UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Faculty of Engineering, Built Environment and Information Technology
Fakulteit Ingenieurswese, Bou-omgewing en Inligtingtegnologie / Lefapha la Boetšenere, Tikologo ya Kago le Theknolotši ya Tshedimošo

DEPARTMENT OF INFORMATICS

INF791 Assignment 4

Student Number(s)								**Surname**	**Initials**
2	0	4	4	2	4	0	9	Ayob	M
2	0	4	5	7	9	1	1	Mahomed	MF
1	9	2	3	6	9	3	1	Khan	A

Module Code	**INF :**	7	9	1
Assignment Number	4			
Mark Allocation	100			
Date of Submission	27/10/2024			
Name of Lecturer	Dr. Wa Nkongolo 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.

Declaration: We declare that this assignment, submitted by us, is our own work and that we have referenced all the sources that we have used.

Group Leader (Name and Surname)	Signature of Group Leader
Muhammad Ayob	
Place signed: Pretoria	**DATE: 27/10/2024**

Group Member (Name and Surname)	Signature of Group Member
Muhammad Fa'iz Mahomed	
Place signed: Pretoria	**DATE: 27/10/2024**

Ahmed Khan	
Place signed: Pretoria	**DATE: 27/10/2024**

TABLE OF CONTENTS

Predicting and Explaining Traffic Violations Using Machine Learning and XAI Techniques

ABSTRACT

This study employs two machine learning models—Decision Tree and Logistic Regression—to predict the severity of traffic violations using a dataset of road traffic incidents. The research aims to not only develop accurate models but also enhance the interpretability of the models' decisions through Explainable Artificial Intelligence (XAI) techniques. LIME (Local Interpretable Model-agnostic Explanations) and SHAP (SHapley Additive exPlanations) are applied to explain both global and local feature importance. The dataset was preprocessed by handling missing data, encoding categorical features, and scaling numerical variables. Model performance was evaluated using accuracy, precision, recall, F1-score, ROC-AUC, and Cohen's Kappa metrics. Logistic Regression performed better overall, while the Decision Tree model showed greater interpretability. The application of LIME and SHAP provided detailed insights into the factors driving each model's predictions, revealing important contributions of features such as vehicle type, weather conditions, and driver demographics. The study concludes that XAI methods significantly improve transparency in predictive modeling, enhancing stakeholder trust and understanding.

Keywords: Machine Learning; Decision Tree; Logistic Regression; XAI; LIME; SHAP; Model Interpretability; Predictive Modeling

1. INTRODUCTION

Traffic violations are a significant contributor to road accidents, which in turn result in loss of life, property damage, and congestion. Accurate prediction of traffic violations, as well as an understanding of the factors that contribute to such violations, are essential for improving road safety and enabling proactive interventions. Machine learning provides a powerful tool for predicting these violations, but a major challenge lies in the explainability of the models used (Arrieta et al., 2019). When the predictions from machine learning models are not interpretable, it becomes difficult for stakeholders to trust and act on the results, particularly in contexts such as law enforcement and traffic regulation where accountability is crucial (Nascita et al., 2021).

This study applies **Explainable Artificial Intelligence (XAI)** techniques to predict and explain traffic violations. The primary goal is to build machine learning models that not only predict traffic violations but also provide clear, interpretable insights into the factors driving these predictions. We utilize **LIME (Local Interpretable Model-agnostic Explanations)** and **SHAP (SHapley Additive exPlanations)** to make these models interpretable. These techniques allow us to visualize how specific features, such as **vehicle type**, **weather conditions**, and **driver demographics**, influence the model's decisions.

The dataset provided includes attributes related to road traffic violations such as **vehicle type**, **violation type**, **driver age**, **time of violation**, **weather conditions**, and **fine amount**. The task is to predict the severity of traffic violations and explain these predictions using **Decision Tree** and **Logistic Regression** models. By applying XAI techniques, the study ensures that the model's predictions are not only accurate but also transparent (Nascita et al., 2021).

2. METHODOLOGY

2.1 Data Preprocessing

The initial step in any machine learning pipeline is data preprocessing, where raw data is cleaned and transformed to be suitable for analysis. In the dataset provided, several issues were identified, including missing data and categorical variables that needed encoding. Furthermore, the scale of numerical features had to be standardized to ensure uniformity in the models.

