Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 62 (2022) 798–805

Transportation Research Procedia

www.elsevier.com/locate/procedia

24th EURO Working Group on Transportation Meeting, EWGT 2021, 8-10 September 2021, Aveiro, Portugal

Road Accident Analysis with Data Mining Approach: evidence from Rome

Antonio Comi*, Antonio Polimeni, Chiara Balsamo

Department of Enterprise Engineering, University of Rome Tor Vergata, 00118 Rome, Italy

Abstract

Nowadays, road accident is one of the main causes of mortality worldwide. Then, measures are required to reduce or mitigate the accident impacts. The identification of the most effective measures requires an effective analysis of accidents able to identify and classify the causes that can trigger an accident. This study uses data mining as well as clustering approaches to analyze accident data of the 15 districts of Rome Municipality, collected from 2016 to 2019. The aim is to find out which data mining techniques are more suitable to analyze road accidents, to identify the most significant causes and the most recurrent patterns of road accidents by means of a descriptive analysis. Besides, a model to foresee road accidents is proposed. Results show that such analyses can be a powerful tool to plan suitable measures to reduce accidents as well as to forecast in advance the areas to be pointed out.

Peer-review under responsibility of the scientific committee of the 24th Euro Working Group on Transportation Meeting (EWGT 2021)

Keywords: road accidents; road safety; accident analysis; cluster analysis; data mining.

1. Introduction

Road accident analysis aims to investigate the main factors that characterize an accident to understand patterns and/or behaviors and, consequently, to identify the appropriate countermeasures to adopt to avoid the accident. Generally, a road accident can be seen as a random event (Prato et al., 2010; Martín et al., 2014; Gupta et al., 2017; Comi et al., 2019) and statistical models have been proposed in order to analyze them, and to foresee the accident and its consequences (Dutta and Fontaine, 2019; Afghari et al., 2020; Kabir et al., 2021). Zeng et al. (2017) proposed an approach based on multinomial logit in order to explore the influence of different attributes (e.g., traffic and road

* Corresponding author. Tel.: +39 06 7259 7061; fax: +39 06 7259 7061.
E-mail address: comi@ing.uniroma2.it

2352-1465 © 2022 The Authors. Published by ELSEVIER B.V.

Peer-review under responsibility of the scientific committee of the 24th Euro Working Group on Transportation Meeting (EWGT 2021)
10.1016/j.trpro.2022.02.099

characteristics) on accident severity. Similarly, Chen et al. (2018) modelled accident severity through attributes related to traffic, weather and road surface conditions, taking their spatial and temporal variability into consideration. Other authors (Elvik, 2013; De Pauw et al., 2018; Park et al., 2021) proposed models that relate the speed of vehicles involved in the accidents by considering also the road characteristics. Another aspect treated in literature are the accidents involving the vulnerable road categories (e.g., pedestrians, bicyclists). With this respect and without the intention of being exhaustive, some works present in the literature are reported below. Focusing on pedestrians, Marcianò et al. (2011) and Ghasedi et al. (2021) developed models able to estimate the probability of a pedestrian to be involved in a road accident and the related probability of to be injured. The accidents suffered by bicyclists are considered in Liu et al. (2020) and Helak (2017), where an analysis on accidents between cars and bicycles is reported. Other works focusing on injury severity in road accidents are Guo et al. (2017), Ling et al. (2017), Li et al. (2021). In addition to the modeling approach, procedures based on data mining features (Samerei et al., 2021; Roland et al., 2021) can be performed by using descriptive (Kumar and Toshniwal, 2015; Li et al., 2017) or predictive techniques (Yassin and Pooja, 2020; Roland et al., 2021). The first ones allow to classify the accidents in relation to a set of variables (e.g., accident type, road characteristics, weather). The latter ones forecast future accidents starting from a set of relevant attributes.

This study aims to analyze road accidents, and to identify and classify them according to some criteria useful for guiding the countermeasure implementation. This aims to develop an analysis and/or a forecasting tool that can be used both at the planning level to define suitable measures to reduce accidents, and at the operational level, to be ready to intervene in those areas where the accident could occur. To achieve the purpose of the research, this study investigates road accidents in the urban area of Rome (Italy), by implementing data mining for descriptive analysis (e.g., *k*-means and Kohonen network; Li et al, 2018; Sanjurjo-de-No et al., 2021) and for forecasting (e.g., artificial neural network, Singh et al., 2020) on data collected from 2016 to 2019.

