2024

Crash injury prediction model for the Virginia Department of Transportation (VDOT)

BIG DATA FOR PUBLIC POLICY
CENTRAL EUROPEAN UNIVERSITY

ESQUENAZI, NATALIA & SARKOZI, TOMAS

WINTER TERM 2024

Table of Contents

Introduction

Car crash injuries pose a major threat to public health. According to Mohammed et al. (2019), 20 to 50 million people are injured in traffic accidents every year, and over a million killed. Moreover, the World Health Organization (WHO) (2019) estimated that car crashes have become the third most common cause of death worldwide, as shown in Table 1. Besides individual-specific losses, car crash injuries pose an economic burden as well. Most countries worldwide spend over 3% of their GDP to cover medical costs, property damage, and loss of productive capacities as a result (Wijnen, 2021).

No.	*1998 Disease or Injury*	*2020 Disease or Injury*
1	Lower respiratory contaminations	Ischemic heart disease
2	HIV/AIDS	Unipolar major depression
3	Perinatal conditions	Road traffic accident
4	Diarrheal diseases	Cerebrovascular diseases
5	Unipolar major depression	Chronic obstructive pulmonary diseases
6	Ischemic heart disease	Lower respiratory contaminations
7	Cerebrovascular diseases	tuberculosis
8	Malaria	War
9	Road traffic accident	Diarrheal diseases
10	Chronic obstructive pulmonary diseases	HIV/AIDS

Table 1. Prediction of the range of traffic accidents

Over the past two decades, the United States have had one of the greatest crash injury rates in the Americas region (Robartes and Chen, 2017). As cities, counties, and departments of transportation invest more resources into road infrastructure, it must be understood what can be done to minimize injuries. Between 2015 and 2023, over 40,000 car crashes with injuries were reported yearly in Virginia. These statistics highlight the importance of understanding the causes behind car crash injuries in Virginia to mitigate the possibility of a higher increase in the future. This research uses supervised machine learning techniques along with car crash data published by the Virginia department of transportation (VDOT) to identify the features that contribute the most to predicting drivers' injuries.

This analysis's contribution is focused on highlighting the importance of evidence in policymaking and opening the floor to informed discussions in road traffic safety management. Through this work, we aim to contribute significantly to the ongoing efforts towards a safer and more resilient transportation system in Virginia and beyond.

Research Question

What is the probability of experiencing a motor vehicle crash injury in the State of Virginia? What are the factors that contribute the most to car crash injuries?

Literature Review

Risk factors of a car crash incorporate driver behavior, environmental conditions, and the behavior of other traffic participants. Literature agrees that wearing seatbelts is one of the most important predictors of driver behavior causing car crashes. The seatbelt usage has been the subject of extensive research and policy focus globally due to its proven effectiveness in reducing the risk of injuries and fatalities in motor vehicle crashes (Evans, 1991). Moreover, Lestina, Williams, Lund, et al. (1991) analyzed motor vehicle crash injury patterns and the impact of Virginia's seat belt law. Their study revealed significant reductions in injuries following the implementation of the seat belt law, emphasizing the importance of legislative measures in mitigating car crash risks and reducing injury severity.

Christophersen and Gjerde (2014) examined the prevalence of alcohol and drugs in blood samples over the decade from 2000 to 2010, and found that alcohol or drugs were present in drivers' blood samples of almost half of the car accidents. In the case of single vehicle accidents, the ratio of intoxicated drivers was 63.8%. The highest proportion of alcohol samples in the blood was found in drivers under 25 years old, while a combination of alcohol and drugs was the most prevalent in the age group from 25 to 35 years old, highlighting the vulnerability of young drivers in traffic.

Research into environmental conditions as car crash severity predictors encompasses various factors, including road, light, and weather conditions, as well as incidents with animals. Malin, Norros, and Innamaa (2019) investigated the impact of road and weather conditions on car crash accidents. Their study, based on data from police-reported accidents, demonstrated that bad road and weather conditions, such as icy rain and slippery surfaces, significantly increased accident severity. Additionally, they identified variations across different road types, with motorways exhibiting higher risks during adverse weather conditions compared to other road types.

In parallel, Liu, Li, et al. (2019) explored factors influencing the severity of night-time vehicle accidents under low illumination conditions. The authors found that the likelihood of fatal night-time accidents on road segments surpasses that at intersections by a factor of 2.387. Moreover, the chances of fatal single-vehicle and vehicle-pedestrian night-time accidents exceed those of fatal vehicle-vehicle night-time accidents by factors of 7.591 and 1.749, respectively. Additionally, the probability of fatal night-time accidents on roads equipped with median dividers is 3.273 times higher than that on roads lacking median dividers.

