1/11/2024

Explainable AI (XAI) on Road Traffic Infringement Data

Assignment 4 Submission

DEPARTMENT OF INFORMATICS

Assignment Cover Page

Student Number(s)								Surname	Initials
1	9	0	1	5	7	2	2	Braga	CLB
2	1	5	6	2	8	3	2	Coetzee	A
2	1	4	7	5	6	5	3	Van der Merwe	ML
0	4	8	6	9	4	6	1	Wood	FS

Module Code	**INF :**	7	9	1
Assignment Number	4			
Mark Allocation				
Date of Submission	1 November 2024			
Name of Lecturer	DR. Mike Nkongolo			

The University of Pretoria commits itself to producing academic work of integrity. By signing this paper, I affirm that I am aware of and have read the Rules and Policies of the University, more specifically the Disciplinary Procedure and the Tests and Examinations Rules, which prohibit any unethical, dishonest, or improper conduct during tests, assignments, examinations and/or any other forms of assessment. I am aware that no student or any other person may assist or attempt to assist another student, or obtain help, or attempt to obtain help from another student or any other person during tests, assessments, assignments, examinations and/or any other forms of assessment.

Table of Contents

Contents

Table of Figures

1. Introduction

Within the current world in which we live in today, numerous road traffic violations can cause significant operational challenges as well as safety challenges. For this reason, it is important to be able to leverage a data-driven approach. By leveraging a data-driven approach allows for underlying factors within these traffic incidents to be understood. Based on the current assignment in which we were given, our focus is on terms of explainable artificial intelligence, and more so the XAI that is used to predict violations in traffic, whilst at the same time looking deeper into numerous insights that these predictions are impacted by. Our main objective is to build a decision tree and logistic regression to make predictive models to utilize several XAI techniques such as partial dependence plots, SHAP, as well as LIME to provide model interpretations. By analyzing this data, the following study can offer model decision making views, which in turn allow for numerous stakeholders to investigate influences of traffic incidents.

2. Literature Background

In the domain of Traffic Safety, the application of machine learning has grown. The importance of this being that the models are interpretable to various involved stakeholders. By looking at the traffic accidents incidents we can identify the challenges that the public face and can use predictive models in order to make sure that the impact is mitigated. (Adeliyi et al., 2023) highlights the importance of using tree-based models to analyze the severity of traffic accidents and further goes on to explain that models are more interpretable when they are understood. By making use of these models, they are able to reduce the complexities, however, at the same time of doing this they are able to keep their accuracies which in turn allows for stakeholders to trust the results.

Recent research has explained that machine learning interpretability is important. (Molnar, 2022) explains that AI systems should be transparent, whilst using XAI tools such as SHAP and LIME to make sure that the bridge between users and complex models is understood. This thus enhances the trust in AI driven tools and decisions.

(Doshi-Velez & Kim, 2017) argues that model predictive performance and interpretability has a tradeoff. He explains that by using decision trees (which are considered simple models) offer a more interpretable outcome, whilst neural networks (which are considered more complex models) offer greater accuracy, but require various rules and XAI to understand and explain it the outcomes. The argument that is posed above is important when deploying various solutions that use AI within real life scenarios.

(Guidotti et al., 2018) investigates numerous XAI techniques, whereby these explain that explanations within certain features will increases model acceptance and the trust of a user. In the study they highlight that visualizations offer a significant contribution when using predictions, thus allowing for stakeholders to have a greater understanding on the outcomes based on the models and data.

The LIME model which (Ribeiro et al., 2016) speaks about explains that this method allows for complex models and behaviors to be interpreted in an understanding manner. By doing so individuals are then able to understand insights thus allowing for Traffic Safety, for example, to be understood. This is especially important where specific incident understanding and cause is vital.

3. Exploratory Dataset Analysis

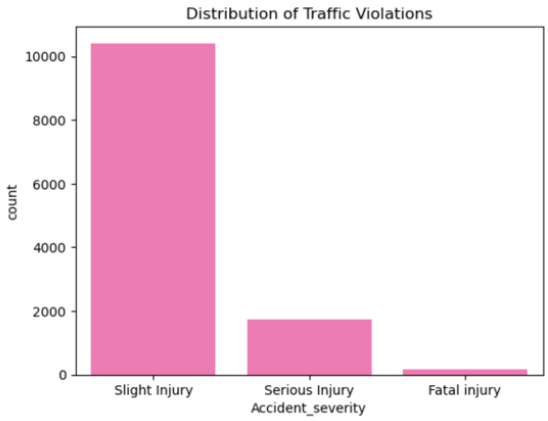

Figure 1 Distribution of Traffic Violations

A bar graph is used to visualize the frequency distribution of injuries resulting from traffic incidents. From Figure 1, we can see that 'slight' injuries is the most common followed by serious and fatal injuries.

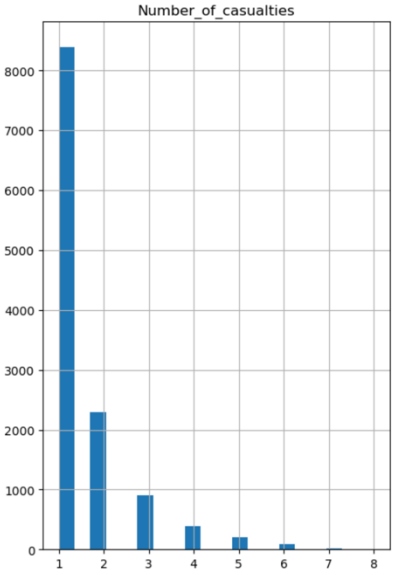

Figure 2 Number of Casualties

In Figure 2, a bar graph is used to display the number of casualties resulting from traffic accidents. From this we can see:

- Most accidents tend to have a low number of casualties amounting to roughly 1-2
- Accidents of high severity resulting in high casualties are a rare occurrence
- The graphs distribution is skewed to the right, suggesting that most accidents result in minor casualties.

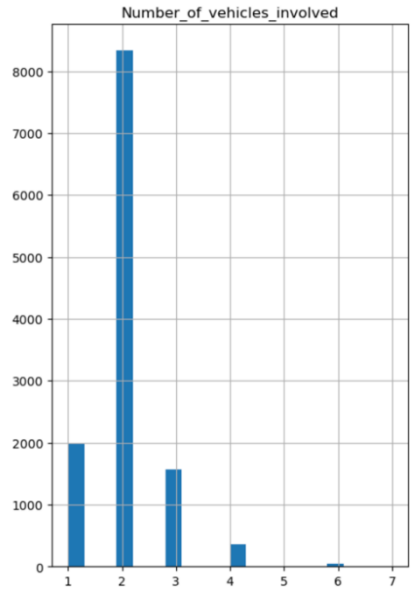

Figure 3 Number of Vehicles involved in Accidents

In Figure 3, a bar graph is used to display the number of vehicles involved in traffic accidents. From this we can see:

- The graph is distributed to the right, suggesting that most accidents involve a low number of vehicles
- Accidents involving a high number of vehicles is a rare occurrence

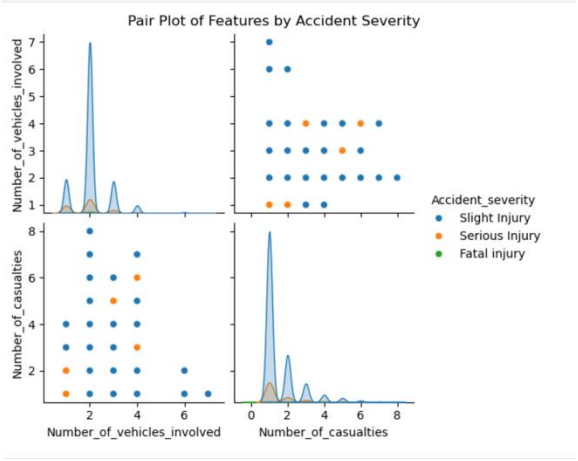

Figure 4 Accident Severity Distribution

A Pair plot is used to identify relationships among datasets. The relationship between the severity of accidents and the number of vehicles involved is displayed in Figure 4. From this we can see:

- Many recorded accidents involve a low number of cars, approximately 1-2 vehicles.
- The most common recorded injuries resulting from accidents are slight injuries.
- Serious and fatal injuries are less common compared to slight injuries, in terms of car accidents, this normally results from a low number of vehicles involved in an accident
- There is a relationship present between high numbers of casualties when more vehicles are involved in accidents.

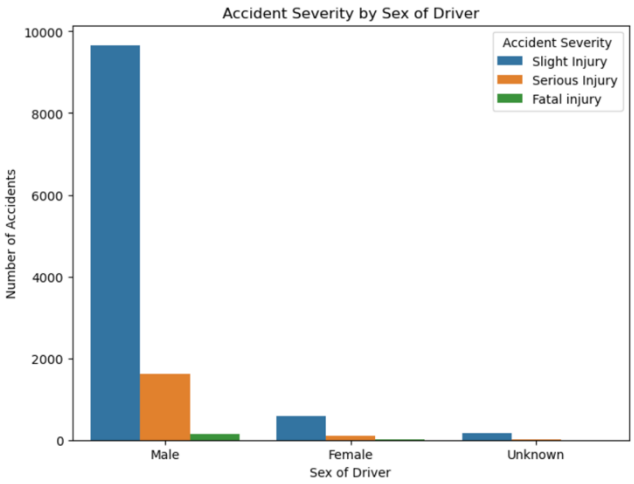

Figure 5 Accident Severity vs Gender

a bar graph is used to display the relationship between accidents and the sex of the driver involved. This relationship is displayed in Figure 5, along with the severity of injuries. This bar graph displays that most accidents involve a male driver, suggesting that accidents are common among male drivers. However, this insight could be a result of a bias representation in the dataset.

