**Project report for the degree of
Masters of Science (Ms) in Business Analytics
at the University of Texas,Dallas**

By Srinath M Krishnan and Asad Moulvi

Table of contents

1.	Executive Summary	3
2.	Background / Context	3
2.1	Domain	3
2.2	Brief Description of the scenario	3
2.3	Decision of Interest	3
2.4	Decision Maker	3
3.	Business Understanding	4
3.1	Business Objective	4
3.2	Situation assessment	4
3.3	Data Mining goals	4
4.	Data Understanding	4
4.1	Data Requirements	5
4.2	Describe data	5
4.3	Sources	6
4.4	Quality	6
5	Describe preparation	6
5.1	Data selection	7
5.2	Prepare data	7
6	Modelling	7
6.1	Data description	9
6.2	Model selection	10
6.3	Justifying the selected models	10
6.4	Model development and output	11
7	Model evaluation	13
8	Discussion	13
8.1	Decision	13
8.2	Limitations	14
8.3	Biases	14
8.4	Enhancements	15

1. EXECUTIVE SUMMARY

The goal of this project is to develop models for predicting the severity of traffic accidents. Both the logistic regression and decision tree models have strengths and weaknesses in this context, with logistic regression slightly outperforming in specificity and AUC. However, logistic regression struggles with complex relationships and assumes independent observations, while decision trees are prone to overfitting and sensitive to data variations. Both models face challenges with missing or erroneous data and are dependent on the quality and quantity of the training dataset, which could lead to biases, particularly in cases of severe accidents. To improve decision-making and mitigate cognitive biases like the availability heuristic and confirmation bias, decision support tools offer objective, data-driven analyses. Enhancements to these models could include real-time data integration, advanced feature engineering, geospatial analysis, a focus on model interpretability, and a user-friendly interface for decision-makers. These improvements would enhance the accuracy and utility of the models in predicting accident severity and in effectively allocating resources.

2. BACKGROUND/CONTEXT

In the year 2022, it is recorded that a staggering 42,795 individuals lost their lives as a result of motor vehicle traffic collisions. This profound issue reverberates through various sectors, profoundly affecting numerous stakeholders, including law enforcement personnel, emergency medical staff, insurance enterprises, and a myriad of others. In the pursuit of this endeavour, our project endeavours to leverage data sourced from the California Highway Patrol to meticulously scrutinise prevailing trends and discernible patterns. Through this meticulous analysis, we aim to furnish invaluable insights and recommendations to pivotal decision-makers, facilitating the formulation and implementation of policies aimed at mitigating the incidence of such tragic events.

2.1 DOMAIN

Our analysis of this dataset is tailored to a governmental perspective, with a specific focus on the California Department of Transportation (Caltrans).

2.2 BRIEF DESCRIPTION OF THE SCENARIO

Our analysis will focus on evaluating the effectiveness of current speed limit policies in California. By studying historical traffic data and accident reports, we aim to determine whether adjustments to speed limits are necessary in specific areas to improve road safety. We will also assess the allocation of law enforcement resources to examine if police patrols properly align with high-risk areas and times. This will provide insights to optimise enforcement deployment. Additionally, we will scrutinise the procedures in place for inclement weather conditions, as California experiences various weather challenges. Our goal is to refine protocols to ensure safety and preparedness during adverse events. Ultimately, we intend to provide Caltrans with data-driven recommendations to enhance transportation policies, improve safety measures, and optimise resource allocation. Key decision-makers within Caltrans, including the Director, division heads, and safety experts, will utilise our analysis to inform critical decisions on speed limit adjustments, police allocation, and inclement weather procedures. Our overarching objective is to deliver actionable insights that promote road safety, effective traffic management, and transportation resilience for the benefit of all Californians.

2.3 DECISION OF INTEREST

Firstly, we intend to evaluate the effectiveness of current speed limit policies. By studying historical traffic data and accident reports, we aim to determine whether adjustments to speed limits are necessary in specific areas to enhance road safety. Secondly, we will assess the allocation of law enforcement resources. This involves examining whether police patrols align with high-risk areas and times, providing valuable insights for optimising resource deployment. Lastly, we will scrutinise the procedures in place for inclement weather conditions. California experiences various weather challenges, and our analysis will help refine protocols to ensure the safety and preparedness of road users during adverse weather events. Our overarching goal is to provide Caltrans with data-driven recommendations to improve transportation policies, enhance safety measures, and optimise resource allocation, ultimately leading to safer and more efficient roadways for the benefit of all Californians.