In addition to environmental conditions, collisions with animals (CWAs) in car crashes have become an increased risk, as researched by Sullivan (2011). Specifically, from 1990 to 2008, the risk in the US of hitting animals doubled. Although not all CWAs incorporated deer, the increased occurrence of accidents follows the pattern of seasonal activity of the deer population. Moreover, fatal car crashes are positively associated with increased speed and decreased visibility, similarly to collisions with pedestrians. As the study by Levulytė et al. (2017) examined the role of pedestrians in road accidents and showed alarming global statistics of road crash fatalities, with an average of 3,287 pedestrian deaths per day, with 69% of all fatal pedestrian fatalities occurring inside urban areas, underscoring the vulnerability of pedestrians.

Furthermore, Das (2021) investigated the characteristics of motorcyclists involved in accidents between motorcycles and automobiles, pointing out the huge risk of motorcycle-car crashes as bike riders, unprotected by the exterior of a vehicle, do not have any alternative way to avoid traffic, unlike cyclists, for example, who can use cycle roads. This results in a 28 times higher chance of fatality during a crash with a car compared to a car's passenger. In addition, the study by Oliveira et al. (2015) shed light on the demographics and experience levels of motorcyclists, providing insights into the role of human behavior as a predictor of car crash risk. Particularly, men aged 20 years or older, with complete secondary education, and experienced in driving both motorcycles and cars, indicating that recklessness while driving the motorcycle is the main cause of traffic accidents.

Collectively, these studies underscore the multifaceted nature of car crash injury predictors, encompassing environmental, behavioral, and animal-related factors. Understanding and addressing these predictors is essential for developing effective strategies to improve road safety and reduce car crashes severity.

Study design

This study aims to predict the probability of getting injured in a car accident in the Virginia traffic system, as defined by the Virginia Department of Transportation (VDOT) (2022) – any injury other than fatal resulting in exposure of tissue, muscles, organs, at least second-degree burns covering a minimum of 10% of the body, unconsciousness after the accident, and broken extremities. The techniques chosen to analyze the data fall into the supervised machine learning domain, which intends to either accurate predict the outcome variable for future observations (prediction) or gain deeper understanding of how the outcome variable and predictors are related (inference) (James et al., 2015:26). The models fitted in this study are logistic regression, decision trees and random forest, the major objective is to predict whether a car accident is likely to involve injuries or not based on a combination of features such us weather conditions, pedestrians and bicycles involved, road ownership, date time and more.

Data source description

The data for this study was collected from the VDOT open data portal: Smarter Roads. The VDOT data collection procedures are through networks of sensors and systems, and the data portal continuously connects to different data sets and makes the latest data available to subscribers. The motor vehicle laws of Virginia under §46.2-373 require law enforcement officers to submit a police crash report to the VDOT for all reportable crashes within 24 hours after the completion of the crash investigation. This analysis uses time series data from motor vehicle car crashes reported between 2015 and 2023 and contains 1,037,647 car accident records.

As the VDOT focuses on reportable cases, any privacy concern is mitigated. The dataset does not provide personally identifiable information that could compromise individual privacy. The privacy implications are further reduced by nature of data collection related to characteristics of crash incident and environment conditions, rather than specifics about the participants of the accident. In addition, the dataset is anonymized. The anonymization process is applied to the crash data during the collection and reporting stage.

Input variables and outcome

This research study uses a combination of 20 qualitative and quantitative features that are described and classified below:

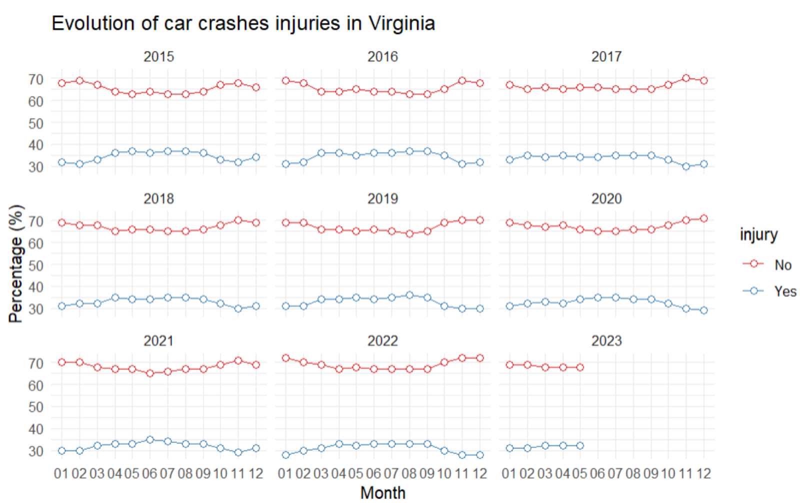
Variable name	Description	Type
Injury	Yes/No	Qualitative
Weather_co	1= No Adverse Condition 3=Fog 4=Mist 5=Rain 6=Snow 7=Sleet/Hail	Qualitative
Light_cond	1=Dawn 2=Daylight 3=Dusk 4=Darkness – road lighted 5= Darkness – road not lighted 6= Darkness – unknown 7=Unknown	Qualitative
K_people	Number of people killed	Quantitative
Veh_count	The number of vehicles involved in the crash	Quantitative
Alcohol_no	0=No 1=Yes	Qualitative
Belted_unb	0=Yes 1=No	Qualitative
Bike_nobi	0=No 1=Yes	Qualitative
Collision	1=Rear end 2=Angle 3=Head on 4=Sideswipe – same direction 5=Sideswipe – opposite direction 6=Object in road 7=Train 8=No collision 9=Object off road 10=Deer 11=Other animal 12=Pedestrian 13=Bicyclist 14=Motorcyclist 15=Backed into	Qualitative
Distracted	0=No 1=Yes	Qualitative
Animal	0=No 1=Yes	Qualitative
Drowsy_not	0=No 1=Yes	Qualitative
Drug_nodru	0=No 1=Yes	Qualitative
Motor_nonm	0=No 1=Yes	Qualitative
Ped_nonped	0=No 1=Yes	Qualitative
Speed_nots	0=No 1=Yes	Qualitative
Senior_not	0=No 1=Yes	Qualitative
Young_noty	0=No 1=Yes	Qualitative
Ownership	1=State road 2=County road 3=City road 4=Federal road 5=Toll roads 6=Private road	Qualitative
Year	Year and month	Date time

Table 2: Features descriptive statistics

```
injury         K_PEOPLE           VEH_COUNT        WEATHER_CO      LIGHT_COND ALCOHOL_NO BELTED_UNB BIKE_NONBI
No :695883   Min.   :0.000000   Min.   : 1.00   1        :858553   1: 28142   0:978027   0:995399   0:1032508
Yes:341762   1st Qu.:0.000000   1st Qu.: 1.00   5        :137029   2:681827   1: 59618   1: 42246   1:   5137
             Median :0.000000   Median : 2.00   6        : 16170   3: 30134
             Mean   :0.006935   Mean   : 1.84   4        : 14339   4:138477
             3rd Qu.:0.000000   3rd Qu.: 2.00   3        :  5106   5:154692
             Max.   :6.000000   Max.   :75.00   7        :  4161   6:  2480
                                                (Other):  2287     7:  1893
  COLLISION_      DISTRACTED ANIMAL     DROWSY_NOT  DRUG_NODRU  MOTOR_NONM  PED_NONPED  SPEED_NOTS SENIOR_NOT YOUNG_NOTY
1       :321315   0:841872   0:973587   0:1008539   0:1028018   0:1021146   0:1024636   0:831504   0:866121   0:843819
2       :266381   1:195773   1: 64058   1:  29106   1:   9627   1:  16499   1:  13009   1:206141   1:171524   1:193826
9       :202090
4       : 81706
10      : 48467
16      : 30590
(Other): 87096
OWNERSHIP       year            month
1:674176   2018   :131848   10      : 96649
2: 32591   2016   :128525   11      : 94998
3:322013   2019   :128172   05      : 94257
4:  2274   2017   :127374   12      : 88742
5:  1623   2015   :125799   01      : 86583
6:  4968   2022   :122434   03      : 85106
           (Other):273493   (Other):491310
```

The dependent variable in this study is injury, a binary two-level dummy variable denoting whether a car accident resulted in injuries or not. The descriptive statistics presented in Table 2 reveal a cumulative count of 341,762 car crash injuries reported in Virginia over the years. Furthermore, it is observed in Graph 1 that the incidence of car crash injuries has maintained a stable proportion, accounting for approximately 30% of the total recorded crashes throughout the study period. In other words, the VDOT has been unsuccessful in reducing crash injuries.

Graph 1: Evolution of car crashes in Virginia



As can be seen in Table 2, the outcome variable injury shows a class imbalance. This study implements an under-sample technique over the train dataset to improve models' performance and reduce any potential biases caused by the majority class. The down sampling reduces the number of observations in the majority class to finally match the number of observations in the minority class. The under sample reduced the number of crashes without injuries to 239,233 and matched the number of crashes with injuries in the training dataset.

