National College of Ireland

# Using Explainable AI LIME to Improve the understanding of Machine Learning Predictions for Traffic Accidents

MSc Research Project
MSc AI

Kevin Heagney
Student ID: x14120488

School of Computing
National College of Ireland

Supervisor: Sheresh Zahoor

**National College of Ireland**
**Project Submission Sheet**
**School of Computing**

National College of Ireland

| | |
|---|---|
| **Student Name:** | Kevin Heagney |
| **Student ID:** | x14120488 |
| **Programme:** | MSc AI |
| **Year:** | 2024 |
| **Module:** | MSc Research Project |
| **Supervisor:** | Sheresh Zahoor |
| **Submission Due Date:** | 16/09/2024 |
| **Project Title:** | Using Explainable AI LIME to Improve the understanding of Machine Learning Predictions for Traffic Accidents |
| **Word Count:** | 5,148 |
| **Page Count:** | 19 |

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

**ALL** internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

| | |
|---|---|
| **Signature:** | |
| **Date:** | 16th September 2024 |

## PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

| | |
|---|---|
| Attach a completed copy of this sheet to each project (including multiple copies). | ☐ |
| **Attach a Moodle submission receipt of the online project submission**, to each project (including multiple copies). | ☐ |
| **You must ensure that you retain a HARD COPY of the project**, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer. | ☐ |

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

| **Office Use Only** | |
|---|---|
| Signature: | |
| Date: | |
| Penalty Applied (if applicable): | |

# Using Explainable AI LIME to Improve the understanding of Machine Learning Predictions for Traffic Accidents

Kevin Heagney
x14120488

**Abstract**

The purpose of this research is to determine the suitability of the Explainable AI tool LIME for explaining the output of Machine Learning prediction classifier tools, for traffic accident data. Many papers have involved the use of XAI LIME for other areas, for example, health care, but I have not found any cases where LIME was used for traffic accident data. The combination of a Machine Learning model and LIME can explain and increase the understanding of what features in a dataset are the most important. This information could be very helpful with accident prevention. In the research, LIME was found to produce a very clear explanation of the Machine Learning model predictions for any selected instances or records in a large dataset.

## 1 Introduction

The World Health Organization (WHO) in a recent study emphasized the importance of road traffic accidents, which account for more than 1.3 million deaths per year Aboulola et al. (2024). Machine Learning (ML) methods can be used to predict road traffic accidents based on the parameters or features that apply. ML methods can also determine the relative importance of the features. This data can then be used to focus on the key causes of accidents and so reduce the severity and frequency of accidents. Also, Explainable Artificial Intelligence (XAI) can be used to explain the results provided by the ML methods. One XAI method that can be used for classification prediction methods is Local Interpretable Model-agnostic Explanation method (LIME) which is used in the study Vijayvargiya et al. (2023).

LIME has been used in health care studies, for example, in Cervantes and Chan (2021), CNN and LIME are used to analyse Covid-19 image data. CNN carries out an analysis of the image data, and the LIME image output of the individual instances can help explain and sometimes reveal shortcomings for a particular ML instance. Also, ML methods and LIME can be used on a combination of datasets. In health care research, Kamal et al. (2021), a dataset containing 6,400 images from Kaggle was used in combination with gene expression data containing 18,234 genes. In this case the methods used were CNN, kNN, LIME, SpinalNet, SVC, Xboost, and CNN provided an accuracy of 97.2%.

XAI methods have been used in studies on traffic accidents, but I have not found any papers that use LIME for studies on traffic accidents. LIME is an XAI method that is model-agnostic and so can be used with a number of ML methods.

The Research Question is: How can Explainable AI LIME be used to Improve the understanding of Machine Learning Predictions for Traffic Accidents?

The use of ML methods and LIME XAI for traffic accident data can help to develop a better understanding of the causes of traffic accidents and so reduce deaths and injuries on the roads.

This paper consists of a number of sections. The Related Work section is made up of reviews of academic literature that address similar work. The Methodology section contains a description of the research methods that are used, the type of equipment, and configurations used. The gathering and processing of the data is also described. Design Specification describes the architecture and methods that are used in the application of the research. The Implementation section describes the processed data, the programming language, code, the methods used and the output. In the Evaluation section, the results are assessed and evaluated both from an academic point of view and from a practical application point of view. The results are illustrated in the form of plots and graphs. In the the Discussion section, a critical analysis of the case studies is carried out. This analysis will examine both the strengths and weaknesses of the case studies, and compare them to earlier research as described in the Related Work section. The Conclusion and Future Work section describes the objectives, the research question, and the research carried out. The achievements with regard to the research question and the goals are described here. The contribution and usefulness of the research is described. Also, suggestions for future work is described here.

# 2 Related Work

In the Literature Review, the various studies have been grouped into subsections:

- Studies with Large Numbers of Models
- Studies with Small Numbers of Models
- Studies with Large Datasets
- Studies with Small Datasets
- Studies with Explainable AI
- Studies with Explainable AI and Image Data
- Summary of Related Work

## 2.1 Studies with Large Numbers of Models

In some studies a large number of ML models are used, and their results are measured to determine the model with the best performance.

In the study Kumeda et al. (2019) the ML models used were Random Forest (RF), Naïve Bayes, Multilayer Perceptron (MLP), Hierarchal LVQ and RBF Network (Radial Basis Function Network). This study used the XAI tool Fuzzy-FARCHD (FF) and a large number of ML methods to determine that the key features were lighting conditions, road class, and number of vehicles. FF provided an explainable model. The dataset used was obtained for 2016 from data.gov.uk. Also, a large number of evaluation metrics were

used and included recall, accuracy, F1-score, precision, kappa, absolute error root, mean, mean squared error, and confusion matrix.

In Aboulola et al. (2024) interpretability is addressed with the SHapley Additive exPlanations (SHAP) model. SHAP is an XAI that provides a global explanation. The ML models used are Long Short-Term Memory (LSTM), MLP, Residual Networks (ResNET), Convolutional Neural Network (CNN), InceptionV3, InceptionV3, Extreme Inception (Xception), EfficientNetB4, MobileNet (MN), AlexNet, and Visual Geometry Group (VGG19). SHAP is used in Aboulola et al. (2024) and it's output is illustrated in Fig. 1. For example, for feature Drug_involve, the long red line on the right indicates that drug consumption contributes to more critical road traffic accidents.

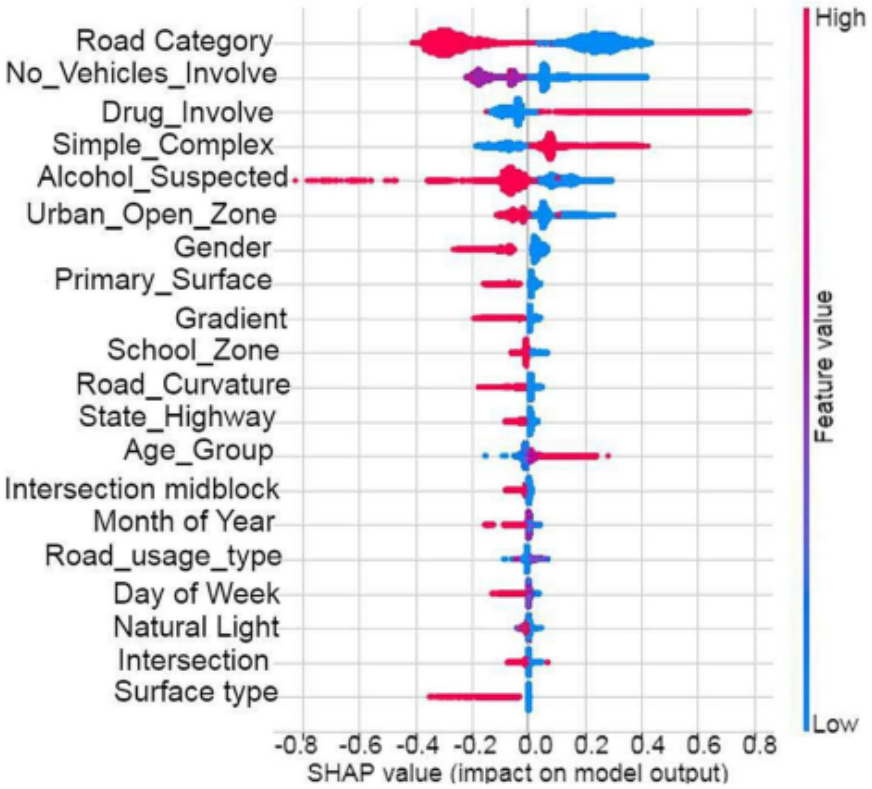


Figure 1: SHAP output

An accuracy of 98.17% is delivered by the MN method, in this research Aboulola et al. (2024). The accident data is from New Zealand for 2016-2020, and contains 378,820 records and has 101 features. The performance metrics used were recall, precision, accuracy and F1 score. The key features were Road Category, Number of vehicles involved, Drug consumption. In this study, a large number of models and a large amount of data used. This study addresses explainability with the SHAP XAI model.