2.4 DECISION MAKERS

Key decision-makers within Caltrans, including the Director, division heads, and safety experts, will utilise our analysis to inform critical decisions. For speed limits, our data-driven insights will guide evidence-based adjustments in areas where necessary, considering historical accident data and traffic patterns. In terms of police allocation, our analysis will help optimise law enforcement deployment for more effective traffic management. Lastly, our findings on inclement weather procedures will inform adjustments to ensure road user safety. Ultimately, our goal is to provide actionable insights that enhance road safety, traffic management, and transportation resilience for all Californians.

3. BUSINESS UNDERSTANDING

The city of San Francisco has implemented a road safety policy known as the Vision Zero SF. Every year, over 30 individuals are killed and at least 500 more are badly injured while travelling on city streets in San Francisco. These fatalities and injuries are both inexcusable and avoidable, and San Francisco is dedicated to preventing similar tragedies. This was the key objective behind the Vision Zero SF.

3.1 BUSINESS OBJECTIVE

Traffic accidents and fatalities have been a major public health concern across California. In 2021 alone, there were over 3,000 traffic fatalities statewide. The leading causes include impaired driving, speeding, and failure to yield right-of-way.
To address this crisis, the state of California has set a goal to reduce traffic fatalities and serious injuries by 20% by 2030 through the implementation of a statewide traffic safety strategic plan.

3.2 SITUATION ASSESSMENT

The key objective is to leverage data analytics to gain insights into the factors contributing to traffic accidents and fatalities across California. This will inform targeted education and awareness campaigns, enforcement measures, and engineering improvements to make streets safer.
The program will analyse available traffic accident data from cities and counties across California looking at factors like time of day, road conditions, demographics, and driver behaviours. Visualisations and statistical models will identify high-risk areas, times, and populations to prioritise interventions.

3.3 DATA MINING GOALS

Preliminary analysis shows weekends, nighttime hours, impaired driving, speeding, and young drivers are major factors in fatal crashes statewide. Weather and road conditions also play a role. Focusing education and enforcement on times, locations, and behaviours associated with severe crashes can have an outsized impact on reducing fatalities and meeting safety goals.

Differences between northern and southern California highlight the need for location-specific interventions. Rural highways see more single-vehicle crashes while urban areas have more crashes at intersections. Solutions will need to be tailored based on insights from the data analysis.

4. DATA UNDERSTANDING

The dataset for this project was sourced from the Statewide Integrated Traffic Records System, which collects and processes collision data from across California. This centralised database is populated using reports filed by both local and state agencies, including the California Highway Patrol (CHP).

4.1 DATA REQUIREMENTS

While the original dataset encompasses collisions statewide, this project focuses specifically on those occurring in several counties in California namely San Mateo, SanFrancisco, Alameda, Santa Clara, Solano, Sonoma and Amador.

Additionally, the data was scrubbed to remove any rows containing null values. Since all attributes being analysed are categorical variables like weather condition, road surface type, etc., it was important to discard incomplete records to avoid skewing the models

4.2 DESCRIBE DATA

The data contains 9 Columns and 2690 Rows. The Features include Weekday, ViolCat, Month, CrashType, ClearWeather, Highway, Daylight. The target class is Severity

VARIABLE NAME	DESCRIPTION OR POSSIBLE VALUES
ID	A unique ID for each motor vehicle accident
County	Actual county name
City	Actual city name
Weekday	1 - Monday 2 - Tuesday 3 - Wednesday 4 - Thursday 5 - Friday 6 - Saturdav 7 - Sunday
Severity	1 - Fatal or severe injury 2 - Others
ViolCat	01 - Driving or Bicycling Under the Influence of Drug 02 - Unsafe Speed 03 - Following Too Closely 04 - Improper Passing 05 - Unsafe Lane Change 06 - Improper Turning 07 - Automobile Right of Way 08 - Pedestrian Right of Way 09 - Pedestrian Violation 10 - Traffic Signals and Signs

