African Conference on Information Systems and Technology

THE 8TH ANNUAL ACIST PROCEEDINGS (2022)

Aug 26th, 11:45 AM - 12:10 PM

Road Traffic Accidents Severity Modelling in Malawi

Chikumbutso Felix Phiri
University of Malawi, phirichiku@gmail.com

Kondwani Munthali
University of Malawi, kmunthali@unima.ac.mw

Yamikani Phiri
University of Malawi, yphiri@unima.ac.mw

Road Traffic Accidents Severity Modelling in Malawi

Chikumbutso Felix Phiri , University of Malawi, phirichiku@gmail.com
Kondwani Munthali , University of Malawi, kmunthali@unima.ac.mw
Yamikani Phiri, University of Malawi, yphiri@unima.ac.mw

ABSTRACT

One of the significant problems Malawi faces today is the rate at which road traffic accidents and deaths are happening on the roads of Malawi. It is very crucial to effectively address such a problem with a limited budget considering that Malawi is a developing country. To supplement the current safety measures, traffic accidents data mining using machine learning models was considered. Being able to predict the severity of an accident as well as determining the weight each attribute contributes to the severity could help authorities make informed decisions. Therefore, this research aimed at modeling the severity of road accidents in Malawi to help reduce traffic accidents or the severity with limited budgetary resources. Using python, three classification algorithms were employed to model the severity of an accident. The algorithms included; Decision trees, Logistic regression and Support Vector Machines. These models were evaluated using accuracy, precision, recall, and F1-score. The logistic regression performed better than the other two and after fitting the model it was discovered that the top three attributes that contributed to fatal accidents were accidents involving a moving vehicle and a pedestrian, accidents that occurred at Dawn or Dust, and accidents involving a moving vehicle and a bicycle

Keywords

Road Traffic accidents, Data mining, Decision trees, Logistic regression, Support Vector Machine.

INTRODUCTION

According to World Health Organization (2020), approximately 1.35 million people die every year on the roads, making road accidents the eighth leading cause of death worldwide. Consequently, these fatalities and injuries lead to significant social and economic losses. In Malawi, about 20 road traffic collisions occur every day, and about 1,000 people are killed in traffic accidents every year WHO, (2020). Across the world, transportation agencies have taken measures and have made significant investments to improve travel safety. In Malawi for example, the Roads Authority introduced rumble strips in areas where accidents frequently occur, like at Linthipe 1. These are used to alert inattentive drivers through noise and vibrations. Again, due to increase in accidents involving motorbikes, the Directorate of Road Traffic Safety and Services (DRTSS) has been conducting civic education and enforcing laws to reduce the number of accidents among other interventions. To supplement the current safety measures, an analysis of road traffic accidents using data mining techniques would help significantly avoid some of these accidents and reduce their severity. The formulation of traffic safety control policy can also benefit from understanding of these results. Considering that Malawi is a developing country there should be limited financial resources for traffic safety, therefore, in order to achieve the greatest impacts of feasible accidents prevention and reduced accidents severity, it is very crucial that decisions are made from a scientific analysis. Aggarwal, (2015) defines data mining as the collection, cleaning, processing, analysis, and gaining of valuable insights from data. Krishnaveni et al (2011) observed that it is a powerful technology with great potential to help companies focus on essential details in their data warehouses. Over the years, this has proven to be true as we have seen data mining being applied in various industries that are able to collect large amounts of data. Throughout the world, road traffic departments have been collecting data about accidents happening in their respective countries. The data has accumulated so much that it cannot be easily analyzed using traditional methods to extract knowledge. This has allowed researchers to apply data mining techniques in an attempt to reduce the number of traffic accidents or the severity of accidents. Shetty et al., (2017), Taamneh et al., (2017) and many other researchers have used different data mining techniques by analyzing data from previous accidents. Information derived from these studies has been used to come up with measures of improving traffic safety.

RELATED WORK

Kumar (2016), confirmed that data mining is a reliable technique for analyzing road accidents to get productive results. Different data mining techniques and algorithms have been since been applied in various research studies. The type of data collected and the main goals influence the choice of methods used.