While there are a handful of studies on the theoretical approaches to take for assessing road accidents at the disaggregate/aggregate level focusing on specific study area (Russo and Comi, 2013 and 2017), the research on addressing the problem by combining disaggregate data coming from different city areas is limited. The few studies in this area are mostly focused on descriptive analysis of the traffic conditions in the study areas. However, to the best of the authors' knowledge, few studies in the literature address to merge accident data from different districts of a city. Therefore, this study contributes to the literature in improving the accuracy findings both in terms of descriptive and predictive analysis, using data from 15 districts of Rome. There are several aspects of transportation operation and planning that have benefited from analyzing and predicting the road accidents in order to improve city sustainability and livability towards to zero-accident goal.

The paper is organized as follows. Section 2 presents the data and methodology implemented for analyzing accident data, while Section 3 presents the results obtained in the 15 districts of Rome. Finally, conclusions and the road ahead are given in Section 4.

2. Data and Methodology

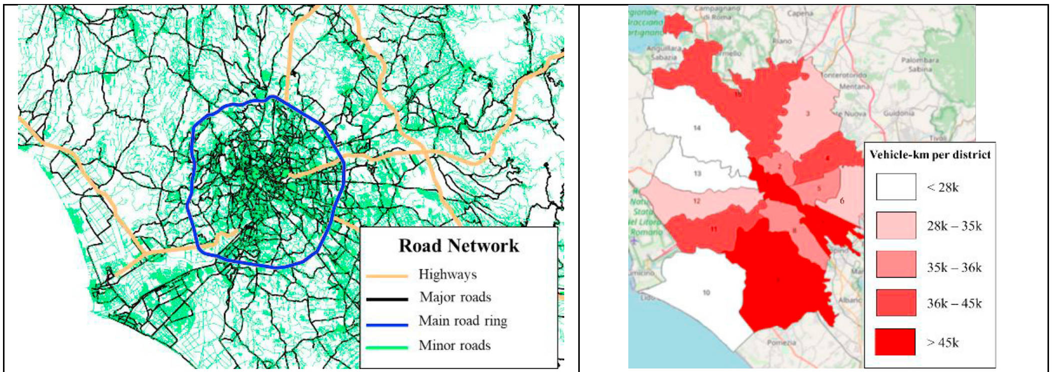
2.1. The study area and the accident dataset

The considered dataset (Open Data - Roma Capitale, 2020) consists of 97,297 road accidents occurred from 2016 to 2019 in Rome (Italy), divided into 15 districts, which has an area of 1285 km^2 and a resident population of nearly three million inhabitants. The recorded accidents are those that required police intervention. The information available for each accident are the type (e.g., rear-end collision), the number of involved people, the time and date, the location, the severity (unharmed, confidential prognosis, died, other) and the type and the number of involved vehicles. Other information relies with the road characteristics, the weather, the lighting and the traffic conditions.

The fleet composition can be summarized in five major classes: vulnerable users (pedestrians and bicyclists), four-wheeled vehicles (e.g., cars), two-wheeled vehicles, public service vehicles and heavy vehicles (e.g., freight transport vehicles). The road network (Fig. 1 a) is composed by nearly 12,000 km of roads, of which nearly 2,200 km compose the major roads, in which every day over two million of vehicles (1.6 million of these vehicles are cars) travel (Fig. 1b).

The average number of vehicles involved in accidents is 1.16, a number that rises to 1.74 excluding the types of accidents in which there is necessarily only one vehicle involved (e.g., collision with a fixed obstacle). Table 1 presents the fleet composition of accidents. To facilitate the representation of the aforementioned table, only accidents occurring between a maximum of two vehicles have been considered: this approximation is possible because accidents involving three or more vehicles represent less than 5% of the total accidents. The majority consists of accidents among four-wheel vehicles (i.e., 52%), while the relevance of accidents involving two-wheel vehicles (25%) and pedestrian and cyclists (10%) is significant. In fact, in Rome about the 7% of trips are by two-wheel vehicles (Rapporto Mobilità Roma, 2019).