## 2.2 Studies with Small Numbers of Models

A small number of models were used in Elawady et al. (2021) and these models consisted of Support Vector Machine (SVM), Artificial Neural Networks (ANN), and RF. The best prediction overall with an accuracy of 98% was delivered by Radial basis function (RBF) SVM model. The data used was from TranStar Houston's TMC, from 2004 to 2013. This

contained in excess of 119,000 records and over 53 features. In this case the performance metrics used were training time and accuracy. The value of the features was calculated using Weka's Gain Ratio. Here a high level of accuracy was achieved with the (RBF) SVM model on a large dataset with a large number of features.

In Boonserm and Wiwatwattana (2021) the ML models used are oversampling with SMOTE, oversampling with SMOTE and Random Undersampling, Random Undersampling, and RF. The best prediction of 83% recall and accuracy was achieved by the RF model that has been fitted with random undersampling. The dataset consists of New Year Festival traffic accidents from 2008 to 2015, and contains 17 columns and 214,950 records. Performance was measured using accuracy, recall, precision, confusion matrix and F1 score. The main features were determined to be delivering method, age, and alcohol drinking.

## 2.3 Studies with Large Datasets

An example of a work with a large dataset is Manzoor et al. (2021) where the dataset includes 49 states in the USA, from February 2016 to June 2020, and contains 49 columns and 4.2 million records. A combination of RF and CNN was used to form a solution called RFCNN. Other ML models used were RF, Gradient Boosting Machine (GBM), AdaBoost Classifier (AB), Extra Tree (ET), and Voting Classifier of ML models. In Manzoor et al. (2021) the process is illustrated in Fig. 2. The best results were delivered by RFCNN with 0.974 precision, 0.991 accuracy, an F1-score of 0.980 and 0.986 recall. Performance was determined by precision, accuracy, F1-score and recall. Feature importance was determined by RF, and the most important features were distance between vehicles, temperature, wind chill, humidity, visibility, and wind direction. In Manzoor et al. (2021) this is illustrated in Fig. 3. In this study a large number of ML models were used on a large dataset to result in high levels of accuracy.

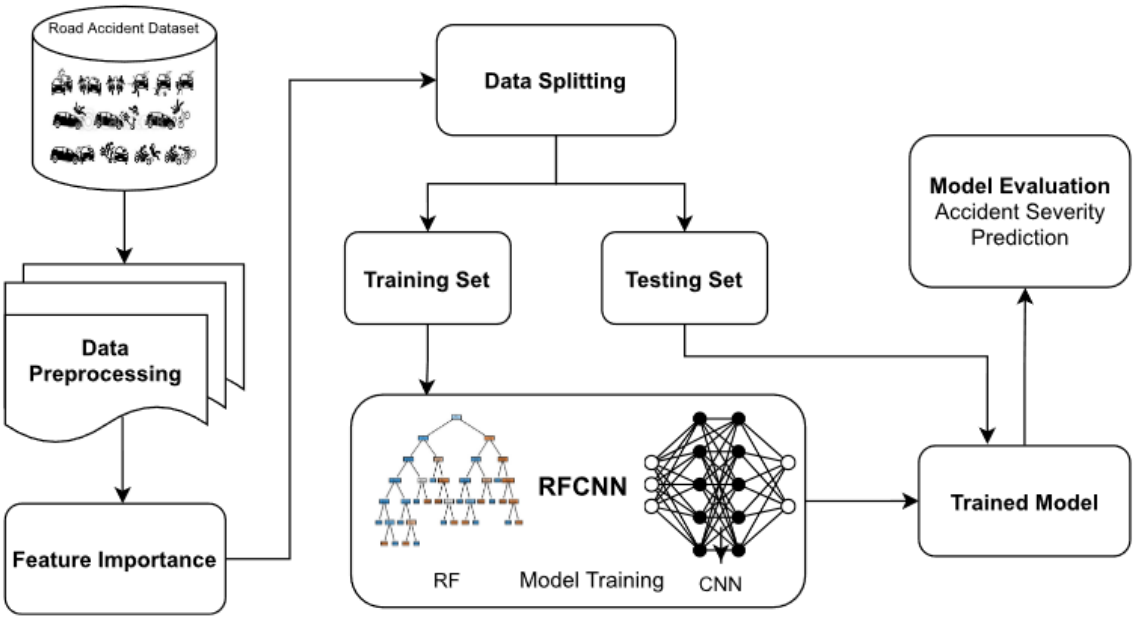


Figure 2: RFCNN Proposed Methodology

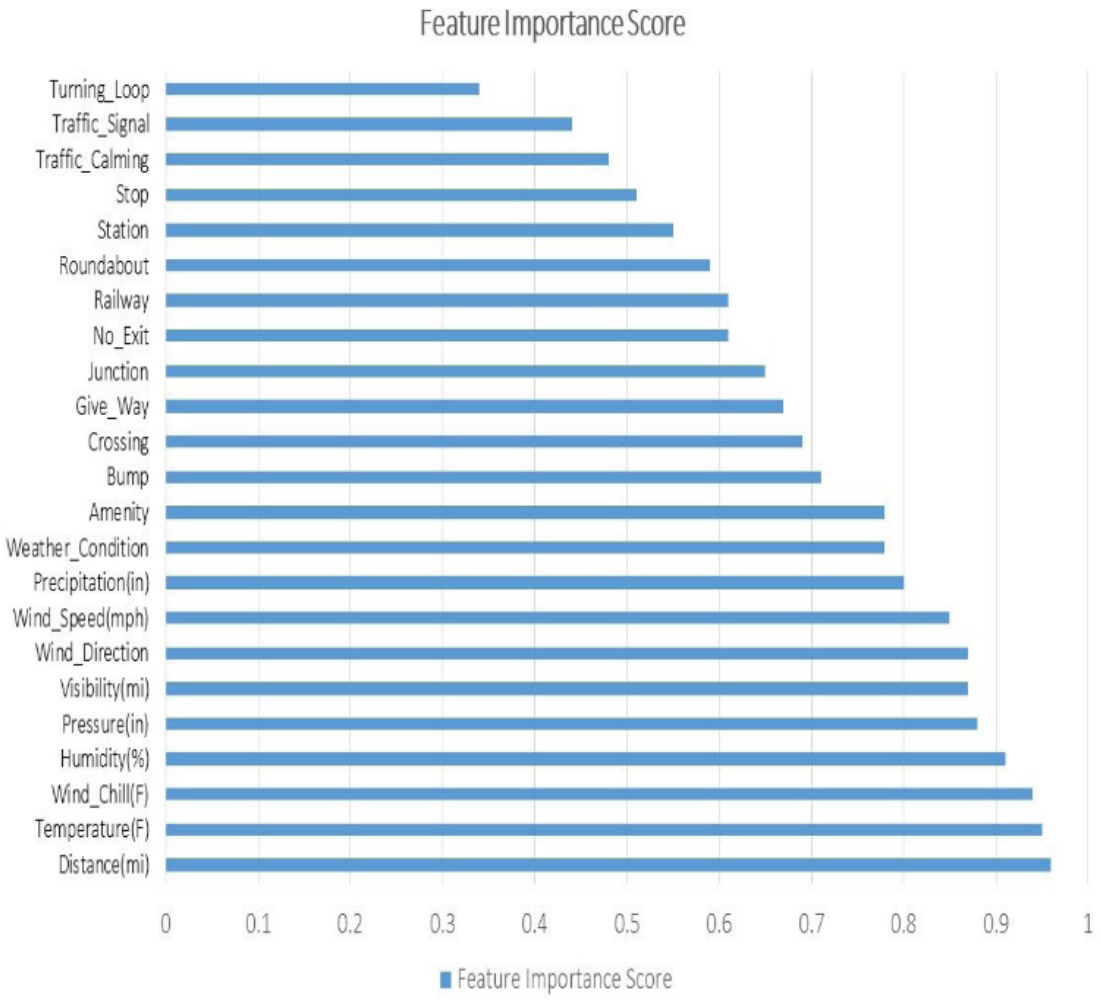


Figure 3: RFCNN Feature Importance

Another example of work with a large dataset is AlMamlook et al. (2019) where the data was from records of accidents in Michigan from 2010 to 2016, and contains 271,563 traffic accidents. The ML models used were RF, NB, Logistic Regression (LR), and AB. Data imbalance is managed with SMOTE. RF delivered the best results with 75.5% accuracy. There evaluation metrics used were recall, F1-score, precision, Area under Receiver Operating Characteristic Curve (AUC), and Receiver Operator Characteristic (ROC). In this study a large number of ML models were used on a large dataset but this did not result in a very high level of accuracy.

