exposed to some related road accidents with severe consequences. The road victims can be categorized by several intrinsic characteristics, such as age and gender. Road traffic injuries are the leading cause of death for children and young adults aged between 5 and 29. Additionally, the sex of young age victims shows clear evidence that young males are more likely to be involved in road traffic crashes than females, with about three quarters (73%) of all road traffic deaths among young males under 25 years [4].

Besides, several risk factors are associated with a road accident can be considered [5]. First, how to accommodate human error while ensuring ensure a safe transport system for all road users. This approach should take into consideration people's vulnerability to severe injuries in road traffic crashes. By recognizing that the road systems must be designed to be forgiving of human error, serveral interventions can be put in place, such as creating safe roads and roadsides, safe speeds, safe vehicles, and safe road users to eliminate fatal crashes and reduce serious injuries.

Speeding is directly related to the likelihood of a crash occurring, and the severity of the consequences of the collision [6]. It increases the chance of death risk for pedestrians, car-to-car side collisions, impacting the fatality risk for car occupants, and severely critical when driving under the influence of alcohol and other psychoactive substances. Also, the non-use of motorcycle helmets, seat belts, and child restraints leads to increased road accidents. Nowadays, the massive use of mobile phones and other distraction gadgets is a growing concern for road safety.

Concerning the road infrastructures, their design combined with the surrounding environment and auxiliary devices can have a considerable impact on their safety [7]. It is essential to ensure adequate pedestrian, cyclists, and motorcyclists facilities in all weather and illumination conditions. Measures such as footpaths, cycling lanes, safe crossing points, illumination, water drainage combined with other traffic calming measures can be critical to reducing the risk of injury among these road users.

Effective interventions can be set in practice to address the identified problems to create safer infrastructure. Incorporating road safety features into land-use and transport planning, raising public awareness, and enforcing laws can reduce road crashes and ultimately reduce fatal victims.

With this in mind, we perform a detailed exploratory analysis to identify the main causes that lead to the occurrence of fatal accidents by employing visualization techniques supported by correlative analysis of the accident severity, supported by the analysis of the association rules concerning the accident severity fatal and main consequent, enabling its interpretability and explainability, since the indented rules can establish a causal relationship between the antecedent and the consequent.

As complementary work, we evaluate several predictive models to predict the number of fatal victims given a set of relevant features for a particular area. This enables authorities to predict the number of fatalities given a specific day and region characteristics, allowing them to act in conformity to reduce deaths.

The rest of the paper is structured as follows: Section 2 discusses related work. After that, Section 3 describes the the proposed method in detail and Section 4 present the dataset and main characteristics, details the exploratory analysis performed by visualization, and association rules with a brief discussion of the main findings, present the predictive modeling objective and model performance evaluation for the task to predict the number of fatal victims, supported by a clustering analysis to identify abnormal patterns. After we discuss our evaluation and findings results in a broader form on Section 6. Finally, Section 7 concludes with a summary and outlines interesting directions for future work.

[1]

2. RELATED WORKS

Accident analysis and prediction are novel research with a strong impact in society, such as developing traffic simulators to identify risk indicators and simulate realistic safety-critical events without making unnecessary and unrealistic behavioral assumptions, [8].

Road design and critical sections such as rail-road crossing are commonly associated with fatal accidents, and statistical models can be employed to examine the relationships between crossing accidents and features of these crossings that lead to a large number of casualties, [9].

With the availability of large volumes of real-time traffic flow data along with traffic accident information, there is a renewed interest in the development of models for the real-time prediction of traffic accident risk. However, the available data is frequently noisy, complicated, and even misleading to be used directly. This fact raises the question of selecting the most important explanatory variables to achieve an acceptable level of accuracy for real-time traffic accident risk prediction.

A solution is proposed by [10] by using Frequent Pattern tree (FP tree) based variable selection method to identify the relevant features to be used by several predictive models for accident risk prediction. Significant limitations in real-time crash prediction models are identified in [11], and constraints are overcome by the use of a random multinomial logit model to identify the most important predictors as well as the most suitable detector locations to acquire data to build such a model.

Afterward, it applies a Bayesian belief net (BBN) to build the real-time crash prediction model. Short-term traffic turbulence significantly affects the likelihood of a crash or crash potential.

To identify the crash potential, [12] proposes a probabilistic real-time crash prediction model relating crash potential to various traffic flow characteristics that lead to the crash occurrence.

Clustering approaches are also common for traffic accident pattern discovery, [13], specifically, by the use of a community detection algorithm to cluster the data, enabling the reduction of the inherent heterogeneity, combined with association rule learning algorithm to discern meaningful patterns within each, related particularly to high accident frequency locations and incident clearance time. The statistical learning model [14] is also commonly used, enabling evaluating real-time crash risk, with classification and regression tree models being also commonly used to select the most important explanatory variables to be applied to Bayesian logistic regression and Support Vector Machine (SVM) models for accident risk prediction.

3. METHODOLOGY

Concerning the methodology, we start by performing an exploratory analysis by pre-processing the dataset and making a detailed exploratory analysis of the distributions and correlations among these variables concerning several aspects that may lead to road accidents.

Association rules to determine the leading causes of the significant severe accidents are employed to determine the key elements and locations where these situations mostly occur, enabling authorities to mitigate these problems. Also, we will perform clustering to assess some patterns. Concerning predictive tasks, we aim to predict the number of accidents for a given day using available features. This task encompasses the assessment of feature relevance and engineering, model performance comparison, and hyperparameter tuning.

