Predicting and Preventing Road Accidents in Chicago

Building a predictive model to identify primary causes of road accidents in Chicago, providing actionable insights for the Traffic Safety Board to enhance road safety through targeted interventions.

Authors : Eugene Maina,
Beryl Okelo,
Beth Nyambura.

Business Problem and Stakeholders

Understanding Contributing Factors to Road Accidents in Chicago

Identify top accident contributors

The analysis aims to pinpoint the primary factors leading to road accidents in the city.

Fatality causes in accidents

Determining which types of accidents result in the highest fatality rates to improve safety measures.

Analysis of accident timing

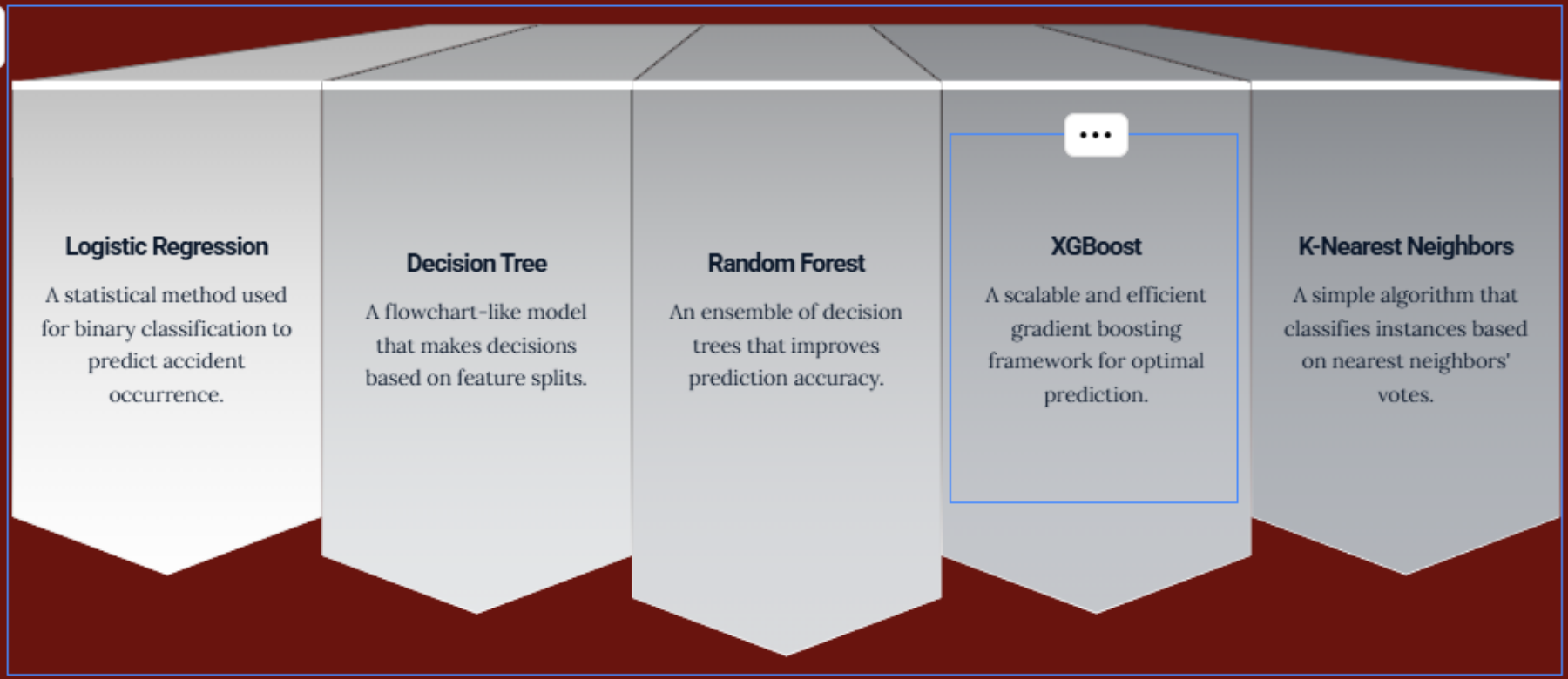
Investigating the monthly distribution of accidents to pinpoint critical times for intervention.