## 2.4 Studies with Small Datasets

An example of a study using a small dataset is Li et al. (2023), where the data was from records of a city in Guangdong Province, China in 2016, and it consisted of 1,045 records and 23 features. Here, a number of ML methods were used to achieve a good level of accuracy, and the XAI tool SHAP was used for explainable AI. The models used are LightGBM, SHAP as well as NB, Xgboost, RF and SVM. The best results were provided by LightGBM with an accuracy of 0.857. The metrics used were precision, accuracy, recall, confusion matrix and F1-score. The main features were calculated to be visibility,

road physics isolation and road alignment.

## 2.5 Studies with Explainable AI

In Vijayvargiya et al. (2023) the XAI tool LIME is used with a health care dataset. RF, LR, SVM, DT, and k Nearest Neighbour (kNN) are the classification models used. The health care dataset has 400 patients and 25 features. RF is the best performing model with precision of 98.5%, recall of 100%, F1-score of 99.28%, and accuracy of 99.17%. The accuracy is high. However, the dataset is quite small, and the study does not go into much detail on XAI.

The XAI model used in P et al. (2022) is LIME. The classification models used are SVM, RF, DT, and XGB. The health care dataset is made up of 769 rows and 9 features for Diabetes. However the dataset it very small. The best performing ML model is RF with an accuracy of 77%.

Two XAI models, LIME and SHAP are used in this work, Rao et al. (2022). However, NB is the only classifier is used. The health care dataset is very small and the test results are not described in detail. However, this study concludes that both LIME and SHAP are the most suitable XAI models for disease datasets.

LIME is used in Tiwari et al. (2024). The ML models are Adaboost, RF, DT, BGX, LR, kNN, NB, MLP and XGB. The training dataset has 45,211 rows and 18 columns, and the test dataset has 4521 rows and 18 columns. The best performing models are as follows. Adaboost has best precision at 97.9%. RF has best recall at 97%. Adaboost has best F1-score at 96.8%. Adaboost has best accuracy at 97.2%. The dataset is a good size, and the accuracy is good especially for Adaboost.

Pre-trained CNN modes obtained from the Tensorsflow implementation were used for a Transfer Learning model in Chayan et al. (2022). This study also uses LIME.

The work, Alodibat et al. (2023) uses DT, Secml and LIME. Secml is a Python library that is suitable for secure ML. DT achieved an accuracy of 94%.

A very large number of ML models and XAI methods are used in the area of health care. For example in Tjoa and Guan (2021), a number of ML models and XAI methods are surveyed. Models used include CNN, GDM, GAM, PCA (Principal Component Analysis), CCA (Canonical Correlation Analysis), TCAV (Testing with Concept Activation Vectors). XAI methods include Class Activation Map (CAM) which can produce heat maps, SHAP, Layer-wise Relevance Propagation (LRP), and Automatic Concept-based Explanations (ACE).

SHAP and Occlusion maps were used in Dissanayake et al. (2021). Occlusion maps help to understand how input data affects a correct prediction. In this case the methods used were Occlus (Occlusion map), Mel-Frequency Cepstral Coefficients (MFCC), LSTM, RNN, CNN, AdaBoost, XGB, SVM.

The work Hamilton et al. (2022) uses a particular version of LIME called Sub-model Stabilized and Sub-grid Superimposed LIME (SubLIME). SubLIME improves the stability and resolution of the XAI explanations. The ML method used is CNN.

It was found in Eriksson and Grov (2022), that ML methods have been used many times in Security Operation Centres (SOCs). However, due to limited interpretability and understanding of the output of the ML methods, there is a lack of trust, and therefore ML methods are not used as much as they could be. XAI tools can help with this, and in this study it was found that both XAI tools LIME and SHAP were found to be useful.

Another type of XAI tool called Model-agnostic SHAPley value explanations (MASHAP)

is assessed in Messalas et al. (2020). This paper says that LIME is slow and therefore might not be appropriate for industrial grade jobs, and MASHAP was found to be faster and more suitable.

Another work which uses both LIME and SHAP is Ashraf et al. (2024). The classification models used are RF, SVC, DT, kNN and NB. kNN provided the best performance. The health care dataset was obtained from Kaggle and has 132 parameters linked with 42 different diseases, with 133 columns in the dataset. Both LIME and SHAP were found to be useful. However, a more diverse dataset would be better.

### 2.6 Studies with Explainable AI and Image Data

A lot of work using LIME with image data has been carried out. In Anand et al. (2024) LIME and Gradient-weighted Class Activation Mapping (GradCAM) are used for XAI. ML models used include CNN, RNN, and LSTM. The dataset consists of 990 images and 2782 videos. The test subset contains 101 images, and the train subset contains 791 images. Both CNN and RNN are found to be useful, but further work is needed to improve the accuracy.

In this study, Sahay et al. (2021), ML methods are used for image captioning, and XAI is then used to explain the ML output instances. In this case the LIME is the XAI method used to explain the image output from the ML methods. LIME is very popular because it is model-agnostic.

As described in Ng et al. (2022), one of the main advantages with LIME is that it is model-agnostic. Also, LIME can be used on image data, tabular data or text data. However, there are stability issues with LIME.

LIME can be used to explain ML classification of images. In this work, Nikith et al. (2022), the images are flowers, and The LIME output consists of images for data instances that explain the predictions of the CNN model.

### 2.7 Summary of Related Work

The studies vary a great deal with regard to number and type of ML models, size of datasets, and whether they use XAI or not. Also, XAI can be used on images. The studies that work on health datasets often use SHAP or LIME. But no paper was found that uses LIME on a traffic accident dataset. Therefore this paper addresses the use of ML classification methods and LIME on traffic accident data. The Research Question is: How can the Explainable AI LIME be used to Improve the understanding of Machine Learning Predictions for Traffic Accidents?

## 3 Methodology

### 3.1 The Dataset and Pre-processing

The dataset contains traffic accident data for the state of Victoria in Australia. The dataset has 152,445 records and 15 features, and is 22.9 Mbyte in size. The name of the dataset is "Victoria Road Crash Data (2012-2023)", and the data file is a CSV file called Vic_Road_Crash_Data.csv. For the Python code, the dataset was renamed to vic_traffic_accident_data.csv. The dataset was sourced from Kaggle as described in Kaggle (2024).

The method for processing the data used in this research is Knowledge Discovery in Databases (KDD). There are a number of steps in the KDD process. Selection is the first step and this selects a set of data. The next step is Pre-processing which involves data cleaning and applying consistency to the data. The Transformation step modifies or transforms, where needed, the dimensionality of the data. The Data mining step involves examination of the data for patterns of interest, and an example of this could be prediction. The final step is Evaluation/Interpretation and this consists of evaluation and interpretation of the mined data. The diagram, Fig. 4, from study Tan et al. (2015), illustrates the KDD process.