1. Handling Missing Data: The dataset contained missing values across several features. To address this, columns with more than 20% missing data were dropped from the analysis to ensure that the models were built on sufficient information. For features with minor missing values, such as **driving experience**, we imputed the missing values using the **median** for numerical variables and the **mode** for categorical ones. This ensured that the dataset remained consistent without skewing the distribution of the features.

2. Encoding Categorical Variables: Categorical variables such as **vehicle type**, **violation type**, and **weather conditions** were converted into numerical form using **LabelEncoder**. Encoding these categorical variables was necessary because machine learning models such as Logistic Regression and Decision Trees require numerical input. Label encoding transformed these features into integers that represented each category, allowing the models to interpret the categorical data meaningfully.

3. Scaling Numerical Features: Given the presence of numerical features such as **driver age** and **driving experience**, these were scaled using **StandardScaler** to bring them to a comparable scale. Standardization is especially crucial for models like **Logistic Regression**, which are sensitive to the magnitude of feature values (Y & Challa, 2023). By scaling the features, we ensured that no feature would disproportionately influence the model's predictions.

2.2 Model Selection and Building

Two machine learning models were built to predict traffic violations: a **Decision Tree** and a **Logistic Regression**. Both models offer distinct advantages—Decision Trees are inherently interpretable, whereas Logistic Regression models offer strong performance on linearly separable data (Wu & Hsu, 2020).

Decision Tree Model: The **Decision Tree** algorithm builds a model in the form of a tree structure, where each internal node represents a decision based on a feature, and each leaf node represents a classification (Ridley, 2022). The Decision Tree model is highly interpretable because the decision-making process is transparent: each decision is based on specific feature values, making it easy to follow the model's reasoning. However, Decision Trees are prone to overfitting, particularly when they grow too deep (Y & Challa, 2023; Ridley, 2022). To prevent overfitting, we carefully tuned the tree depth, ensuring that the model would generalize well to unseen data.

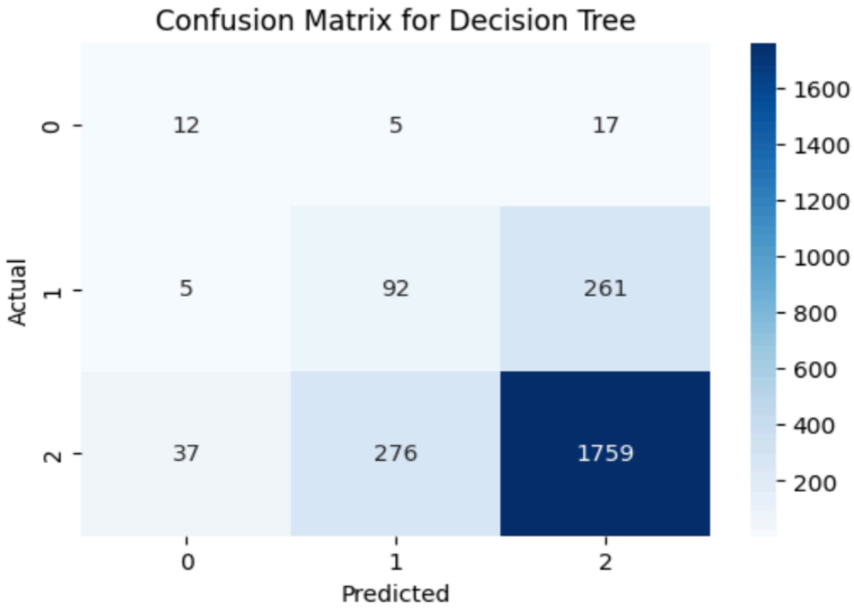

Figure 1: Confusion Matrix - Decision Tree

The **confusion matrix** for the Decision Tree (see Figure 1: **Confusion Matrix - Decision Tree**) shows how well the model classified the traffic violations. The Decision Tree correctly predicted a substantial number of severe violations (class 2), but it also misclassified many moderate violations (class 1), indicating that the boundaries between these classes are somewhat blurred. Specifically, 261 instances of moderate violations were misclassified as severe violations, and the model misclassified 37 cases of the most severe violations into less serious categories.