Several studies in literature focus on predicting the severity of accidents or identifying factors that affect severity (Beshah & Hill, 2010; Krishnaveni et al., 2011; Taamneh et al., 2017). In these studies, different algorithms were tested on the data, and the best-performing one was chosen for prediction. Chong et al. (2005) and Krishnaveni et al. (2011) agreed that accurately predicting severity can lead to a greater understanding of the relationship between the factors, which could provide crucial information for road accident prevention policies. Chong et al. (2005) used accident data from 1995 to 2000 to investigate the performance of support vector machines, neural network, decision tree (DT), and a hybrid DT. Using accuracy as a performance evaluation metric, the hybrid DT and DT performed better. DTs do not require a predefined relationship between the dependent and independent variables and have also proven to be very useful in handling prediction and classification problems in general (Chong et al. 2005). Beshah & Hill, (2010) studied the role of road-related factors on accident severity in Ethiopia. Various classification models were built using a DT, Naive Bayes (NB), and K-nearest neighbor classifiers. Accuracy was also used as an evaluation metric in this research, all three classifiers produced similar accuracy. The receiver operating characteristics (ROC) curve was then used to evaluate the same models and the results were also similar, it was then concluded that all three classifiers performed similarly well in predicting accident severity.

Similarly, Taamneh et al. (2017) used accidents data from Abu Dhabi to explore the performance of different data mining techniques in predicting the severity of accidents from 2008 to 2013. The algorithms used were DT (J48), NB, Rule Induction (PART), and Multilayer Perceptron (MLP). Using accuracy and area under the curve (AUC) to compare the performance of the algorithms, the results indicated that the DT (J48) classifier, PART classifier, and MLP classifier performed similarly well. NB on the other hand demonstrated low accuracy. Krishnaveni et al. (2011) studied traffic accident records of 2008, which had 34,575 cases. This study classified the type of injury severity of various traffic accidents by applying and comparing NB classifier, AdaBoostM1 Meta classifier, PART Rule classifier, J48 DT classifier, and Random Forest (RF) Tree classifier. To reduce the dimensionality of the dataset, a Genetic Algorithm was used for feature selection and the outcome showed that the RF outperformed the other four algorithms based on their accuracy levels. RF must have performed better because it runs well on large databases, it handles thousands of input variables without variable deletion, and also the learning is fast. In addition to this, it has an effective method for estimating missing data while maintaining accuracy. Yuan et al., (2017) obtained motor vehicle crash data from the Iowa Department of Transportation containing crash records from 2006 to 2013. To predict traffic accidents, four classification models were evaluated, namely, Support Vector Machine (SVM), RF, DT, and Deep Neural Network (DNN). SVM was used because it has an efficient library for large-scale data classification. At the same time, classification and regression trees were used because of their ability to handle both numerical and categorical data. Using accuracy, precision, recall, F-Score, and area under the curve (AUC) it was concluded that RF and DNN generally perform better than the other models.

Road traffic Accidents in Malawi

From the recent report issued by the Directorate of Road Traffic and Safety Services, out of 8,194 accidents that occurred in 2015, 888 were fatal, 706 were serious injuries, 2,632 resulted in minor injuries, and 3,944 in property damage. Compared to the number of accidents that occurred in 2014, there was an increase of 11.4%. 1,068 individuals died as a result of these traffic incidents. In 2014, there were 1,060 fatalities, an increase of 1% from the previous year. According to WHO, (2020) in Malawi, about 20 road traffic collisions occur every day, and about 1,000 people are killed in traffic accidents every year.

METHODOLOGY

Data Collection

This research used secondary data. Malawi's road traffic accident records captured over a period of five years (2016 to 2020) was used. The data was initially stored in their road safety database and was then extracted and shared with the researchers in Microsoft Excel format. The following attributes were captured in these accidents' records: accident number, severity, date, time, district, nearest police station, road number, the section of the road, noticeable physical feature close to the accident scene, accident type, road geometry, surroundings, surface type, road condition, weather, other factors, whether an animal was involved in the accident, whether there were any obstructions, whether there was a speed limit sign, speed limit and lighting condition

Data Preprocessing

In the data preprocessing phase, some of these attributes were left out as they brought noise to the dataset. Additionally, some input variables containing irrelevant information, redundant information, and attributes with over 80% unknown values were removed. This phase mainly involved data cleaning, transformation, and reduction. This was done to ensure that the data is of good quality before feeding it into a model. In this research, data was considered good if it had no missing values, no duplicate data, and no irrelevant fields. The data set was carefully reviewed for the issues mentioned above, and processed as follows: removing the invariant attributes (e.g., police station), removing the descriptive and wordy attributes (e.g., noticeable physical features close to the accident scene and road number), removing irrelevant attributes (e.g., accident ID), removing the attributes with over 80% unknown values (e.g., other factors, speed limit, obstruction, animal and section) and removing the redundant information (e.g., date, time). The final list of the attributes is presented in Table 1.