4. Methodology

4.1 Data Collection and Preprocessing

The dataset provided is populated with the following features:

- Infringement ID: a unique identification for each violation.
- Vehicle Type: the type of vehicle (e.g., car, truck, motorcycle).
- Violation Type: the type of traffic violation (e.g., speeding, parking, signal jump).
- Drivers Age: the age of the driver.
- Location: the violation incident location
- Time of Violation: the time that the violation is recorded.
- Weather Condition: the description oof the weather at the time of the violation.
- Fine Amount: the amount allocated as a penalty fee due to the violation

The pre-processed Excel dataset is loaded onto a jupyter notebook file for analysis and data cleaning (preprocessing). Within the preprocessing, a label encoder is used to transform categorical features into numerical features for accurate machine learning interpretations.

The numerical features are standardised using the *StandardScaler* function from the *sklearn.preprocessing* module to make sure that each category has a mean of 0 and standard deviation of 1. This is a critical step for machine learning model to perform accurately.

The preprocessing phase ensures that the data quality is clean and reliable. Below are the steps that were used in order to make sure that the data for the road traffic infringement was prepared:

- **Handling Missing Data:**

In order to maintain data integrity, any columns that were missing more than 50% of the values were dropped. Any roads that have missing values in critical fields such as time and accident severity were removed in order to avoid inconsistencies when the model was conducting training.

- **Imputation:**

Categorical variables were imported with values of their most frequent value, in other words the mode. Numerical variables were inputted with the median, and this was done in order to make sure that the data distortion was minimized.

- **Encoding Categorical Variables:**

in order to convert categorical data into numerical form label encoding was applied. This was done in order to make sure that it was fit for machine learning compatibilities.

- **Feature Scaling:**

StandardScaler was used in order to make sure that the numerical features were standardized, thus ensuring that the variables were able to equally contribute to the performance of the model.

- **Train-test Split:**

The data set that we were provided with was divided into testing and training sets using an 80/20 split. This is done to evaluate the performance of the models on unseen data.

4.2 Model Building:

The dataset is split into training and testing sets following an 80:20 split. This is done to ensure accurate model validation.

Two models are used to predict traffic violations. The following models were trained and compared:

- Decision Tree Model: a model that uses tree-based splitting for feature importance to identify non-linear relationships
- Logistic Regression Model: a linear model that makes predictions via planes through data

The models are evaluated through the following metrics, to determine the predictive performance of each model:

- Accuracy

- Precision
- Recall
- F1 Score
- Cohen's Kappa
- computational time

the Precision, Recall and F1 Score are metrics used to determine the performance of models against each class in terms of identifying true positives. Cohen's Kappa I a quantitative metric that determines the reliability between the predictions and classes.

4.3 Explainable AI (XAI):

INF:ollowing XAI techniques are used to explain the machine learning models more understandable:

- Local Interpretable Model-Agnostic Explanations (LIME): a technique used to explain each prediction individually
- SHapley Additive exPlanations (SHAP): this technique assigns a score of importance to features to provide local and global explanations
- Partial Dependence Plot (PDP): this technique explains how each feature affects a model's predictions and explains the global behaviour of the model

These techniques are used to gain deeper insights into the decisions made by each machine learning model.

4.4 Results Analysis:

A combination of visualisations are used to display the results in an insightful manner (Khan & Khan, 2011). This allows for the data to be accurately communicated in a meaningful manner (Franconeri et al., 2021).

Additionally, the following tools and Libraries are used in this study:

- Python Library: coding language used for data handling and data analysis

- Scikit-learn: used for the preprocessing and implementation of different machine learning
- XAI libraries: used to implement the XAI techniques
- Matplotlib with Seaborn: used to create different visualisations of the data.

5. Results

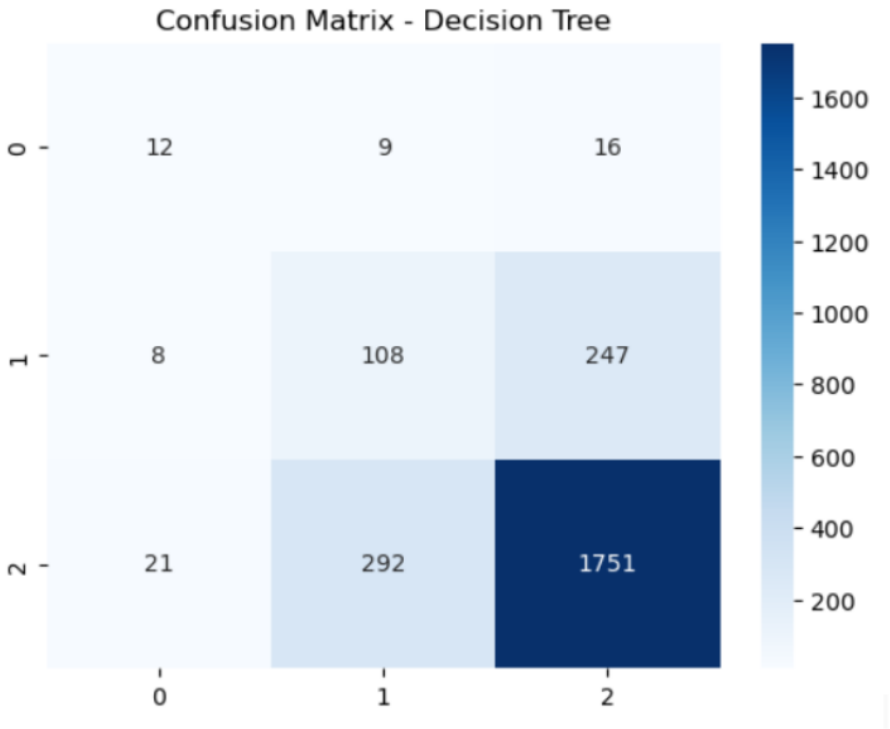

Figure 6: Confusion matric- Decision Tree

The figure above shows the confusion matrix for the decision tree model. The confusion matrix displays how well the model does in predicting class labels across the 3 classes.

- Class 0:
  - Correctly identified 12 instances
  - Misclassified 9 instances as class 1
  - Misclassified 16 instances as class 2

    - The model struggles with correctly identifying class 0 as there are more misclassifications than correct predictions
- Class 1:
    - Correctly identified 108 instances
    - Misclassified 8 instances as class 0
    - Misclassified 247 instances as class 2
    - The model struggles with differentiating between class 1 and class 2 as can be seen from the high amount of misclassifications
- Class 2:
    - Correctly identified 1751 instances
    - Misclassified 21 instances as class 0
    - Misclassified 292 instances as class 1
    - The model has the strongest performance as it correctly identified a high number of class 2 instances however there is also quite a high number of misclassifications as class 1/

The decision tree model performed best in correctly classifying class 2 instances. The model however struggled to classify class 0 and class 1 as many of the misclassifications fell into class 2.

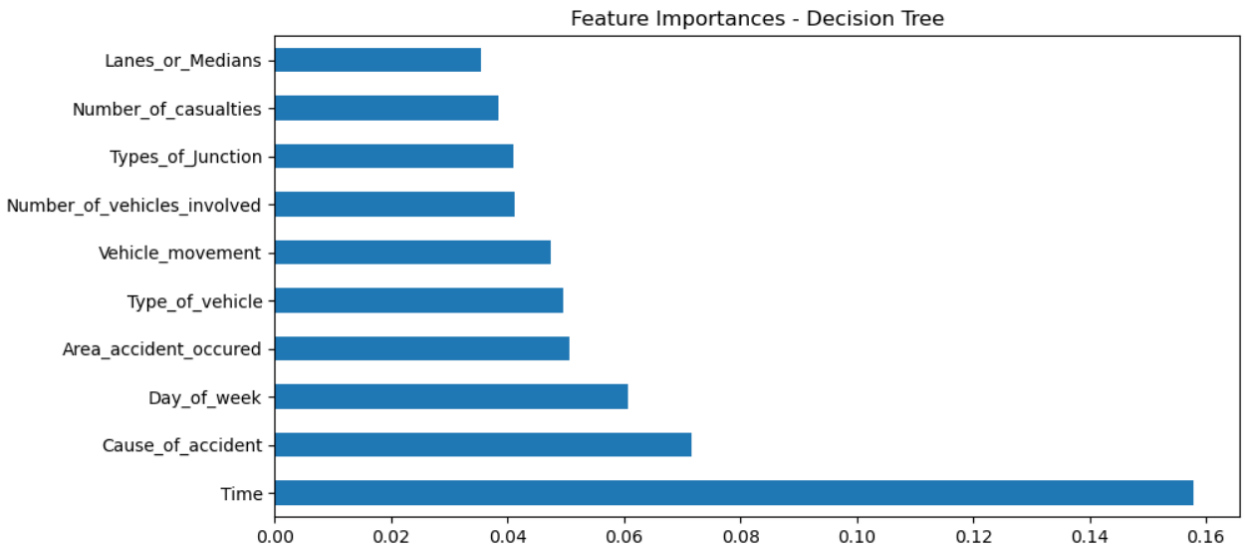

Figure 7: Feature importance - Decision Tree

The figure shows the decision tree model's feature importance, which is the relative importance of different features in making predictions. This can show the most influential variables and help understand the model's decisions.

- Time: The most influential feature of the decision tree model is Time with an importance rating of 0.16. This means that the time at which an accident occurs plays a big part in determining the severity of the accident.
- Cause of accident: The second most important feature is the cause of an accident with an importance score of 0.09.
- Day of Week: The feature with the third highest importance in the model is the day of week with an importance score of 0.06.
- Area accident occurred, type of vehicle and vehicle movement: These three features show similar importance values between 0.04 and 0.05 scores, indicating they have an impact but are not of the highest importance.
- Number of vehicles involved, types of junctions, number of causalities and lanes or medians: These features show a low level of importance ranging around 0.02 and 0.03, indicating these features contribute very little to the model predictions.