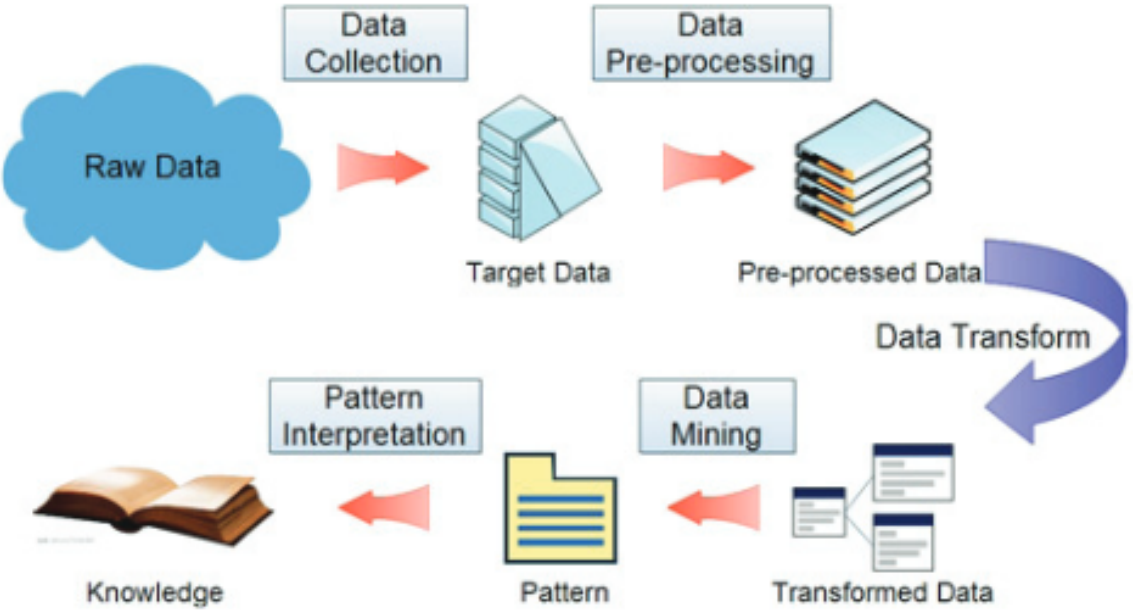


Figure 4: Knowledge Discovery in Databases (KDD) Process

The data in the dataset was cleaned and pre-processed as described in the dataset table in Fig. 7.

Histograms were plotted to visualize the data and they illustrated that there were outliers especially in the case of the features, LIGHT_CONDITION in Fig. 5 and ACCIDENT_TYPE in Fig. 6. Overfitting is where a model delivers good results on training data but does not perform well on the validation data. Removal of the outliers, where appropriate, could help to prevent overfitting. But there was no data available on the importance of these outliers, so I did not remove them from the data. However, this could reduce the accuracy of the ML and LIME results.

A large number of the features had categorical values. So pandas get_dummies is used to convert these to multiple features with 0.0/1.0 values, and this technique contributes to improved accuracy. This is described in the table in Fig. 7. Because of the large number of categorical values and the need to use pandas get_dummies to convert these, this resulted in a total of 129 columns. And so, to avoid having a much larger number of features, I did not take the option to process DATE or TIME as categorical values. Another option was to make DATE categorical and define it by month, and this might have improved accuracy. This would provide data on the time of the year and seasonal factors for example and might have improved accuracy. Also, TIME could be set to

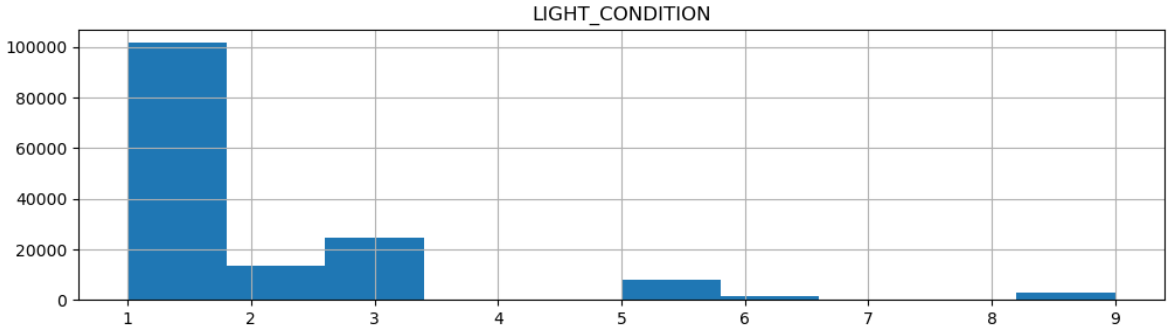


Figure 5: Histogram of LIGHT_CONDITION

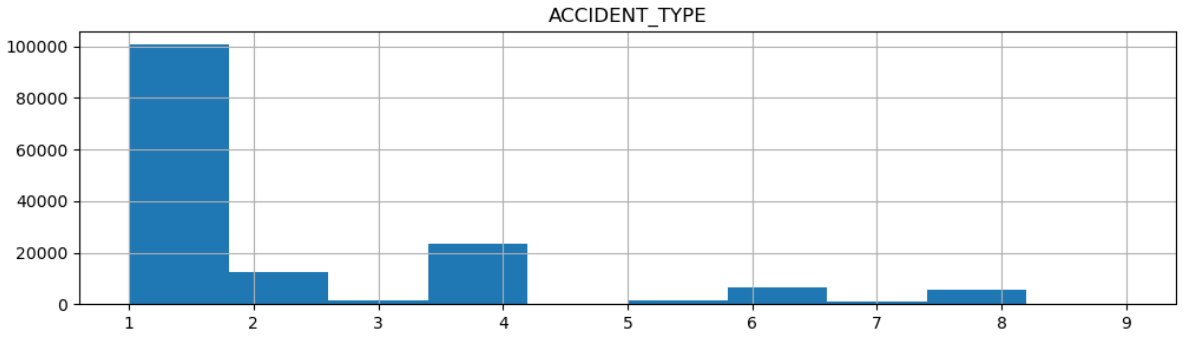


Figure 6: Histogram of ACCIDENT_TYPE

categorical by hour of day. I did not take the option to process DATE or TIME as categorical values, however this could be done in future work.

Pre-processing the large number of categorical values also resulted in creating very long feature names, which can be seen in the LIME graphical output, in Fig. 15.

The feature SEVERITY represents the severity of the accident, and has 4 separate values, 1, 2, 3, and 4. Of a total of 152,445 records, the value "4" occurs four times, and the value "1" occurs 2,599 times. Therefore, the records for values 1 and 4 are removed. This leaves two values 2 and 3, and these values are then converted to 0.0 and 1.0. A pie chart is drawn to display the balance between the two target values for the SEVERITY feature, in Fig. 8. Checks were made for Null values, using isnull.sum. StandardScalar from sklearn, in the Python code, is used to transform the data so that all the values fall between 0 and 1. This improves the accuracy by setting all the features to the same scale.

## 3.2 Machine Learning Classification Methods

The ML classification methods used for prediction of the SEVERITY of the traffic accident data are RF, LR, and kNN.

## 3.3 LIME XAI

The method used for XAI is LIME. The study Vijayvargiya et al. (2023) illustrates the LIME process in Fig. 9. First of all, an ML classification method, for example LR, is used to predict traffic accident severity using the traffic accident data in the dataset. LIME

| Feature | Action | Description |
|---|---|---|
| ACCIDENT_NO | drop | This is just an ID number. |
| ACCIDENT_DATE | drop | |
| ACCIDENT_TIME | | Convert from hours:minutes:seconds to seconds. |
| ACCIDENT_TYPE | drop | |
| ACCIDENT_TYPE_DESC | | Categorical. Use pandas get_dummies to convert. These values are unbalanced as can be seen in "LIGHT_CONDITION" histogram. |
| DAY_OF_WEEK | drop | |
| DAY_WEEK_DESC | | Categorical. Use pandas get_dummies to convert. |
| DCA_CODE | drop | |
| DCA_DESC | | Categorical. Use pandas get_dummies to convert. |
| LIGHT_CONDITION | | Categorical. Change numerical values to text values, and then use pandas get_dummies to convert. These values are unbalanced as can be seen in "LIGHT_CONDITION" histogram. |
| NODE_ID | drop | |
| ROAD_GEOMETRY_DESC | | Use pandas get_dummies to convert from categorical to 0/1. |
| SEVERITY | move | This is the target value. Move column to end. |
| SEVERITY | | Initially there are 4 classifications. Classification 1 and 4 are 2,599 records out of a total of 152,445 records. Reduce number of classifications to 2 by deleting records with class 1 or 4. Change values 2/3 to 0.0/1.0. |
| SPEED_ZONE | | Categorical. Change numerical values to text values, and then use pandas get_dummies to |
| RMA | | Use pandas get_dummies to convert from categorical to 0/1. |

Figure 7: Dataset Features

is then used to explain the prediction made by the ML method for selected instances or records. An example of LIME graphical output is illustrated in Fig. 11. In this way, LIME helps to understand why the ML method has made a given prediction. LIME can be used to explain the prediction of any selected instance or record, and so if the classifier makes an incorrect prediction, the LIME output can be used to help understand why the ML method made an incorrect prediction. The LIME output can be in the form of a graphical output, for example, a barchart, or in the form of an image.

