

Traffic Crash Cause Prediction

Empowering data-driven road safety decisions

Presented by: Gerald Wanjala

Date: June 2025

Problem Statement

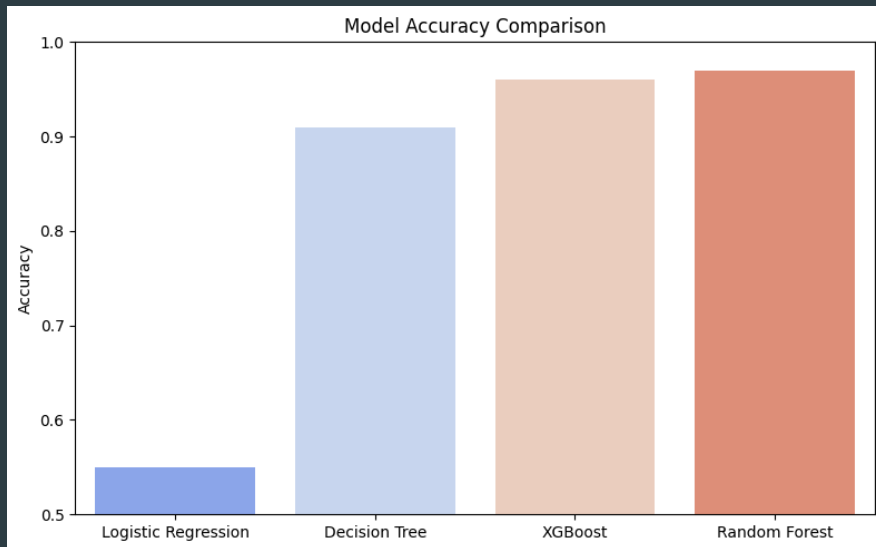
- ▶ Road accidents remain a major public safety and economic concern.
- ▶ Identifying primary contributory causes of accidents is critical for prevention.
- ▶ Traditional analysis methods are time-consuming and less accurate.

Business Objectives

- ▶ Build a predictive model to classify the primary contributory cause of traffic crashes.
- ▶ Provide actionable insights to support road safety decision-making.
- ▶ Enable data-driven allocation of traffic enforcement and prevention resources.
- ▶ Improve the efficiency and accuracy of crash cause analysis.
- ▶ Support long-term strategies to reduce accidents and save lives.

Machine Learning Models Evaluated

Visual



Model accuracy

- ▶ Random Forest 96.8%
- ▶ XGBoost 95.7%
- ▶ Decision Tree 91.4%
- ▶ Logistic Regression 55.2%

Best Model - Random Forest

Why It Wins:

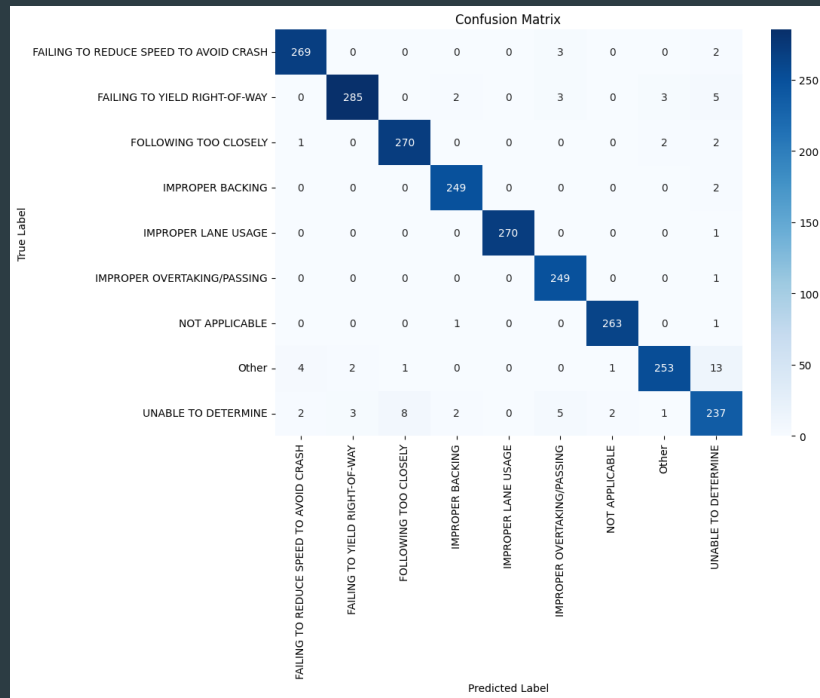
- ▶ High overall accuracy (96.8%)
- ▶ Consistent performance across all classes
- ▶ Handles data imbalance and noise well

Business Impact:

- ▶ Reliable forecasts of crash causes
- ▶ Drives actionable safety measure

Prediction Quality by Class

Confusion Matrix



Bullets:

- ▶ Most predictions accurate across classes
- ▶ Slight confusion in class 7 and 8
- ▶ Strong performance in critical categories (0-6)

Business Benefits

- ▶ Reduce accident rates with targeted prevention
- ▶ Optimize law enforcement efforts based on top causes
- ▶ Save costs related to crash response and investigations
- ▶ Inform road safety policies with data-driven evidence

Recommendations

- ▶ Adopt Random Forest model for real-time cause prediction □
- ▶ Develop dashboards to visualize crash causes by time, location, and type □
- ▶ Periodically retrain the model to improve accuracy as new data comes in

Next Steps

- ▶ Integrate model into a dashboard or app interface.
- ▶ Pilot test with historical crash reports.
- ▶ Roll out to transport authorities for operational use.
- ▶ Monitor model performance and refine over time