The number of accidents in the four-year period is higher in the inner district, while the lowest values are in the North-Western districts. Besides, the yearly trend remained constant (Fig. 2a, Table 2): the months in which the number of accidents is higher are March and November, while August is the month in which this number is lower. These results are due to the different levels of traffic throughout the year: in the summer months there are fewer vehicles in circulation. The main types of accidents (Table 2) occurred over the years are the collision with an obstacle (from 23% to 27%), the side collision (from 23% to 24%) and the head-on collision (from 25% to 27%).



a) Distribution of road type b) number of vehicle-km (vkm) per district

Fig. 1. Road network data.

Table 1. Fleet composition.

	Pedestrians & Cyclists	Two-wheeled vehicles	Four-wheeled vehicles	Public services vehicles	Heavy vehicles	TOT
Single vehicle accidents	91	2,918	11,703	319	561	**15,592**
Pedestrians & Cyclists	23	1,790	5,825	163	394	**8,195**
Two-wheeled vehicles		895	17,568	258	1,182	**19,903**
Four-wheeled vehicles			41,745	1,904	7,339	**50,988**
Public services vehicles				23	142	**165**
Heavy vehicles					347	**347**
TOTAL	**114**	**5,603**	**76,841**	**2,667**	**9,965**	**95,190**

Table 2. Types of road accidents.

Type of accident	2016 [%]	2017 [%]	2018 [%]	2019 [%]
Collision with obstacle	26.19	23.06	27.42	25.93
Rear-end collision	15.37	15.94	14.99	15.33
Side collision	22.54	23.65	22.64	23.62
Head on collision	25.72	26.86	24.85	24.99
Rollover	1.02	0.97	0.89	0.92
Pedestrian hit	6.38	6.66	6.46	6.65
Sudden braking	1.27	1.18	1.3	1.07
Vehicle fall	0.67	0.76	0.64	0.58
Run-off roadway	0.84	0.92	0.81	0.91

By comparing the data related to Rome with the national data (ISTAT, 2020), it emerges that the trend is very similar. Considering two descriptive indices (Table 3), as the rate between the number of injured and the accidents

(IL) and the rate between the number of deaths and the population (TM), it can be seen that these values for Rome are a bit lower than the national one (the IL decrease over the years, ranging from 128.1 to 129.3, while the TM range from 4.4 to 5.1 with a peak in 2018). Furthermore, the traffic levels of the city are not homogeneous between the different districts as plotted in Fig. 1b. It causes that the majority of accidents are recorded in the historic center (I and II districts) and in mainly residential areas (V, VI and VII districts), which are the most populated zones in the city. However, as shown in Table 4, the district X, although less populated than residential areas, presents the higher value of accidents and deaths for vkms. In fact, pointing out accidentality rate per people living/visiting the district (calculated as the number of accidents over the number of workers and residents in each district), the most dangerous districts are the VI and the X (Fig. 2b) with an accident rate ranging from 11% to 13%.

Table 3. Road accident indices: Rome *vs* Italy.