Supervised machine learning

I. Logistic regression

This study uses multiple logistic regression to estimate the true probability of having a crash injury in Virginia. The logistic model estimates the relationship between predictor variables and the binary outcome, representing the probability of a positive outcome (James et al., 2015:26). By interpreting the coefficients, it is expected to determine the impact of each feature on the log odds of the positive outcome. A logit regression model is implemented in this study, which uses a logistic cumulative distribution function. The equation can be written as follows, where p represents the probability of a car crash injury:

$$log\left(\frac{p}{1-p}\right) = \beta 0 + \beta 1x1\ (weather\ conditions) + \beta 2x2(pedestrians\ involved) + \cdots + \beta nxn$$

The odds ratio shows five co-variables as most relevant for car crash injuries: pedestrians, bikes, motorcycle, seatbelt and collision type. If pedestrians are involved in a car accident, the odds of experiencing a crash injury are 4304% times higher compared to incidents where pedestrians are not involved. The odds for bikes involved in the crash rise to 3306%, and for motorcycles 891%. In terms of drivers not using seatbelt, the probability of a crash injury increases by 354% compared to drivers with seatbelt. Finally, a collision type 14 (crash with a motorcycle) and 3 (head on crash) also show positive odds of 142% and 115% respectively.

Table 3: Logistic regression odds ratio in percentage differences

PED_NONPED1	BIKE_NONBI1	MOTOR_NONM1	BELTED_UNB1	COLLISION_14	COLLISION_3	COLLISION_8	DRUG_NODRU1
4304.789288	3306.602880	891.007008	354.745545	142.860382	115.924407	82.794876	70.054586
OWNERSHIP3	VEH_COUNT	COLLISION_9	COLLISION_16	DROWSY_NOT1	ALCOHOL_NO1	SENIOR_NOT1	WEATHER_CO8
33.749231	31.766385	30.711387	30.291549	28.481977	22.341473	20.657600	19.246452
COLLISION_2	SPEED_NOTS1	WEATHER_CO4	LIGHT_COND2	WEATHER_CO11	WEATHER_CO9	LIGHT_COND3	COLLISION_12
18.364243	13.957787	10.202624	10.158232	9.525953	8.972109	8.597945	7.097183
WEATHER_CO3	LIGHT_COND5	DISTRACTED1	year2016	COLLISION_13	LIGHT_COND4	ANIMAL1	year2017
6.744261	4.911623	2.832019	1.976238	-2.494843	-2.713015	-3.865224	-4.749257
year2018	WEATHER_CO10	LIGHT_COND6	year2019	YOUNG_NOTY1	COLLISION_7	year2020	COLLISION_5
-4.817165	-5.535889	-5.580949	-6.893703	-6.980659	-7.519908	-9.212477	-9.557109
WEATHER_CO5	year2021	year2023	year2022	COLLISION_6	OWNERSHIP5	OWNERSHIP6	OWNERSHIP2
-11.664453	-12.603513	-14.383263	-14.638772	-15.192553	-22.622571	-26.024859	-27.700516
WEATHER_CO7	OWNERSHIP4	WEATHER_CO6	COLLISION_11	COLLISION_4	COLLISION_15	COLLISION_10	LIGHT_COND7
-30.210751	-31.227413	-40.902296	-49.010902	-50.925709	-56.226386	-71.422518	-80.216387

