CHICAGO CAR CRASH PROJECT - ML

AGENDA

INTRODUCTION

SUMMARY

OUTLINE

BUSINESS PROBLEM

DATA

TECHNIQUES

RESULTS AND FINDIGS

RECOMMENDATION

MODELLING

REMARKS

PROBLEM STATEMENT

- The primary objective of this study is to identify and understand the main contributing factors to car accidents in Chicago, with a specific focus on driver behavior.

- By analyzing various modeling techniques and their results, the aim is to determine the primary causes of car accidents and provide insights that can contribute to the development of effective strategies for accident prevention and road safety improvement.

DATA

Data in this case is obtained from https://data.cityofchicago.org/Transportation/Traffic-Crashes-Crashes/85ca-t3if

TECHNQUES

Data Collection

Data Preprocessing:

Exploratory Data Analysis

Feature Selection

Modeling Techniques:

Regression Analysis:

Classification Analysis:

Clustering Analysis

Model Evaluation

Interpretation and Recommendations

RESULTS

RESULT

INTRODUCTION

SUMMARY

OUTLINE

BUSINESS PROBLEM

DATA

TECHNIQUES

RESULTS AND FINDIGS

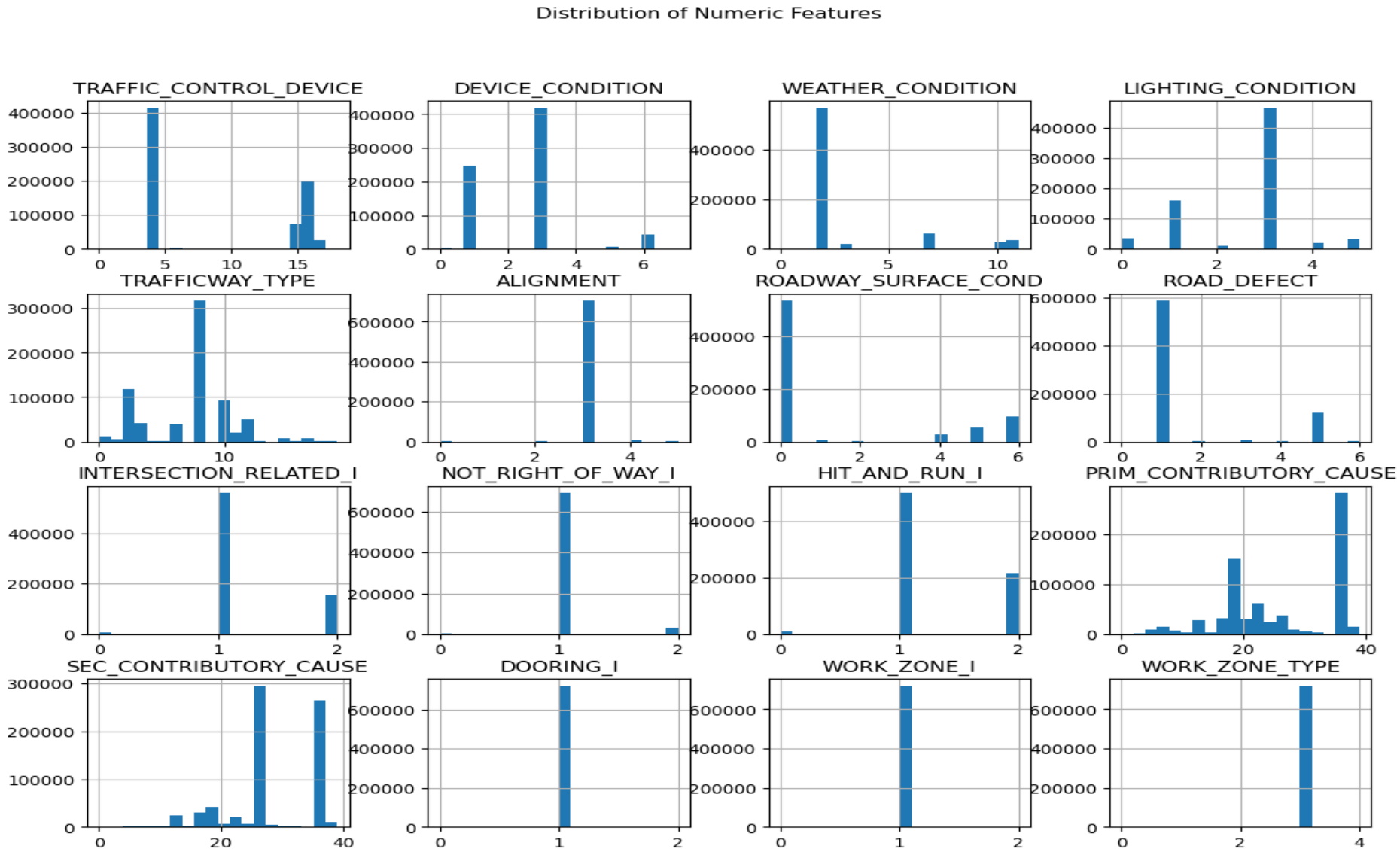
RECOMMENDATION

MODELLING

REMARKS

VISUALISATIONS

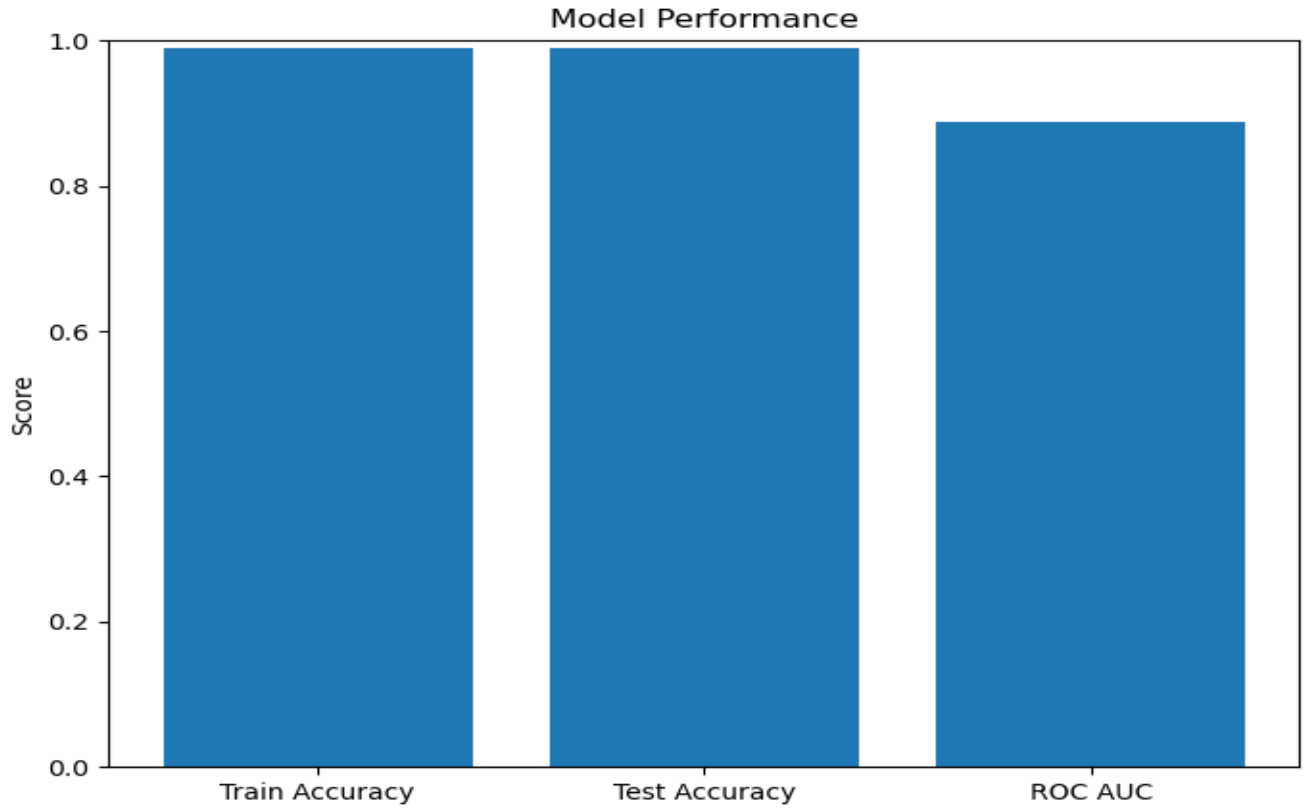
Some of the visualizations done



MODELLING: RANDOMCLASSIFER

RandomClassifer has been used to train and test
Train Accuracy: 0.9907520128378939
Test Accuracy : 0.9907520128378939
ROC Curve: 0.8871933925078359