To find relations between the road conditions and the fatal victims, we will employ the Apriori algorithm to find interesting rules that can correlate the severe accident with the

[1]hhttps://github.com/HSOFEUP/FEUP-TAECAC$_2020.git$

road's predominant aspects such as cross-walks, traffic lights, weather.

To sustain the main findings, associative rules are analyzed in combination with conditional plots and other relevant information to identify the several aspects that lead to severe accidents.

4. DATASET DESCRIPTION AND EXPLORATORY ANALYSIS

For the task, we explore the Road Safety Data provided by the UK Road Administration department, containing road safety data about the circumstances of personal injury road accidents in GB, the types of vehicles involved, and the consequential casualties combined with relevant discriminative variables.

Additionally, other auxiliary files contain the code mapping of the several variables in the dataset and some supplemental information.

A summary of the variables is present in table 1

Table 1: Brief description of variables types

Variable Name	Type
Accident Index	char
Location Easting OSGR	int
Location Northing OSGR	int
Longitude	num
Latitude	num
Police Force	int
Accident Severity	int
Number of Vehicles	int
Number of Casualties	int
Date	char
Day of Week	int
Time	char
Local Authority District	int
Local Authority Highway	char
X1st Road Class	int
X1st Road Number	int
Road Type	int
Speed limit	int
Junction Detail	int
Junction Control	int
X2nd Road Class	int
X2nd Road Number	int
Pedestrian Crossing Human Control	int
Pedestrian Crossing Physical Facilities	int
Light Conditions	int
Weather Conditions	int
Road Surface Conditions	int
Special Conditions at Site	int
Carriageway Hazards	int
Urban or Rural Area	int
Did Police Officer Attend Scene of Accident	int
LSOA of Accident Location	char

Some columns have a file containing a table of correspondence to map each of the code numbers (int) into mean full information. We will integrate this information recursively over the columns to fill the data frame with the complete report, mapping the codes to the corresponding information.

The traffic accident dataset encompasses 140k entries with a neglective number of incomplete examples. Some preprocessing such as column renaming for easy access and decoding the column codes into meaningful factors/categories, is performed using an auxiliary file that acts as a look-up table for the dataset.

4.1. Exploratory Analysis

A preliminary analysis of the road type, number of victims, and the degree of severity of the accidents are summarized in Figure 1.

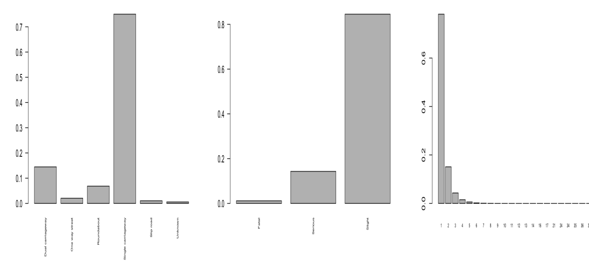

Fig. 1: Accidents histogram arranged by road type, severity and number of casualties

Single carriage accounts for most car accidents, with the majority of the casualties being not severe and the mean number of victims are 1. Correlation analysis between the degree of causality and the number of vehicles involved is exhibited in Fig 2.

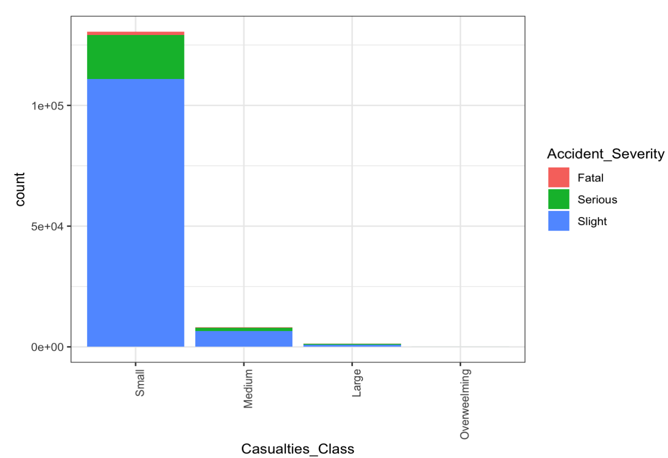

Fig. 2: Correlation between accident severity and number causalities

Road traffic has its dynamics and is possible to observe that Sunday account for a reduced number of accident and fatalities since Sunday is the day of the week people spend most at home, Fig. 3. Also, afternoon accounts for many accidents, a fact that is strongly related with people returning home from work using congested traffic roads, and often the in the afternoon period where the visibility is reduced, (Fig. 4).

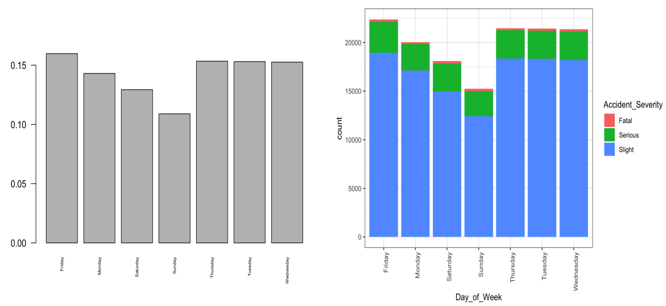

Fig. 3: Accident by day of week and correlative severity analysis

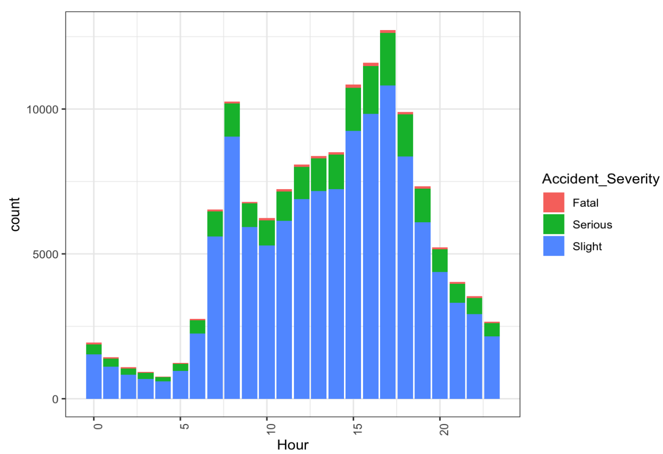

Fig. 4: Accident by hour period

The road surface has a considerable influence on the number of accidents, and evidence shows that rainy periods account for a percentage of accidents. However, the vast number occur in dry conditions (Figs. 5) in the afternoon when people return home.