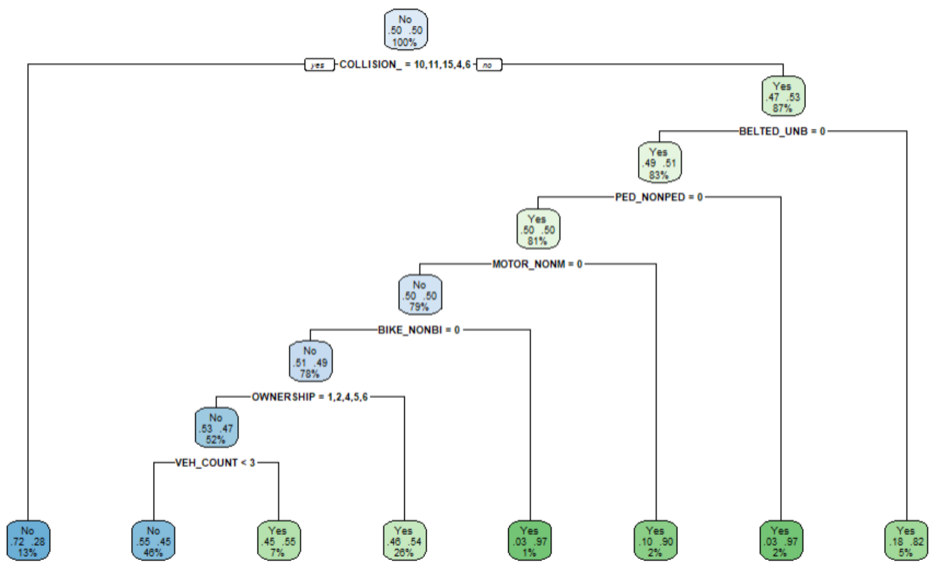
II. Decision tree and Pruning

Classification trees are implemented in this study to predict crash injuries in Virginia. The algorithm selects the best predictors and the optimal split point to classify the data between crash injury and no crash injury. Following this, the key aspect in decision trees is that each observation predicted belongs to the most commonly occurring class of training observations (James et al., 2015:311). The structure contains a combination of root and leaf nodes, which has a majority class assigned as the predicted class label for the observations. The main advantage of classification trees is related to its interpretation. The decision tree rules and the splitting criteria make it easier to understand and explain the predictions.

The results observed in the decision tree show that collision type is the root node. The left side of the decision tree indicates that for collision types such as crash with animals, a deer or a fixed object in the road, there is less probability of experiencing a car crash injury at 72%. The right side of the decision tree includes seven relevant variables: collision type, seatbelt, pedestrians, motorcycle, bicycle, road ownership and vehicle count. When the collision type is classified as rear end, angle or head on, the probabilities can be interpreted as follows:

- If the driver was not using a seatbelt, the probability of crash injury rise to 82%.
- If pedestrians were involved in the crash, even though the driver was using seatbelt, the probability of injuries is 97%.
- If a motorcycle was involved in the crash, even though the driver was using seatbelt and pedestrians were not involved, the probability of injuries is 90%.
- If a bicycle was involved in the crash, even though the driver was using seatbelt and pedestrians and motorcycles were not involved, the probability of injuries is 97%.
- If the car accident happened on a city road, even though the driver was using seatbelt and pedestrians, motorcycles and bicycles were not involved, the probability of injuries is 54%.
- If more than 3 vehicles were involved in the crash, even though the driver was using seatbelt, pedestrians, motorcycles and bicycles were not involved, and the crash happened either on state, county, federal or private roads, the probability of injuries stands at 55%.

Graph 2: Decision tree plot



In decision trees models, it is important to consider overfitting issues in the data that might produce a poor performance on the test set (James et al., 2015:311). There are several strategies to overcome overfitting, pruning is an alternative to optimize the decision tree for a better performance. In pruning, the complexity parameter controls the size of the decision tree and removes branches that have weak predictive power. By using a complexity plot (graph 3), this study found the optimal parameter (0.01). From the table 4, we can see that the error does reduce significantly beyond the first split, which suggests that pruning might be beneficial to prevent overfitting. However, the decision tree plot with pruning, achieved the same results as without pruning.

Graph 3: Complexity parameter plot

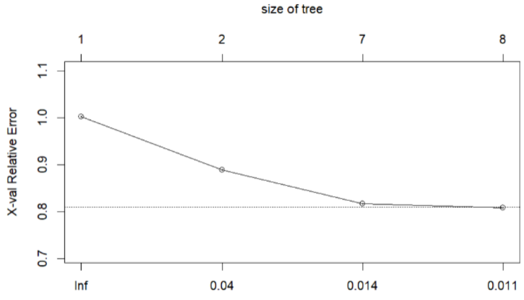


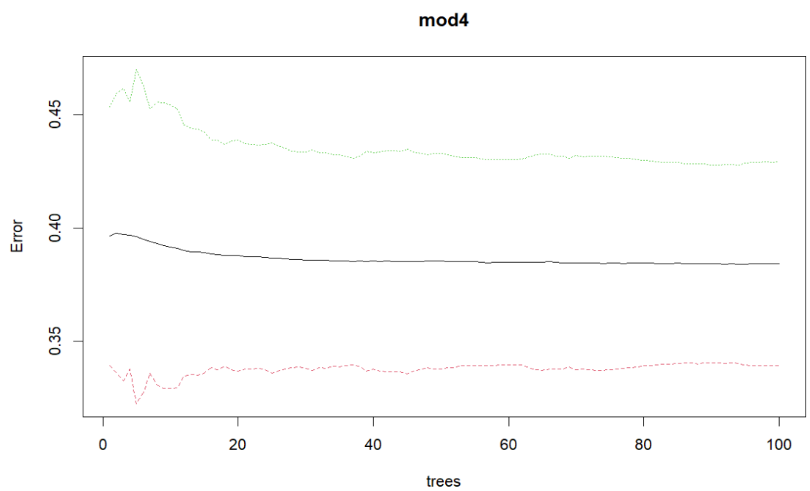
Table 4: Complexity parameter table

```
Classification tree:
rpart(formula = injury ~ VEH_COUNT + ALCOHOL_NO + BELTED_UNB +
    BIKE_NONBI + OWNERSHIP + COLLISION_ + WEATHER_CO + LIGHT_COND +
    DISTRACTED + ANIMAL + DROWSY_NOT + DRUG_NODRU + MOTOR_NONM +
    PED_NONPED + SPEED_NOTS + SENIOR_NOT + YOUNG_NOTY + year,
    data = train, method = "class", control = rpart.control(maxdepth = 10,
        minsplit = 10))

Variables actually used in tree construction:
[1] BELTED_UNB BIKE_NONBI COLLISION_ MOTOR_NONM OWNERSHIP  PED_NONPED VEH_COUNT

Root node error: 239233/478466 = 0.5

n= 478466

        CP nsplit rel error  xerror      xstd
1 0.112037      0   1.00000 1.00310 0.0014457
2 0.014123      1   0.88796 0.88923 0.0014368
3 0.013125      6   0.81735 0.81743 0.0014214
4 0.010000      7   0.80422 0.80922 0.0014191
```