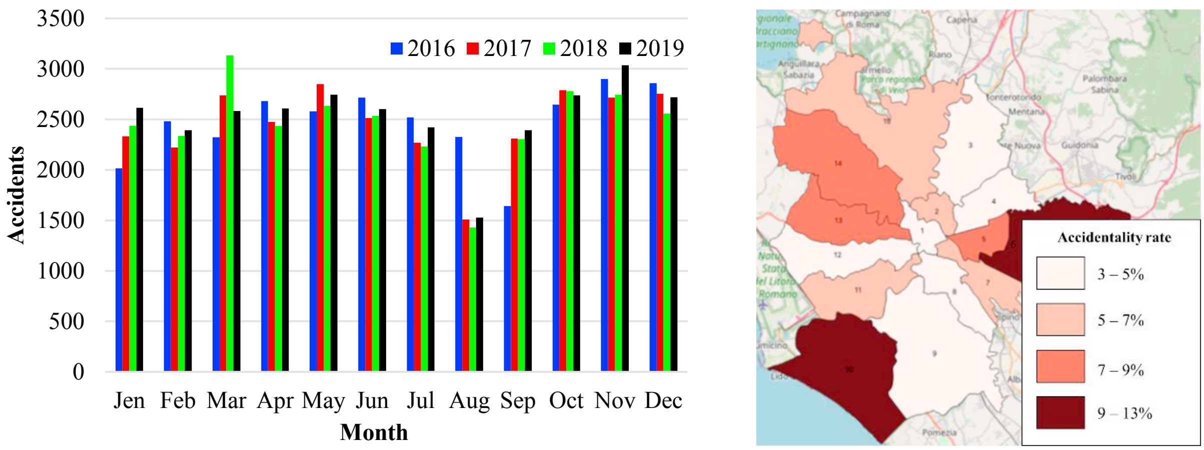
Index		**2016**	**2017**	**2018**	**2019**
IL	Rome	129.3	128.9	128.7	128.1
	Italy	141.7	141.1	140.8	140.2
TM(*)	Rome	4.4	4.9	5.1	4.6
	Italy	5.2	5.1	5.5	5.3

(*) *for 100,000 inhabitants*

Table 4. Number of accidents, injuries and deaths per km of road and veh-km in each district.

		Values per km of roads			***Values per vkm***		
District	**Population**	**Accident**	**Injuries**	**Deaths**	**Accident**	**Injuries**	**Deaths**
I	185,435	18.6	9.5	0.074	0.0282	0.0145	0.00011
II	168,354	15.0	8.0	0.063	0.0223	0.0119	0.00009
III	205,019	8.2	3.9	0.024	0.0165	0.0079	0.00005
IV	176,981	8.4	4.1	0.046	0.0143	0.0070	0.00008
V	247,302	15.4	7.6	0.063	0.0226	0.0111	0.00009
VI	257,534	7.5	4.2	0.029	0.0257	0.0144	0.00010
VII	308,076	10.9	5.7	0.057	0.0179	0.0094	0.00009
VIII	131,180	8.3	3.9	0.047	0.0133	0.0062	0.00008
IX	182,026	5.0	2.1	0.016	0.0115	0.0048	0.00004
X	231,723	5.1	2.8	0.027	0.0262	0.0142	0.00014
XI	155,586	6.7	3.5	0.023	0.0090	0.0047	0.00003
XII	141,104	8.9	4.4	0.045	0.0135	0.0067	0.00007
XIII	134,147	10.8	4.6	0.062	0.0182	0.0078	0.00011
XIV	191,776	6.8	3.1	0.019	0.0189	0.0085	0.00005
XV	159,984	4.6	2.2	0.019	0.0134	0.0065	0.00006
MEAN	**2,877,215 ***	**8.5**	**4.2**	**0.037**	**0.0176**	**0.0088**	**0.00008**

* *Total*



a) Number of accidents b) Accidentality rate per district

Fig. 2. Accidents.

2.2. Methodology analysis

This study employs several models to explore their performance in describing and predicting road accidents inside Rome. Below, each of these models is described briefly. Interested readers may refer to Jonshon and Kuhn (2016) and Bonaccorso (2017) for more details on these algorithms. In particular, R-project software were used for mining data, while spreadsheet software program and geographic information system (GIS) were used for data cleaning and preprocessing.

Descriptive analysis is used to uncover groups or clusters of data objects based on similarities among these objects occurring as a result of interactions among independent variables. Predictive analysis is used to forecast future events or behaviors based on mapping a set of input values to an output value.

In particular, the *descriptive analysis* was implemented through *k*-means clustering and Kohonen network, which is useful to uncover clusters (i.e., groups) of data objects that are more similar to each other. K-means clustering is the most popular non-hierarchical clustering algorithm: “K” refers to the number of clusters used to reduce the dimensionality of the problem, while “means” refers to the cluster being represented by the mean of observations on selected variables. This algorithm is convenient because of its computing speed: it is, in fact, a heuristic algorithm and converges quickly to a local optimum. Because of the extremely high number of variables in the dataset, through the Principal Components Analysis (PCA) it was possible to reduce the size of the dataset by selecting the most significant variables. Kohonen networks are based on the idea of self-organized learning but, even though they are a particular type of neural networks, they are suitable for clustering, because the algorithm does not predict values of a target variable. The strength of this algorithm lies in the fact that it is not influenced by local optimum, returning in output a more robust clustering, exploring more precisely the search space.