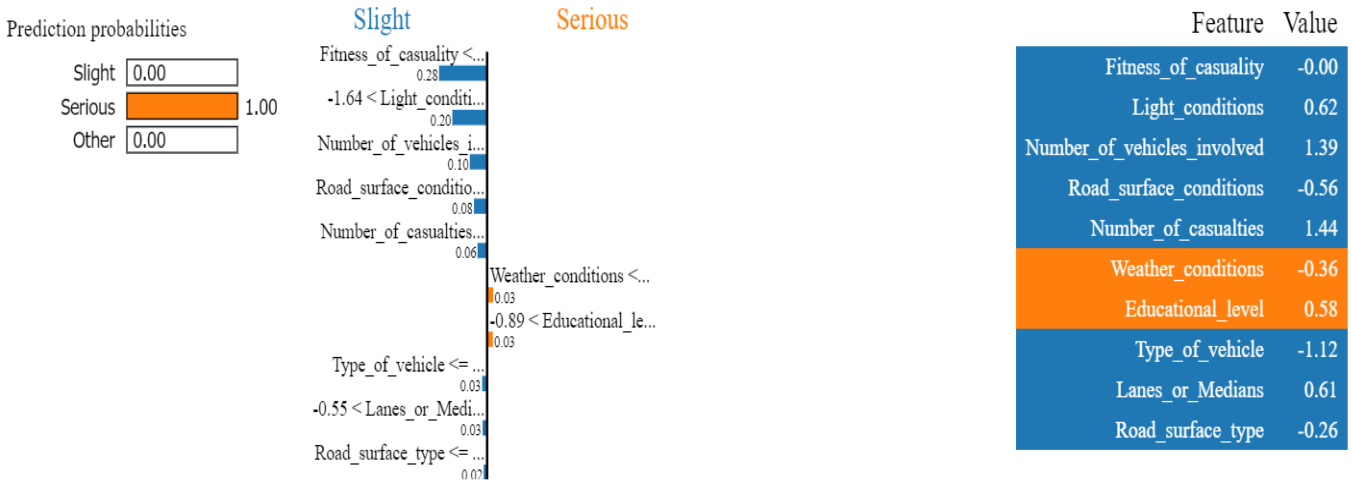

Figure 8: Prediction probabilities and features

The figure shows the prediction probability breakdown for a certain instance where the model has classified the severity of the accident as “Serious” with a probability of 1.00 or 100%. The analysis shows how different features contribute to that prediction.

- Fitness of casualty and light conditions
    - These have minimal effects on the predictions, with values of 0.00 and 0.62 respectively.
- Number of vehicles involved and road surface conditions
    - The number of vehicles involved shows a positive contribution of 1.39 showing an increase in the number of cars involved increases the likelihood of a serious classification. Road surface conditions have a negative contribution of -0.56 showing that better road surface conditions decrease the likelihood of a serious classification.
- Number of casualties
    - The number of casualties has a positive contribution of 1.44 implying a higher number of casualties suggests a higher likelihood of a serious classification.
- Weather conditions and education level
    - Weather conditions have a negative contribution of -0.36 showing that good weather conditions decrease the likelihood of a serious classification. While education level has a positive influence with a score of 0.58.
- Type of vehicle and lanes or medians
    - Type of vehicle has a negative contribution of -1.12 suggesting specific types of vehicles are less likely to be classified as serious. Lanes or medians contribute positively with a value of 0.61 to show more complex lane structures could increase the possibility of a serious classification

The figure shows that a “Serious” accident is driven by factors such as the number of casualties, number of vehicles involved and lanes or medians. These factors push the model towards predicting a serious classification. However, factors such as weather conditions and type of vehicle reduce the likelihood of a serious classification.

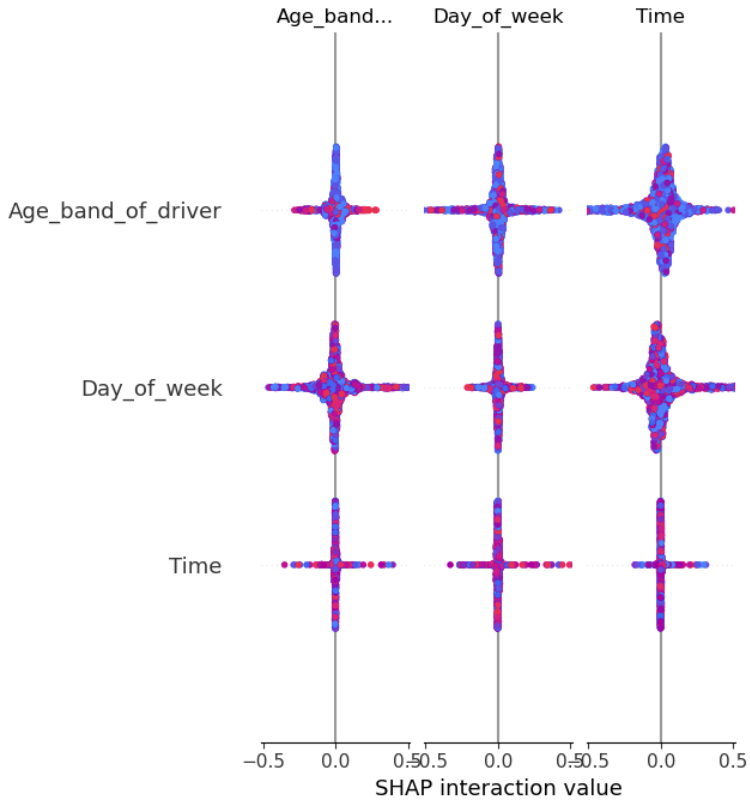

Figure 9: SHAP interaction value

The figure shows a SHAP interaction plot which shows the interaction effect between the three factors of age band of driver, day of the week and time. The colour ranging from blue to red represents the value of the features by the SHAP interaction values on the x-axis. This indicates how the combination of these features impacts the prediction.

- Age band of driver interaction
    - There is minimal interaction between the age band of driver and day of week and time as can be seen by the SHAP interaction value being close to 0. This suggests that these factors have a low impact on the model.
- Day of week interaction
    - Day of the week also shows a minimal interaction between the age band and driver and time as the values are concentrated around 0. This suggests the interaction has a low impact on the model.

- Time interaction
    - The time feature shows a stronger interaction with age band of driver and day of week as compared to the other factors. This can be seen with the SHAP interaction values being more spread out suggesting that time of the accident when combined with other features, specifically age band of driver and day of week have a bigger influence on the model.

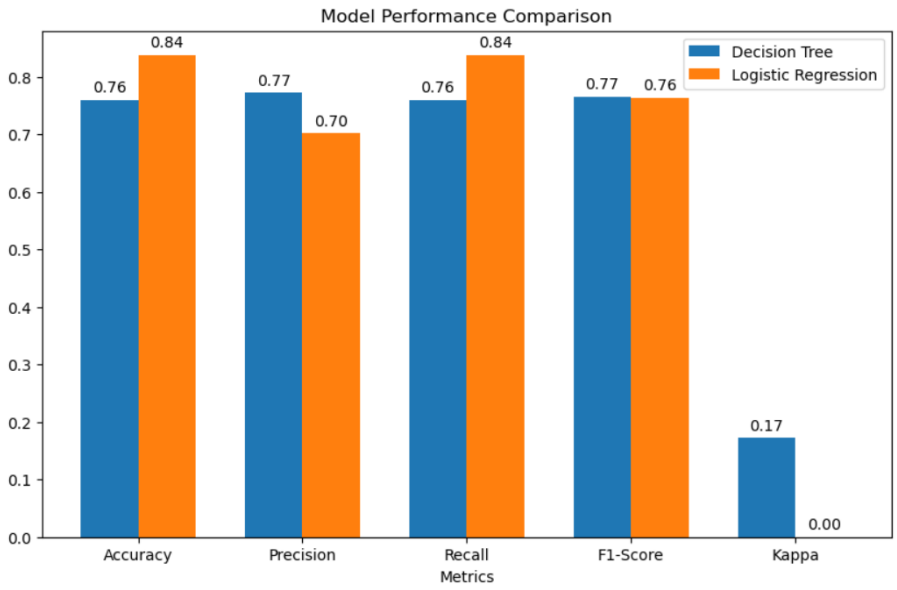

Figure 10: Model performance comparison

The figure above shows the difference in performance of the two models, Decision Tree and Logistic Regression. These two models were evaluated on the key metrics of Accuracy, Precision, Recall, F1-Score, and Cohen's Kappa.

- Accuracy: The logistic regression model has a higher accuracy score of 0.84 as compared to the accuracy score of the decision tree which is 0.76. This means that the logistic regression model has more correct predictions than the decision tree model.
- Precision: The decision tree model performs better than the logistic regression model for the precision score. The decision tree model has a score of 0.77 compared to 0.70

for logistic regression model. This means that the decision tree model minimizes the number of false positives and making sure that the positive predictions are correct.

- Recall: The logistic regression model performed better than the decision tree model for the recall score as the logistic regression has a score of 0.84 and the decision tree model has a score of 0.76. This means that the logistic regression is better at identifying actual positive cases and makes fewer false negative predictions.
- F1-score: The F1-score balances the precision and recall and is 0.77 for the decision tree and 0.76 for the logistic regression model. This means that there is a balanced performance for both models with only a small difference in which logistic regression is scored a bit lower.
- Cohen's Kappa: Cohen's Kappa measures the level of agreement between the predicted classifications and the actual values. The decision tree model scored 0.17 which shows there is some reliability in the model's predictions versus the actual outcomes, but the score is very low. The logistic regression model however shows a score of 0.0 which could either show that the performance is not any better than random chance or it could present a null value, indicating the model's predictive ability is not aligned with the actual values at all.

Overall, the logistic regression model shows better performance in accuracy and recall which is important for cases where identifying the true positive cases is the highest priority. However, the data tree model performs better in precision and Cohen's Kappa, showing that it might be better in cases where the quality of positive predictions is the priority.

