Python task – Danesh kar

Road Accidents Data -2022

MohammadAmin Omidzadehnik

Contents

1 Introduction

The automobile industry is one of the developing industries in the world. Various areas are directly or indirectly related to this industry. In all industries, the main goal is to reduce costs and increase revenues. The automobile industry is no exception to this. One of the important issues in the automobile industry is the discussion of accidents and incidents that occur over a period of time. In the upcoming report, the situation of road accidents in 2022 will be examined and evaluated. Also, the situation of road accidents and its effects on the society and different industries in each country have diverse and different effects and consequences.

The importance of this dataset in Iran comes from the fact that road accidents directly affect various industries. In the field of road and urban planning, the situation of road accidents is significantly related to the condition of roads, traffic signs, route length, etc. In the military and law enforcement field, the situation of accidents leads to more presence of traffic police, dealing with road accidents, the need for more monitoring (such as imperceptible control, smart speed control cameras, etc.). In the field of automobile manufacturing and production, it is also possible to examine the causes of accidents based on the type of car, the way it is made, the quality of construction, and other issues. In this specialized and important field, it is necessary to investigate more. Another area that is directly related to this topic; The insurance industry. Since insurance companies issue insurance policies in various fields. Car-related damage coverage is also one of the activities that can be seen in the insurance industry. Therefore, the data in this dataset can dramatically change the performance of the insurance industry. For example, the amount each car pays for body insurance or third party insurance. Regarding these insurance policies, the amount of damage paid by the insurance, the duration of the insurance, the probability of an accident for the insured person, and the possibility of him using the insurance policy must be carefully examined and determined. Among the other things that should be considered in the discussion of driving accidents, we can mention the importance of this issue for the healthcare sector of the country. The amount of driving accidents, the level of accidents, the amount of injuries, casualties, and deaths, etc., are among the criteria that have a significant impact on the policies adopted in the fields related to health and treatment. Therefore, it can be said that road accidents have a direct relationship with the medical field of a country. Also, among other departments that have an indirect relationship with this case, we can mention the fire brigade, municipality, radio and television, etc. Although each of these bodies is not directly related to the discussion of cars and its accidents, they can have tangible effects in this field.

2 Research literature

There have been many domestic and foreign researches and articles on the topic of accidents and traffic accidents. For example, psychological, social, cultural, educational, and other influential issues in this field have been analyzed as a case study or comprehensively.

From the point of view of personality traits, it is suggested that a person has more accidents based on his characteristics. These characteristics refer to the five dimensions of neuroticism, extroversion, flexibility, responsibility and adaptability. These dimensions are emphasized as

contributing factors to identify the personality characteristics of the accident perpetrators as well as to predict the driving behavior and its related consequences. In a post-event study, the number of 40 drivers who were at fault in the accident who were hospitalized in Al-Zahra Hospital in Isfahan in 2017 were compared with a sample of 40 drivers who were not involved in the accident. The questionnaire used in the research was NEO-PI-R. The comparison of the two groups of accident agents and witnesses in this research leads to the possible conclusion that personality factors can be one of the important factors in the occurrence of accidents (Kanani et al., 1390).

Valuable researches have been done in the field of car production and machine learning approach. Concepts such as advanced driver assistant system and accident prediction system were proposed in the path of autonomous driving development and increasing the level of road safety. In an accident prediction system, comparing the importance of reducing error values, we prioritize a value that increases the level of safety, saves the lives of vehicle occupants, and prevents accidents. Therefore, a group learning model was proposed to predict the accident. In addition, a network scenario is simulated that shows the expected time of a crash in a network (Salary & Hoseyni seno, 2023).

Among other researches, we can mention those articles that cover many scientific aspects. Estimating the probability of occurrence and the consequences of attacks is one of the important points regarding the occupational health of drivers who have the potential of sudden disability attacks due to illness. In other words, determining the probability of an accident following a sudden incapacitation attack while driving, the presence of various diseases, including heart diseases (such as myocardial infarction, arrhythmia, etc.), strokes, epilepsy, diabetes, etc. ... can be accompanied by sudden attacks of disability while driving. Of course, in most cases, the driver finds enough time to steer the car to the side of the road. Although some medical conditions, such as syncope, do not provide this opportunity for the driver and can cause the unfortunate consequences of these attacks in the drivers (Attarchi et al., 2021).

The drivers reported whether they had fallen asleep some time whilst driving, and what the consequences had been. Sleep or drowsiness was a contributing factor in 3.9% of all accidents, as reported by drivers who were at fault for the accident. This factor was strongly over-represented in night-time accidents (18.6%), in running-off-the-road accidents (8.3%), accidents after driving more than 150 km on one trip (8.1%), and personal injury accidents (7.3%). A logistic regression analysis showed that the following additional factors made significant and independent contributions to increasing the odds of sleep involvement in an accident: dry road, high speed limit, driving one's own car, not driving the car daily, high education, and few years of driving experience (Sagberg, 1999).