III. Random Forest

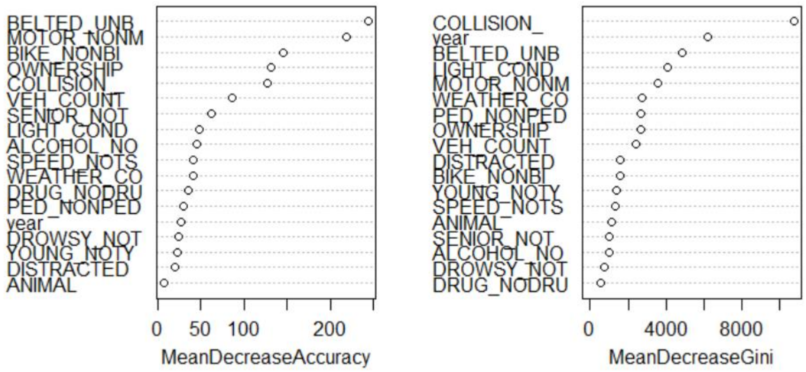
The last model implemented in this study is random forest. Considering the existence of high variance and correlation in decision trees, random forest is helpful to overcome this issue by decorrelating the trees to get a less variable and more reliable result (James et al., 2015:320). In this model, the selected predictor subset size is 4 and the number of trees is 100. In Graph 4, it is possible to see a decrease in the error along with an increase in the number of trees that means that quality will improve with an increase in the number of trees.

Graph 4: Random forest plot



Regarding variable's relevance, Graph 5 shows the mean decrease accuracy and the mean decrease Gini plot. The first one refers to what will happen to accuracy if we remove one feature from the model. In this case, the variables that will result in a larger decrease in accuracy are seatbelt and motorcycle. The second one has a different concept and measures the relative decrease in Gini index or impurity as a result of each variable. Following this, the most important variable to the model is collision type.

Graph 5: Variable importance plot



Findings and model evaluation

In this chapter, we present the main findings and evaluation results of the machine learning models implemented to predict the probability of an injury in car accidents within the Virginia traffic system. We discuss the insights gained from logistic regression, decision trees, and random forest models trained on a dataset collected from the VDOT.

I. Logistic regression

Based on our logistic regression model we identified the most important predictors of crash injury: collision with pedestrians, cyclists and motorcyclists, drivers not using a seatbelt and collision type. The first two predictors increase the likelihood of injury by more than 3000%, pointing out that non-motorists are at greatest risk in road accidents. The first instrument to measure model performance is the roc curve, which shows a curve that is not very close to the Y axis (true positive rate). The second model performance measurement is a confusion matrix, which indicates that we correctly predicted 22,207 car crash injuries in comparison with 10,280 prediction of type I error. The number of car crashes without injury predicted was 198,485 compared to 80,322 predictions of type II error as depicted in table 5. The overall accuracy result of the logistic model is 71%.