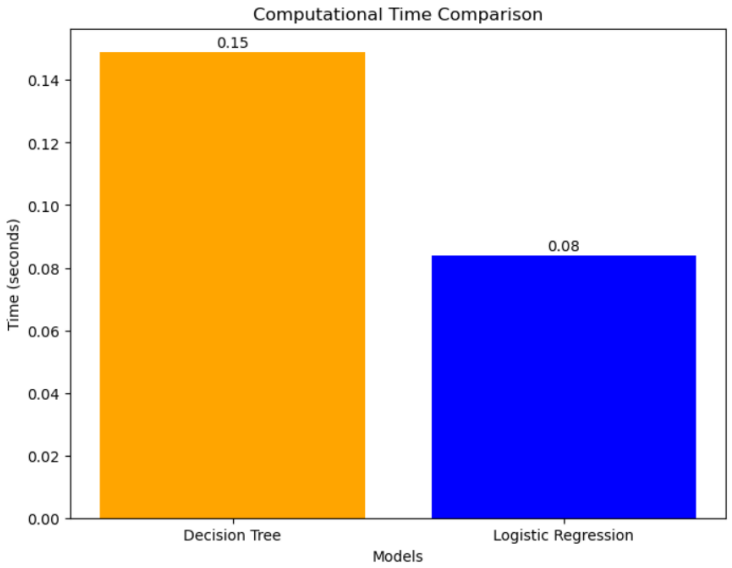

Figure 11: Computational time comparison

The computational time required for training the decision tree and logistic regression model was compared. The decision tree model required 0.15 seconds to train which was significantly longer than the logistic regression model which took 0.08 seconds to train. The difference in computational time could be due to the complexity of the decision tree model which involves recursively partitioning the feature space as compared to the logistic regression model which uses a fixed number of iterations to optimize its parameters. It is necessary to compare these models when deciding which one to use as the logistic regression model is better suited for computational efficiency however, the computational time and model performance should be taken into consideration.

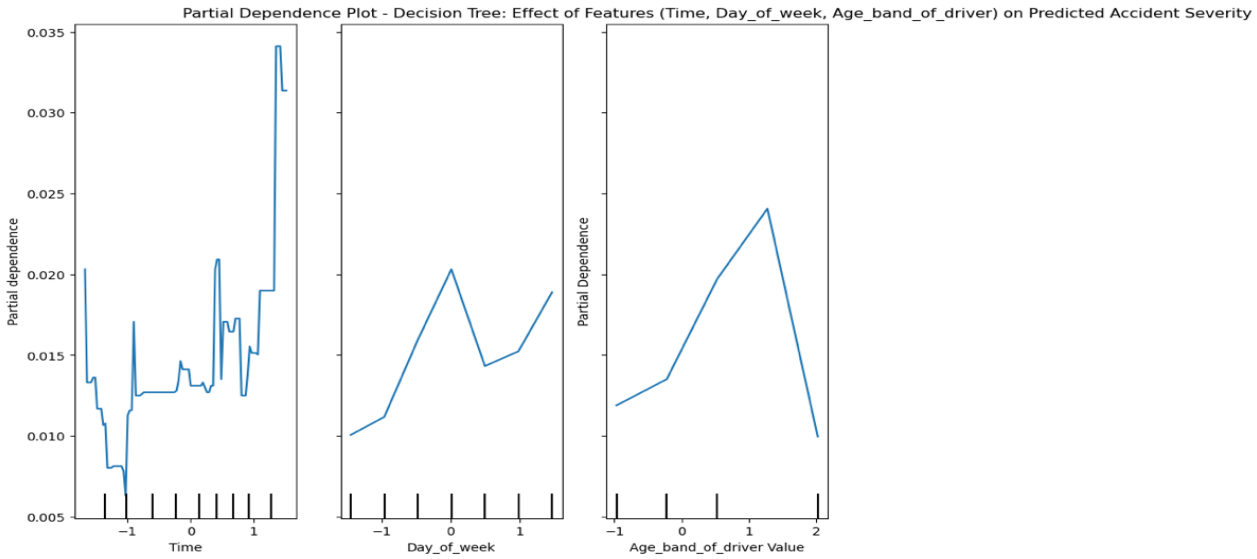

Figure 12: Partial dependence plot - Decision Tree

In the above diagram, a Partial Dependence Plot (PDP) is used to highlight the effect of features on the accident severity predictions for the Decision Tree model. The three metrics and their insights are as follows:

- Time:

The first section of the plot measures the partial dependence against the 'Time' feature. This highlights how accident severity changes over different time periods. As time increases, so does the partial dependence with an upward trend. This suggests that accidents tend to happen later in the day, along with increased levels of severity. The fluctuations could suggest that certain times are higher risk, potentially linking to high traffic times and late-night hours.

- Day_of_week:

The second section of the plot measures the partial dependence against the 'Day_of_week' feature. Here we see how the accident severity changes across different days of the week. The spike displayed suggests that certain days tend to have a higher accident severity level, potentially meaning that accidents tend to occur more on certain days of the week.

- Age_band_of_driver:

The last section of the plot measures the partial dependence against the 'Age_band_of_driver' feature. this displays the effect of the drivers age on accident severity predictions, where we see an increase in the partial dependence around specific age groups. This suggests that accidents are more severe within certain age groups, and the frequency of accidents occurring is also more common around certain age groups.

This PDP highlights the dependency of the 'Time,' 'Day_of_week,' and 'Age_band_of_driver' features on accident severity predictions resulting from the Decision Tree model. Each feature displays a relationship with accident severities, highlighting their importance in predictions.

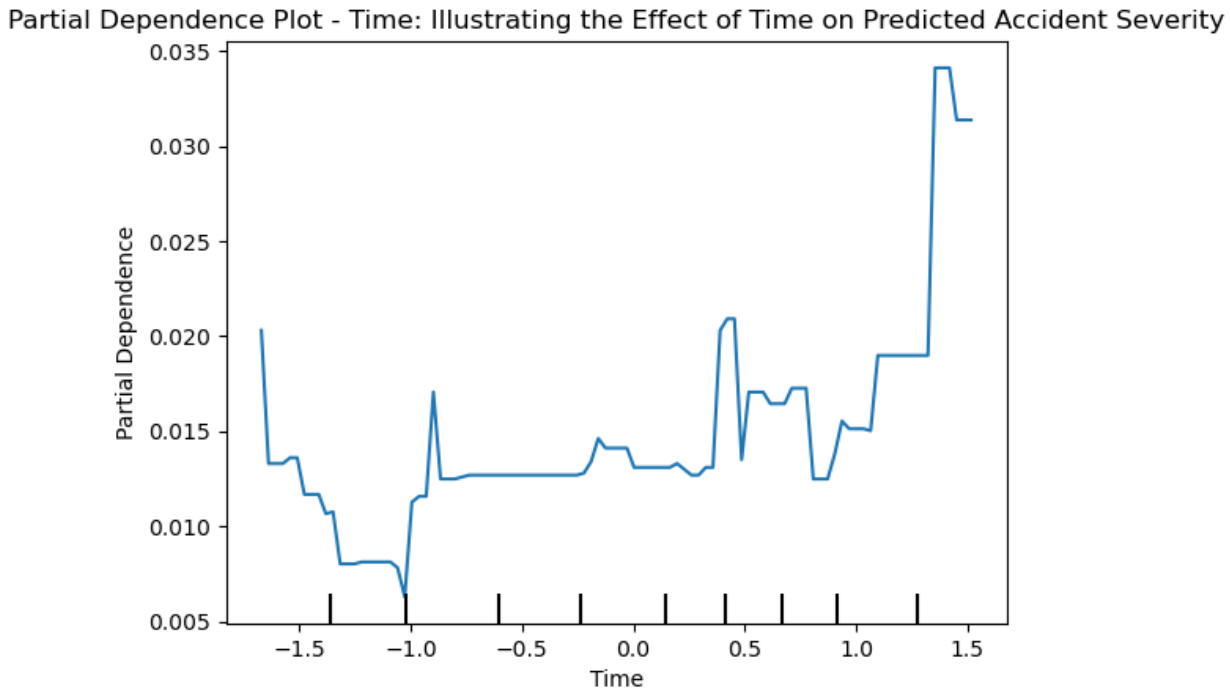

Figure 13: Partial Dependence plot: Time

The Partial Dependence Plot (PDP) shows the effect of the time variable on the Accident_severity variable. From the plot, the following can be seen:

- Partial dependence: The y-axis indicates the average change in the predicted accident severity. A higher partial dependence value shows that there is a higher chance of getting a more severe accident.
- Time: the x axis shows the normalized values for time. The time shows the different hours during the day when the accidents are recorded. Early hours represent hours early in the day, while positive hours are later in the day.

Once this is understood, the following was found:

Lower Accident Severity: Lower accident severity was found around -1 on the time scale and there is a dip in the partial dependence values. This shows that the accidents that occur around this time are less severe.

Fluctuations in Middle Values: Between -0.5 and 0.5, the values for the partial dependencies fluctuate but tend to stay low. This suggests that the accident severity levels during the day stay relatively stable as there is no significant increase or decrease.

Increase in Severity for Positive Values: The partial dependencies increase in values around 1 on the x axis and onwards with the max value peaking at 0.03. This shows that accidents that occur at these times are predicted to have higher severity on average.

From the graph, the accident severities follow a U-shape pattern. This is concluded as accident severity is lower around -1 and higher at both the beginning and later times of the day.