LIME provides an explanation for each separate instance of an ML prediction, but LIME does not provide a global explanation for the ML predictions. (SHAP, on the other hand, does provide a global explanation). LIME can consume a lot of processing power, especially if the dataset contains a lot of categorical values which results in many features. The graphical output of LIME may need to be interpreted by an expert in the subject matter, to understand how the separate features interact. When there are many features, especially when categorical values must be transformed, the number of features that can be plotted is limited by the space available for the plot. However, while keeping

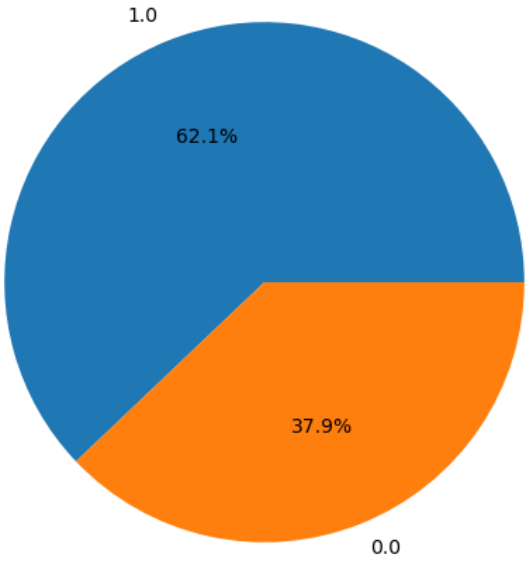


Figure 8: Pie Chart for SEVERITY Target Value

these limitations in mind, LIME does provide explanations of any selected range of ML prediction instances and so can provide valuable insight and understanding of the data.

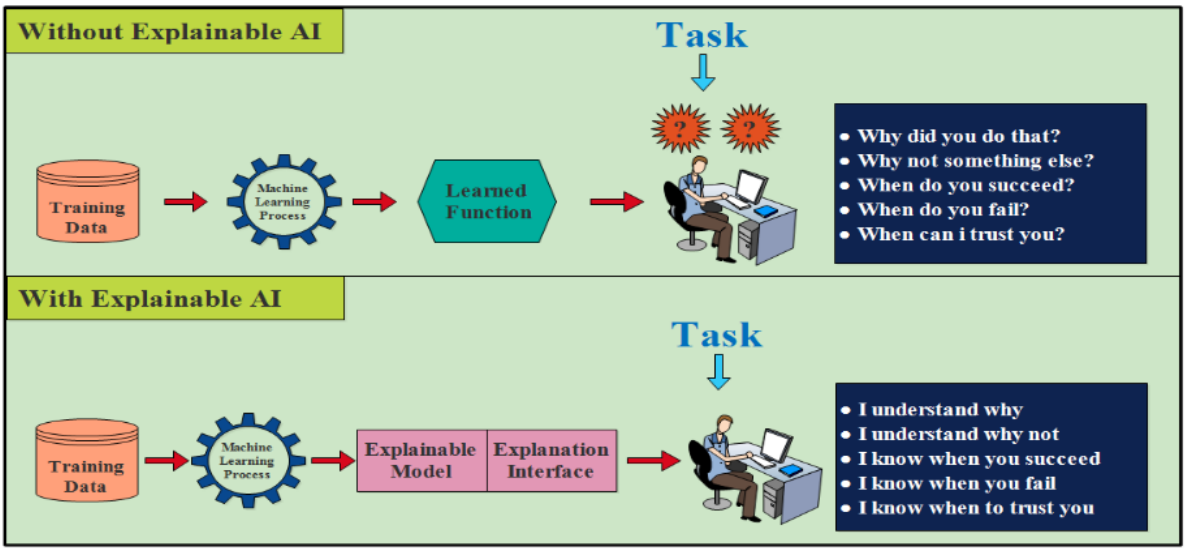


Figure 9: Flow of LIME XAI

# 4 Design Specification

The ML classification methods LR, kNN, and RF, are used to predict traffic accident severity using the traffic accident data in the dataset. The XAI method LIME is then used to explain the predictions made by the ML methods for selected instances or records.

# 5 Implementation

Python 3.9 is the programming language used in this research. The development environment used is Jupyter Notebook 7.0.8 with Anaconda Navigator 2.6.2. The Python software libraries used are pandas 2.1.4, numpy 1.26.4, matplotlib 3.8.4, scikit-learn 1.4.2, and lime 0.2.0.1. The computer used is a Dell laptop with 16 GB ram, with an Intel i7 cpu, running 64-bit Windows 10.

The accuracy of the results can be affected by overfitting. Overfitting occurs where a model delivers good results on training data but does not perform well on the validation data. Bagging and Feature Selection are two methods that can be used to reduce overfitting. Bagging is a method that carries out estimates on a number of random subsets and then combines all of these predictions to arrive at a prediction. Feature Selection could be used to remove the least important features, and so reduce overfitting.

## 5.1 Case Study: Logistic Regression and LIME

In this case the ML prediction method used is LR. There are limitations to LR. If the target feature is imbalanced then LR can be biased towards the larger value. However, in this case the target value, SEVERITY, is quite well balanced. This is illustrated in the Pie chart for SEVERITY in Fig. 8. Also, outliers in the feature values can affect LR performance. For LR, Feature Selection could be used to remove the least important features, and so reduce overfitting, but they were not used, however this method could be used in future work.

The recall for LR is 0.8664. Further details on ML evaluation can be seen in Fig. 10. Then LIME is used to explain the predictions in graphical format. Output for LR and LIME is displayed in the diagrams in Fig. 11 and Fig. 12. The LIME plot illustrates the relative importance of the various features.

| | Logistic Regression | k Nearest Neighbour | Random Forest |
|---|---|---|---|
| **Recall** | 0.8664 | 0.7311 | 0.6790 |
| **F1-score** | 0.7468 | 0.6892 | 0.6620 |
| **Precision** | 0.6562 | 0.6519 | 0.6459 |
| **Accuracy** | 1.0 | 1.0 | 0.5686 |
| **AUC-ROC** | 0.5593 | 0.5449 | 0.5329 |

Figure 10: Machine Learning model performance values

## 5.2 Case Study: k Nearest Neighbour and LIME

In this case the ML prediction method used is kNN. kNN has some limitations. Performance can be affected when the number of features is large. kNN can be affected by outliers. kNN is usually regarded as a black-box model; however, LIME helps to address this. Bagging and Feature Selection could reduce overfitting, but they were not used, however these methods could be used in future work.

The recall for kNN is 0.7311. Further details on ML evaluation can be seen in Fig. 10. Then LIME is used to explain the predictions in graphical format. Output for kNN and LIME is displayed in the diagrams Fig. 13 and Fig. 14.

### 5.3 Case Study: Random Forest and LIME

In this case the ML prediction method used is RF. There are some limitations to RF. RF is an ensemble of decision trees, and this reduces overfitting. However, overfitting can occur for a big dataset. Bagging and Feature Selection could reduce overfitting, but they were not used, however these methods could be used in future work.

Like LR, if the target feature is imbalanced then RF can be biased towards the larger value. However, in this case the target value, SEVERITY, is quite well balanced and this is illustrated in the Pie chart for SEVERITY in Fig. 8. RF is usually regarded as a black-box model; however, LIME helps to address this. Again, like LR, outliers in the feature values can affect RF performance.

The recall for RF is 0.6790. Further details on ML evaluation can be seen in Fig. 10. Then LIME is used to explain the predictions in graphical format. Output for RF and LIME is displayed in the diagrams Fig. 15 and Fig. 16.

