Road Vehicle Accident Severity Prediction in Seattle, WA

Gabriel Siqueira Kakizaki

October 15, 2020

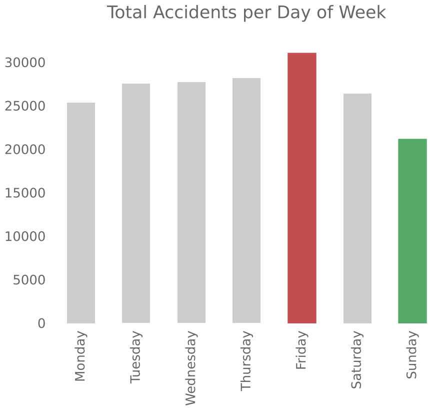
Predicting accident severity is important for policymakers

- Road vehicle accidents are a problem that in 2019 caused more than 38 thousand estimated deaths, and injuries in about 4.4 million people, only in the USA.
- Policymakers need information on what factors cause road accidents when creating or improving on existing preventive policies.
- Data analysis can help extract the needed insights.

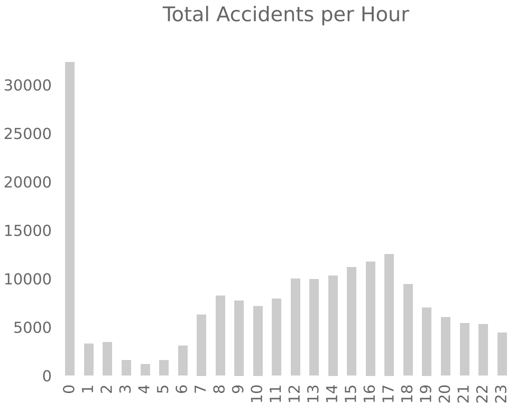
Data sources

- Open data from the city of Seattle data-seattlecitygis.opendata.arcgis.com.
- Approximately 195 thousand vehicle collisions from 2004 to May 2020.
- We cleaned the dataset and prepared for analysis.
  - Useless columns (e.g., ids) and the ones missing more than 10% of values were dropped
  - Missing values were imputed with the most common value.
  - Redundant information was removed and multicollinearity addressed.
- The dataset is imbalanced with 70% low, and 30% high severity.

Most accidents happen on Friday, and the least on Sunday

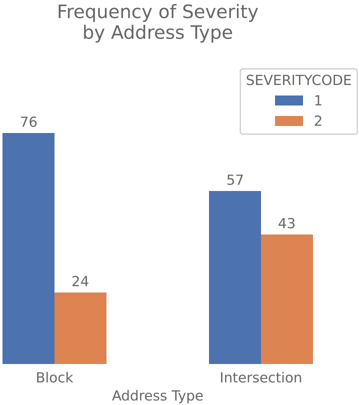


Accidents happen more on peak hours

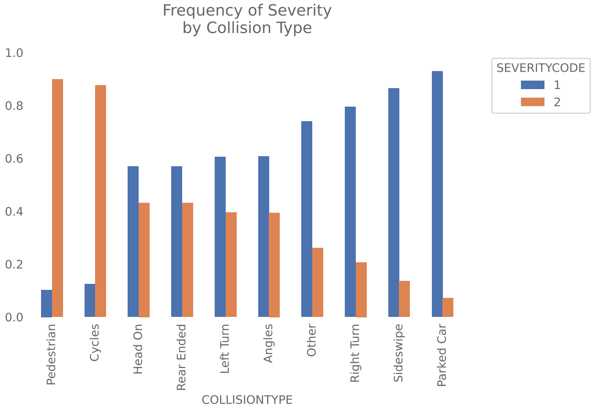


The large amount of accidents at midnight (0 hour) should be missing values.

Accidents at intersections are more likely to be severe



Collision types influence severity



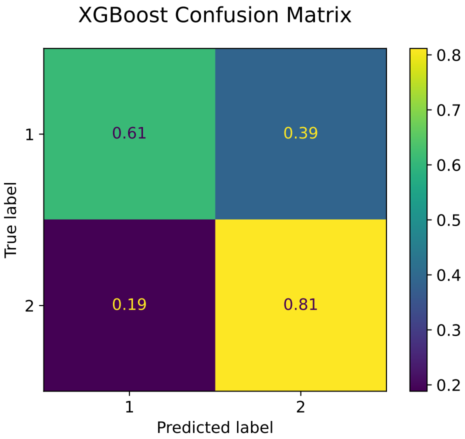
- Collisions with pedestrians and bicycles are the most severe.
- Hitting a parked car almost always means no injury.

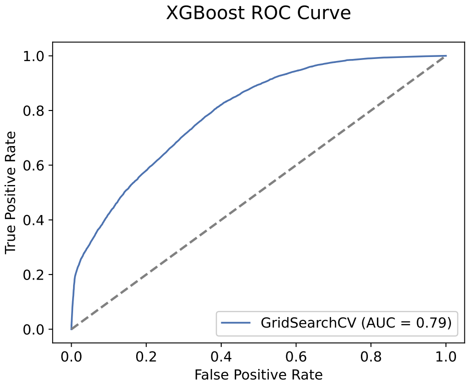
Model performance

Imbalanced Models	Precision	Recall	F1-score	AUC
Logistic Regression	0.75	0.75	0.71	0.79
Random Forest	0.73	0.75	0.72	0.77
XGBoost	0.75	0.76	0.72	0.79
Balanced Models				
Logistic Regression	0.75	0.67	0.68	0.79
Random Forest	0.76	0.67	0.69	0.79
XGBoost	0.76	0.67	0.68	0.79

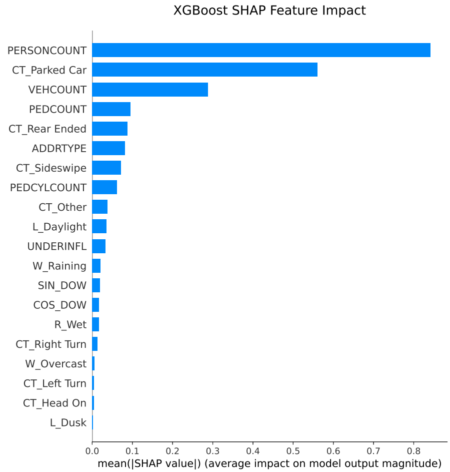
Table: Weighted average precision, recall, f1-score and AUC for the models.

XGBoost performance

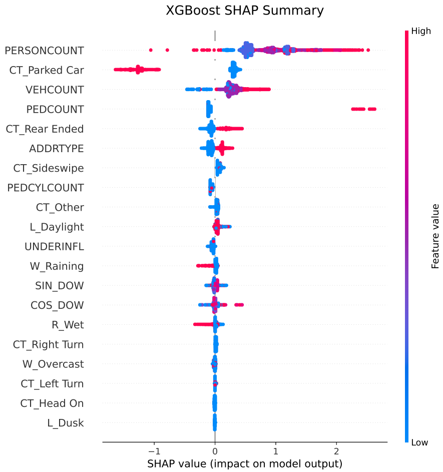




XGBoost most important features



How feature values impact XGBoost model output



Conclusion

- We analyzed which factors influence accident severity.
- Machine learning models can predict severity based on open data.
- For future research:
  - Use weather, traffic and data available in real time to predict the risk of accident.