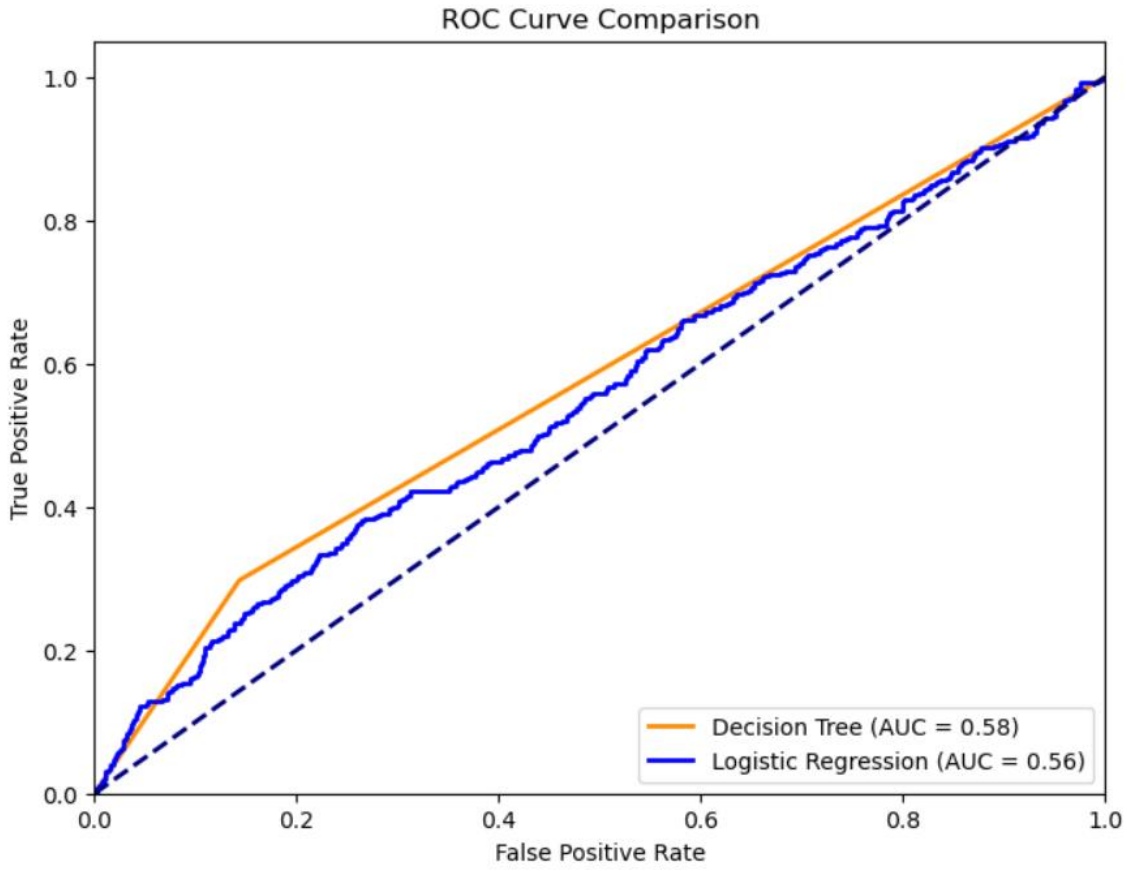

Figure 14: ROC Curve Comparison

The ROC curve is plotted to compare the decision tree model and the logistic regression model. The comparison is to test the difference in ability to distinguish between classes, the AUC (Area Under the ROC Curve) is calculated to show this.

- The AUC for the decision tree model is 0.58 which is slightly better (by 0.5) which shows the model has slightly better performance in correctly identifying between positive and negative classes.
- The AUC for the logistic regression model is 0.56 which suggests that the model might struggle a bit to distinguish between positive and negative classes as compared to the decision tree model.

Both the decision tree model and the logistic regression model have slightly higher ROC curves than the diagonal line (random chance) which could indicate that the models are not strong in being able to identify positive from negative classes.

```
                                Accident_severity_Serious Injury  \
Accident_severity_Slight Injury                         -0.950360
Number_of_vehicles_involved                             -0.085872
Age_band_of_driver_Unknown                              -0.033082
Day_of_week_Monday                                      -0.023003
Time_11:10:00                                           -0.026438
...                                                           ...
Time_22:04:00                                            0.054375
Time_20:55:00                                            0.057067
Time_07:49:00                                            0.058734
Time_23:36:00                                            0.062792
Accident_severity_Serious Injury                         1.000000

                                Accident_severity_Slight Injury  \
Accident_severity_Slight Injury                         1.000000
Number_of_vehicles_involved                             0.095386
Age_band_of_driver_Unknown                              0.041312
Day_of_week_Monday                                      0.028454
Time_11:10:00                                           0.027819
...                                                          ...
Time_22:04:00                                          -0.051676
Time_20:55:00                                          -0.053245
Time_07:49:00                                          -0.055818
Time_23:36:00                                          -0.059675
Accident_severity_Serious Injury                       -0.950360
...
Time_23:36:00                                -0.028166
```

Figure 15: Linear regression model results

A correlation matrix was performed to see which variables in the dataset had a strong Correlation. The focus was put on the following variables:

- Accident_severity_Serious Injury
- Accident_severity_Slight Injury
- Casualty_severity

The results were as follows:

1. Correlation between Accident_severity_Serious Injury and Accident_severity_Slight Injury

There was a strong negative correlation between these two variables with a score of -0.950360. This showed that the results were working as expected as these are two distinct variables that cannot be classified as both a serious and slight injury at the same time.

2. Correlation between Casualty_severity and accident severities

when looking at Accident severity with slight injury, there is a weak negative correlation that shows that these two variables do not strongly influence each other.

When focusing on Accident severity with serious injury, it can be seen that there is a weak positive correlation. This shows that there is a slight tendency for casual severities to increase when accidents are classified as a serious injury.

3. Number of vehicles involved

With both slight and serious injury, there is a weak positive and negative correlation. This shows that the number of vehicles involved in the accident has a small association with the severity of the accident.

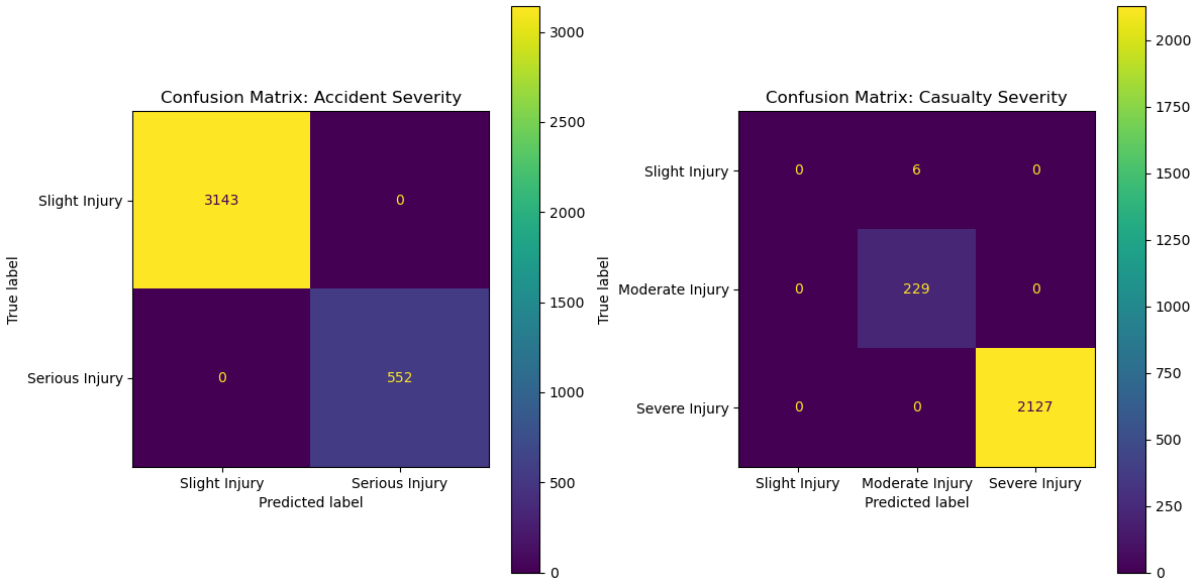

Figure 16: Confusion matrix focusing on Accident Severity and Causality Severity

The confusion matrix focuses on the variable Accident Severity where the following was investigated:

- True labels: Slight Injury and Serious Injury
- Predicted labels: Slight Injury and Serious Injury

The results were as follows:

- 3143 true positives: These values were included when the model could successfully predict slight injury.
- 552 true positives: These values were included when the model correctly predict serious injuries.
- 0 false positives or false negative: The model did not incorrectly predict accident severities.

The second confusion matrix focused on the variable Casual Severities that included the following labels:

- True labels: Slight Injury, Moderate Injury and Severe Injury
- Predicted labels: Slight Injury, Moderate Injury and Severe Injury

The results were as follows:

- 6 false negatives: The model incorrectly predicted 6 cases of slight injury as moderate injury.
- 229 true positives: The model was able to predict 229 results for moderate injury cases.
- 2127 true positives: The model was able to predict 2127 results for severe injury cases.

- 0 false positives or false negative: There were no misclassifications that were made between the different categories.

The causalities had some difficulty being classified as there were 6 instances where causalities were incorrectly classified. However, the overall performance of the model was good as it was able to classify most of the causalities.