Model Evaluation Summary

MODEL	Accuracy	Precision	Recall	F1 Score	Comments
LOGISTIC REGRESSION	0.438	0.360	0.417	0.366	Baseline model, struggles with multi-class complexity and imbalanced classes.
Decision Tree	0.410	0.348	0.410	0.342	Slight improvement, interpretable but prone to overfitting.
Random Forest	0.393	0.334	0.392	0.349	Handles non-linearity and feature interactions better, more robust.
XGBoost (default)	0.417	0.353	0.417	0.366	Strong performance, handles class imbalance and complex patterns.
XGBoost (GridSearch)	0.433	0.365	0.433	0.373	Hyperparameter tuning yields marginal gains, best overall performer.
KNN	0.395	0.322	0.395	0.337	Sensitive to feature scaling and high dimensionality, not ideal for this dataset.

Modeling Process for Accident Prediction

Overview of Predictive Models Used



Analyzing Accident Patterns

Real Accident Data

Utilized **real accident data** from the City of **Chicago** to gain insights into traffic incidents and improve predictive accuracy.

Prediction Models

Tested multiple **prediction models** to identify patterns within the accident data, enhancing our understanding of traffic safety and risk factors.

Visualizations and Data insights

Visuals were created to analyze distributions..

Model Performance Analysis

39%-44%

Modest accuracy range

This indicates that our models work **modestly** in predicting causes of **accidents**, impacting **legal** and **policy** actions.

39%-41%

Recall indicates modest prediction

The recall shows that models can modestly predict the **real crash** and its cause, aiding in **preventive measures**.

32%-36%

Low precision levels

A **low precision** indicates the model struggles to identify the **real cause** of crashes, leading to **misclassification**.

34%-37%

Overall F1 score rating

The F1 score provides a balanced view of the model's performance in **identifying crashes** and their **causes**.

Traffic Accident Findings

01 **Accident Severity Analysis**

No fatalities and minimal injuries reported.

02 **Weekend Crash Trends**

Higher crashes on Fridays and Saturdays.

03 **Seasonal Accident Patterns**

More accidents in spring and summer months.

04 **Primary Causes of Accidents**

Tailgating and right-of-way violations are prevalent.

Alignment with Business Questions

Key insights from predictive models on road accidents

01

Top 10 Primary Contributors

Focus on 'PRIM_CONTRIBUTORY_CAUSE' with feature importances from XGBoost and Random Forest.

02

Accidents Causing Most Fatalities

Analyze features related to injury counts to understand severe outcomes.

03

Accident Occurrence by Month

Learn temporal patterns from the 'CRASH_MONTH' feature to identify trends.

RECOMMENDATIONS

Focus on high-risk behaviors during peak hours.

Direct law enforcement efforts to monitor and reduce risky driving behaviors, especially during busy times like weekends and holidays.

Enhance patrols on weekends and holidays.

Increase the presence of law enforcement during weekends and holidays to deter violations and ensure public safety on the roads.

Implement stiffer penalties for violations.

Introduce stricter fines or penalties for the most frequent traffic violations to encourage compliance and reduce accidents.

Launch educational campaigns targeting violations.

Design and run campaigns that specifically address the common causes of traffic violations to educate the public and promote safer driving.

Redesign dangerous intersections and roads.

Improve signage and signal timing at hazardous intersections, and add speed bumps or pedestrian crossings in critical areas.

CONCLUSION

Predictive Accuracy

This project shows that **machine learning models**, especially **XGBoost**, can predict **primary causes** of road accidents with modest accuracy.

Insights into Patterns

While not perfect, these predictions provide valuable **insights** into the **underlying patterns** and behaviors leading to crashes.

Accident Details

The models highlight **where**, **when**, and **why** accidents occur, providing crucial information for **preventive measures**.

Informed Decisions

This information can directly inform **road design improvements**, smarter allocation of **patrol units**, and policy decisions.

Targeted Interventions

The models enable the City of Chicago to identify leading **accident causes**, high-risk periods, and contributing factors.

Future Exploration

Next Steps: Try different models like **K-Clusters** and **Neural Networks** for enhanced predictions.