Subsequently, *predictive analysis* was developed through decision trees and neural networks. They were used to forecast future events by mapping a set of input values to an output value. Decision trees are a predictive modelling approach in which information is divided into classes and is used to generate rules with the aim to classify previous and new cases in a punctual way. In the following section, the results obtained with two different algorithms of classification trees, respectively C5.0 and CHAID (acronym of Chi-squared Automatic Interaction Detection) are presented: both ones have the advantage of returning an accurate predictive model within a short span of time. They are also quite simple to understand and both ones perform well with large datasets. Neural networks are a powerful computing tool, based and inspired by the human mind elaboration process. This approach allows the creation of a highly precise predictive model, nevertheless, because neural networks are a black box once they are trained, the model returned in output is quite difficult to interpret.

3. Results and discussion

3.1. Descriptive analysis

K-means algorithm requires as input the number of clusters into which divide the dataset, but this parameter is not known *a priori*: through the “elbow rule”, it is possible to find, in a heuristic way, a number that allows good diversification between the various groups. Although the best performing number of clusters should be 10, a significant value of percentage of variance explained is obtained by six clusters.

The number of records contained in the various clusters (Table 5) is, respectively, 12765, 2893, 24565, 23351, 6974, 26749 elements, with a dissimilarity between clusters of 82.7%.

It is interesting to highlight the cluster in which the predominant type of accident is the pedestrian hit (cluster 1): for these accidents, in fact, the number of injuries and deaths is the highest and, moreover, it is the only cluster in which the vehicle concerned is not (at least explicitly) a car, but an *unknown vehicle*. In the other clusters, it is noted that the diversification occurs mainly at a temporal level and that the predominant type of accident is that of the side collision: these accidents occur above all in high traffic conditions, in which the severity of the accidents is lower (as confirmed, in fact, the number of deaths and injuries drastically lower than in the first cluster).

In order to better compare the results, 6 clusters were also considered for the Kohonen network algorithm, whose results are reported in Table 6. A comparison between the two algorithms is given in Table 7.

Table 5. K-means clustering.

CLUSTER	MONTH	TYPE OF ACCIDENT	CHARACTERISTICS OF ROAD SECTION	VEHICLE TYPE	INJURIES	DEATHS	DISTRICT
1	June	Pedestrian hit	Slope section	Unknown	1.015	0.025	7
2	March	Side collision	Slope section	Four-wheeled	0.421	0.003	11
3	June	Rear-end collision	Flat section	Four-wheeled	0.616	0.012	7
4	April	Side collision	Slope section	Four-wheeled	0.465	0.002	3
5	October	Rear-end collision	Straight section	Four-wheeled	0.459	0.002	11
6	October	Side collision	Slope section	Four-wheeled	0.475	0.003	5

Table 6. Kohonen clustering.

CLUSTER	MONTH	TYPE OF ACCIDENT	INJURIES	DEATHS	DISTRICT
1	January	Pedestrian hit	High	High	14-15
2	December	Run-off the roadway	Medium	Low	14-15
3	December	Vehicle fall	High	Low	5-6
4	December	Fall from an overpass	High	High	1-2
5	January	Side collision	Low	Low	1-2
6	January	Rear-end collision	Medium	Low	3-4

Table 7. K-means and Kohonen network comparison.

K-means	Kohonen network
• Clustering is highly influenced by local optima. The data analyzed were classified chiefly on the basis of the most frequent types of accident (e.g., side collision) • The temporal clustering is more uniform throughout the year: accidents are also classified in the central months. • The data are processed rapidly.	• The data have been classified, especially in relation to the type of the accident, more precisely, also showing the less frequent types. • The temporal clustering, however, is mainly concentrated in the months of January and December. • Higher data processing time, nevertheless, this approach is preferable.