	11 - Fell Asleep
ClearWeather	1 - Clear weather 0 - Not clear weather
Month	1 - January 2 - February 3 - March 4 - April 5 - May 6 - June 7 - July 8 - August 9 - September 10 - October 11 - November 12 - December
CrashType	A - Head-On B - Sideswipe C - Rear End D - Broadside E - Hit Obiect F - Overturned G - Vehicle/Pedestrian
Daylight	1 - Daylight 0 - No Daylight
Highway	1 - Highway 0 - No Highway

4.3 SOURCES

The data as previously mentioned was extracted from the Statewide Integrated Traffic Records System, which collects and processes collision data from across California. The Internet SWITRS program uses this database to allow California Highway Patrol (CHP) personnel, members of its Allied Agencies, researchers, and members of the public to obtain various sorts of statistics information in electronic format. The program allows the user to create custom reports based on several categories such as, but not limited to, locations, dates, and collision kinds.

4.4 QUALITY

The data is filtered for the state of california and thus does not contain any null values. Further all the feature columns are categorical so dummy variables are required to convert them into the numerical format though in this case it does not look like a requirement.

DATA PREPARATION

The process of data preparation is foundational to the analytical workflow. Initially, the script initiates by importing essential libraries and loading data from an Excel file ('SF.xlsx'). The dataset undergoes thorough exploration via visualisation techniques, providing insights into crucial attributes such as crash types, severity distributions, and interrelationships between variables like ViolCat and Severity. Following this exploratory phase, data preprocessing commences. Key steps include the transformation of weekdays into a binary format and filtering the dataset based on specific crash types, thereby refining the dataset for subsequent analysis. Subsequently, the dataset is partitioned into distinct subsets for training and testing, ensuring that model performance can be accurately assessed. The identification and selection of relevant features for modelling, alongside partitioning the data into training and validation sets, set the groundwork for subsequent stages of model development and evaluation. These preparatory measures establish a robust and structured dataset, laying the groundwork for subsequent analytical processes.

5.1 DATA SELECTION

In this context, the data selection process involves a deliberate focus on specific attributes within the California dataset ('SF.xlsx'). Filtering the dataset to include particular crash types ('B', 'C', 'D', 'E') enables a concentrated analysis on incidents of higher severity or specific classifications. This selective approach aims to narrow the scope of analysis to relevant and impactful data points, facilitating a more targeted exploration of crucial patterns and relationships within the California dataset. Furthermore, the dataset's representation of California incidents allows for a focused examination of localised factors that might influence crash severity, providing insights into trends and contributing variables specific to this region. This concentration on the California dataset ensures that the subsequent analysis and modelling are tailored to the characteristics and patterns inherent to this specific geographic area, potentially yielding insights that are particularly pertinent to the state's traffic safety landscape.

5.2 DATA CLEANING

The data cleaning process includes identifying rows which may have any null values and using techniques such as imputation and omission to deal with null values. In this context, the dataset did not contain any null values and the overall data was relatively clean with the exception of a few repeating rows which were removed.

5.3 PREPARE DATA

To prepare the data, we apply transformation on the categorical predictor variables such as Weekday where we reduce it from an ordinal scale to dummy variables.

0: 1(Monday) to 5(Tuesday)

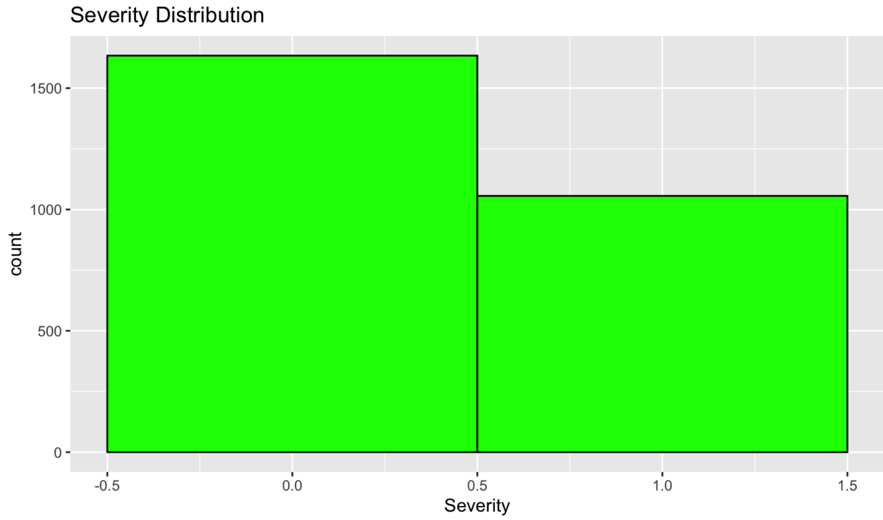
1: 6 (Saturday)to 7(Sunday)

The other predictor variables do not require any changes and hence can be plugged into the models directly.