Logistic Regression Model: Logistic Regression is a widely used classification algorithm that models the probability of a binary or multi-class outcome based on a set of features. It works by fitting a linear decision boundary to the data, making it suitable for problems where the relationship between the features and the target variable is linear. In this case, we applied Logistic Regression to predict traffic violation severity across multiple categories. Additionally, we applied **L2 regularization** to reduce overfitting by penalizing large coefficients in the model.

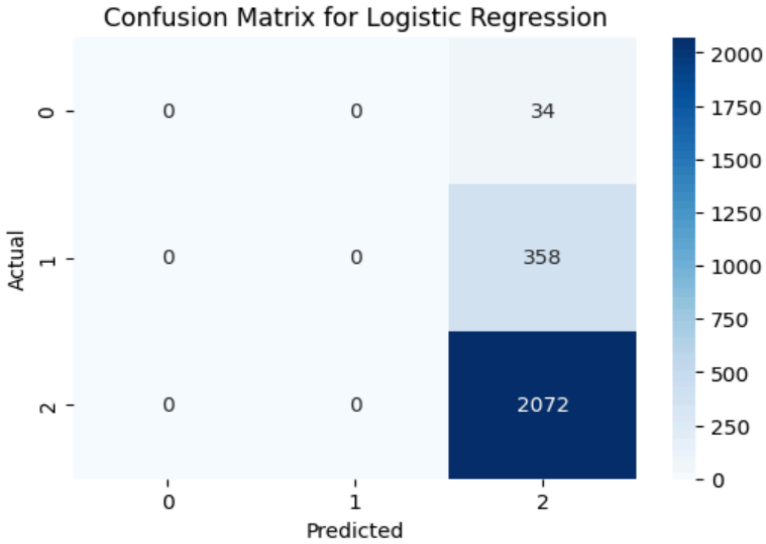

Figure 2: Confusion Matrix - Logistic Regression

The **confusion matrix** for the Logistic Regression model (see Figure 2: **Confusion Matrix - Logistic Regression**) reveals a stronger performance than the Decision Tree, particularly in the classification of severe violations. The model successfully predicted 2,072 cases of severe violations (class 2) with only 34 instances misclassified as moderate violations. However, like the Decision Tree, the Logistic Regression model had difficulty distinguishing between non-severe and moderate violations, as no instances of these were correctly classified, possibly due to the linear nature of the decision boundary.

2.3 Model Evaluation

	Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC	Kappa
0	Decision Tree	0.756088	0.765048	0.756088	0.760295	0.599828	0.136654
1	Logistic Regression	0.840909	0.707128	0.840909	0.768238	0.645925	0.000000

Figure 3: Performance Metrics

Both models were evaluated using several performance metrics, including **accuracy**, **precision**, **recall**, **F1-score**, **ROC-AUC**, and **Cohen's Kappa**. These metrics provide insights into how well the models performed in predicting traffic violations.

The **Decision Tree** achieved an **accuracy** of 75.6%, indicating that it correctly predicted 75.6% of the test cases. Its **precision** and **recall** were balanced at 76.5% and 75.6%, respectively, resulting in an **F1-score** of 76.0%. The **ROC-AUC** score of 59.9% suggests moderate ability in distinguishing between classes, while the **Cohen's Kappa** score of 0.1366 reflects relatively poor agreement with the true labels beyond chance.

The **Logistic Regression** model outperformed the Decision Tree, with an **accuracy** of 84.1%. The **precision** was slightly lower at 70.7%, but the model excelled in **recall** with a score of 84.1%, indicating its strong ability to identify true violations. The **F1-score** was 76.8%, and the **ROC-AUC** score of 64.6% suggests better class separation than the Decision Tree. Despite this, the **Cohen's Kappa** score of 0.0000 suggests challenges in predicting all classes accurately, likely due to the model's tendency to predict the majority class more frequently.