Attribute	**Values**
Severity	1: Fatal, 2: Serious injury, 3: Slightly/ minor injury, 4: Damages only, 5: Animal only
District	1: Chitipa, 2: Karonga, 3: Mzuzu, 4: Rumphi, 5: Mzimba, 6: Nkhatabay, 7:Kasungu, 8:Nkhotakota, 9:Ntchisi, 10:Dowa, 11:Mchinji, 12:Lilongwe, 13:Salima, 14:Dedza, 15:Ntcheu, 16:Mangochi, 17:Balaka, 18:Machinga, 19:Zomba, 20:Mwanza,

	21:Neno, 22:Blantyre, 23:Chiradzulu, 24:Mulanje, 25:Phalombe, 26:Chikwawa, 27:Thyolo, 28:Nsanje
Accident type	1: Moving+moving head-on, 2: Moving+moving rear end, 3: Moving+moving side, 4: Moving+moving overtake, 5: Moving+moving turn, 6: Single moving rollover, 7: Single moving collision, 8: Moving+pedestrian, 9: Moving bicycle, 10:Moving+controlled animal, 11:Moving+uncontrolled animal, 12:Moving+other
Road geometry	1: Straight Road, 2: Curve, 3: Roundabout, 4: T-junction, 5: Y-junction, 6: +-junction, 7: X-junction, 8: Bridge, 9: Road/Rail crossing
Surroundings	1: Rural area, 2: Urban area, 3: Peri/ urban, 4: Farm / compound
Surface type	1: Bitumen, 2: Gravel, 3: Earth
Road condition	1: Good/ Fair, 2: Potholes, 3: Corrugated, 4: Slippery
Weather	1: Dry, 2: Rain/Wet, 3: Mist, 4: Windy, 5: Dust
Posted speed limit	1: Speed limit posted, 2: Speed limit not posted
Light condition	1: Day light, 2: Night, 3: Dawn/Dusk

Table *1* Attributes and value description

Machine Learning Modelling

Modeling the severity of road accidents was achieved through classification methods. As a means of doing cross-validation, the data was split into test and training sets. Using the Sci-kit learn library in python, 80% of records were used to train each model and 20% to test the performance of the model. The results were analyzed and the models were compared on how they performed in predicting accident severity. Figure 1 below illustrates the framework described.