However, some odd number of accidents are identified. Some road section accounts for a large number of vehicles involved with a significant number of victims. To detail this find we sub-setting the dataset, to attain only occurrences that contain only a large number of victims $geq9$, on Fig. 6 is possible to identify three prominent locations where this genre of an accident have occurred.

Close analysis on google maps of the nearby Chesterfield M1 (Fig. 7) section reveals that these locations correspond mostly to high-ways areas that contain curves and entry points, mostly with high-density traffic and several

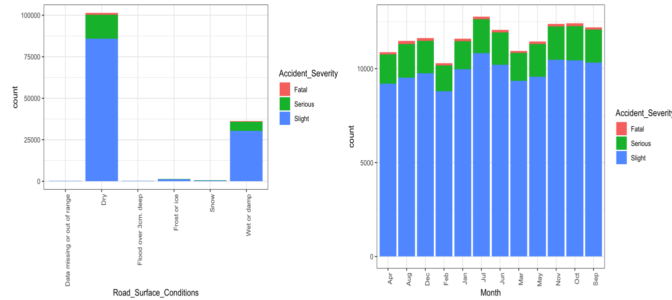

Fig. 5: Accident by road conditions, and month period

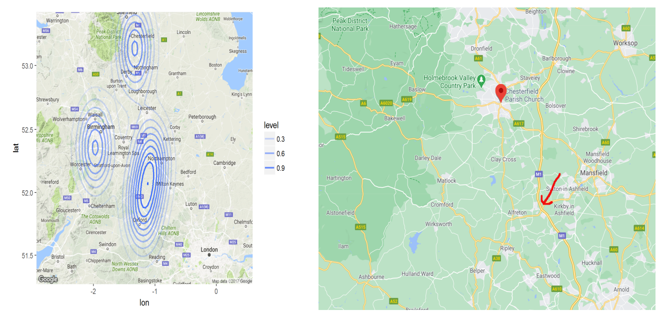

Fig. 6: Accidents involving a large number of victims.

trucks. Also, M1 connects major cities such as Nottingham, Sheffield, and Leeds, where supplies arrived for London harbors to be distributed along with the country.

Fig. 7: Google Street location of the site (Chesterfield).

Junctions intersections account for a large percentage of fatal victims, with most of the occurrences occurring nearby.

The vast majority of accidents occur in urban environments (Figs. 8 ,9) in crossing-junctions without control systems with pedestrian involved. Majority respect the STOP sign. The lack of crosswalks in the nearby also assumes a considerable impact on the number of victims since pedestrians are the most fragile ones.

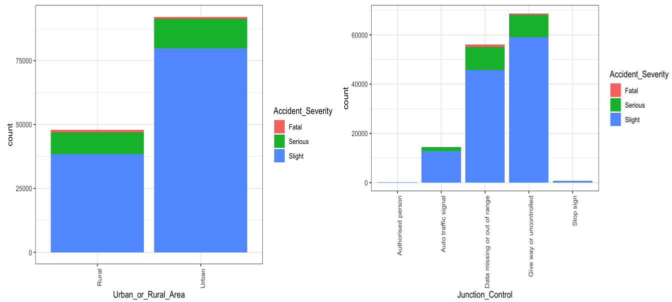

Fig. 8: Accidents by urban/rural environments, junctions.

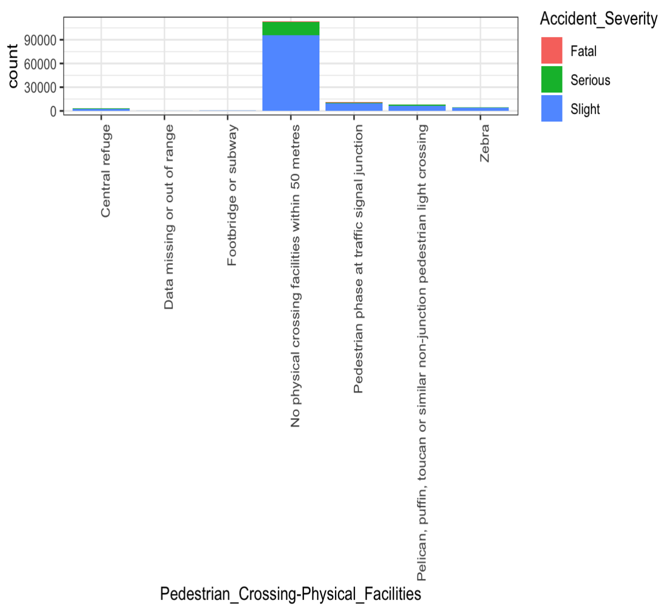

Fig. 9: Accidents with pedestrian facilities conditions.

Police interventions occur mostly in urban scenarios, corroborated by the number of police interventions (Fig. 10, that have a significant incidence in the metropolitan area, where primary police resources are available also.