3.2. Predictive analysis

Decision trees, association rules and artificial neural networks (ANNs) examine the data and estimate the outcome values of a dependent variable, thus they perform predictive analysis. Considering the severity of the accident as the dependent variable (i.e., target variable), the decision trees generated with the C5.0 and CHAID algorithms are similar: the most significant variables, in both cases, consisted of the type of accident, the allowed speed, the traffic conditions, the road conditions and the number of vehicles involved.

Before training the introduced models, the dataset is divided into two sections: training/validation and test sets. Each model is trained and validated on the training set (i.e., years 2016, 2017 and 2018). Then, the models are compared in their prediction performance on the test set (i.e., 2019).

The confusion matrices in the Table 8 and Table 9 provide insight into the predictive ability of the two algorithms, which is 77% for the C5.0 and 76.1% for the CHAID algorithm, in training phase, and equals to 76.3% and 77.4%, respectively, in testing phase. The rows represent the observations, while the columns the predicted values: the elements in the diagonal indicate the correct predictions for accidents occurred in 2019 (test set) on the basis of trees generated from data about accidents occurred between 2016 and 2018 (training set).

Also, for training the neural network, the variable considered as target was the severity of the accident. The most significant variables, in this case, were the time window in which accident occurred, the number of vehicles involved and the location of the accident (i.e., intersection, roundabout). Results (Table 10) are similar to the ones obtained with the CHAID decision tree algorithm. The estimated precision in training phase is equal to 76.6%, while the prediction accuracy reaches 77.8%.

The significant higher computational costs do not produce any advantage, allowing us to conclude that decision trees are the best method, for this specific case, to predict accidents. In the light of the above results, it can be concluded that most road accidents in Rome end up in light or no injury. The accuracy of the forecasting models presented so far is strongly influenced by the presence of a predominant category of accidents, i.e., non-serious accidents, showing a rigid behavior. Besides, in the test set the accidents in Rome are in contrast whit what happened in the previous years and the predictive analysis fails in capturing patterns not observed in the training set.

Table 8. C5.0 confusion matrix (test set).

Observed injuries \ Predicted injuries	0	1	2-3	4-5	>5
0	7113	101	0	0	0
1	1608	1171	108	1	0
2-3	267	345	87	3	0
4-5	13	42	19	0	0
>5	1	3	2	0	0

Table 9. CHAID confusion matrix (test set).

Observed injuries \ Predicted injuries	0	1	2-3	4-5	>5
0	7188	26	0	0	0
1	1651	1185	52	0	0
2-3	269	380	52	1	0
4-5	14	49	11	0	0
>5	1	4	1	0	0

Table 10. ANN confusion matrix (test set).

Observed injuries \ Predicted injuries	0	1	2-3	4-5	>5
0	7121	93	0	0	0
1	1604	1253	31	0	0
2-3	408	196	98	0	0
4-5	61	9	4	0	0
>5	5	1	0	0	0

4. Conclusions

This paper had the purpose to find out which data mining techniques are more suitable to analyze road accidents. The selection of this techniques is based on the review of the state of the art: many studies have shown that k-means and Kohonen network algorithms are advantageous for descriptive analysis; on the other hand, it has been shown that decision trees and neural network are useful for predictive analysis. Clustering analysis highlighted the safety problems for the vulnerable users (pedestrians, bicyclists and motorcyclists), and pointed out that the high dangerous accidents occur with unknown vehicles, putting a serious issue in sensibilizing vehicle drivers when involved in accidents. Thus, it is intended to apply data mining techniques that will enable to identify characteristics of the roads, as well as the time of the year, which can have a higher impact on crashes. Once main factors related to accidents are known, the road safety authorities could implement preventive and corrective measures in road sections that require them (e.g., update street signs, maintenance/modification of critical road points).

Accident severity seems to be influenced mainly by type of vehicles involved. Obviously, the involvement of vulnerable road users and heavy commercial vehicles causes the outcome results to be more severe than when only private vehicles are engaged. Also relevant for the severity of the accident the road infrastructure conditions are. In particular, the number of rear-end collision on flat section can be reduced assuring an enforcement of vehicle speed control and avoiding distraction in driving.