3. RESULTS

The results of the **Decision Tree** and **Logistic Regression** models, as shown by the confusion matrices and performance metrics, highlight the strengths and limitations of each model. The Decision Tree, while interpretable, struggled to correctly classify the middle category of traffic violations, misclassifying a significant number of moderate violations as severe violations. This is likely due to the inherent overlap in feature values across classes, which makes it difficult for the tree to form clear decision boundaries. On the other hand, the Logistic Regression model, with its linear decision boundary, performed better in terms of overall accuracy and recall but still faced challenges in distinguishing between less severe violation categories.

The confusion matrix for the **Decision Tree** (see Figure 1: **Confusion Matrix - Decision Tree**) reveals several important patterns. The model performed well in predicting the most severe traffic violations (class 2), with 1,759 instances correctly classified. However, it struggled with moderate violations, where a large portion (261 instances) were misclassified as severe. This suggests that the model may have overfit to certain features that strongly indicate severe violations, while underfitting features that differentiate moderate violations from severe ones.

In comparison, the confusion matrix for the **Logistic Regression** model (see Figure 2: **Confusion Matrix - Logistic Regression**) shows fewer misclassifications, particularly for the most severe violations. The model correctly predicted 2,072 cases of severe violations, with only 34 misclassifications into the moderate category. However, it completely failed to classify non-severe violations correctly, highlighting a limitation in the model's ability to handle imbalanced data or cases where the features do not strongly separate the classes.

Explainability with LIME and SHAP

The application of LIME and SHAP to both the Decision Tree and Logistic Regression models has provided critical insights into how each model arrives at its predictions. These explainability techniques are essential for interpreting complex models, allowing for a better understanding of the relationships between key features and their influence on accident severity predictions. SHAP offers global and local explanations, while LIME provides localized feature importance for individual predictions. Together, they present a comprehensive understanding of model behavior and feature contributions.