An interesting fact concerns the fatal accidents in rural areas, presenting some evidence that the lack of crossing facilities and lack of pedestrian infrastructures may increase fatal victims. Geo-referencing these particular cases (Fig. 11) suggests some critical locations where pedestrian and crossing infrastructures should be improved.

To assess the relevance of the features, the entropy gain concerning the response variable of interest (accident severity) is used. In combination with a correlation filter selection, a final set of features is obtained, summarized in Table 2.

Some of the most identified aspects and features that lead to fatal victims are also present in the table, supporting our exploratory analysis.

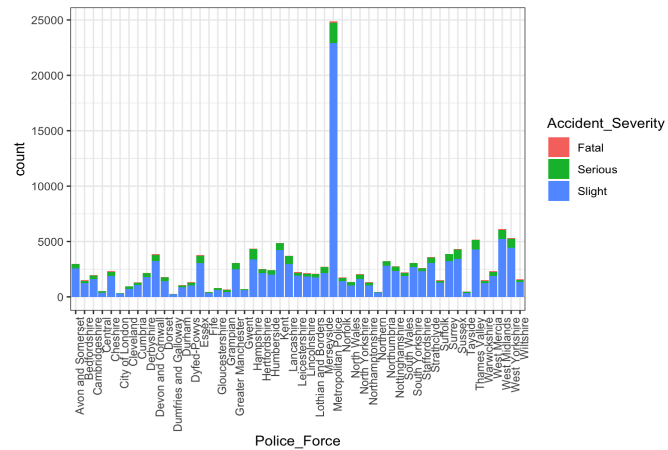

Fig. 10: Accidents where police force was present.

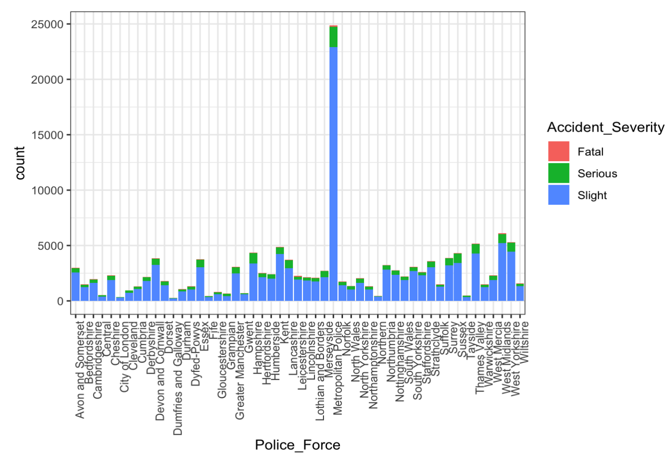

Fig. 11: Accident example in rural areas .

Table 2: Final variable set

Variable Name
Pedestrian Crossing Physical Facilities
Local Authority District
Police Force
Did Police Officer Attend Scene of Accident
Speed limit
Urban or Rural Area
Junction Detail
Second Road Class
Junction Control
Light Conditions
Day Period Casualties Class

4.2. Causalities by Association Rules

Considering the initial objective of determining the leading causes of severe accidents, association rules are employed to find causality relations between the precedents that may lead to severe fatal accidents.

As a brief formulation of the rules measures, quality can be summarized as:

- **Support:** Importance of a set (Number of Transactions containing the set S).

- **Confidence:** Measures strong of the rule (Percentage of Transactions that having an set of items also have other item) $conf(A->B) = \frac{sup(AB)}{sup(A)}$.

- **Improvement:** Minimal difference between confidence rule and immediate simplification's

- **Lift:** Measures the ration between confidence and support of the item-set appearing on the consequent $lift(A->B) = \frac{sup(A \cup B)}{/} sup(A) * sup(B)$ where < 0 Negative Correlated, 0 Independent and > 1 Positive Correlated) **NOTE:** Only measure co-occurrence (not implication!!)

- **Conviction:** Sensitive to rule direction. $Convi(A->B) = 1 - \frac{sup(A)}{1-conf(A->B)} = 1$ Independent, $>> 1$ Consequent depends on the antecedent in a strong way.

With this in mind, the rules are generated based on the transactions. According to Fig. 12 there is clear evidence that most of the accident are slight and the number of casualties involved is small, and occur mostly in locations where no physical crosswalks are present nearby, meaning that some percentage of the accident is due to absence of adequate traversal facilities for pedestrians in the surroundings,

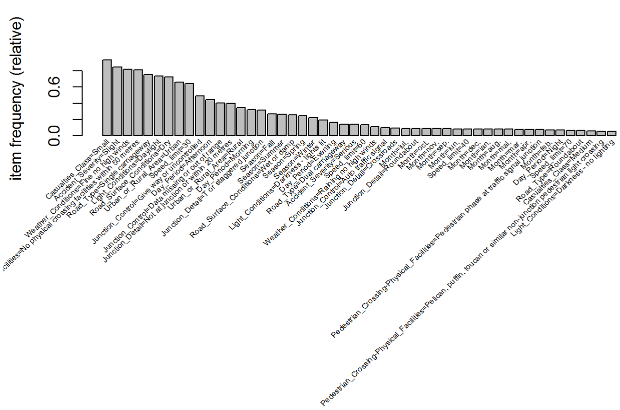

Fig. 12: Item-set occurrence histogram (ordered).

Filtering the cases to those that only contain fatalities (Fig. 13), the no crossing fatalities appear at a high rank, corroborating the previous analysis.

