Fatal Injury Prediction with Machine Learning

Enhancing Safety and Planning in Traffic Infrastructure

Robert Sheynin

Business Motivation

Why modeling Traffic management matters

- Without a data-driven approach
  - preventative measures are reactive rather than proactive.
- With Modeling
  - Analyze systemic risk factors in infrastructure.
  - Informed Urban planning leads to improvements in public safety.
  - Fewer severe accidents reduces the burden on municipal services.

Data Understanding

Dataset Overview

- The dataset includes **crash reports** with factors like weather, road conditions, vehicle type, and driver behavior.
- Data is sourced from **official traffic records** and includes thousands of incidents.

Key Variables

- **Target Variable:** Fatal vs. Non-Fatal Injury.
- **Predictors:** Driver decisions, Intersection Type, Contributory Cause, Crash Type, etc.

Data Understanding

Limitations & Assumptions

Assumptions

- There are correlated features with predictive power over the target.
- The data sufficiently represents a traffic situation.

Limitations

- Feature list is highly saturated.
- Many categorical, heterogenous elements.

Data Preparation

- **Imputation**: Use median for numerical features, mode for categorical.
- **Categorical Encoding**: Transformed text-based factors into numerical values.
- **Numerical Scaling**: Regularize numerical features by scaling
- Reduce feature density:
  - Removing features with too many unique values
  - **Significance Testing**.

Modeling

Model Selection

- **Logistic Regression** for its interpretability, efficiency, and coefficients.
- **Decision Tree** for its interpretability, categorical and interaction feature handling.

Performance Metrics

- Classification report - specifically recall.
- ROC-AUC - for discriminatory strength.
- Confusion Matrix - tracking progress between model iterations.

Recommendations

Speeding (strongest predictor)

- Implement stricter speed enforcement mechanisms.

Pedestrian and cyclist crashes

- Increase vehicle lane isolation from pedestrian crossings and bicycle lanes.

Weather and road conditions and road defects

- Increase investment in road maintenance and winter pre-treatment programs.

Recommendations

Continued

Intersection design

- Convert high-risk intersections into less hazard prone intersection types.
- Implement AI-based traffic signal optimizations and merge assistance.

Driver behavior

- Enforce stricter penalties for reckless driving, expand public awareness campaigns, and increase police patrols in high-risk areas.

Recommendations

Summary

Impact on City Planning & Traffic Management

- Provides a **predictive framework** for infrastructure and development.
- Provides an **analysis of contributing features** to fatal injury.
- Helps cities **prioritize high-risk roads** for improvements before accidents happen.

How This Model Can Be Used in Cities

- **Used by urban planners** to design safer roadways.

Next Steps

Future Enhancements

- Incorporate real-time traffic & weather feeds to refine predictions.
- Expand model to pedestrian risk assessment for walkability planning.
- Include **Geospatial** data to determine whether certain areas are at higher risk.

Conclusions

- Machine learning enables cities to move from reactive to proactive traffic management.
- Data-driven insights lead to safer roads, reduced congestion, and reduced burden on municipal spending.
- **Next Steps:** Integrate this model into **urban planning software**.

Thank you! Questions?