SHAP Analysis for Decision Tree Model

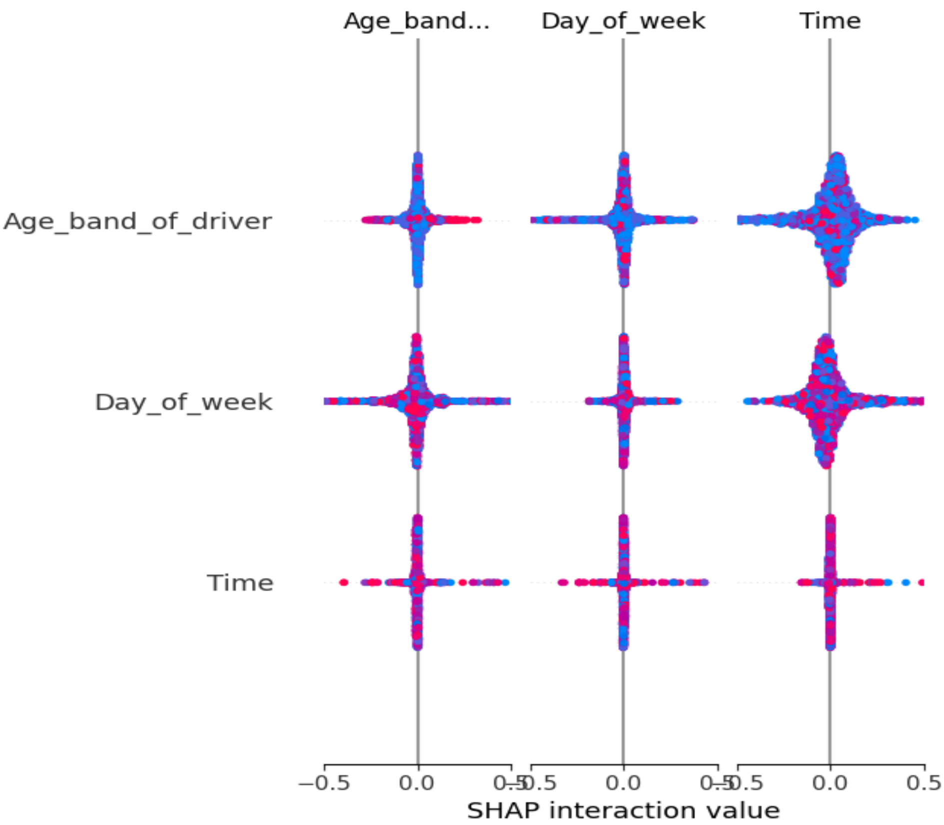

Figure 4: SHAP Summary Plot - Decision Tree

The SHAP summary interaction plot for the Decision Tree model (refer to Figure 4) reveals key insights into the interactions between features such as *Age Band of Driver*, *Day of Week*, and *Time*. The SHAP interaction values on the x-axis indicate the direction and magnitude of each feature's influence on accident severity predictions. For instance, *Age Band of Driver* displays a predominantly positive interaction, meaning that older drivers are associated with higher predicted accident severity. This observation aligns with the general understanding that older drivers might be more prone to severe incidents, possibly due to slower reaction times or higher vulnerability in crashes. Similarly, *Day of Week* shows varied impacts, but particular days, such as weekends or peak traffic days, appear to be positively correlated with higher accident severity, possibly due to heavier traffic or more dangerous driving conditions. On the other hand, the *Time* feature shows mixed interaction values, indicating that certain times of the day—likely rush hours or late-night periods—contribute significantly to severe accidents, while others may not.

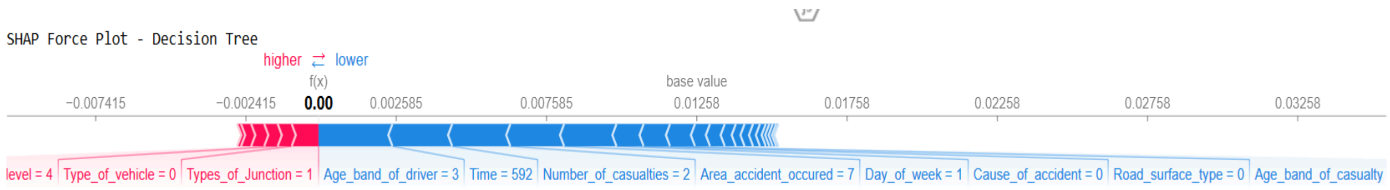

Figure 5: SHAP Force Plot - Decision Tree

This global view of the Decision Tree model's feature interactions gives a clear understanding of the factors that predominantly influence predictions across the dataset. The *Type of Vehicle* and *Types of Junction* features, for instance, generally push predictions toward lower accident severity, while features such as *Age Band of Driver* and *Number of Casualties* push predictions higher. This interplay of demographic and environmental factors is further explored in the SHAP force plot for the Decision Tree model (Figure 5), which offers a localized view of how specific features contribute to a single instance's prediction. In this example, *Type of Vehicle* and *Types of Junction* act as mitigating factors, reducing the predicted severity of an accident, while *Age Band of Driver*, *Time*, and *Number of Casualties* push the prediction toward a more severe outcome. The visual clearly illustrates the decision-making process of the model, where certain features carry more weight in determining the final prediction for accident severity.