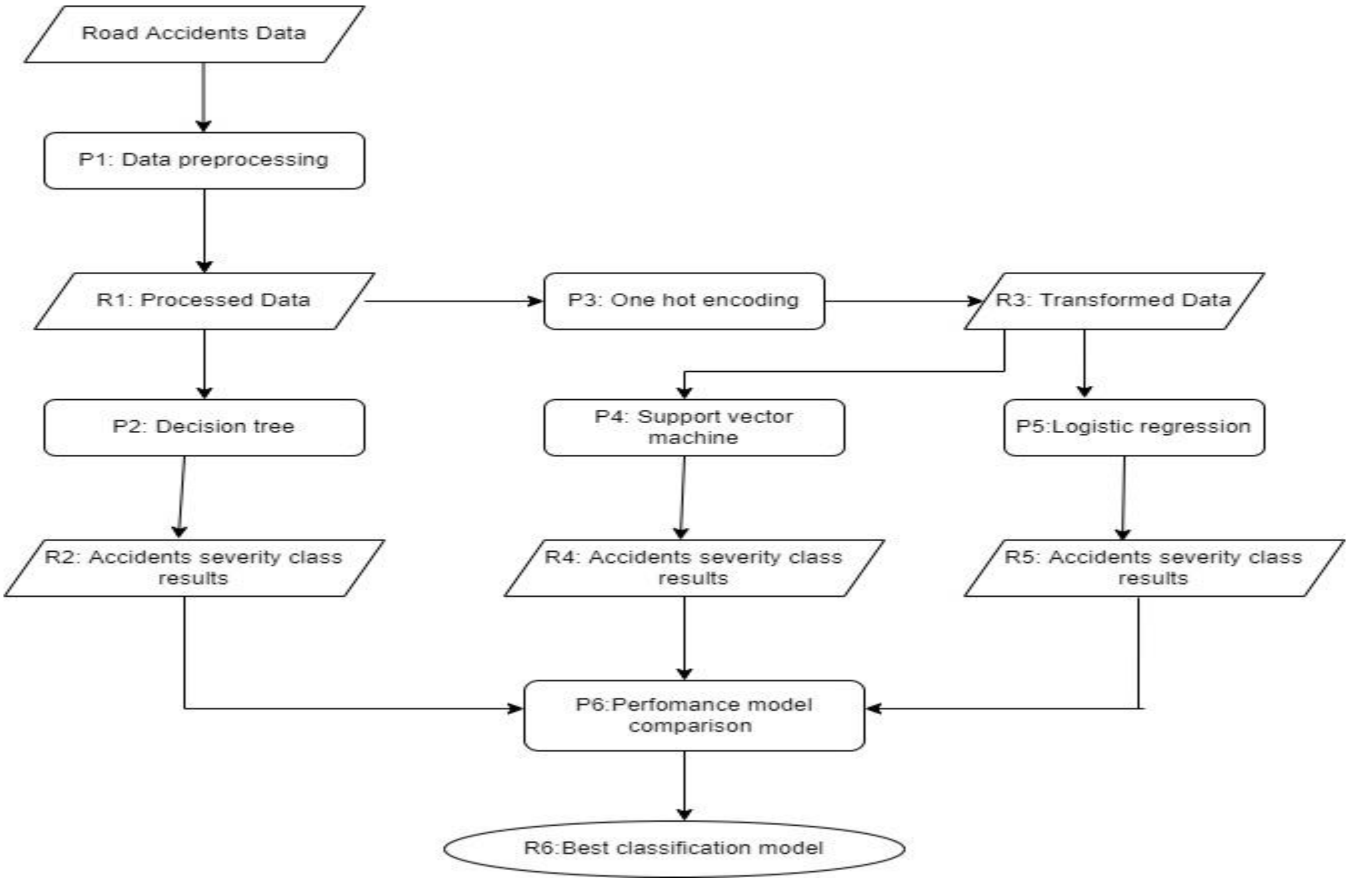

Figure 1 Modelling summary

Decision Trees

A decision tree was used as one of the classification algorithms. Since a DT model, specifically a classification and regression tree (CART), can be trained directly from categorical data, the output of the preprocessing stage R1, was used as input for the DT. A maximum depth of 9 was used after looping through a range of 1 to 25 to choose the best value. Unlike Yuan et al., (2017) who used a maximum depth of 13, our model began to overfit when the maximum depth went beyond 9. The maximum depth of 9 maximized the accuracy without overfitting. A criterion of entropy was used considering that attributes were in classes and not continuous. For the parameter Splitter, 'best' was used since it considered all features and it chose the best split, furthermore, the random split was also tested but it produced a slightly lower accuracy. Default values for minimum samples split and minimum samples leaf parameters were used. This was so because any changes to the values had no impact on the performance of the model.

The steps described above were repeated starting from process P1, this time around the severity classes were grouped into 2: fatal and non-fatal accidents. What was categorized as fatal initially remained fatal whereas serious injury, slightly/ minor injury, damages only, and Animal only were categorized as non-fatal accidents. In process P2 only the maximum depth parameter changed, while the rest remained the same. Maximum depth took the value 6, beyond this the model was overfitting.

Logistic Regression

Logistic regression (LR) was also used for predicting accidents severity. The output from the general preprocessing had to go through one more process before feeding into the algorithm. Since all variables were categorical, and logistic regression only handles categorical variables after they have undergone some transformation, one hot encoding was used to transform the data. By using grid search cross-validation for logistic regression the following parameters emerged as the best parameters for the model, C=5.79 and penalty = l2. Since the value that is being predicted (severity) is in 5 categories, this problem had to be solved as a multinomial logistic regression with a corresponding solver that supports multinomial classification. The model was trained using these parameters.