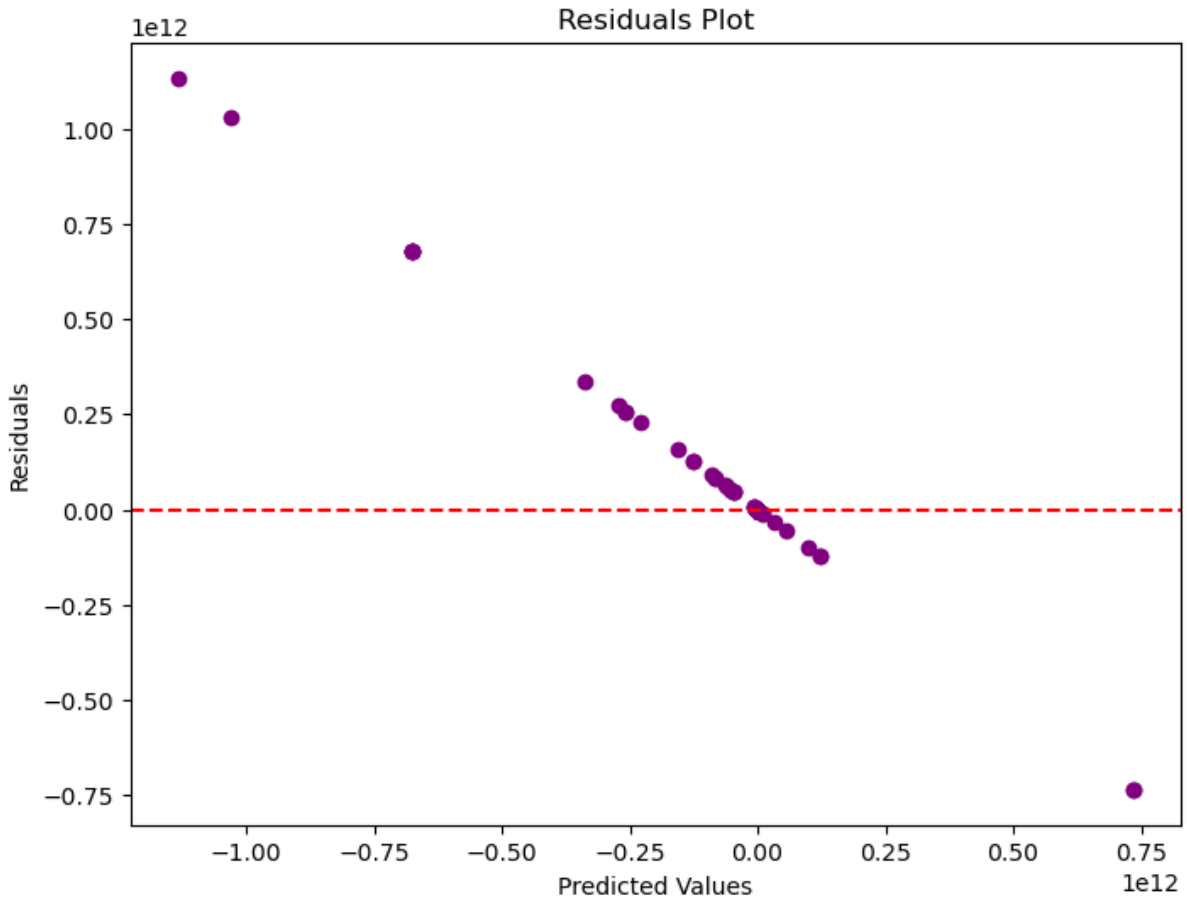

Figure 17: Residual plot results

The above graph is a residual plot that shows the difference between the predicted values and the actual values. This includes:

1. Residuals

The y axis represents the residuals, the closer the values tend towards 0, the better the model should be performing. The residuals should be scattered randomly around the x axis to ensure that the model is correctly performing. However, from the model we created, there are clear outliers, particularly around the x axis.

From the result, the following can be concluded:

1. Extreme outliers: The residual graph shows a series of extreme residuals (both negative and positive values). The values can be seen near -1.0 and 0.75. This shows that the model produced erroneous predictions for some of the data points.

Non-Random Pattern: The plot does not show the residuals randomly distributed around the horizontal zero line, but it also shows that there is a downward trend from left to right. This indicates that the model is consistently over or under predicting for different ranges of predicted values.

Linear Trend: The residual graph shows a clear linear trend that suggests that the model is missing some important relationships in the data. A linear regression model was performed on the data. From the results we can see that the data has a more complex and non-linear relationship.

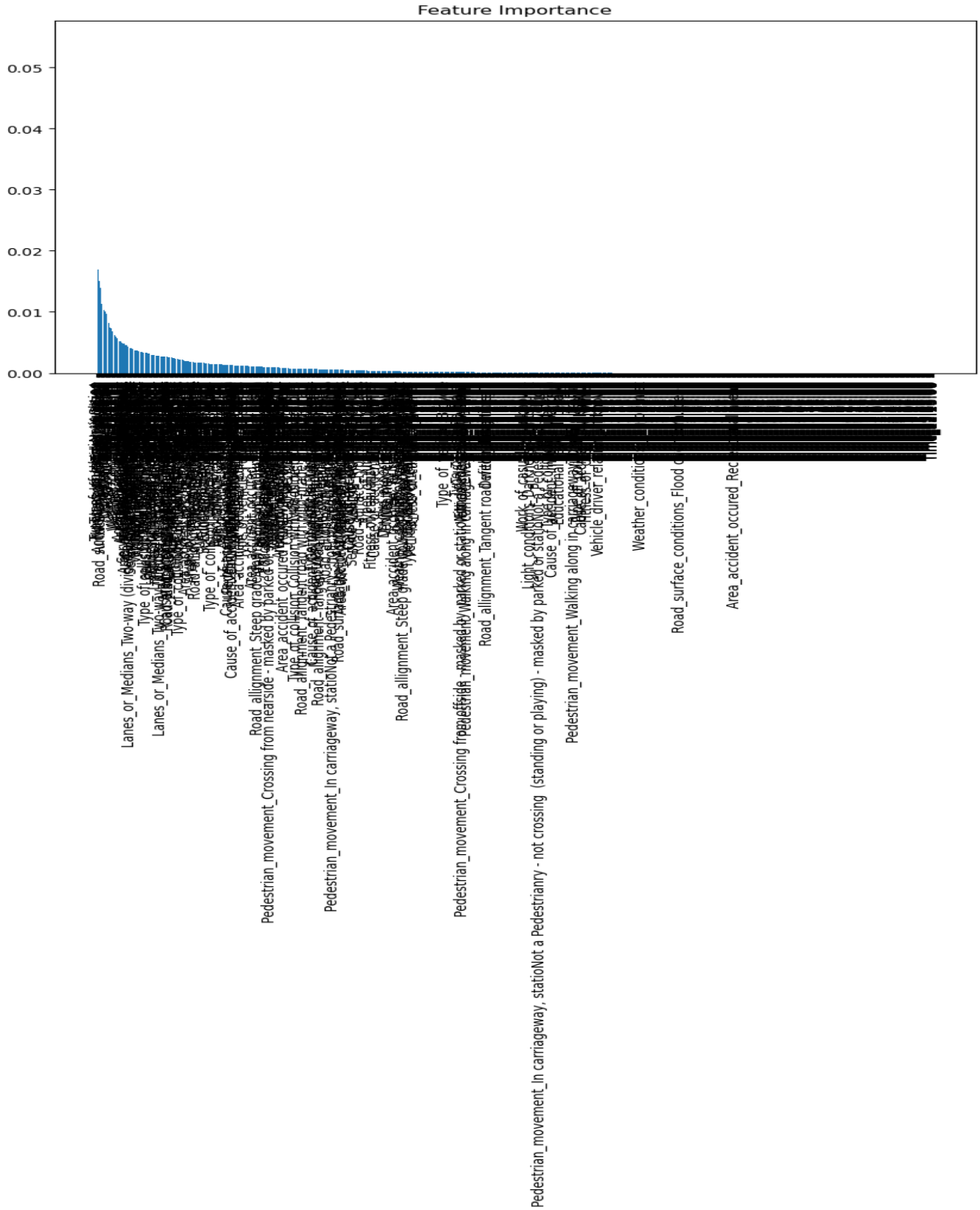

Figure 18: Feature importance plot

When looking at the above graph the following can be seen:

X-Axis (Feature Names): This axis contains the names of all the features used in the model.

Y-Axis (Feature Importance Scores): The Y-axis represents the importance scores of each feature. The higher the bar, the more significant that feature is in predicting the target variable.

Skewed Importance: A small subset of features have significantly higher importance scores compared to the rest. This suggests that only a few features are critical to the model's performance, while most features contribute very little.

Long Tail: The plot shows a "long tail" of features with near-zero importance. These features are not contributing much to the prediction and could potentially be removed to reduce the dimensionality of the model and improve performance.