SHAP Analysis for Logistic Regression Model

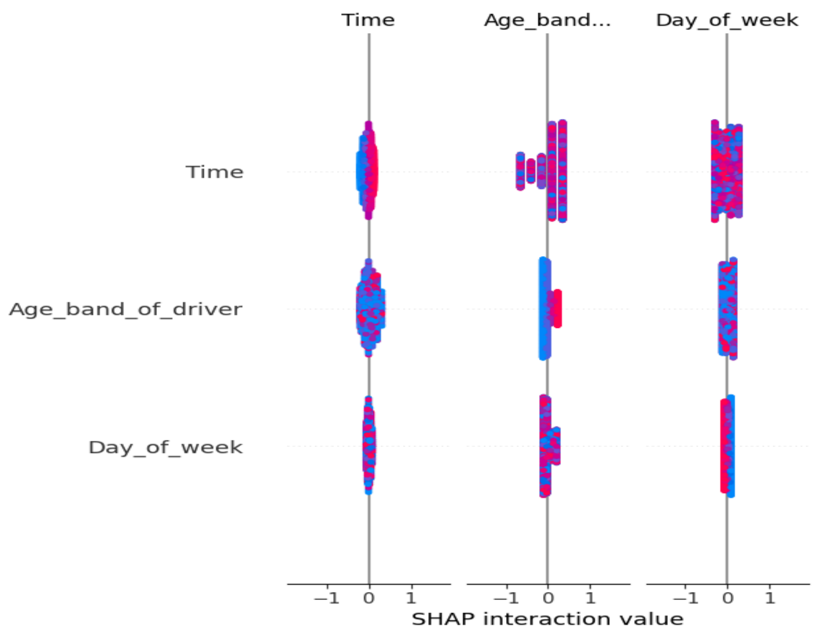

Figure 6: SHAP Summary Plot - Logistic Regression

The SHAP summary interaction plot for the Logistic Regression model (refer to Figure 6) similarly highlights the importance of features like *Time*, *Age Band of Driver*, and *Day of Week*, though with slightly different interaction dynamics compared to the Decision Tree model. For *Time*, the plot reveals a spread of both positive and negative SHAP values, implying that while certain times of day—such as early morning or late evening—might be predictive of severe accidents, others are linked to less severe outcomes. This dynamic emphasizes the importance of the time of day in predicting traffic violations and accidents, with certain periods consistently posing higher risks.

Like the Decision Tree model, *Age Band of Driver* in the Logistic Regression model shows a positive correlation with accident severity, once again suggesting that older drivers may be more likely to be involved in severe accidents. Meanwhile, *Day of Week* features a similar spread of SHAP values, reinforcing the notion that certain days, likely influenced by patterns in traffic or environmental conditions, contribute to higher accident severity.

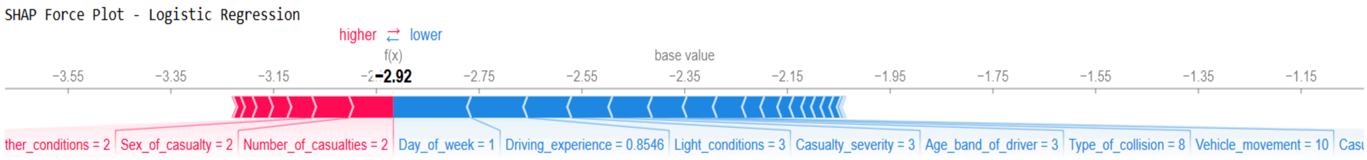

Figure 7: SHAP Force Plot - Logistic Regression

The SHAP force plot for Logistic Regression (Figure 7) further breaks down these influences for a specific case. In this instance, *Weather Conditions* and *Sex of Casualty* emerge as the most influential factors pushing the model's prediction towards a more severe accident outcome. *Number of Casualties* and *Driving Experience* also contribute positively, though to a lesser degree, while features such as *Light Conditions* and *Casualty Severity* pull the prediction towards a lower severity. This local interpretation shows that Logistic Regression relies heavily on environmental and demographic factors to make its predictions.