The processes described above were repeated starting from P1. The severity classes were grouped into fatal and non-fatal accidents. The data was split into training and test sets using the same ratio as in the initial run. The parameters were changed and a few more were introduced, multi_class parameter was omitted since the problem became a binary classification upon grouping severity into two classes. After looping through several options, the following parameter values were chosen to train the model; maximum iteration was increased to 1000, C was changed to 2, and solver was changed to 'saga'.

Support Vector Machine

Just like logistic regression, as some of the input data was categorical, it had to undergo one-hot encoding as well for an SVM. Nonetheless, a few more decisions had to be made before feeding the data into an SVM. Deciding which kernel function to go with to succeed in classification is a difficult task in SVM. Polynomial, linear, and RBF were all tested on the dataset, and RBF emerged to be the best choice for this dataset as it was able to separate the classes with higher accuracy. Just as it was with the case of Chong et al., (2005) where they found RBF to be the best choice. Consequently, the RBF kernel was used to train an SVM in this research, and two extra parameters were used. The parameter C, which is common to all SVM kernels took the value 10. And the parameter gamma took the value of 0.0001. Grid Search cross-validation was used to choose the optimal values for C and gamma after supplying a range of values to the function Chong et al., (2005) and Yuan et al., (2017).

Starting with process P1, the stages were repeated, but this time around the severity classes were divided into two groups. Fatal and non-fatal accidents. The data was split into training and test sets using the same ratio as in the initial run. The parameters for tuning the model remained the same as those in the first run.

RESULTS AND DISCUSIONS

In this research, the decision tree, logistic regression, and support vector machines were compared on how they performed in predicting the severity of accidents. The first run had severity categorized into five: Fatal, Serious injury, Slightly/ minor injury, Damages only, and Animal only. The performance measurements of the three models are listed in **Table 2**. DT and LR performed similarly well in predicting the severity of an accident compared to SVM. **Table 2** indicates that the DT, LR, and SVM accurately predicted the severity by 70.37%, 70.33, and 64.64 respectively.

	Accuracy	Precision	Recall	F1-score
Decision tree	70.37	67.29	70.37	67.64
Logistic regression	70.33	67.38	70.35	67.66
Support vector Machine	64.64	50.33	64.63	56.36

Table 2 results of classifying accidents severity into 5 categories

When severity was categorized into two classes, fatal and non-fatal accidents, the models predicted severity with the metrics shown in **Table 3**. This time around all three algorithms produced almost the same accuracy. With all accuracies above 85%, Decision trees, logistic regression, and support vector machines can all be used in the prediction of severity and one would expect an outcome with high accuracy. However, for precision and F1-score, the SVM did not perform as good as the DT and LR. Reducing the severity classes significantly helped improve the performance of the models as is evident by the increased accuracy, precision, recall and F1-score. Table 3 indicates that DT, LR, and SVM accurately predicted the severity by 86.79%, 87.55%, and 85.50% respectively.

	Accuracy	Precision	Recall	F1-score
Decision tree	86.79	84.49	86.79	84.48
Logistic regression	87.55	86.02	87.55	86.33
Support vector Machine	85.50	73.09	85.40	78.81

Table 3 results of classifying accidents severity into 2 categories

In this research, both the LR and DT performed well in predicting accident severity. However, the LR performed slightly better than the DT, because of the binary nature of the problem and the ability of the model to train on categorical data. This, however, differs from what most researchers found in traffic accidents analysis using data mining techniques. In Ethiopia and Abu Dhabi DTs were found as the best performing model for accident severity prediction largely because LR was not considered among the models (Beshah & Hill, 2010, Taamneh et al. 2017). Using accuracy and area under the curve (AUC) to compare the performance of the algorithms, the results indicated that the DT was among the best-performing algorithms. On the other hand, LR is rarely used in literature for accident severity prediction using data mining, but for this research, it is slightly better than the DT. Lastly, the support vector machine performed well, but not as good as the other two models, especially in terms of precision. The case is similar to when (Yuan et al., 2017) analysed motor vehicle crash data from the Iowa Department of Transportation. To predict traffic accidents, the Support Vector Machine was not among the best performing models.

Fitting a Logistic Regression

The logistic regression was then chosen as the best model for our data and was then fit to get more insights from the data. All the attributes contributed to determining the severity of an accident. However, some had a bigger impact compared to others. Using the model's coefficients, we were able to determine each attribute's contribution to determining the severity of an accident. The higher the coefficient implied the more likely an accident will be fatal and vice versa. With reference to the Table 4 below, the top 3 attributes that had a higher chance of causing a fatal accident than a non-fatal one are.