Graph 6: Roc curve

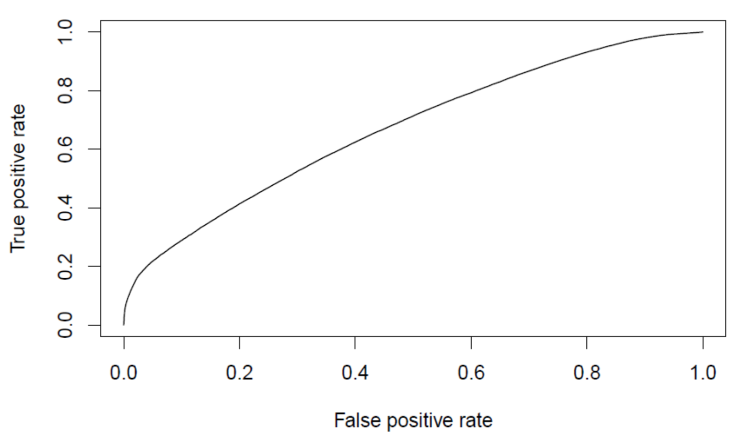


Table 5: Logistic regression confusion matrix

	FALSE	TRUE
No	198485	10280
Yes	80322	22207

II. Decision tree and pruning

In the decision tree model, the most important predictors of a car crash injury are similar to the logistic model: motorcycle, bicycle and pedestrians. Although, decision tree model enriches our perspective by collision type level. The decision tree identified specific collision types, such as rear end, angle, and head-on, as critical factors associated with varying probabilities of injury. Notably, the model also highlighted the importance of seatbelt usage and impact of road ownership, placing city owned road to most car crash injury prevalent.

In our decision tree model, we correctly predicted 52,621 car crash injuries in comparison with 49,908 prediction of type I error, while 142,781 car crashes without injury are compared to 65,984 predictions of type II error as depicted in table 6. The model shows an accuracy rate of 62%, which is lower than the logistic regression. In addition, after the pruning method we did not find any performance improvements in the model.

Table 6: Decision tree confusion matrix

```
          Actual
Predicted     No    Yes
      No  142781  49908
      Yes  65984  52621

               Accuracy : 0.6277
                 95% CI : (0.626, 0.6294)
    No Information Rate : 0.6706
    P-Value [Acc > NIR] : 1

                  Kappa : 0.1896

 Mcnemar's Test P-Value : <2e-16

            Sensitivity : 0.6839
            Specificity : 0.5132
         Pos Pred Value : 0.7410
         Neg Pred Value : 0.4437
             Prevalence : 0.6706
         Detection Rate : 0.4587
   Detection Prevalence : 0.6190
      Balanced Accuracy : 0.5986

       'Positive' Class : No
```

III. Random Forest

Based on the random forest model, we identified the following variables as the most important predictors of injury during car crashes: not using a seatbelt, collision with motorcyclists and collision type. The confusion matrix shows that the model correctly predicted 58,724 car crash injuries in comparison with 43,787 prediction of type I error. The number of correct injuries prediction increased compared to the decision tree previous result. Also, 137,613 car crashes without injury were predicted in comparison with 71,152 predictions of type II error as depicted in table 7. The accuracy rate is 63%, and is also higher than in the decision tree.

Table 7: Random forest confusion matrix

```
Confusion Matrix and Statistics

          Actual
Predicted     No    Yes
      No  137613  43787
      Yes  71152  58742

               Accuracy : 0.6308
                 95% CI : (0.6291, 0.6325)
    No Information Rate : 0.6706
    P-Value [Acc > NIR] : 1

                  Kappa : 0.2173

 Mcnemar's Test P-Value : <2e-16

            Sensitivity : 0.6592
            Specificity : 0.5729
         Pos Pred Value : 0.7586
         Neg Pred Value : 0.4522
             Prevalence : 0.6706
         Detection Rate : 0.4421
   Detection Prevalence : 0.5827
      Balanced Accuracy : 0.6161

       'Positive' Class : No
```

Final recommendations

Based on the findings of our machine learning models we identified the most common predictors of car crash injuries in Virginia, the following recommendations are therefore proposed for the VDOT as an initiative to decrease the probability of such injuries:

1. *Enhance Enforcement of Seatbelt Laws*: Implement stricter enforcement measures, including higher fines or penalties, for drivers who fail to wear seatbelts. Public awareness campaigns emphasizing the importance of seatbelt usage should also take place in public champagnes to promote traffic security.
2. *Improve Infrastructure for Cyclists and Pedestrians*: Allocate resources to build dedicated cycling lanes and pedestrian-friendly pavements in high-traffic areas. By providing safe and separate pathways for cyclists and pedestrians, the risk of collisions with motor vehicles can be significantly reduced.
3. *Implement Visibility Measures for Pedestrian*s: Mandate the use of high-visibility clothing or accessories for pedestrians, especially during low-light conditions or in areas with limited visibility. Wearing reflective vests or bands can enhance the visibility of pedestrians to drivers, thereby decreasing the likelihood of accidents.
4. *Enforce Strict Helmet Regulations for Motorcyclists*: Enforce regulations requiring motorcyclists to wear high-quality helmets that meet stringent safety standards. Implementing mandatory helmet laws and conducting regular inspections to ensure compliance cannot prevent car-motorcycle accidents, although still, help mitigate the severity of head injuries in motorcycle accidents.
5. *Implement Traffic Calming Measures*: Introduce traffic calming measures, such as speed bumps, traffic islands, and roundabouts, in areas prone to car accidents. These measures can help reduce vehicle speeds and enhance road safety for all road users.
6. *Promote Public Education and Awareness Campaigns*: Launch comprehensive public education campaigns to raise awareness about the risks associated with car accidents and the importance of complying with road safety regulations. Use various channels, including social media, TV, and community events, to spread safety messages effectively.

