Predicting Severity of Road Accidents

Team: Tristan Freiberg, Supriyo Ghosh, Elizabeth Kelley, Jack Wagner

Github | Dataset

Objectives:

- Create a model to predict the severity of a road traffic accident based on features of the accident.
- Create an application that allows stakeholders to predict the severity of a real or hypothetical traffic accident.

Data:

- We used a dataset of 300,000 road traffic accidents in Montreal that were reported to police between 2011 and 2022. This dataset was provided by the Société de l'assurance automobile du Québec to the Government of Québec.
- We translated the original dataset from French into English.
- Accident severity was grouped into three categories: material damage only, minor, and serious/fatal. Serious/fatal and minor accidents were underrepresented, comprising only 1% and 20%, respectively, of the recorded accidents.

Stakeholders: Emergency call takers, urban planners.

Models:

- We explored several models, including: SVM, XGBoost, LightGBM, RandomForest, and Logistic Regression.
- We selected XGBoost for our final model. This selection was made due to: the fact that the dataset is imbalanced; the large size of the dataset; the large number of accident entries with incomplete data; and the existence of feature interaction. Because serious/fatal and minor accidents were severely underrepresented, we upsampled these accidents to improve model performance.
- As a KPI, we used a weighted average class-wise F_2 score. This prioritized accurate identification of high-severity accidents.

Results:

- With hyperparameter tuning, our best weighted average class-wise F_2 score was approximately 43%.

Web Interface:

- We created an app using Streamlit, available in our GitHub repository. This app could be used either as an assistive tool for emergency call takers or by urban planners when making decisions about road design.
- This web app allows stakeholders to directly interact with our model by inputting accident data and obtaining a severity prediction for that accident.

Future Work:

- Other datasets with additional features (such as car model, driver information, etc) may provide better predictive power. With appropriate processing, these datasets could be fed into our existing pipeline to provide models for other geographic locations.