1. An accident involving a moving vehicle and a pedestrian
2. An accident that occurred at dawn or dusk
3. An accident involving a moving vehicle and a bicycle.

The bottom 3 attributes had a lower chance of causing fatal accidents and these were

1. An accident involving two vehicles moving side by side
2. An accident that occurred in an urban area
3. An accident involving a moving vehicle and an uncontrolled animal

Below are the coefficients of the attributes that contributed to accident severity.

Attribute	Description	coefficient
accident type_8	Moving + Pedestrian	5.405831
light condition_3	Dawn/Dusk	3.251403

accident type_9	Moving + Bicycle	3.246903
surroundings_4	Farm/Compound	2.732928
accident type_12	Moving + Other	2.620153
accident type_1	Moving + Moving	2.340571
accident type_6	Single moving rollover	2.274108
road geometry_7	X-Junction	2.022932
weather_5	Dust	1.789397
surroundings_1	Rural area	1.767158
road geometry_2	Curve	1.425653
accident type_7	Single moving collision	1.376577
road geometry_8	Bridge	1.37016
weather_1	Dry	1.344641
road condition_1	Good/Fair	1.343218
road condition_2	Potholes	1.228312
surface type_3	Earth	1.170019
road condition_3	Corrugated	1.104156
surface type_1	Bitumen	1.095167
road geometry_5	Y-Junction	1.073332
posted_speed_limit_1	Speed limit posted	1.065488
road geometry_1	Straight road	1.057091
surroundings_3	Peri/Urban	0.982609
weather_2	Rain/Wet	0.963796
accident type_4	Moving+Moving	0.951089
road geometry_9	Road/Rail crossing	0.950392
posted_speed_limit_2	Speed limit not posted	0.910609
road geometry_4	T-Junction	0.813645
road geometry_6	+ Junction	0.799293
weather_3	Mist	0.76422
light condition_2	Night	0.760634
surface type_2	Gravel	0.757195
accident type_10	Moving + Controlled animal	0.68572
accident type_5	Moving + Moving turn	0.559661
weather_4	Windy	0.547475
road condition_4	Slippery	0.532592
light condition_1	Day light	0.392314
road geometry_3	Roundabout	0.350126
accident type_2	Moving + Moving rear end	0.314506
accident type_3	Moving + Moving side	0.306897

surroundings_2	Urban	0.204454
accident type_11	Moving + Uncontrolled animal	0.081729

Table 4 coefficients of the attributes that contributed to accident severity

CONCLUSION

DTs and LR yielded the most accurate severity predictions compared to support vector machines. These models were evaluated using accuracy, precision, recall, and F1-score. However, the LR performed slightly better than the DT hence it was chosen to fit the data and get more insight. The coefficients of the attributes were calculated to get a picture of how much an attribute contributes to the severity of an accident. In the end, the type of accident turned out to be an attribute that has a higher chance of determining if an accident is fatal or not. Light conditions and surroundings also showed a higher impact on the severity. In terms of accident type, it was discovered that the top three attributes that contributed to fatal accidents were accidents involving a moving vehicle and a pedestrian, accidents that occurred at Dawn or Dust, and accidents involving a moving vehicle and a bicycle. From these findings we recommend having functional streetlights to avoid accidents at dawn or dusk, improved road infrastructure to include sidewalks or walkways to avoid vehicle and human contact and enforce proper regulations on bicycles.

In terms of the methodology used in this research, it is clear that using three methods of analysis is sufficient for comparing model performance. On the other hand, repeating each method while reducing the classes greatly improved the performance of our models hence improving the reliability of our results.

Study Limitations

The main limitation of this research was the data itself. The data available did not contain human factors and exact locations that may also contribute to accidents. For example, whether the driver was drunk or not, age group, and geographic coordinates of the accident scene. The data collected at road traffic only contained environmental factors, such as road geometry, surface type, road condition, etc. Spatial analysis of these accidents would also be a great area to look into so that high risk areas could be prioritized in traffic safety improvement. If the data had both human, spatial and environmental factors, the results generated would have provided a more realistic picture of what happens on our roads for these accidents to occur.