When rearranging the rules that lead to fatal occurrences, the lack of crossing facilities nearby is a significant factor that leads to deadly accidents, represented on rules of Fig. 14 and the relation Graph on Fig. 15.

and when a fatal accident occurs, there always the presence of the police. The majority of the fatal accidents occur with no significant weather disturbances, meaning that the

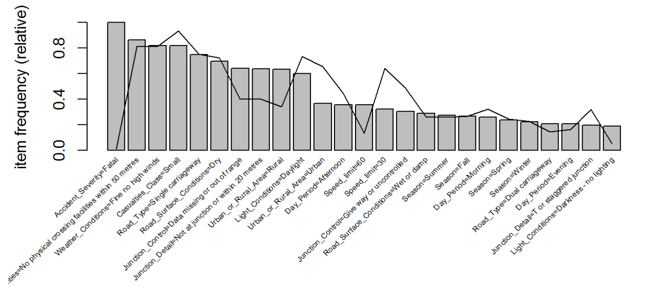

Fig. 13: Item-set occurrence histogram with fatal cases (ordered).

```
##    lhs                                                                                  rhs
support confidence lift
## [1] {}                                                                               => {Accident_Se
verity=Fatal} 1.0000000         1    1
## [2] {Pedestrian_Crossing-Physical_Facilities=No physical crossing facilities within 50 metres} => {Accident_Se
verity=Fatal} 0.8620050         1    1
```

```
##    lhs                                                                                  rhs
support confidence lift
## [1] {}                                                                               => {Casualties_
Class=Large} 1.0000000         1    1
## [2] {Pedestrian_Crossing-Physical_Facilities=No physical crossing facilities within 50 metres} => {Casualties_
Class=Large} 0.8605183         1    1
## [3] {Weather_Conditions=Fine no high winds}                                          => {Casualties_
Class=Large} 0.8170732         1    1
```

```
##     items                                                      support
## [1] {Accident_Severity=Fatal,
##      Did_Police_Officer_Attend_Scene_of_Accident=Yes}       0.9801980
## [2] {Accident_Severity=Fatal,
##      Casualties_Class=Small}                                 0.8193069
## [3] {Did_Police_Officer_Attend_Scene_of_Accident=Yes,
##      Casualties_Class=Small}                                 0.8001238
```

Fig. 14: Rules with fatal as consequent (snapshot).

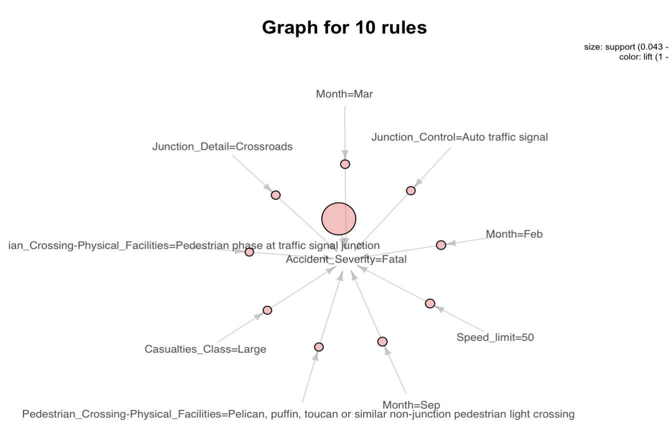

Fig. 15: Graph Rules with fatal as consequent.

weather is not the main responsible for fatal victims' occurrence in a large percentage.

Many other non-relevant rules are also found, some of them are obvious such as evening and season fall, lead to darkness conditions, Fig 16. However, these are not in the scope of the defined objective to determine the main factors that lead to fatal accidents.

```
## [2]  {Speed_limit=60,
##       Day_Period=Evening,
##       Season=Fall}                    => {Light_Conditions=Darkness - no lighting} 0.0041
```

Fig. 16: Fall Season and Evening day period lead to dark conditions

5. PREDICTIVE MODELING AND CLUSTERING

Considering the main objective of predicting the number of casualties given a set of features, a detailed analysis is performed to correctly handle missing values and rearrange the number of levels or categories into broader groups to be handle by models. Several models with different hyperparameters were trained and evaluated using the corresponding training set and validation set, with final predictions made on a reserved test set.

Considering that the number of longitude and latitude factors are very discriminating, only two decimal points were kept, enabling a more abstract representation of the location. Low feature importance features with many missing values are removed to avoid numerical errors. Variables with too many categories were aggregated to reduce the granularity since some models have difficulties handling such various classes in a single feature.

The final set of features is summarized in Table 3.

140k entries form the final dataset, and a test set containing 30% is kept apart to assess final model performance. Hyperparameter tuning and model evaluation is made on the training set using a holdout technique with 70/30 for training and validation. The optimal model is finally trained using all training data.

Several models with several hyper-parameter are evaluated, such as Trees, SVM, Naive Bayes (NB), Random Forest (RF), using the Mean Square Error (MSE). Fig. 17 present the error distribution of the Tree model with several parameters.

Using a lower number of Standard Errors (SE) to use in the post-pruning of the tree lad to better models.

Concerning ensemble model, the RF Fig. 18 present the error distribution lower than Tree model in general with several parameters.

Several kernels on SVM are evaluated, with error distribution among the different parameters and kernels presented in Fig. 19.

By comparing several models (Fig. 20) the RF present the lower error.