6 MODELLING

6.1 DATA DESCRIPTION

The California state data consists of the following counties: San Mateo, SanFrancisco, Alameda, Santa Clara, Solano, Sonoma and Amador. The data contains 9 columns and 2690 Rows. The Features include Weekday, ViolCat, Month, CrashType, ClearWeather, Highway, Daylight. The target class is Severity.



The following plot gives us a distribution of the target class in the dataset. The target class contains 1633 “0” class members and 1057 “1” class members respectively. Since the datapoint distribution is not imbalanced, we don't require any sampling algorithms to artificially adjust the distribution.

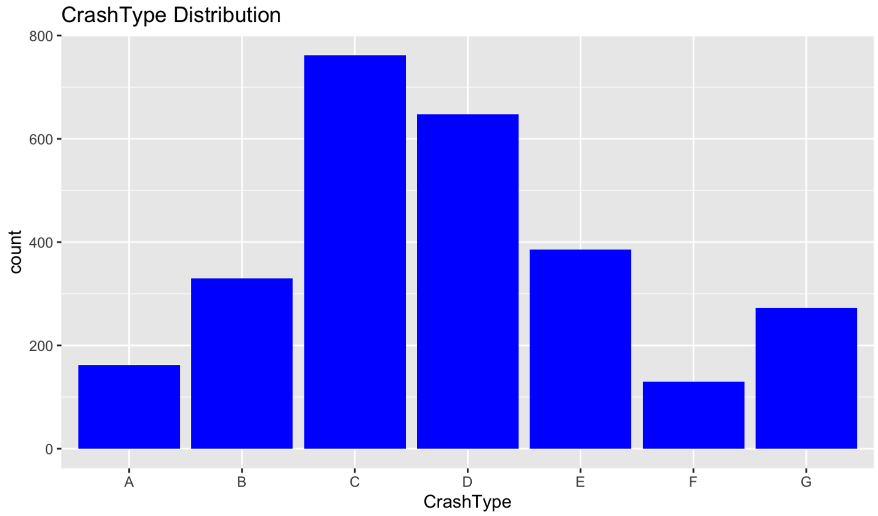
We can do a further analysis on the data including the importance of features and how significant they will be for the target class. For this we will use R.

```
Boruta performed 99 iterations in 12.0755 secs.
 5 attributes confirmed important: CrashType, Daylight, Highway, Month,
ViolCat;
 1 attributes confirmed unimportant: Weekday;
 1 tentative attributes left: ClearWeather;
```

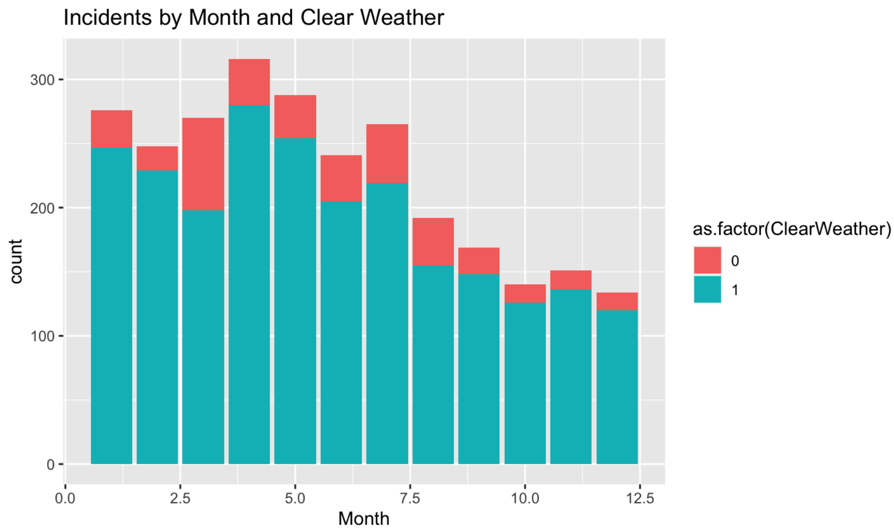
On further analysis we can bring out the exact percentage of significance of each important feature.

```
> importances <- attStats(rough_fix_mod)
> importances <- importances[importances$decision != "Rejected", c("meanImp", "decision")]
> importances[order(-importances$meanImp), ]
           meanImp  decision
CrashType 71.18818 Confirmed
Month     62.74755 Confirmed
ViolCat   41.24359 Confirmed
Highway   37.54696 Confirmed
Daylight   5.99167 Confirmed
```