REFERENCES

1. Agrawal, R. (2016). Mining Association Rules between Sets of Items in Large Databases. January 1993. https://doi.org/10.1145/170035.170072
2. Beshah, T., & Hill, S. (2010). Mining Road Traffic Accident Data to Improve Safety: Role of Road-related Factors on Accident Severity in Ethiopia.
3. Chong, M., Abraham, A., & Paprzycki, M. (2005). Traffic Accident Analysis Using Machine Learning Paradigms. 29, 89–98.
4. Das, Subasish, Anandi Dutta, Raul Avelar, Karen Dixon, X. S. & M. J. (2019). Supervised association rules mining on pedestrian crashes in urban areas: identifying patterns for appropriate countermeasures. *International Journal of Urban Sciences*, *23*(1), 30–48.
5. Ihueze, C. C., & Onwurah, U. O. (2018). *Road tra ffi c accidents prediction modelling : An analysis of Anambra State , Nigeria*. *112*(December 2017), 21–29. https://doi.org/10.1016/j.aap.2017.12.016
6. Kassu, A., & Anderson, M. (2018). Determinants of Severe Injury and Fatal Traffic Accidents on Urban and Rural Highways. *International Journal for Traffic and Transport Engineering*, *8*(3), 294–308. https://doi.org/10.7708/ijtte.2018.8(3).04
7. Krishnaveni, S., Hemalatha, M., Professor, A., & Head, &. (2011). A Perspective Analysis of Traffic Accident using Data Mining Techniques. In *International Journal of Computer Applications* (Vol. 23, Issue 7).
8. Kumar, S., & Toshniwal, D. (2016). A data mining approach to characterize road accident locations. *Journal of Modern Transportation*, *24*(1), 62–72. https://doi.org/10.1007/s40534-016-0095-5
9. Li, L., Shrestha, S., & Hu, G. (2017). Analysis of road traffic fatal accidents using data mining techniques. Proceedings - 2017 15th IEEE/ACIS International Conference on Software Engineering Research, Management and Applications, SERA 2017, 363–370. https://doi.org/10.1109/SERA.2017.7965753
10. Martín, L., Baena, L., Garach, L., López, G., & Oña, J. De. (2014). Using data mining techniques to road safety improvement in Spanish roads. *Procedia - Social and Behavioral Sciences*, *160*(Cit), 607–614. https://doi.org/10.1016/j.sbspro.2014.12.174
11. Melecki Khayes. (2020). *Violence and Injury Prevention*. https://www.who.int/violence_injury_prevention/road_traffic/countrywork/mwi/en/
12. Saba Momeni Kho, Parham Pahlavani, B. B. (2021). Classification and association rule mining of road collisions for analyzing the fatal severity, a case study. *Journal of Transport & Health*, *23*.
13. Sajjad, S., Ehsan, T., & Sara, Z. (2017). THE EFFECT OF DRIVERS' DEMOGRAPHIC CHARACTERISTICS ON ROAD ACCIDENTS IN DIFFERENT SEASONS USING DATA MINING. *Scientific Journal on Traffic and Transportation Research*, *29*.
14. Shetty, P., C, S. P., Kashyap, S. V, & Madi, V. (2017). Analysis of road accidents using data mining techniques. In *International Research Journal of Engineering and Technology*. www.irjet.net

15. Szalay, Z. (2019). Analysis of traffic accident black spots: an application of spatial clustering segmentation method. https://www.researchgate.net/publication/340062435
16. Taamneh, M., Alkheder, S., & Taamneh, S. (2017). Data-mining techniques for traffic accident modeling and prediction in the United Arab Emirates. *Journal of Transportation Safety and Security*, *9*(2), 146–166. https://doi.org/10.1080/19439962.2016.1152338
17. WHO. (2020). *Road Traffic Imjuries*. World Health Organization. https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries
18. Yuan, Z., Zhou, X., Yang, T., Tamerius, J., & Mantilla, R. (2017). *Predicting Traffic Accidents Through Het-erogeneous Urban Data: A Case Study* (Vol. 9). https://doi.org/10.475/123_4
19. Zegeer, C. V., & Bushell, M. (2010). Pedestrian Crash Trends and Potential Countermeasures from Around the World. *Accident Analysis and Prevention*, *44*(1), 3–11.