6. Use Cases

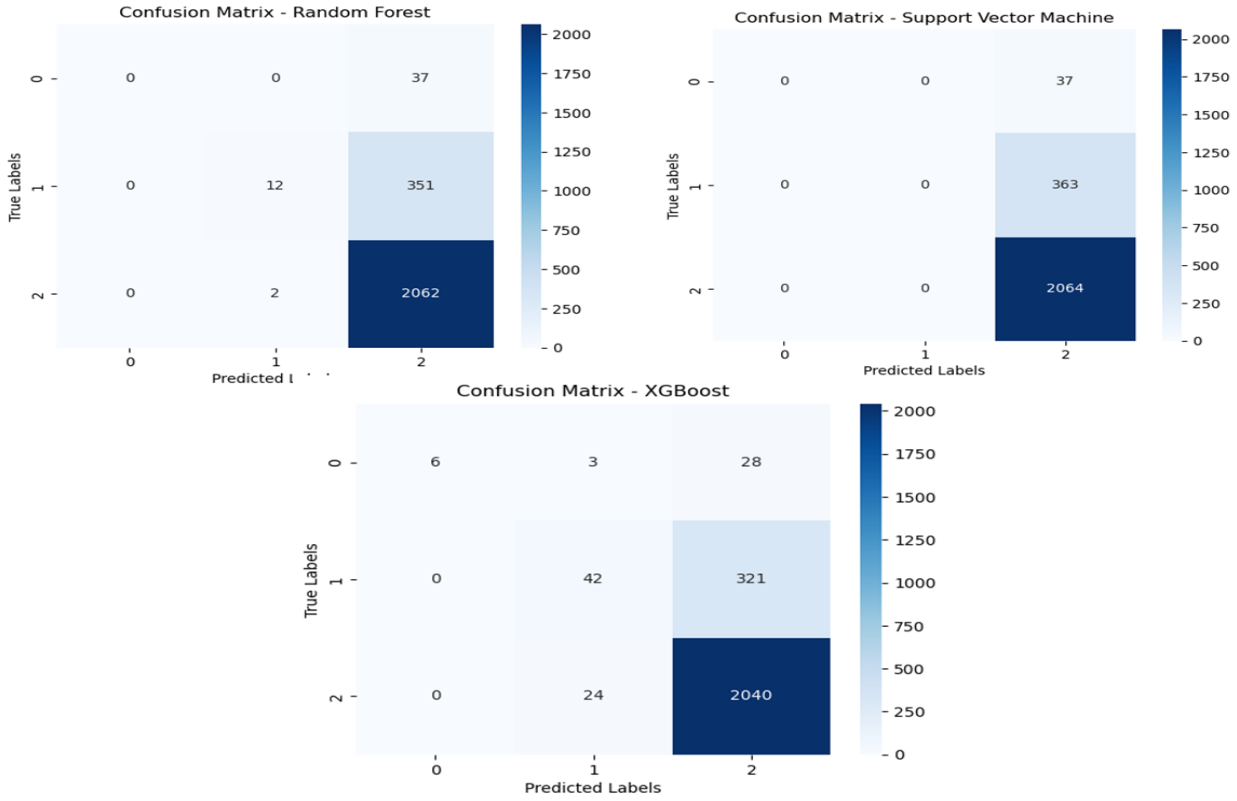

Figure 19: Comparison of confusion matrix

The current use case displays the confusion matrices for the random forest, support vector machine and XGBoost models and allows for a representation of predicted variables versus the true labels. Each of these models can highlight the points where the models are predicting correctly as well as where they are making mistakes. The random forest model has been able to correctly identify most Class 2 accidents – of which this is 2062 instances. However, this model struggled with classes 0 and 1, and this can be seen by the low values for each category which emphasizes the poor classification in each of those classes.

The XGBoost model can predict better for class 0 and 1 in comparison to the random forest model but has challenges in differentiating between clause 1 and 2 which may be an indication of poor decision boundaries.

The support vector machine model displays a higher accuracy in class 2, however, it struggles in classification of classes 0 and 1. This is evident as each of these classes display zero correctly classified predictions and indicates it only predicts for Class 2.

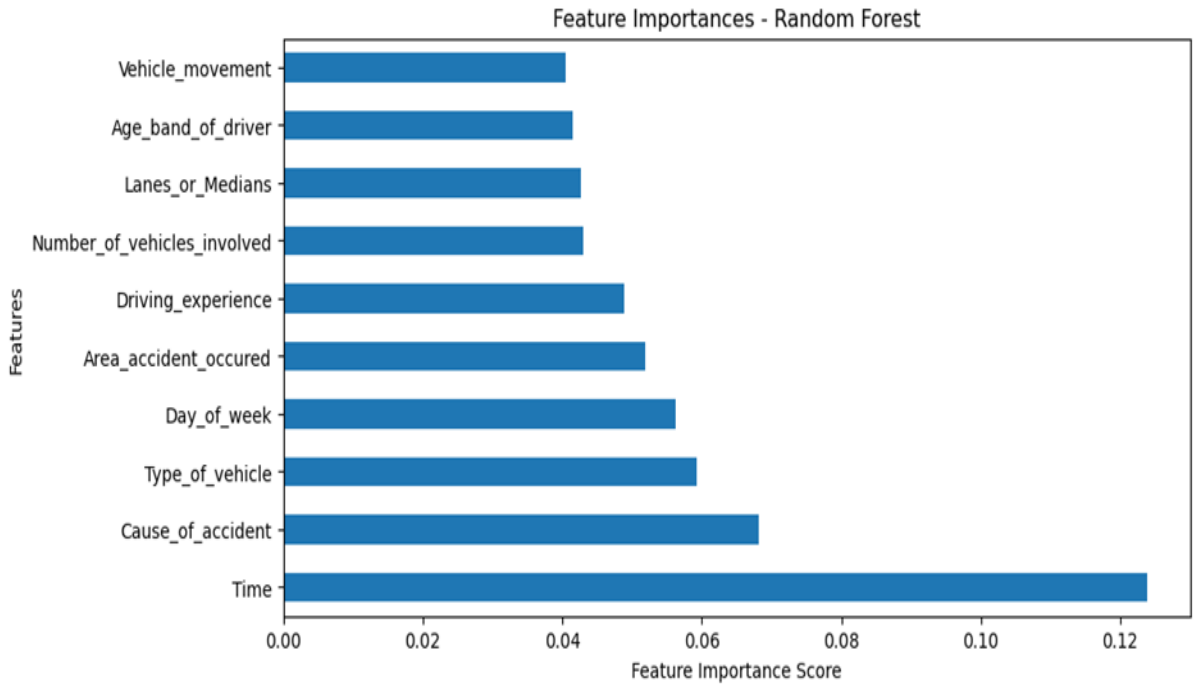

Figure 20: Feature importance - Random Forest

The feature importance graph as displayed above is indicating the features that are considered the most influential in being able to predict the severity of an accident. According to the results it is

shown that time is considered the most critical feature when looking at the severity of accidents. This is then followed by the cause of the accident and then the type of vehicle. By looking at the age band of a driver and the vehicle movement, it is evident that these also have a role but not as much as the two mentioned above.

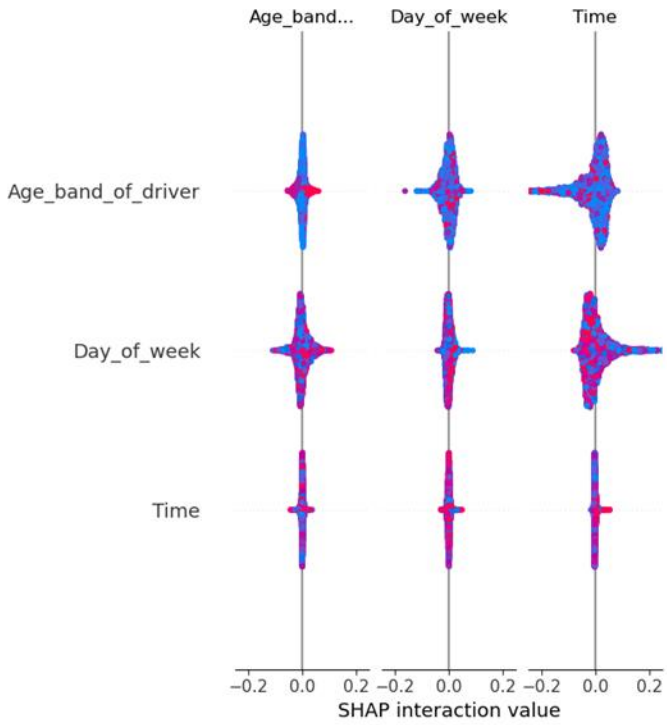

Figure 21: SHAP interaction value

The following SHAP plot shows how the predictions are impacted by each feature in the data set. If we look at the accident severity, we notice that the ones that have the most influence are the time, day of the week, and the age band of the driver. By looking at the interactions between each of the features we can notice that there is a relationship between the day of the week, and the time, and how it is impacting the severity.

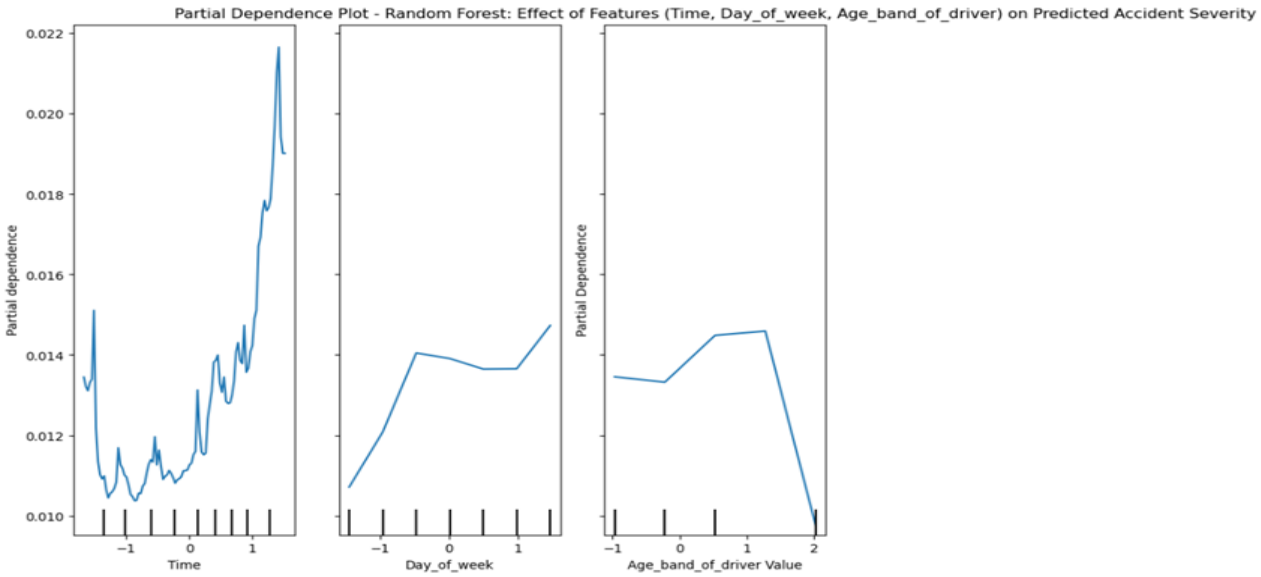

Figure 22: Partial dependence plot: Random Forest

This graph shows how the effect of three features impacts the severity of an accident. The time shows an increasing trend which means that the model predicts a higher severity for an accident on specific hours of the day. Accidents fluctuating on days of the week would suggest that there are days that are more prone to severe accidents. Furthermore, some age bands show an increase in accident severity while others have a decrease and the far drop in the graph based on the age is an indication that there is a particular group that is less likely to be involved in severe accidents.