The analysis indicates that CrashType is a crucial predictor variable. Thus a distribution analysis can be performed.



From the bar plot, we can tell category “C” is the most common CrashType and category “F” is the least common.



From the above plots, we get a clearer vision of the significance of the predictor variables in use, especially that of the Month which we could use to come up with conclusions as to why certain months have more accidents compared to others.

6.2 MODEL SELECTION

The task at hand is that of classification so we will only be looking at supervised models that can help with classification. When choosing the right model we usually look at :

Size and Complexity of Data: If the dataset is large and complex, models like Random Forests or GBMs might perform better. For smaller datasets, simpler models like decision trees or KNN might suffice.

Interpretability: If you need to explain the model's decisions, simpler models like decision trees or logistic regression are more interpretable.

Performance Metrics: Consider what metrics are important for your task (accuracy, precision, recall, etc.) and choose a model that tends to perform well on those metrics.

We will be using Classification Trees (Decision trees for classification tasks) and logistic regression for our task.

6.3 JUSTIFYING THE SELECTED MODELS:

In the context of predicting crash severity using the provided dataset, the selection of Decision Trees and Logistic Regression is strategically advantageous. Decision Trees are particularly effective due to their ability to handle both categorical and numerical data, as seen in the varied features of the dataset, such as 'ViolCat' and 'Month'. Their interpretability is a key strength, offering clear, visual decision-making paths, crucial in contexts where explaining model decisions is important. Moreover, they adeptly accommodate both binary and multi-class targets, making them versatile for different classification scenarios, and are well-suited to datasets with non-linear feature-target relationships.

On the other hand, Logistic Regression is an excellent choice for binary classification tasks, assuming the 'Severity' target variable is binary. It stands out for its computational efficiency, a significant advantage for moderate-sized datasets. The model's ability to provide probabilistic outputs adds a layer of depth to the prediction, making it not just a classifier but also a tool for assessing prediction confidence. Furthermore, as a baseline model, Logistic Regression sets a performance benchmark, allowing for an effective comparison with more complex models. The combination of these two models offers a balanced approach, leveraging the strengths of both to provide a comprehensive understanding of the factors influencing crash severity.

6.4 MODEL DEVELOPMENT AND OUTPUT:

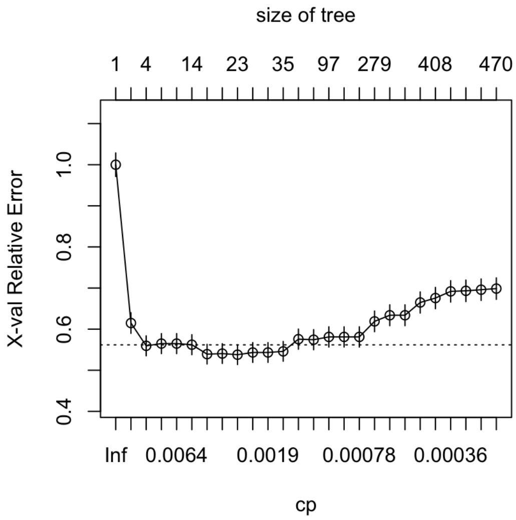
Decision Trees:

For the Decision trees there are three key components we need to look out for while developing the model.

1. Choosing the Right Algorithm: Select an appropriate Decision Tree algorithm (like CART . The choice depends on how the algorithm handles different data types and splits.In this case we will be using a Classification tree since our target variable is a binary value.