## 6 Evaluation

The ML methods used are RF, LR and kNN. LIME is used for XAI. The recall for LR is 0.8664, the recall for kNN is 0.7311, and the recall for RF is 0.6790. The performance of the three ML models varied, with LR providing the best recall score. The performance of RF was not very good. Further details on ML evaluation can be seen in Fig. 10. The XAI tool LIME generated plots to explain the predictions of the ML methods. Examples of these plots are the outputs for LR and LIME which are displayed in the diagrams Fig. 11 and Fig. 12. So, in this case the most important two features are "LIGHT_CONDITION_Nine" and "DCA_DESC_OFF END OF ROAD/T-INTERSECTION.". Also, for this instance, the "SPEED_ZONE_Thirty_kph" is plotted near the bottom of the list.

These plots, in the form of bar charts, illustrate the relative importance of the features of the dataset. LIME can give these explanations for any selected set of instances or records in the dataset. This information can then be used to determine what features have the greatest impact on a traffic accident, and so this can greatly help actions to mitigate accidents in the future. For example, if the most important feature illustrated by LIME was the road type, then plans could be made to upgrade the road.

### 6.1 Case Study: Logistic Regression and LIME

In this case the ML prediction method used is LR. The recall for LR is 0.8664. Further details on ML evaluation can be seen in Fig. 10. Then LIME is used to explain the predictions in graphical format. Output for LR and LIME is displayed in the diagrams Fig. 11 and Fig. 12.

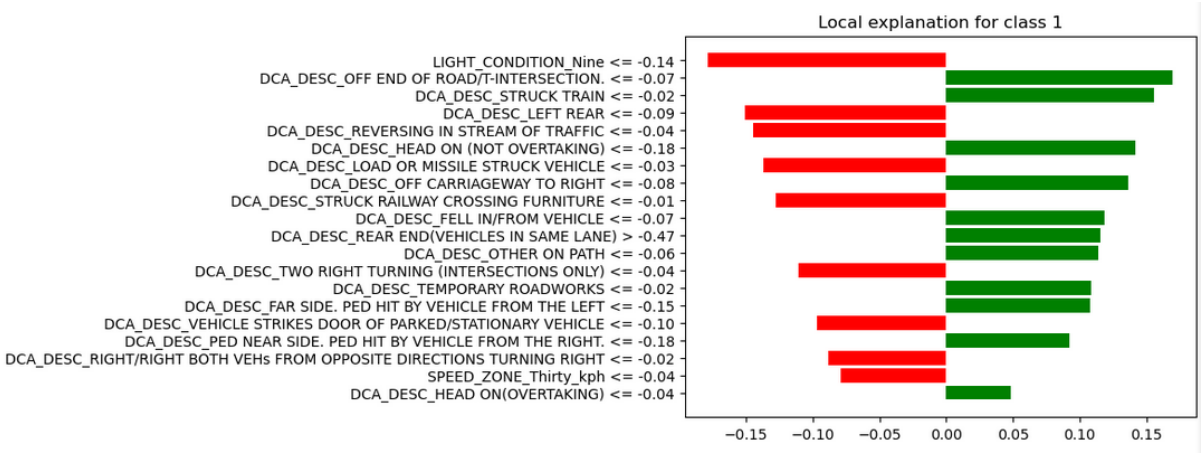


Figure 11: Local Explanation A, using LIME, for LR Prediction

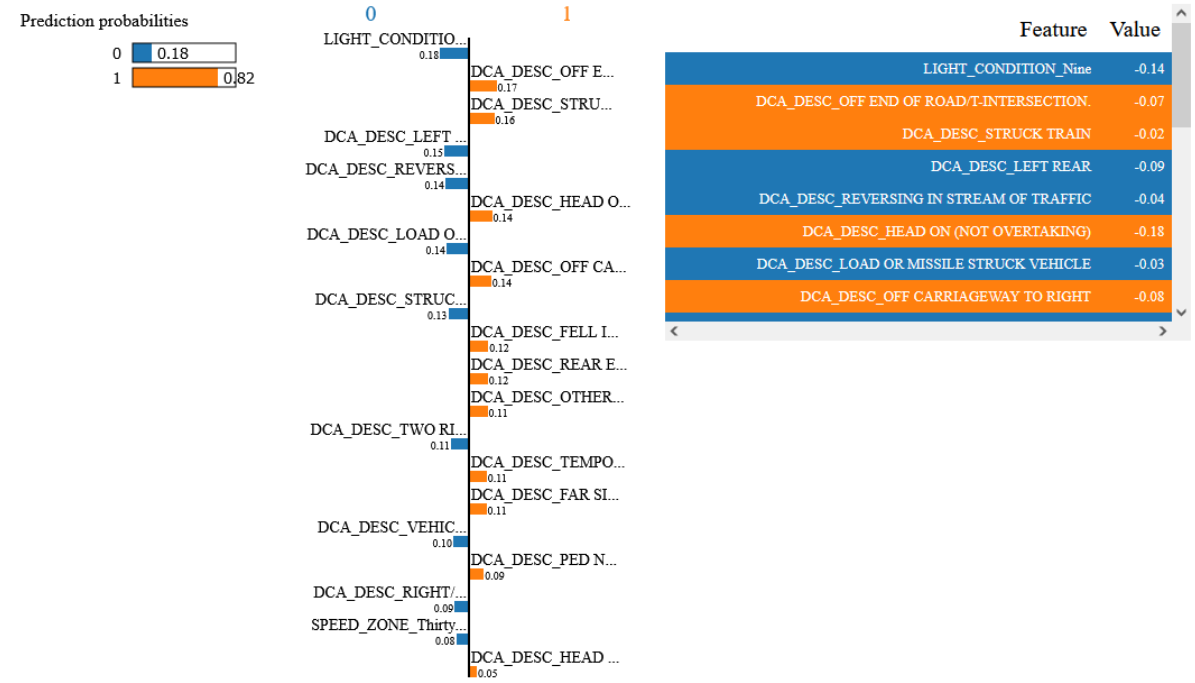


Figure 12: Local Explanation B, using LIME, for LR Prediction

## 6.2 Case Study: k Nearest Neighbour and LIME

In this case the ML prediction method used is kNN. The recall for kNN is 0.7311. Further details on ML evaluation can be seen in Fig. 10. Then LIME is used to explain the predictions in graphical format. Output for kNN and LIME is displayed in the diagrams Fig. 13 and Fig. 14.

## 6.3 Case Study: Random Forest and LIME

In this case the ML prediction method used is RF. The recall for RF is 0.6790. Further details on ML evaluation can be seen in Fig. 10. Then LIME is used to explain the predictions in graphical format. Output for RF and LIME is displayed in the diagrams Fig. 15 and Fig. 16.