Table 3: Brief description of variables in use

Variable Name
Longitude
Latitude
Police Force
Accident Severity
Number of Vehicles
Number of Casualties
Police Force
Number of Vehicles
Number of Casualties (target)
Local Authority District
First Road Class
Road Type
Speed limit
Junction Detail
Junction Control
Pedestrian Crossing-Human Control
Pedestrian Crossing-Physical Facilities
Light Conditions
Weather Conditions
Road Surface Conditions
Special Conditions at Site
Carriageway Hazards
Urban or Rural Area
Did Police Officer Attend Scene of Accident
Season
Month
Hour
Day Period
Casualties Class
Number Vehicles Class

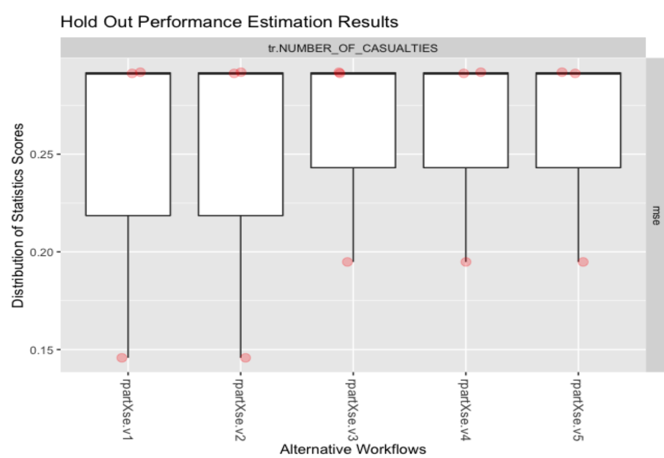

Fig. 17: Box-plot of MSE of Trees models with different hyper-parameters.

Using several models with default parameters(Fig. 20) the RF also present the lower error, confirming the performance of ensemble models.

The RF models account for the best performance among all models, corresponding to expectations that ensembles commonly present a higher accuracy since the enable to explain more variance without overfitting. NB it's not a proper regression model. For a final model training, we reuse the best model RF with corresponding optimal parameters and

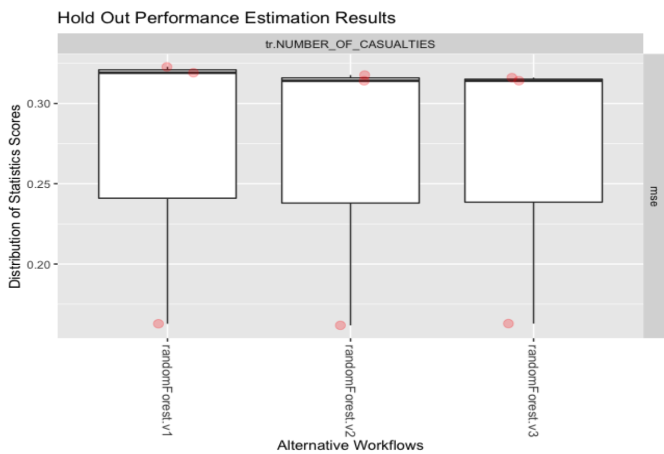

Fig. 18: Box-plot of MSE of RF models with different hyper-parameters.

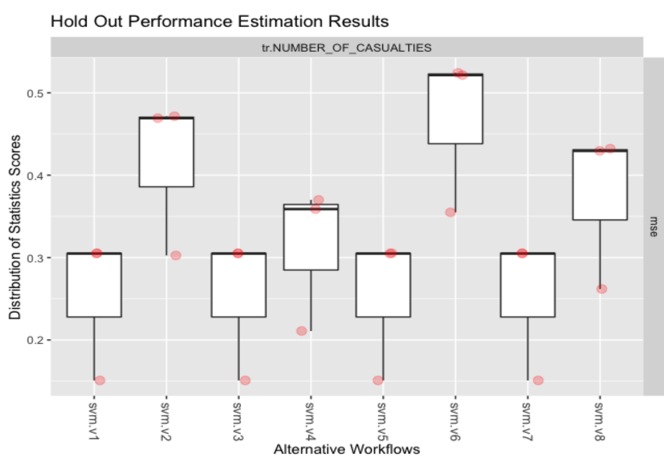

Fig. 19: Box-plot of MSE of SVM models with different hyper-parameters.

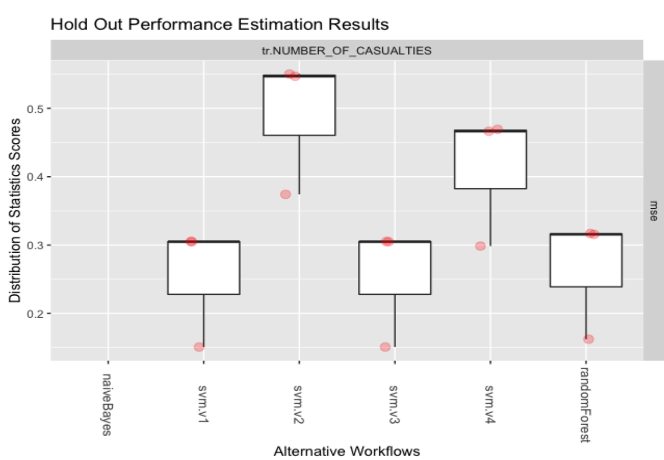

Fig. 20: Box-plot of MSE of several models with different hyper-parameters.

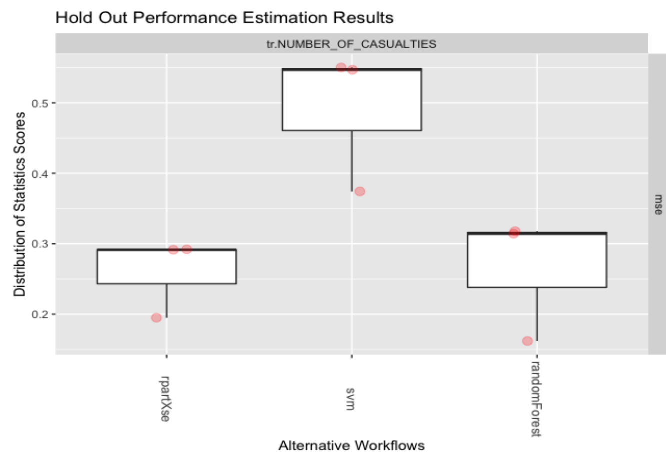

Fig. 21: Box-plot of MSE of several models with default parameters.

retrain the model using the training and validation set simultaneously and evaluate in the unseen test set. The results show a Mean Absolute Error (MAE) of 0.3 on the test set.