RandomClassifier Model Bar Graph



MODELLING:KNN XGBOOST

Produces the poorest model results

```
from sklearn.metrics import roc_auc_score
from sklearn.preprocessing import LabelBinarizer

# Convert the target arrays to one-hot encoded format
lb = LabelBinarizer()
y_test_one_hot = lb.fit_transform(y_test)
y_test_pred_one_hot = lb.transform(y_test_pred)

# Calculate ROC AUC score
roc_auc = roc_auc_score(y_test_one_hot, y_test_pred_one_hot, multi_class='ovr')
print(f'ROC AUC Test: {roc_auc}')
```

✓ 0.8s Python

```
ROC AUC Test: 0.49985988434849027
```

MODELLING: KNN:GRIDSEARCH

- The accuracy score represents the proportion of correctly classified instances out of the total number of instances in the test set. In this case, the best KNN model achieved an accuracy score of approximately 0.9805, which means that it correctly predicted the class labels for approximately 98.05% of the instances in the test data.

```
# Evaluate the model
accuracy = accuracy_score(y_test, y_test_pred)
print(f"Accuracy Score (Best Model): {accuracy}")
```

```
Accuracy Score (Best Model): 0.9805079821818886
```

CLASSIFICATION REPORT

```python
from sklearn.metrics import classification_report

# Calculate and print the classification report
classification_rep = classification_report(y_test, y_test_pred)
print("Classification Report (KNN):\n", classification_rep)
```

```
Classification Report (KNN):
              precision    recall  f1-score   support

           0       0.97      0.92      0.94       303
           1       1.00      0.21      0.35        43
           2       0.85      0.86      0.86       480
           3       0.87      0.83      0.85       767
           4       0.89      0.77      0.82       451
           5       0.96      0.98      0.97      4061
           6       0.96      0.99      0.97      7033
           7       0.91      0.23      0.36       127
           8       0.93      0.95      0.94      2528
           9       0.94      0.88      0.91      1512
          10       0.97      0.54      0.70       162
          11       0.96      0.81      0.88      1935
          12       0.94      1.00      0.97     11726
          13       0.96      0.89      0.92      2269
          14       0.95      0.77      0.85       690
          15       0.95      0.77      0.85      1002
          16       0.90      0.57      0.70       893
          17       0.96      0.96      0.96     15259
          24       0.98      0.97      0.98     12011
          25       0.00      0.00      0.00         9
          26       0.99      1.00      0.99     18816
          27       1.00      0.15      0.26        34
          28       0.98      0.97      0.97      4589
          29       1.00      0.21      0.35        38
          30       0.94      0.97      0.95      2228
          31       0.99      0.43      0.60       171
          32       0.94      0.80      0.87       832
          33       0.89      0.88      0.89       927
          34       0.92      0.47      0.63       160
          35       0.91      0.53      0.67       241
          36       0.99      1.00      1.00    139472
          37       0.97      0.77      0.86      1793
          38       0.96      0.86      0.91      2099
          39       0.97      0.98      0.98      5685

    accuracy                           0.98    361430
   macro avg       0.93      0.76      0.81    361430
weighted avg       0.98      0.98      0.98    361430
```

CONCLUSION

- Based on the different models conducted and the analysis of their results, one of the primary causes for car accidents in Chicago could be identified as the "Driver's Behavior."- Several factors related to driver behavior, such as speeding, reckless driving, distracted driving, and impaired driving, consistently appeared as significant features in the models and had a strong impact on crash outcomes.
- These findings suggest that driver-related factors play a crucial role in contributing to car accidents in Chicago..

RECOMMENDATION

- - By analyzing the models' feature importance and coefficients, it can be inferred that addressing driver behavior through targeted interventions, awareness campaigns, and stricter enforcement of traffic regulations could potentially help mitigate the occurrence of car accidents in Chicago.

- - However, it's important to note that the primary cause of accidents can vary based on various factors such as the dataset used, the modeling techniques employed, and the specific context of the analysis.

- - It is recommended to consider a comprehensive approach that takes into account multiple factors, including road infrastructure, weather conditions, and other external influences, to gain a holistic understanding of the primary causes of car accidents in Chicago.