2. Hyperparameter Tuning: This is crucial for optimising the Decision Tree. Grid Search or Random Search can be used to find the optimal values for parameters like max_depth, min_samples_leaf, and criterion. We will primarily use min_split and min_bucket.(both set at 2 and 1 respectively.)
3. Pruning: To avoid overfitting, prune the tree by setting a maximum depth or minimum number of samples per leaf. Once we construct our full tree, the following model summary can be obtained.

	CP	nsplit	rel error	xerror	xstd
1	0.38513514	0	1.00000	1.00000	0.028641
2	0.02702703	1	0.61486	0.61486	0.025102
3	0.00810811	3	0.56081	0.55946	0.024286
4	0.00675676	9	0.51216	0.56486	0.024370
5	0.00608108	11	0.49865	0.56486	0.024370
6	0.00472973	13	0.48649	0.56216	0.024328
7	0.00405405	15	0.47703	0.53919	0.023963
8	0.00270270	19	0.46081	0.54054	0.023985
9	0.00236486	22	0.45270	0.53784	0.023941
10	0.00202703	26	0.44324	0.54324	0.024029
11	0.00180180	28	0.43919	0.54324	0.024029
12	0.00135135	34	0.42568	0.54595	0.024072
13	0.00108108	67	0.37973	0.57568	0.024535
14	0.00101351	83	0.36216	0.57432	0.024514
15	0.00096525	96	0.34865	0.58108	0.024616
16	0.00090090	126	0.30676	0.58108	0.024616
17	0.00067568	149	0.28514	0.58108	0.024616
18	0.00054054	278	0.19730	0.61892	0.025158
19	0.00050676	297	0.18514	0.63378	0.025360
20	0.00049787	305	0.18108	0.63378	0.025360
21	0.00045045	324	0.17162	0.66486	0.025763
22	0.00038610	407	0.13108	0.67568	0.025896
23	0.00033784	421	0.12568	0.69189	0.026091
24	0.00030030	452	0.11351	0.69324	0.026107
25	0.00022523	461	0.11081	0.69595	0.026139
26	0.00000000	469	0.10811	0.69865	0.026171



We must find the tree to the left of the one with minimum error whose cp value lies within the error bar of one with minimum error. In this case the ideal cp value for the pruned tree is **0.00405405.**

Logistic Regression:

1. Regularisation Technique: We traditionally switch between different regularisation techniques to prevent overfitting choosing between L1, L2, or Elastic Net regularisation .
2. Solver Selection: Different solvers can be used depending on the size of the dataset and whether or not the problem is binary or multi-class.
3. Class Weight Balancing: If the dataset is imbalanced, adjust the class weights to prevent bias towards the majority class.

In this case we will stick with a binary classification without any class weight balancing since the data is not imbalanced. Keeping the regularisation and solver at default we obtain the following confusion matrix for the model given the cutoff is 0.5.

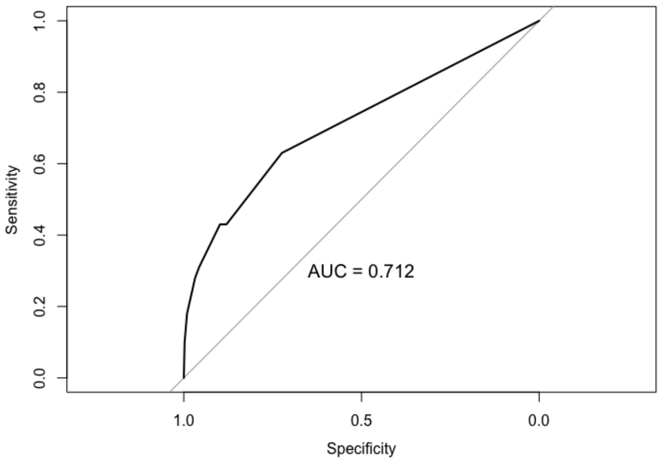
We apply 10 fold cross validation to the logistic regression model which will help with a more accurate measure of how well the Logistic Regression model is likely to perform on unseen data. By training and testing the model on different subsets of the dataset, it gives a better indication of the model's ability to generalise beyond the specific data it was trained on. It also helps in avoiding overfitting of the model.