From the data mining perspective, further research should investigate whether limiting the analysis to fatal accidents would simplify the task of data mining techniques in recognizing accident patterns without the "noise" probably created by considering also severe and light injury accidents. A limitation of the approach (as emerged in the predictive analysis) is that data mining fails in capturing patterns not observed previously or the impacts due to the implementation or relaxation of enforcement. In fact, as emerged by the time series analysis, in the 2019 (test set) the accidents in Rome raise in opposition on what occurred in the previously years. Therefore, it emerges the opportunity to combine different approaches for capturing such a pattern. In fact, up to date, although different studies have considered the periodic/timely feature in accident prediction, very few studies comprehensively evaluate the impact of periodic/timely component on statistical and machine learning prediction models. Therefore, the findings in this paper suggest that a hybrid prediction approach should be investigated, which could be effective for both statistical and machine learning models in accident prediction.

Acknowledgements

The authors want to thank the anonymous reviewers for their valuable suggestions in improving the paper.

References

Afghari, A.P., Haque, M.M., Washington, S., 2020. Applying a joint model of crash count and crash severity to identify road segments with high risk of fatal and serious injury crashes. Accident Analysis & Prevention, 144.

Bonaccorso, G., 2017. Machine learning algorithms. Packt Publishing Ltd.

Chen, F., Chen, S., Ma, X., 2018. Analysis of hourly crash likelihood using unbalanced panel data mixed logit model and real-time driving environmental big data. Journal of Safety Research 65, 153–159.
Comi A., Persia L., Nuzzolo A., Polimeni A., 2019. Exploring Temporal and Spatial Structure of Urban Road Accidents: Some Empirical Evidences from Rome. In: Nathanail E., Karakikes I. (eds) Data Analytics: Paving the Way to Sustainable Urban Mobility. CSUM 2018. Advances in Intelligent Systems and Computing, vol 879. DOI: 10.1007/978-3-030-02305-8_18, Springer, Cham, pp. 147-155.
De Pauw, E., Daniels, S., Franckx, L., Mayeres, I., 2018. Safety effects of dynamic speed limits on motorways. Accident Analysis & Prevention, Road Safety on Five Continents 2016 - Conference in Rio de Janeiro, Brazil. 114, 83–89.
Dutta, N., Fontaine, M.D., 2019. Improving freeway segment crash prediction models by including disaggregate speed data from different sources. Accident Analysis & Prevention, 132.
Elvik, R., 2013. A re-parameterisation of the Power Model of the relationship between the speed of traffic and the number of accidents and accident victims. Accident Analysis & Prevention 50, 854–860.
Ghasedi, M., Sarfjoo, M., Bargegol, I., 2021. Prediction and Analysis of the Severity and Number of Suburban Accidents Using Logit Model, Factor Analysis and Machine Learning: A case study in a developing country. SN Appl. Sci. 3, 13.
Guo, Y., Zhou, J., Wu, Y., Chen, J., 2017. Evaluation of Factors Affecting E-Bike Involved Crash and E-Bike License Plate Use in China Using a Bivariate Probit Model. Journal of Advanced Transportation 2017, e2142659.
Gupta, M., Solanki, V. K. and Smith, V. K., 2017. Analysis of Datamining Technique for Traffic Accident Severity Problem: A Review. Second International Conference on Research in Intelligent and Computing in Engineering, ACSIS 10, 197–199.
Helak, K., Jehle, D., McNabb, D., Battisti, A., Sanford, S., Lark, M.C., 2017. Factors Influencing Injury Severity of Bicyclists Involved in Crashes with Motor Vehicles: Bike Lanes, Alcohol, Lighting, Speed, and Helmet Use. South Med J 110, 441–444.
Kabir, R., Remias, S.M., Lavrenz, S.M., Waddell, J., 2021. Assessing the impact of traffic signal performance on crash frequency for signalized intersections along urban arterials: A random parameter modeling approach, Accident Analysis & Prevention, 149, 105868.