We firmly believe that implementation of evidence-based policy making, such as the recommendations provided to the VDOT to enhance road safety, is an example of good governance principles. By relying on empirical evidence and data-driven analysis, VDOT can make informed decisions that prioritize the well-being and safety of the public.

References

Das, S. (2021). Identifying key patterns in motorcycle crashes: findings from taxicab correspondence analysis. Transportmetrica A: transport science, 17(4), 593-614. Retrieved from: Identifying key patterns in motorcycle crashes: findings from taxicab correspondence analysis: Transportmetrica A: Transport Science: Vol 17, No 4 (tandfonline.com)

Evans, L. (1991). Traffic Safety and the Driver. Solomon Islands: Van Nostrand Reinhold.

Christophersen, A. S., & Gjerde, H. (2014). Prevalence of alcohol and drugs among car and van drivers killed in road accidents in Norway: an overview from 2001 to 2010. Traffic injury prevention, 15(6), 523-531.

Lestina, D. C., Williams, A. F., Lund, A. K., Zador, P., & Kuhlmann, T. P. (1991). Motor vehicle crash injury patterns and the Virginia seat belt law. *Jama*, *265*(11), 1409-1413. Retrieved from: Motor Vehicle Crash Injury Patterns and the Virginia Seat Belt Law | JAMA | JAMA Network

Levulytė, L., Baranyai, D., Sokolovskij, E., & Török, Á. (2017). Pedestrians' role in road accidents. *International Journal for Traffic and Transport Engineering*, *7*(3), 328-341. Retrieved from: 6d5cf703-9db3-e0b2ijtte.2017.7(3).04.pdf

Liu, J., Li, J., Wang, K., Zhao, J., Cong, H., & He, P. (2019). Exploring factors affecting the severity of night-time vehicle accidents under low illumination conditions. *Advances in Mechanical Engineering*, *11*(4), 1687814019840940. Retrieved from: Exploring factors affecting the severity of night-time vehicle accidents under low illumination conditions - Jing Liu, Jingyu Li, Kun Wang, Jianyou Zhao, Haozhe Cong, Ping He, 2019 (sagepub.com)

Malin, F., Norros, I., & Innamaa, S. (2019). Accident risk of road and weather conditions on different road types. *Accident Analysis & Prevention*, *122*, 181-188. Retrieved from: Accident risk of road and weather conditions on different road types - PubMed (nih.gov)

Mohammed, A. A., Ambak, K., Mosa, A. M., & Syamsunur, D. (2019). A review of traffic accidents and related practices worldwide. The Open Transportation Journal, 13(1). Retrieved from: A Review of Traffic Accidents and Related Practices Worldwide (opentransportationjournal.com)

Oliveira, A. L. D., Petroianu, A., Gonçalves, D. M. V., Pereira, G. A., & Alberti, L. R. (2015). Characteristics of motorcyclists involved in accidents between motorcycles and automobiles. Revista da Associação Médica Brasileira, 61, 61-64. Retrieved from: SciELO - Brazil - Characteristics of motorcyclists involved in accidents between motorcycles and automobiles Characteristics of motorcyclists involved in accidents between motorcycles and automobiles

Sullivan, J. M. (2011). Trends and characteristics of animal-vehicle collisions in the United States. *Journal of safety research*, *42*(1), 9-16. Retrieved from: Trends and characteristics of animal-vehicle collisions in the United States - ScienceDirect

Wijnen, W. (2021). Socio-economic costs of road crashes in middle-income countries: Applying a hybrid approach to Kazakhstan. IATSS research, 45(3), 293-302. Retrieved from: Socio-economic costs of road crashes in middle-income countries: Applying a hybrid approach to Kazakhstan - ScienceDirect

World Health Organization. (2019). Global status report on road safety 2018. World Health Organization. Retrieved from: Global Status Report on Road Safety 2018 - World Health Organization - Google Books

Robartes, Erin, and T. Donna Chen. (2017). The Effect of Crash Characteristics on Cyclist Injuries: An Analysis of Virginia Automobile-Bicycle Crash Data. Accident Analysis & Prevention 104 (July): 165–73. https://doi.org/10.1016/j.aap.2017.04.020.

"SmarterRoads." n.d. Accessed April 10, 2024. https://smarterroads.org/login

James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. (2015). An Introduction to Statistical Learning with Applications in R. Second edition. USA: Springer.