For the analysis of the information, the different algorithms employed to make predictions about road accidents are listed and compared, as well as their applicability depending on the types of data being analyzed, along with the results obtained and their ease of interpretation and analysis. The best results reported by the authors are obtained when two or more analytic techniques are combined, in such a way that analysis of the obtained results is strengthened. Among the future challenges in road traffic forecasting lies the enhancement of the scope of the proposed models and predictions by the incorporation of heterogeneous data sources, that include geo spatial data,

information from traffic volume, traffic statistics, video, sound, text and sentiment from social media, that many authors concur that can improve the precision and accuracy of the analysis and predictions (Gutierrez-Osorio & Pedraza, 2020).

Official estimates of road accident costs from 1990 or later were compiled from easily accessible sources for twelve countries. Estimates of the gross national product were taken from OECD publications. On the average, the total costs of road accidents, including an economic valuation of lost quality of life, were estimated to about 2.5% of the gross national product. Excluding the valuation of lost quality of life, road accident costs on the average amounted to 1.3% of the gross national product. When valuation of lost quality of life is included, costs ranged from 0.5 to 5.7% of GNP. When valuation of lost quality of life is disregarded, costs ranged from 0.3 to 2.8% of GNP (Elvik, 2000).

Influence of driver sex on road accidents is assessed in this article. Accident records for 3 years and for three different income regions were analyzed. Annual distance traveled, social and economic participation, and effect of public vehicle accidents were considered. Effects of environmental factors and driver age were also included. Driver faults analysis identified possible reasons for accident differences. Analysis of accident severity was used to assess degree of harm. Statistical analysis at the 5% significance level was used to evaluate all differences. The results show that male accident rates are significantly higher. This trend is consistent through all the analyses. Accident differences are significant only in normal driving conditions. Drivers over age 50 had the lowest accident rates. Accident rate differences were caused by lack of attention and impatience among male drivers. Appropriate means of communication should alert concerned populations to these findings (Al-Balbissi, 2003).

Road accidents sustained at work represent between 20% and 40% of work fatalities in most industrialized countries, yet few data on occupational road accident risk factors have been published. A case control study was performed to assess the role of work-related risk factors in the occurrence of occupational road accidents. A preliminary qualitative study was carried out to identify possible occupational factors in occupational road accidents, and to draw up the case control study. Cases were recruited from the Rhône road trauma registry (France), controls from voting lists. A telephone interview was performed. Exposure to road risk was measured as a percentage of work time. One hundred and forty-six cases and 440 matched controls were interviewed. Accident risk was found to increase with exposure. Driving was associated with more difficult working conditions than found in jobs not involving driving. These difficulties, however, were not systematically associated with increased occupational road accident risk. Among factors which still emerge after adjustment for road risk exposure, there are scheduling issues (inflexible schedule organization, lack of consecutive rest-days, lack of flexibility in performing the work), difficulties of communication with superiors, low seniority in the activity, low educational level and physical constraints at work. This study highlights some possible occupational road accident risk factors. Given the chosen case/control methodology, the findings may be considered as advancing our knowledge of the subject, but need confirmation by further studies (Fort et al., 2010).

3 Analysis description

3.1 Data preprocessing

In data science, the most important part of any project is data preprocessing. In this project, I have analyzed data consequently. For preparing data to run ML algorithm, you should utilize many techniques such as data integration, data cleaning, data feature engineering, data validation, etc.

At first, I check all possible side of problem in data. Data redundancy is one of the most important parts which you should check on your data. In addition, you need to know about missing values. It is very important to check records do not values. Fortunately, there is no NULL value in dataset. Checking unique values is another step in data cleaning. In this moment you are going to figure out not only one or two columns make your dataset unique, but also you should consider all parts of structure to have multiple columns to find distinct value. accident_index, accident_year, accident_reference, vehicle_reference, casualty_reference, casualty_class are features make your dataset unique. It means you should consider these columns to reach unique record.

3.2 Descriptive Statistics

When you finalize preparation, you are able to analyze data descriptively. In other means, you can have an overall insight of the dataset. For instance, in road accident, the age average of Severity 1 is more than 40. While in class 2 they are younger than 40. Also, class 3 is the youngest group as well. The vast majority of casualty severity belongs to male in each class. Furthermore, Drivers have the most share in any casualty classes. Pedestrians are more than passengers in class 2. This is the exact opposite in class 3. In all of classes, urban areas have significant difference among other zones.