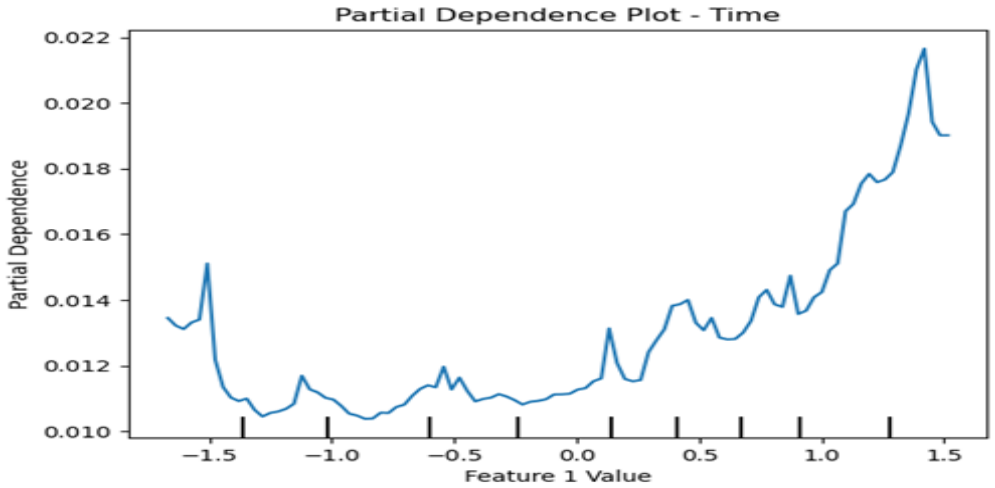

Figure 23: Partial dependence plot- Time

This graph shows an upward trend which indicates that as the time increases, potentially on later hours day, more severe accidents are viewed. This can be due to a multitude of reasons such as reduced visibility being a dark, traffic congestion and the fatigue of a driver.

7. Discussion

Within this assignment we applied the machine learning models of decision tree, logistic regression and random forest to predict the severity of traffic accidents. The dataset we used comprises of characteristics such as driver details, road conditions, weather conditions and other factors to train our models on. Additionally, we added the XAI techniques of SHAP, LIME and Partial Dependence Plots (PDP) to gain insights into how the model predictions worked.

Model performance and confusion matrices

The decision tree model had an overall accuracy of 76% and the logistic regression had an overall accuracy of 84% which indicated that logistic regression is more accurate at correctly predicting severity overall. However, the decision tree model and logistic regression model showed different performances with the key metrics such as accuracy, precision, recall and F1-score. The decision tree model even outperformed the logistic regression model in precision, making it better at minimising false positives. On the other side, the logistic regression model performed better in terms of recall making it more effective at identifying true positives. This could be important for use cases where all instances of a class need to be identified. The final key metric, the F1-score scored similarly at 77% for both models.

The Cohen's Kappa showed that both models are not reliable as the decision tree scored 0.17 and the logistic regression scored 0.00 which could also be a null value. This shows that the models' predictions and the actual classifications do not match very well, and the model might be no better than random selection.

When evaluating the random forest classifier for accident severity, the confusion matrix showed the model performed well when distinguishing between less serious injury and serious injury with no false positives or false negatives. The model did however struggle with the classification of moderate injuries and distinguishing moderate injuries from less serious injuries, and the outcome of this was more misclassifications as moderate injuries.

These misclassifications show that the model might be prone to errors when predicting subtle differences and this is also a common issue with other machine learning models where the model is better at predicting the majority class but struggles to accurately predict the minority class.

Residual Analysis and Feature Importance

The residual plot for the Random Forest was used to predict the number of causalities and it revealed significant issues with the model's fit. The residuals should be scattered along the horizontal zero line however the plot shows a linear trend suggesting the model was under or over predicting. There were also extreme outliers present near -1.0 and 0.75 which shows large prediction errors. The results of this could show that the Random Forest is not capturing the complexity of the relationship between input features and number of causalities. The low R2 value confirms this as only 17% of the variance in the target variable.

The feature importance plot for the Random Forest showed that only a few features such as Time and Cause of the accident were of higher importance, but most features had almost zero importance. This shows that only a small subset of features contributed to the model’s performance. The time of accident was the most important feature with a score of 0.16 followed by a score of 0.09 for cause of accident. Day of week, area of accident and type of vehicle had medium impact in terms of importance. Features including number of vehicles involved, junction type and road surface conditions contributed very little to the model’s predictions.

The fact that there were many features with a near zero importance shows that these features might be redundant or irrelevant to the model's predictive task. By removing these less important features, the model's performance could improve, and the computation efficiency would improve as well.

Correlation Analysis

Correlation analysis was done to further understand the relationship between different variables. An emphasis was put on Accident_Severity and Casulity_Severity and the results were as follows:

- Accident_severity_Serious Injury and Accident_severity_Slight Injury: had a strong negative relationship which was to be expected as these two variables are mutually exclusive events. It was primarily used to understand that the model created was working as expected.
- Casualty_severity and both serious injury and slight injury: The correlation between these two variables were found to have a positive and negative weak relationship. This shows us that causality severity does not have a strong influence on accident severity. This indicates that other factors might play a more important role when determining causality outcomes.

The correlation matrix between Accident_Severity and Accident_severity did not show a strong relationship which shows the level of complexity of the dataset. Because a linear regression model was created to analysis the data, it did not show us strong results which could mean the data follows a more non-linear relationship.

Explainable AI (XAI) Techniques

By making use of tools such as LIME, SHAP, and Partial Dependence plots, XAI (Explainable Artificial Intelligence) could be performed. The models were created so that we could have a deeper understanding of how individual features influence the model's decisions. For instance:

SHAP: These plots revealed that features such as the number of causalities, the number of vehicles involved, and the lanes or medians involved had a strong positive import on the model's predictions of serious accidents. However, variables such as weather conditions and the type of vehicle reduced the chances of a severity being classified as serious.

LIME: LIME models were created so that specific instances where the model had to predict serious accidents showed that various featured had to be combined so that there could be an influence on the predictions. It was seen that the number of vehicles involved had a strong positive influence on the predictions, however, road surface had a strong negative influence on the predictions.

The results above provided transparency in the model's decision-making process. This allows stakeholders to further understand why particular predictions were made. This is important, especially in the context of road safety where AI systems are trusted as decisions must be explainable and justifiable to all stakeholders (such as law enforcement agencies and public safety organisations) involved.

Model Improvement Recommendations

From the results that have been created and analysed, several recommendations can be made to improve the model's performance and its ability to predict:

Address class imbalance: The models created struggled with distinguishing between classes such as slight injury and moderate injury using the variable Accident_Severity. Tools and techniques such as SMOTE (Synthesis Minority Over-Sampling) can be used to ensure that the dataset is more balanced. This is done to ensure that the model's ability to predict minority classes are improved.

Feature selection: From the feature importance analysis, features in the dataset have almost no relationship on each other. By shifting focus on other variables, creating new models, simplifying the model and reducing overfitting, there is a potential that the model's performance can be enhanced.

Model Complexity: By conducting a residual analysis, it was seen that the random forest regressor did not do a good job on capturing the complexity of the relationship between the different variables in the dataset. By making use of tools such as Gradient Boosting Regressors and Neural Networks, the relationships can be better seen.

Hyperparameter Tuning: By tuning the hyperparameter of the models created (this includes models such as the tree depth in Decisions Tress and the number of estimators used in the Random Forest), the model's performance could be better optimised. This can be done by making use of tools such as GridSearchCV and RandomizedSearchCV.

8. Future Research

This study used machine learning models and explainable AI to interpret the different factors which could influence traffic accident severity. The following factors could be areas of future research:

- Diverse Data Sources: Future studies could incorporate additional data such as geospatial data, real-time traffic flow or even driver behaviour to enhance the model and improve accuracy. Factors such as weather conditions could provide better insights into predicting accident likelihood and possible road safety risks.
- Enhanced Explainability Techniques: Future research could focus on refining explainability methods to help simulate scenarios and evaluate potential safety measures. This could also help for evidence-based decisions to reduce accident severity.
- Investigating ethical and privacy implications: Future studies could explore ethical considerations such as privacy protection and how to avoid bias. This is important in maintaining models with predictive accuracy without compromising user privacy.
-

9. Conclusion

In conclusion, this study used three machine learning models (decision tree, logistic regression and random forest) with XAI techniques (SHAP, LIME and PDP) to provide the model interpretations. The Logistic Regression model performed the best with accuracy and recall metrics in terms of identifying accident severity. However, the Decision Tree model performed better with the precision metric which excelled at minimising false positives. The XAI techniques highlighted the influence of the "Time" and "Cause of accident" classes, ultimately providing deeper insights to the results from the machine learning models.

This study highlights the potential of machine learning models mixed with XAI techniques to uncover the key influences on traffic incidents while making reliable predictions on future traffic violations. These tools prove to be valuable in assisting informed decision-making regarding influences contributing to traffic violations.

10. References

1) Adeliyi, T. T., Oluwadele, D., Igwe, K., & ... (2023). Analysis of road traffic accidents severity using a pruned tree-based model. In *International Journal of* .... researchgate.net. https://www.researchgate.net/profile/Oluwasegun-Aroba/publication/372549214_Analysis_of_Road_Traffic_Accidents_Severity_Using_a_Pruned_Tree-Based_Model/links/64bfd75c8de7ed28bac29286/Analysis-of-Road-Traffic-Accidents-Severity-Using-a-Pruned-Tree-Based-Model.pdf

2) Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv Preprint arXiv:1702.08608*. https://arxiv.org/abs/1702.08608

3) Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., & ... (2018). A survey of methods for explaining black box models. *ACM Computing* https://doi.org/10.1145/3236009

4) Molnar, C. (2022). Interpretable machine learning: A guide for making black box models explainable. Christophm. Github. Io/interpretable-ml-book. In ... *neural networks and beyond: A review of methods and* ....

5) Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). ‘ Why should i trust you?’ Explaining the predictions of any classifier. *Proceedings of the 22nd ACM* https://doi.org/10.1145/2939672.2939778