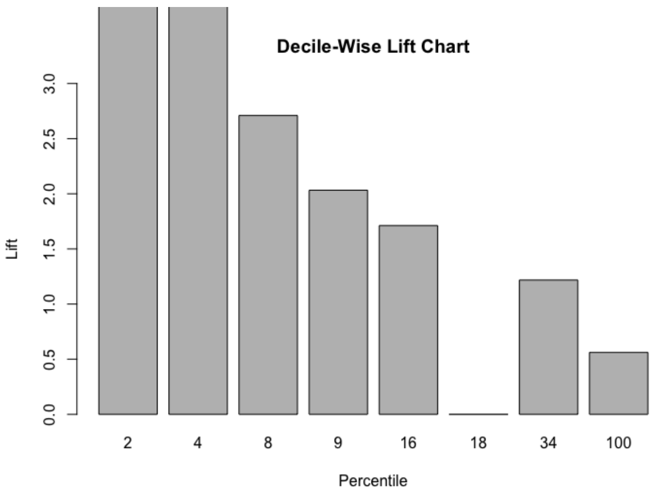
7 MODEL EVALUATION

The dataset has been partitioned into distinct subsets for training and testing purposes, adhering to a division of 70% for training, comprising 1538 instances, and 30% for testing, encompassing 659 instances. This strategic separation allows for robust evaluation and validation of the model's performance.

Decision Trees

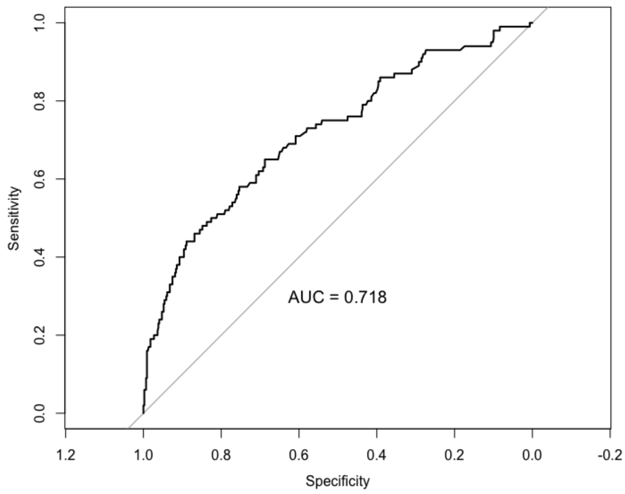
```
               Accuracy : 0.8413
                 95% CI : (0.8078, 0.8711)
    No Information Rate : 0.8155
    P-Value [Acc > NIR] : 0.06544

                  Kappa : 0.3202

 Mcnemar's Test P-Value : 7.923e-10

            Sensitivity : 0.28000
            Specificity : 0.96833
         Pos Pred Value : 0.66667
         Neg Pred Value : 0.85600
             Prevalence : 0.18450
         Detection Rate : 0.05166
   Detection Prevalence : 0.07749
      Balanced Accuracy : 0.62416

       'Positive' Class : 1
```





Logistic Regression

```
               Accuracy : 0.8247
                 95% CI : (0.79, 0.8558)
    No Information Rate : 0.8155
    P-Value [Acc > NIR] : 0.3123

                  Kappa : 0.2219

 Mcnemar's Test P-Value : 5.159e-11

            Sensitivity : 0.20000
            Specificity : 0.96606
         Pos Pred Value : 0.57143
         Neg Pred Value : 0.84221
             Prevalence : 0.18450
         Detection Rate : 0.03690
   Detection Prevalence : 0.06458
      Balanced Accuracy : 0.58303

       'Positive' Class : 1
```



8 DISCUSSION

In this section, we delve into the core questions shaping our decision-making process. Firstly, the model suggests that our optimal decision should be based on comprehensive data analysis and predictive insights. This recommendation stems from the model's ability to process vast amounts of information, leading to more informed and potentially accurate conclusions. However, the model does have its limitations, notably in its reliance on historical data and assumptions, which might not fully encapsulate future complexities or unforeseen variables. Furthermore, cognitive biases such as confirmation bias or availability heuristic could influence the decision-making process. Implementing decision support mechanisms can mitigate these biases by providing alternative perspectives and challenging preconceived notions. To enhance decision support, focusing on improving real-time data integration, refining predictive algorithms for better adaptability, and incorporating ethical considerations could bolster the model's effectiveness in enabling optimal decision-making.