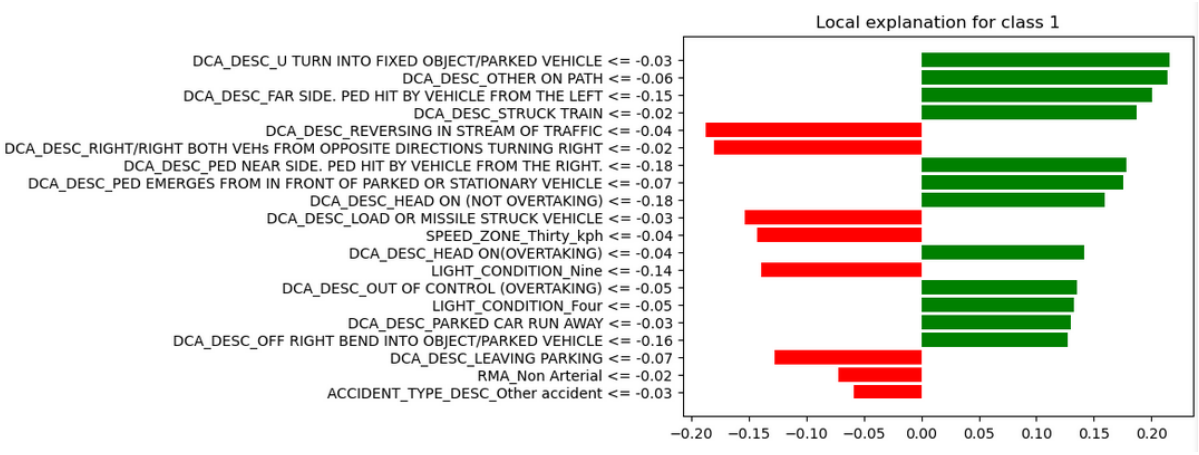


Figure 13: Local Explanation A, using LIME, for kNN Prediction

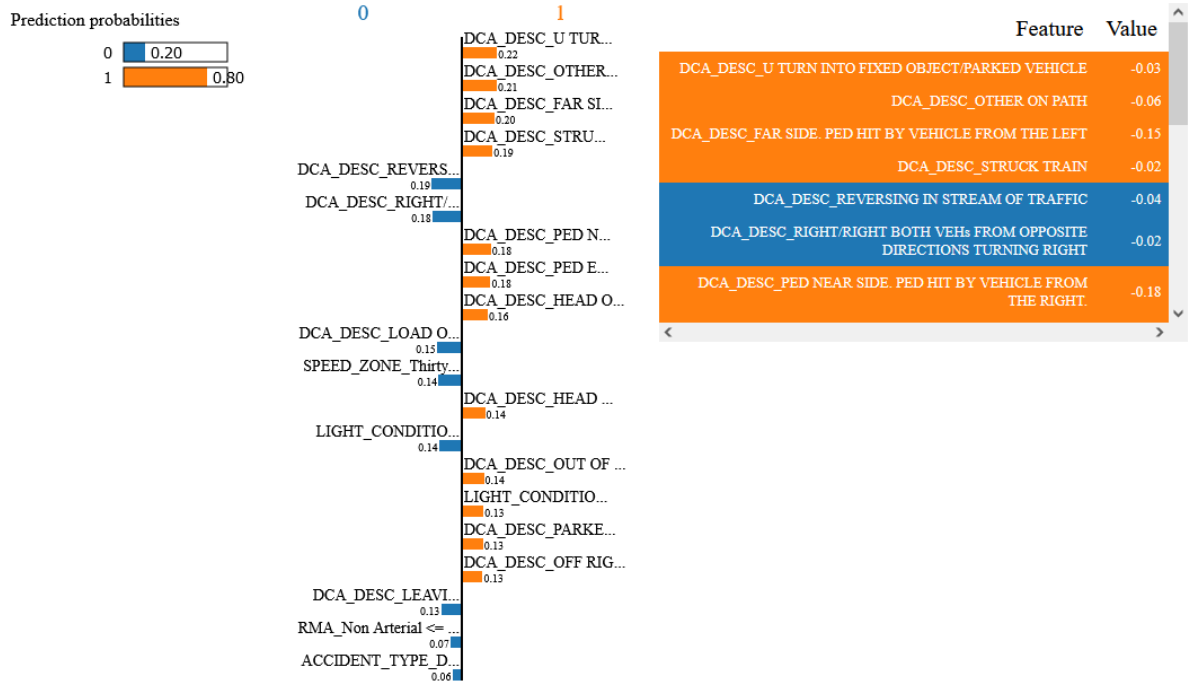


Figure 14: Local Explanation B, using LIME, for kNN Prediction

## 6.4 Discussion

In this case three ML methods were used to demonstrate the capability of the LIME XAI method. However, there are a large number of other ML classification methods that could be used. XAI methods including LIME have been used in other areas, for example health care. However, in my search I did not find any papers that used the XAI tool LIME to help to explain the ML predictions for traffic accident data.

A large dataset was used, which contained 152,000 records. In the case of the SEVERITY feature, there were four values, 1, 2, 3, and 4. However, there were only four instances of the value 4, and the value 2 occurred just 2,500 times out of a total of 152,000 records. Therefore the values 1 and 4 were removed from the SEVERITY feature in the dataset. Also, a number of features were dropped as described in the table in Fig. 7. In future

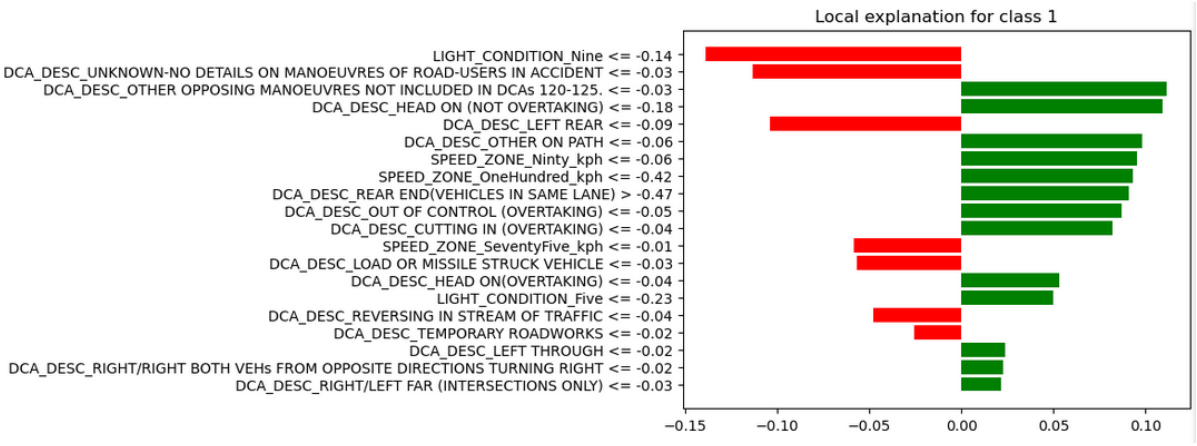


Figure 15: Local Explanation A, using LIME, for RF Prediction

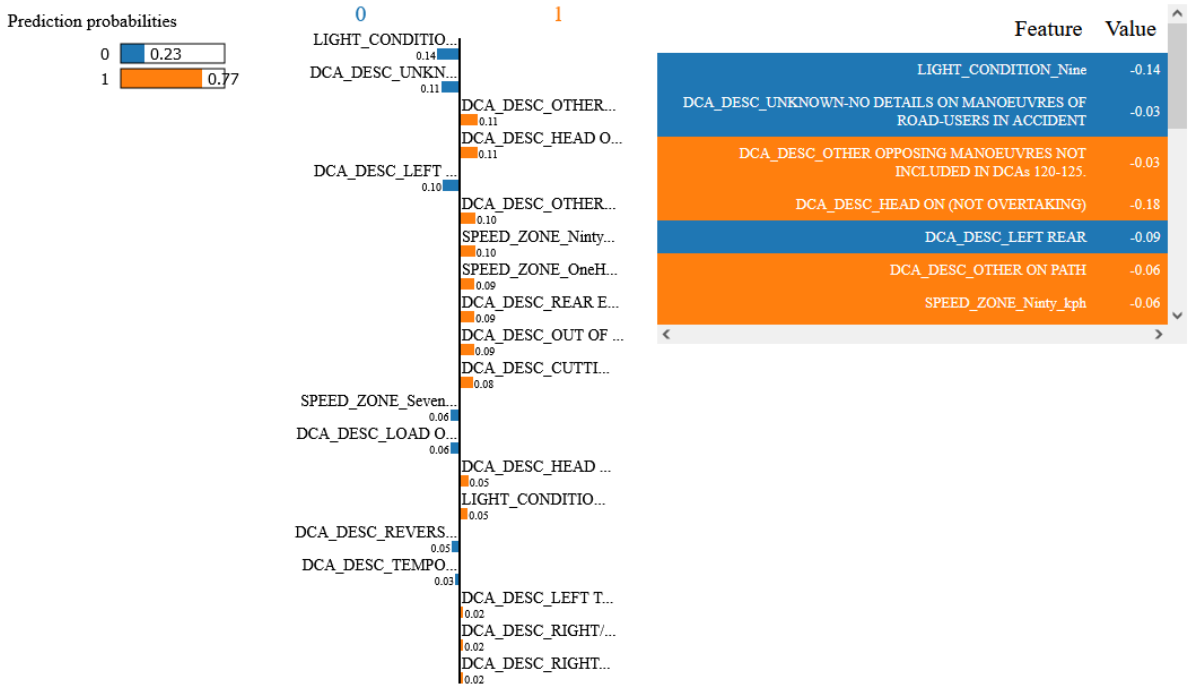


Figure 16: Local Explanation B, using LIME, for RF Prediction

work, maybe some combination of these features could be included.

# 7 Conclusion and Future Work

In a World Health Organization (WHO) study road traffic accidents were found to account for more than 1.3 million deaths per year, Aboulola et al. (2024). ML methods can be used to predict road traffic accidents based on the parameters or features that apply. And ML methods can also determine the relative importance of the features. This data can then be used to focus on the key causes of accidents and so reduce the severity and frequency of accidents. The research question is: How can Explainable AI LIME be used to Improve the understanding of Machine Learning Predictions for Traffic Accidents? XAI can be used to explain the results provided by the ML methods. In Aboulola et al.