Although males are more than females in casualty severity, generally they dad more accident than men. The distribution of people who had accident in 2022 has positive skewness. It means whatever age of people gets older, the number of them is decreasing. Also, people with age of 20 have the highest amount among others. Unfortunately, children who cannot celebrate their first birthday are significantly more than other teenager groups.

3.3 Data transformation

Before running any specific machine learning methods, you need to transform data to the best suitable part to use. In this data, we defined main label, categorical age feature, changing some features in string. Encoding features is another part of shaping data for lunching the final model. Finally, we drop unused column and reconsider new data frame to take the best performance for them.

3.4 Machine learning

Due to the dataset classification is one of the most important parts of decision making among all methods of machine learning. To classify intensity of accident to slight and serious. For running this method, we split data to train and test with 20 percent of test size. Then we define weight of classes. Then we trained data with classification in pool and fit based on model. It is able to predict test model based on train part. As you can see AUC contains almost 70 percent score. Now, you

can define baseline of roc AUC. After classification, you can find TP and TN of classification in confusion matrix.

Finally, we value all effects on classification algorithm. As you can see, sex, passenger, and type of vehicle are very significant fort this model.

4 Analysis result and discussion

According to dataset, you can see that the most important reason of accident in 2022 was related to gender, age, and area. Aso, classification helps us define accidents to slight and serious more accurately.

By looking more carefully in analysis, dataset is crystal and clear. But you should consume several columns as unique records. It has no duplicated and missing values. Men have more impact on casualty severity and accidents. Also, urban areas impact on most of accident in 2022. Youngsters are more dangerous than elderly in automobile incidents.

Although most of accident were slight, it needs to be reduced as much as possible. So, in the following part all solutions will be demonstrated.

5 Suggestion

Age of drivers is one of the critical factors to occurs an accident. So, for preventing more accident, I suggest having more rules to give youths a driving license. In addition, you can have more sentences for those drivers after accident. This is the exact same behavior for men than women.

The other factor to commit an accident is area. I believe more laws and constraints should be ordered in urban zones. It makes people to control their driving more than before. For example, setting more speed cameras or police officers in these areas are very helpful.

Furthermore, type of vehicles is another impact on accident in 2022. Defining special way for buses helps traffic. It also causes less accident as well.

About type of people who have accident on 2022, drivers have more accident than other classes. So, they should learn more things for driving. Aside of education, changing manufactures can be another factor of accidents that need to be reconsidered.

All in all, there are lots of features in accident in data set. But if you can control those factors which have more impacts on chance of accident, you will be successful in reducing of danger of accident and number of automobile incidents.

6 References

Al-Balbissi, A. H. (2003). Role of Gender in Road Accidents. *Traffic Injury Prevention*, *4*(1), 64-73. https://doi.org/10.1080/15389580309857

Attarchi, M. s., Kheyr khah , J., & Bakhshayesh eghbali, B. (2021). Assessing the likelihood of a road accident following a sudden disability [Research]. *journal of medical council of islamic republic of iran*, *39*(3), 164-170. http://jmciri.ir/article-1-3138-fa.html

Elvik, R. (2000). How much do road accidents cost the national economy? *Accident Analysis & Prevention* , .851-849 ,(6)32https://doi.org/https://doi.org/10.1016/S0001-4575(00)00015-4

Fort, E., Pourcel, L., Davezies, P., Renaux, C., Chiron, M., & Charbotel, B. (2010). Road accidents, an occupational risk. *Safety Science*, *48*(10), 1412-1420. https://doi.org/https://doi.org/10.1016/j.ssci.2010.06.001

Gutierrez-Osorio, C., & Pedraza, C. (2020). Modern data sources and techniques for analysis and forecast of road accidents: A review. *Journal of Traffic and Transportation Engineering (English Edition)*, *7*(4), 43 .446-2https://doi.org/https://doi.org/10.1016/j.jtte.2020.05.002

Kanani, K., Aghaei, A., & Abedi, M مقایسه ویژگی های شخصیتی عاملان تصادف رانندگی با رانندگان بدون سابقه . (1390) تصادف. یافته های نو در روان شناسی (روان شناسی اجتماعی).

Sagberg, F .(1999) .Road accidents caused by drivers falling asleep. *Accident Analysis & Prevention*, *31*(6), 639-649. https://doi.org/https://doi.org/10.1016/S0001-4575(99)00023-8

Salary, H., & Hoseyni seno, S. A پیش بین تصادف در شبکه های خودرویی، یک رویکرد یادگیری ماشین مبتنی بر . (2023) لبه.