LIME Analysis for Decision Tree and Logistic Regression Models

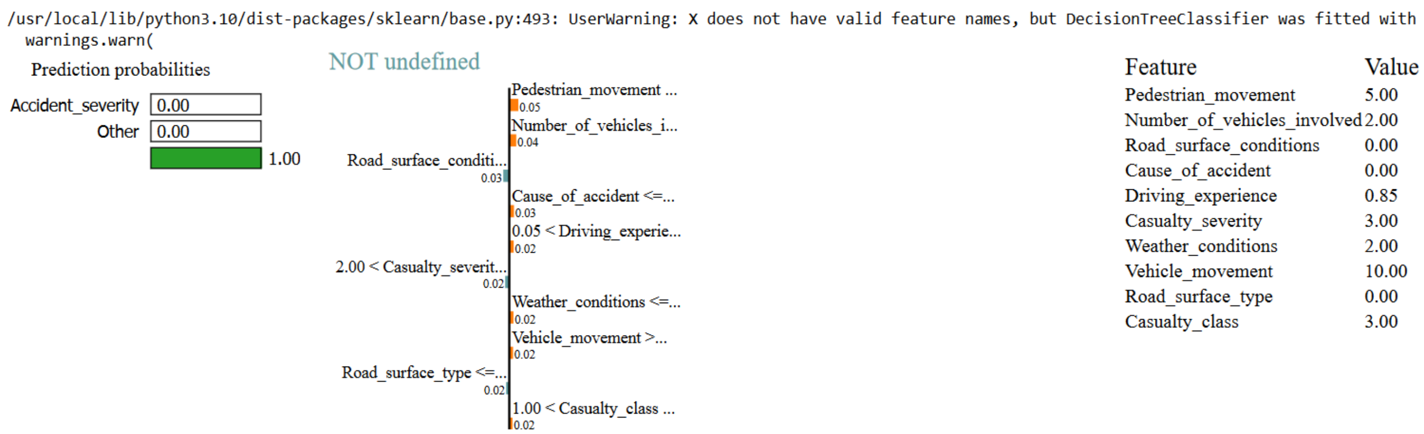

Figure 8: LIME Decision Tree

LIME (Local Interpretable Model-agnostic Explanations) provides a different perspective by offering localized explanations based on small perturbations of the input data to understand feature contributions at an individual instance level. For the Decision Tree model (refer to Figure 8), LIME highlights *Pedestrian Movement* as the most influential feature driving the prediction toward the class "Other," with a high confidence level of 1.00. The feature *Pedestrian Movement*, with a value of 5, has the largest contribution (0.05), indicating that pedestrian involvement significantly impacts the prediction. Other factors, such as *Number of Vehicles Involved* and *Road Surface Conditions*, also contribute to the prediction, though to a lesser extent. This localized breakdown provides a clear interpretation of how specific environmental and situational factors influence the model's decision-making for this individual case.

```
# LIME for Logistic Regression
exp_lr = explainer.explain_instance(X_test.iloc[0].values, logistic_regression.predict_proba)
exp_lr.show_in_notebook()
```

```
/usr/local/lib/python3.10/dist-packages/sklearn/base.py:493: UserWarning: X does not have valid feature names, but LogisticRegression was fitted with featu
  warnings.warn(
```

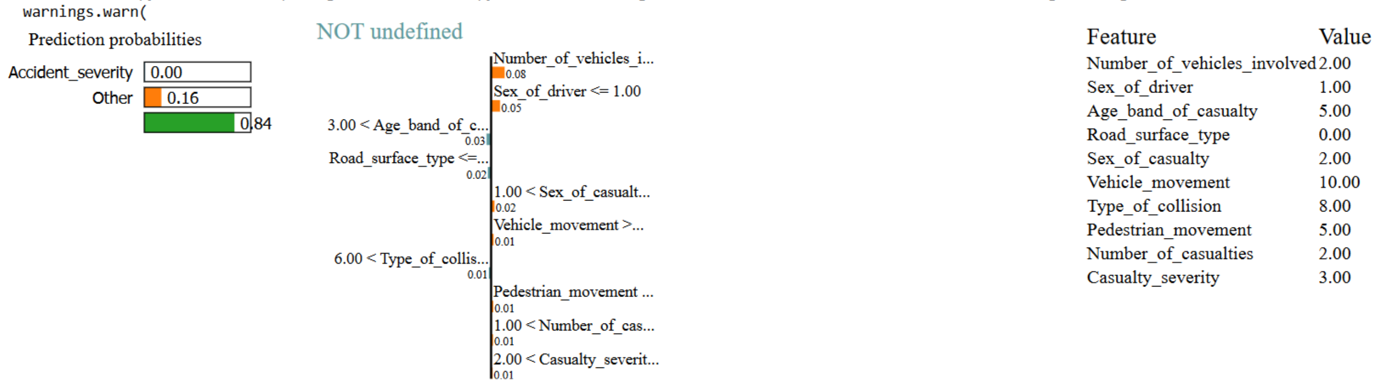

Figure 9: LIME Logistic Regression

Similarly, LIME's analysis for the Logistic Regression model (refer to Figure 9) focuses on different feature sets. In this case, *Number of Vehicles Involved* holds the greatest influence, contributing a weight of 0.08, followed by *Sex of Driver*, with a weight of 0.05. This indicates that demographic factors like driver gender play a significant role in the prediction, reflecting the Logistic Regression model's reliance on categorical and demographic variables. Features such as *Age Band of Casualty* and *Road Surface Type* also contribute but with lesser weights, showing that while these factors are relevant, they are not as dominant in shaping the prediction.