(2024) a large number of ML models are used and interpretability is addressed using the SHAP ML model. A number of studies used LIME for health care data, but not in the area of traffic accidents. So, in this paper I focus on the use of the XAI tool LIME to help to explain the predictions of ML methods. Because LIME is model-agnostic it can be used to explain the predictions of a large number of ML classifiers. In this paper the ML methods used are LR, kNN, and RF. For the traffic accident dataset in this study it was found that there were a large number of categorical features. And after using pandas get_dummies, this resulted in more than 100 features.

Future work needs to be done to explore other XAI methods such as SHAP. SHAP was used in many of the studies including traffic accident studies, and I think that it would be very interesting to compare and contrast how the two XAI methods, LIME and SHAP, contribute to a better understanding of the ML predictions. LIME is an XAI that is local and so can help to explain each selected local instance of an ML prediction. Whereas SHAP is a global XAI and is good at explaining the global values. Perhaps a combination of the two methods, LIME and SHAP, used together would complement each other, with LIME providing local explanations and SHAP providing a global explanation. It would also be interesting to carry out future work on other available datasets and it would be good to have a dataset that also included data on the vehicle, for example the condition of the tyres, and the length of skid marks to indicate amount of braking.

The three ML methods were applied to a large dataset containing more than 140,000 records, and LIME was then able to explain the ML predictions, for any chosen instance or range of instances. With this greater understanding of the ML predictions, the most important factors contributing to traffic accidents can be better understood, and can then be used to mitigate traffic accident risks and reduce death and injury on the roads.

# References


Aboulola, O. I., Alabdulqader, E. A., AlArfaj, A. A., Alsubai, S. and Kim, T.-H. (2024). An automated approach for predicting road traffic accident severity using transformer learning and explainable ai technique, *IEEE Access* **12**: 61062–61072.

AlMamlook, R. E., Kwayu, K. M., Alkasisbeh, M. R. and Frefer, A. A. (2019). Comparison of machine learning algorithms for predicting traffic accident severity, *2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT)*, pp. 272–276.

Alodibat, S., Ahmad, A. and Azzeh, M. (2023). Explainable machine learning-based cybersecurity detection using lime and secml, *2023 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT)*, pp. 235–242.

Anand, I., N, K. T., Charitha, P. S., Kodipalli, A. and Rao, T. (2024). Accident detection using images and videos with cnn, lstm, and interpreting the results using lime gradcam, *2024 IEEE 9th International Conference for Convergence in Technology (I2CT)*, pp. 1–8.

Ashraf, K., Nawar, S., Hosen, M. H., Islam, M. T. and Uddin, M. N. (2024). Beyond the black box: Employing lime and shap for transparent health predictions with

machine learning models, *2024 International Conference on Advances in Computing, Communication, Electrical, and Smart Systems (iCACCESS)*, pp. 1–6.

Boonserm, E. and Wiwatwattana, N. (2021). Using machine learning to predict injury severity of road traffic accidents during new year festivals from thailand's open government data, *2021 9th International Electrical Engineering Congress (iEECON)*, pp. 464–467.

Cervantes, E. G. and Chan, W.-Y. (2021). Lime-enabled investigation of convolutional neural network performances in covid-19 chest x-ray detection, *2021 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*, pp. 1–6.

Chayan, T. I., Islam, A., Rahman, E., Reza, M. T., Apon, T. S. and Alam, M. G. R. (2022). Explainable ai based glaucoma detection using transfer learning and lime, *2022 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE)*, pp. 1–6.

Dissanayake, T., Fernando, T., Denman, S., Sridharan, S., Ghaemmaghami, H. and Fookes, C. (2021). A robust interpretable deep learning classifier for heart anomaly detection without segmentation, *IEEE Journal of Biomedical and Health Informatics* **25**(6): 2162–2171.

Elawady, A., Khetrish, A. and Hamad, K. (2021). Predicting traffic incident severity level using machine learning, *2021 14th International Conference on Developments in eSystems Engineering (DeSE)*, pp. 432–437.

Eriksson, H. S. and Grov, G. (2022). Towards xai in the soc – a user centric study of explainable alerts with shap and lime, *2022 IEEE International Conference on Big Data (Big Data)*, pp. 2595–2600.

Hamilton, N., Webb, A., Wilder, M., Hendrickson, B., Blanck, M., Nelson, E., Roemer, W. and Havens, T. C. (2022). Enhancing visualization and explainability of computer vision models with local interpretable model-agnostic explanations (lime), *2022 IEEE Symposium Series on Computational Intelligence (SSCI)*, pp. 604–611.

Kaggle (2024). Victoria road crash data (2012-2023). `https://www.kaggle.com/datasets/jaspreetkhokhar/victoria-road-crash-data-2012-2023`, Last accessed on 2024-08-11.

Kamal, M. S., Northcote, A., Chowdhury, L., Dey, N., Crespo, R. G. and Herrera-Viedma, E. (2021). Alzheimer's patient analysis using image and gene expression data and explainable-ai to present associated genes, *IEEE Transactions on Instrumentation and Measurement* **70**: 1–7.

Kumeda, B., Zhang, F., Zhou, F., Hussain, S., Almasri, A. and Assefa, M. (2019). Classification of road traffic accident data using machine learning algorithms, *2019 IEEE 11th International Conference on Communication Software and Networks (ICCSN)*, pp. 682–687.

Li, J., Guo, Y., Li, L., Liu, X. and Wang, R. (2023). Using lightgbm with shap for predicting and analyzing traffic accidents severity, *2023 7th International Conference on Transportation Information and Safety (ICTIS)*, pp. 2150–2155.

Manzoor, M., Umer, M., Sadiq, S., Ishaq, A., Ullah, S., Madni, H. A. and Bisogni, C. (2021). Rfcnn: Traffic accident severity prediction based on decision level fusion of machine and deep learning model, *IEEE Access* **9**: 128359–128371.

Messalas, A., Aridas, C. and Kanellopoulos, Y. (2020). Evaluating mashap as a faster alternative to lime for model-agnostic machine learning interpretability, *2020 IEEE International Conference on Big Data (Big Data)*, pp. 5777–5779.

Ng, C. H., Abuwala, H. S. and Lim, C. H. (2022). Towards more stable lime for explainable ai, *2022 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)*, pp. 1–4.

Nikith, B. V., Nikhil, M. T., Sri Siddhartha, M. S. and Murali, K. (2022). Lime explainability on flower classification, *2022 6th International Conference on Computation System and Information Technology for Sustainable Solutions (CSITSS)*, pp. 1–4.

P, N., V, M., A, D., K, B., M, A. and C, R. (2022). A prediction and recommendation system for diabetes mellitus using xai-based lime explainer, *2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS)*, pp. 1472–1478.

Rao, S., Mehta, S., Kulkarni, S., Dalvi, H., Katre, N. and Narvekar, M. (2022). A study of lime and shap model explainers for autonomous disease predictions, *2022 IEEE Bombay Section Signature Conference (IBSSC)*, pp. 1–6.

Sahay, S., Omare, N. and Shukla, K. K. (2021). An approach to identify captioning keywords in an image using lime, *2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)*, pp. 648–651.

Tan, Y., Zhang, C., Ma, Y. and Mao, Y. (2015). Knowledge discovery in databases based on deep neural networks, *2015 IEEE 10th Conference on Industrial Electronics and Applications (ICIEA)*, pp. 1217–1222.

Tiwari, U., Ashwani, S., Tripathy, A. J. and Kumar, K. D. (2024). A two-stage ensemble approach for analysis of optimizing customer churn with lime interpretability, *2024 IEEE International Conference on Computing, Power and Communication Technologies (IC2PCT)*, Vol. 5, pp. 1540–1545.

Tjoa, E. and Guan, C. (2021). A survey on explainable artificial intelligence (xai): Toward medical xai, *IEEE Transactions on Neural Networks and Learning Systems* **32**(11): 4793–4813.

Vijayvargiya, A., Raghav, A., Bhardwaj, A., Gehlot, N. and Kumar, R. (2023). A lime-based explainable machine learning technique for the risk prediction of chronic kidney disease, *2023 International Conference on Computer, Electronics Electrical Engineering their Applications (IC2E3)*, pp. 1–6.