5.1. Agglomerate Analysis (Clustering)

Accidents commonly are associated with patterns that typically occur. Considering the target objective is to predict the number of fatal casualties, Partition Around Medoids (PAM) algorithm enables the search k representative objects (the medoids) among the cases in the given data set.

As with k-means each observation is allocated to the nearest medoid. PAM is more robust to the presence of outliers because it uses original objects as centroids instead of averages that may be subject to the effects of outliers.

Moreover PAM uses a more robust measure of the clustering quality, an absolute error instead of the squared error usually used in k-means, $H(C,k) = \sum_{i=1}^{k} \sum_{X \in c_i} |x - \bar{x}C_i|$ where $\bar{x}C_i$ is the centroid of cluster C_i

To overcome limited computational resources, a 50% stratified sample from the 140k entries is gathered, and columns with a large number of categorical factors were summarized.

Analyzing the elbow graph in Fig. 22, the optimal number of cluster is 5.

using the t-sne for visualization of high dimensional spaces [15], a clear cluster in red (Fig. 23 is evident.

Some clusters are well defined formed, namely considering pedestrian and urban areas and some rural and crossing a bit mixed, but some clusters are a bit intermixed, meaning that the data does not follow a well-defined space, meaning that some fatal accidents are subject to many various factors that don't manifest in conjunction too often.

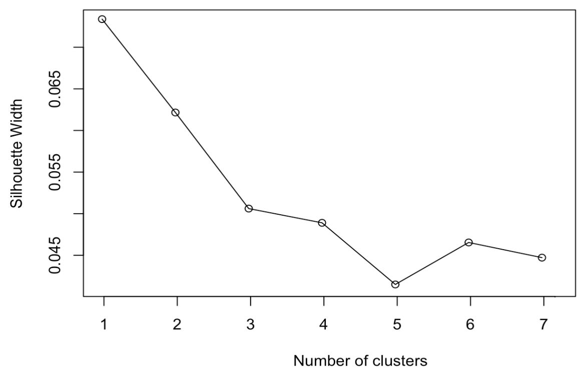

Fig. 22: Elbow graph for cluster selection.

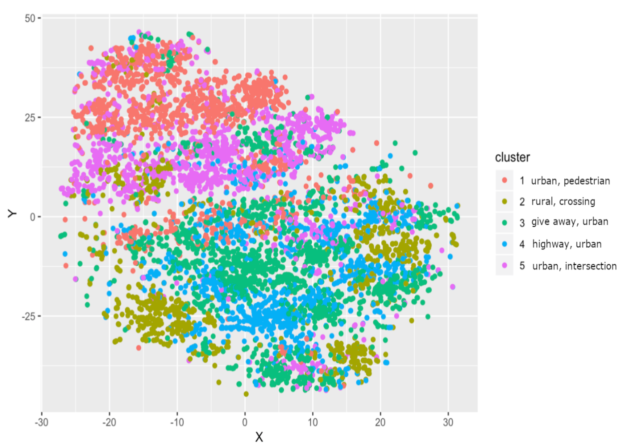

Fig. 23: t-sne cluster visualization.

6. EXPERIMENTAL EVALUATION

In this study, a set of experiments were conducted to find fundamental causes that lead to a fatal accident on traffic roads. The exploratory analysis clearly shows that the crossing and city areas account for many fatal accidents. These locations are mainly characterized by the existence of a large number of cars and pedestrians moving in simultaneously, and that sometimes share the same road space, namely in crossing locations and junctions, requiring a fair amount of attention by both intervenients.

Some cases can be problematic when a particular intervenient doesn't respect the general rules of transit.

The city's large ecosystem imposes difficulties to planing authorities to account for every situation, such as pedestrian that repetitively cross the road out of the predefined crossing areas, leading to an increase in the probability of a fatal accident to occur. Even with significant cities having a large percentage of police resources, this problem remains a problem to solve.

Besides, some heavily dense traffic roads that connect major cities account for many chain accidents, sometimes involving dozens of vehicles. Our exploratory analysis identified three central locations, and the detailed study reveals the main reasons for those to occur. Mostly are highways that interconnect major cities, used by trucks and heavy vehicles to deliver goods. The tricky sections correspond to accentuate high-speed curves near the intersection for vehicle access. The sudden change of speed that some cars may perform when entering or exiting these sections suggests that they are responsible for the occurrence of major vehicle chain crashes.

The association rules show evidence that junctions and crossing facilities or lack of them are responsible for the fatal accidents, corroborating the visual exploration analysis by establishing clear antecedents related to deadly crash occurrences.

For predicting fatal accidents, a feature importance study was performed using Person correlation, enabling to remove highly correlated features that don't contribute to the model generalization capabilities. For the task, several models were evaluated and tuned to determine the top performance predictive model to determine the number of fatal occurrences given a set of features. The several experiments showed that RF exhibited the best results, corroborating the theories that ensembles, in most cases, can improve model precision. An assessment of model features indicates that a fatal crash is more likely to occur on-road sections with crossing facilities and junctions during the peak period. Finally, as expected, the crash potential increases with exposure.