8.1 DECISION/RECOMMENDATION

Both models, logistic regression and decision tree, exhibit varying strengths in predicting accident severity. While the logistic regression model shows higher specificity (0.96606) compared to the decision tree (0.96833), the decision tree performs slightly better in sensitivity (0.28000) than the logistic regression model (0.20000). However, when considering the Area Under the Curve (AUC), logistic regression edges slightly ahead with a value of 0.718 compared to the decision tree's AUC of 0.712. In the context of predicting severity in traffic accidents, I would lean toward the logistic regression model due to its higher specificity, which indicates its ability to accurately identify non-severe accidents. This aspect might be crucial in allocating emergency resources more efficiently by reducing false alarms for less severe incidents.

8.2 LIMITATIONS

Both logistic regression and decision tree models have their limitations. Logistic regression assumes linearity between the independent variables and the log odds of the dependent variable, potentially leading to underperformance if the relationship is more complex. It might struggle with capturing intricate nonlinear relationships within the data. Additionally, logistic regression assumes independence of observations, which might not hold in scenarios like traffic accidents, where spatial or temporal dependencies exist between incidents.

On the other hand, decision trees tend to be prone to overfitting, especially when the tree becomes too complex or when the dataset is noisy. They can create overly detailed structures that perfectly fit the training data but fail to generalise well to new, unseen data. Decision trees are also sensitive to small variations in the data, leading to different trees being generated for similar datasets.

Both models might also face challenges with missing or erroneous data, which could impact their predictive accuracy. Lastly, interpretability might be a limitation for complex decision tree models, as understanding the decision-making process can be challenging when the tree becomes too deep or intricate.

Another notable constraint is their dependence on the quality and quantity of available data. In the context of predicting accident severity, these models rely heavily on the dataset used for training. One specific limitation could be the scarcity of positive instances representing severe accidents within the dataset. Insufficient representation of such cases might lead to biases and reduce the model's ability to accurately predict severe accidents. Additionally, imbalanced datasets with a majority of non-severe accidents could skew the model's predictions toward the dominant class, affecting the overall predictive performance

8.3 BIASES

In the realm of decision-making regarding predicting severity in traffic accidents, several cognitive biases might come into play. The availability heuristic could be influential, where decision-makers heavily weigh recent or vivid examples more than other information. For instance, if there were recent high-profile severe accidents, decision-makers might overestimate the likelihood of such incidents occurring again. The confirmation bias could also impact decisions, leading individuals to seek or interpret information in a way that confirms their preexisting beliefs or hypotheses about accident severity.

Decision support tools mitigate these biases by providing structured, objective analyses. Algorithms used in decision support systems rely on data and statistical methods, reducing the influence of cognitive biases. These systems offer a broader perspective by considering a vast amount of historical data, potentially recognizing patterns and trends that human decision-makers might overlook due to biases. Additionally, decision support tools can facilitate scenario analysis, allowing decision-makers to explore various outcomes based on different parameters without the influence of biases. By relying on data-driven insights and probabilistic models, decision support systems counteract the impact of cognitive biases, enabling more informed and objective decision-making in complex scenarios like predicting accident severity.

8.4 ENHANCEMENTS

To bolster decision support for predicting accident severity, several enhancements could significantly improve the process. Firstly, incorporating real-time data integration would be invaluable. By tapping into live traffic feeds, weather updates, and other dynamic variables, the model could adapt and make predictions based on the most current conditions, enhancing accuracy and responsiveness.

Secondly, feature engineering is crucial. Identifying and incorporating more relevant features that influence accident severity, such as road conditions, visibility, time of day, and driver behaviour, would enrich the model's predictive capabilities. Moreover, integrating geospatial analysis could provide location-specific insights, considering factors unique to particular areas, like accident-prone zones or intersections with higher incidents.

Additionally, focusing on model interpretability is vital. Complex machine learning models might deliver accurate predictions but lack transparency. Implementing models that balance accuracy with interpretability would allow decision-makers to understand the reasoning behind predictions, fostering trust and enabling them to act upon the model's insights more confidently.

Lastly, developing a user-friendly interface for decision-makers is key. Presenting the model's outputs in a clear, intuitive manner, perhaps through visualisation tools or interactive dashboards, would facilitate easier interpretation and utilisation of the model's insights in making informed decisions regarding traffic safety measures and resource allocation.