Kumar, S., Toshniwal, D., 2015. A data mining framework to analyze road accident data. Journal of Big Data 2, 26.
ISTAT, 2020. Incidenti stradali in Italia. https://www.istat.it/it/files/2020/07/Incidenti-stradali-in-Italia-Anno-2019-aggiornamento27ottobre2020.pdf
Li, L., Shrestha, S., Hu, G., 2017. Analysis of road traffic fatal accidents using data mining techniques, 2017 IEEE 15th International Conference on Software Engineering Research, Management and Applications (SERA), 363–370.
Li, X., Liu, J., Zhang, Z., Parrish, A., Jones, S., 2021. A spatiotemporal analysis of motorcyclist injury severity: Findings from 20 years of crash data from Pennsylvania. Accident Analysis & Prevention 151, 105952.
Li, Z., Chen, C., Ci, Y., Zhang, G., Wu, Q., Liu, C., Qian, Z. (Sean), 2018. Examining driver injury severity in intersection-related crashes using cluster analysis and hierarchical Bayesian models. Accident Analysis & Prevention 120, 139–151.
Ling, Z., Cherry, C.R., Dhakal, N., 2017. Factors influencing single-bicycle crashes at skewed railroad grade crossings. Journal of Transport & Health, Road Danger Reduction 7, 54–63.
Liu, J., Khattak, A.J., Li, X., Nie, Q., Ling, Z., 2020. Bicyclist injury severity in traffic crashes: A spatial approach for geo-referenced crash data to uncover non-stationary correlates. Journal of Safety Research 73, 25–35.
Marcianò, F.A., Vitetta, A., 2011. Risk analysis in road safety: an individual risk model for drivers and pedestrians to support decision planning processes. Int. J. SAFE 1, 265–282.
Martín, L., Baena, L., Garach, L., López, G. and de Oña, J., 2014. Using data mining techniques to road safety improvement in Spanish roads. Procedia - Social and Behavioral Sciences 160, 607 – 614.
Open Data - Roma Capitale 2020, available at: https://dati.comune.roma.it/ (accessed 12 January 2021).
Park, E.S., Fitzpatrick, K., Das, S., Avelar, R., 2021. Exploration of the relationship among roadway characteristics, operating speed, and crashes for city streets using path analysis. Accident Analysis & Prevention 150, 105896.
Prato, C., Bekhor, S., Gal-Tzur, A., Mahalel, D., Prashker, J., 2010. Exploring the potential of data mining techniques for the analysis of accident patterns, 12th World Conference on Transportation Research. Lisbon, Portugal.
Rapporto Mobilità Roma, 2019 | Roma Servizi per la Mobilità (www.romamobilita.it)
Roland, J., Way, P.D., Firat, C., Doan, T.N., Sartipi, M., 2021. Modeling and predicting vehicle accident occurrence in Chattanooga, Tennessee. Accident Analysis & Prevention, 149.
Russo, F., Comi, A., 2013. Impacts of accidents involving shopping and restocking vehicles on an urban road network. WIT Transactions on the Built Environment, 2013, 134, pp. 517–528.
Russo, F. and Comi, A., 2017. From the analysis of European accident data to safety assessment for planning: the role of good vehicles in urban area. European Transport Research Review 9.
Samerei, S.A., Aghabayk, K., Mohammadi, A., Shiwakoti, N., 2021. Data mining approach to model bus crash severity in Australia. Journal of Safety Research 76, 73–82.
Sanjurjo-de-No, A., Arenas-Ramírez, B., Mira, J., Aparicio-Izquierdo, F., 2021. Driver Liability Assessment in Vehicle Collisions in Spain. International Journal of Environmental Research and Public Health 18, 1475.
Singh, G., Pal, M., Yadav, Y., Singla, T., 2020. Deep neural network-based predictive modeling of road accidents. Neural Comput & Applic 32, 12417–12426.
Zeng, Z., Zhu, W., Ke, R., Ash, J., Wang, Y., Xu, J., Xu, X., 2017. A generalized nonlinear model-based mixed multinomial logit approach for crash data analysis. Accident Analysis & Prevention 99, 51–65.
Yassin, S.S., Pooja, 2020. Road accident prediction and model interpretation using a hybrid K-means and random forest algorithm approach. SN Applied Sciences 2, 1576.