7. CONCLUSIONS

This paper suggests the rational methods by which fatal crash precursors are determined from experimental results and evaluate the deadly crash prediction model's performance for different assumptions. The findings from this study are summarized as follows:

1. The main criterion for the fatal crash precursors is the lack of surrounding support infrastructures in critical locations that, when crashes occurred, significantly lead to fatal victims, contrary to more intensive roads such as highways in normal traffic conditions.

2. The difference between the speed and the abrupt transition of speed within the road section, that is, the formation and dissipation of a traffic queue, has positive effects on the crash occurrence, particularly in high traffic highways.

3. Rural locations present a relevant percentage of fatal accidents, meaning that the lack of convenient road infrastructures combined with the distributed as dispersion nature of the rural area and their population imposes difficulties to the correct assessment o potential risk locations to authorities.

4. The categorization of crash precursors is determined based on the overall fit of the crash prediction model, the sta-

tistical significance of features, and the consistency of the coefficients with the order of levels of fatal crash precursors. In the analysis, three elements appear to be the most suitable for explaining the differential impacts of fatal precursors in road accidents, mostly the pedestrian crossing facilities, the city location, and nearby road intersections.

Having been calibrated for historical crash data, the proposed the predictive model can be employed to predict fatal crash potentials ahead of time. For the next step, this model needs to be applied to actual traffic conditions. Besides, investigators need to examine how the model's crash potential can help reduce the crash potential and improve freeway traffic safety.

Because the fatal crash potential can be predicted using the gathered data and the predictive model, automated real-time countermeasures, such as variable speed limits, can also be implemented to reduce crash potential. This tool certainly saves manual human intervention and effectively controls the traffic flow to prevent crashes. To systematically implement real-time countermeasures, it is necessary to identify and classify potential risky areas into different levels of risk tolerance. For example, a variable speed limit would be in operation only when the estimated crash potential exceeds the specified threshold value of risk tolerance.

Multi-relational data mining can also be useful to handle a large volume of data, using multiple external sources of data to help the inference by exploring all relevant information from all sources.

Furthermore, the safety benefit of automated or human traffic control in some locations can reduce the number of fatal victims by imposing an active citizen awareness to dangerous situations, namely pedestrians when crossing in inadequate areas in the city. For one thing, the simulation can be performed before and after manual and automated traffic control. This before-and-after study would evaluate whether this human and automated traffic control can effectively reduce the overall crash potential that leads to fatal victims.

8. REFERENCES

[1] UN GA, "Transforming our world: the 2030 agenda for sustainable development," *Division for Sustainable Development Goals: New York, NY, USA*, 2015.

[2] Luke B Connelly and Richard Supangan, "The economic costs of road traffic crashes: Australia, states and territories," *Accident Analysis & Prevention*, vol. 38, no. 6, pp. 1087–1093, 2006.

[3] Vinand M Nantulya and Michael R Reich, "Equity dimensions of road traffic injuries in low-and middle-income countries," *Injury control and safety promotion*, vol. 10, no. 1-2, pp. 13–20, 2003.

[4] Gerald McGwin Jr and David B Brown, "Characteristics of traffic crashes among young, middle-aged, and older drivers," *Accident Analysis & Prevention*, vol. 31, no. 3, pp. 181–198, 1999.

[5] World Health Organization, *Global status report on road safety 2015*, World Health Organization, 2015.

[6] Letty Aarts and Ingrid Van Schagen, "Driving speed and the risk of road crashes: A review," *Accident Analysis & Prevention*, vol. 38, no. 2, pp. 215–224, 2006.

[7] Robert B Noland and Lyoong Oh, "The effect of infrastructure and demographic change on traffic-related fatalities and crashes: a case study of illinois county-level data," *Accident Analysis & Prevention*, vol. 36, no. 4, pp. 525–532, 2004.

[8] Jeffery Archer, *Indicators for traffic safety assessment and prediction and their application in micro-simulation modelling: A study of urban and suburban intersections*, Ph.D. thesis, KTH, 2005.

[9] Jutaek Oh, Simon P Washington, and Doohee Nam, "Accident prediction model for railway-highway interfaces," *Accident Analysis & Prevention*, vol. 38, no. 2, pp. 346–356, 2006.

[10] Lei Lin, Qian Wang, and Adel W Sadek, "A novel variable selection method based on frequent pattern tree for real-time traffic accident risk prediction," *Transportation Research Part C: Emerging Technologies*, vol. 55, pp. 444–459, 2015.

[11] Moinul Hossain and Yasunori Muromachi, "A bayesian network based framework for real-time crash prediction on the basic freeway segments of urban expressways," *Accident Analysis & Prevention*, vol. 45, pp. 373–381, 2012.

[12] Chris Lee, Bruce Hellinga, and Frank Saccomanno, "Real-time crash prediction model for application to crash prevention in freeway traffic," *Transportation Research Record*, vol. 1840, no. 1, pp. 67–77, 2003.

[13] Lei Lin, Qian Wang, and Adel W Sadek, "Data mining and complex network algorithms for traffic accident analysis," *Transportation Research Record*, vol. 2460, no. 1, pp. 128–136, 2014.

[14] Rongjie Yu and Mohamed Abdel-Aty, "Utilizing support vector machine in real-time crash risk evaluation," *Accident Analysis & Prevention*, vol. 51, pp. 252–259, 2013.

[15] Laurens van der Maaten and Geoffrey Hinton, "Visualizing data using t-sne," *Journal of machine learning research*, vol. 9, no. Nov, pp. 2579–2605, 2008.