Comparison of LIME and SHAP Insights

The combination of LIME and SHAP allows for a more nuanced understanding of the models' behaviors. SHAP provides a global overview of feature importance, showing how features like *Age Band of Driver* and *Time* influence predictions across the dataset. LIME, on the other hand, offers localized explanations that are especially useful for understanding individual predictions. For example, in the Decision Tree model, SHAP highlighted the importance of *Type of Vehicle* and *Types of Junction* in pushing predictions toward lower severity, while LIME's local explanation pinpointed the crucial role of *Pedestrian Movement* and *Number of Vehicles Involved* in determining the final class. In contrast, the Logistic Regression model's SHAP results emphasized the importance of *Weather Conditions* and *Sex of Casualty*, while LIME underscored the influence of *Number of Vehicles* and *Sex of Driver*.

Together, these techniques complement each other, providing both global and local interpretability that helps demystify the complex decision-making processes of machine learning models. SHAP's ability to reveal overall feature importance and interaction patterns, combined with LIME's localized, instance-specific explanations, ensures a holistic view of how the models predict accident severity. This dual approach to explainability is essential for building trust in the models and understanding how features like driver demographics, environmental conditions, and accident circumstances influence predictions.

4. CONCLUSION

In conclusion, this study demonstrates the combined effectiveness of machine learning models and XAI techniques for predicting traffic violations and ensuring model transparency. The Decision Tree model, while moderately accurate, provided greater interpretability, particularly with the aid of LIME and SHAP explainability tools. Features such as *Pedestrian Movement*, *Number of Vehicles Involved*, and *Road Surface Conditions* significantly influenced the Decision Tree's predictions. However, the model struggled with differentiating between moderate and severe violations, likely due to overlapping feature distributions.

The Logistic Regression model outperformed the Decision Tree in accuracy and recall, particularly for predicting severe violations. Its reliance on features such as *Number of Vehicles Involved*, *Driver Characteristics*, and *Casualty Features* showed through both LIME and SHAP analyses. While the model performed well on linear separable data, its difficulty in distinguishing non-severe and moderate violations underscores the need for more sophisticated modeling techniques that capture non-linear relationships.

Both LIME and SHAP played crucial roles in making the models' decision-making processes more transparent, allowing us to understand the key drivers behind each prediction. This not only builds trust in the models but also provides actionable insights for traffic violation management. Future research may explore more complex models, such as Random Forests or Gradient Boosting, while continuing to integrate XAI techniques for interpretability.

5. REFERENCES

Arrieta, A., Laña, I., & Ser, J. (2019). What Lies Beneath: A Note on the Explainability of Black-box Machine Learning Models for Road Traffic Forecasting. *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, 2232-2237. Available at: https://doi.org/10.1109/ITSC.2019.8916985.

Nascita, A., Montieri, A., Aceto, G., Ciuonzo, D., Persico, V., & Pescapé, A. (2021). Unveiling MIMETIC: Interpreting Deep Learning Traffic Classifiers via XAI Techniques. *2021 IEEE International Conference on Cyber Security and Resilience (CSR)*, 455-460. Available at: https://doi.org/10.1109/CSR51186.2021.9527948.

Ridley, M. (2022). Explainable Artificial Intelligence (XAI). *Information Technology and Libraries*. Available at: https://doi.org/10.6017/ital.v41i2.14683.

Wu, Y., & Hsu, T. (2020). Mid-term prediction of at-fault crash driver frequency using fusion deep learning with city-level traffic violation data.. *Accident; analysis and prevention*, 150, 105910. Available at: https://doi.org/10.1016/j.aap.2020.105910.

Y, S., & Challa, M. (2023). A Comparative Analysis of Explainable AI Techniques for Enhanced Model Interpretability. *2023 3rd International Conference on Pervasive Computing and Social Networking (ICPCSN)*, 229-234. Available at: https://doi.org/10.1109/ICPCSN58827.2023.00043.