



Car Collision Severity Prediction

Ruth Hashkes
25/09/2020



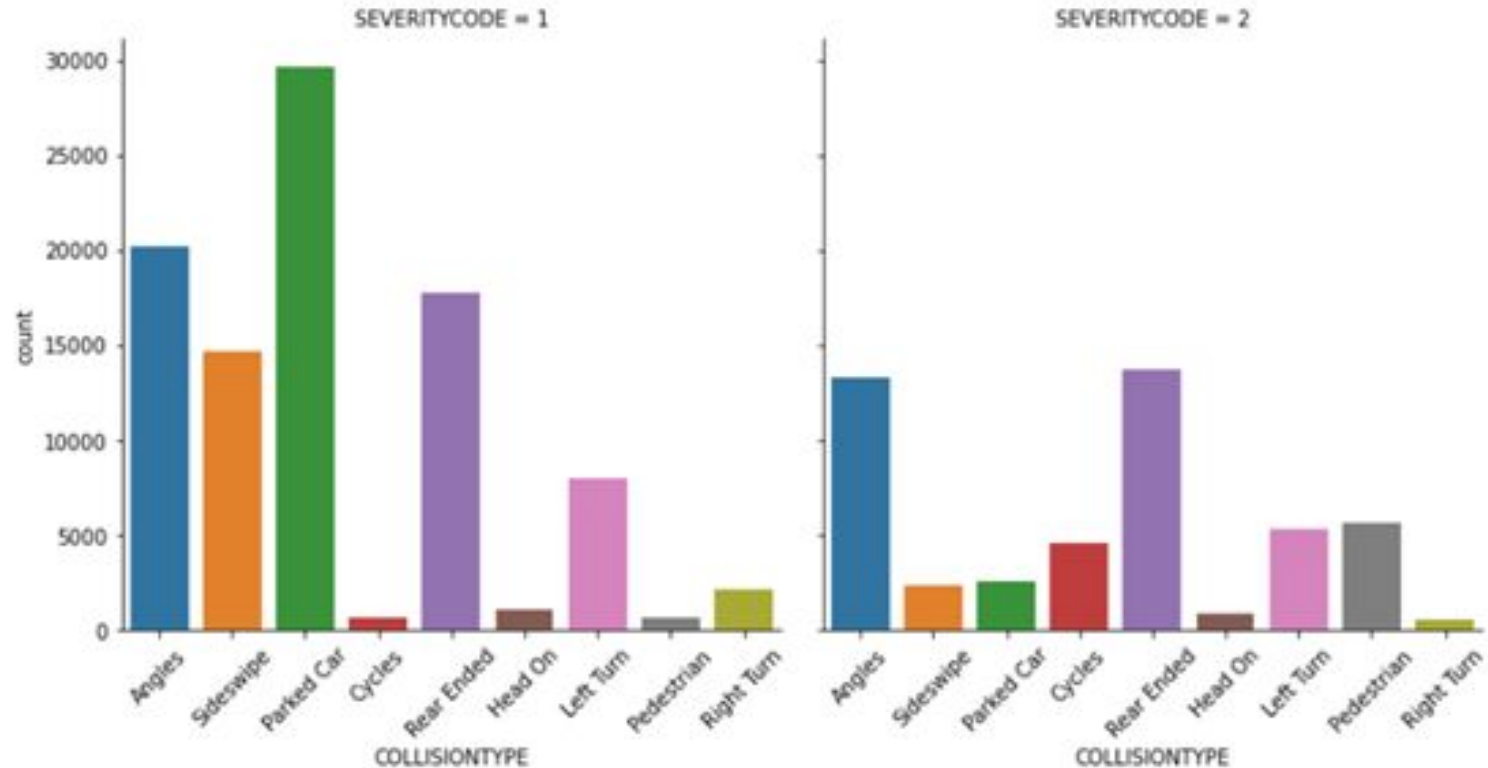
Introduction

- According to the World Health Organization, road traffic injuries caused an estimated 1.35 million deaths worldwide in the year 2016
- Car collision data collected by Seattle Police can be used to learn and predict severity of collisions
- Entities of interest: Drivers, insurance companies, local authorities

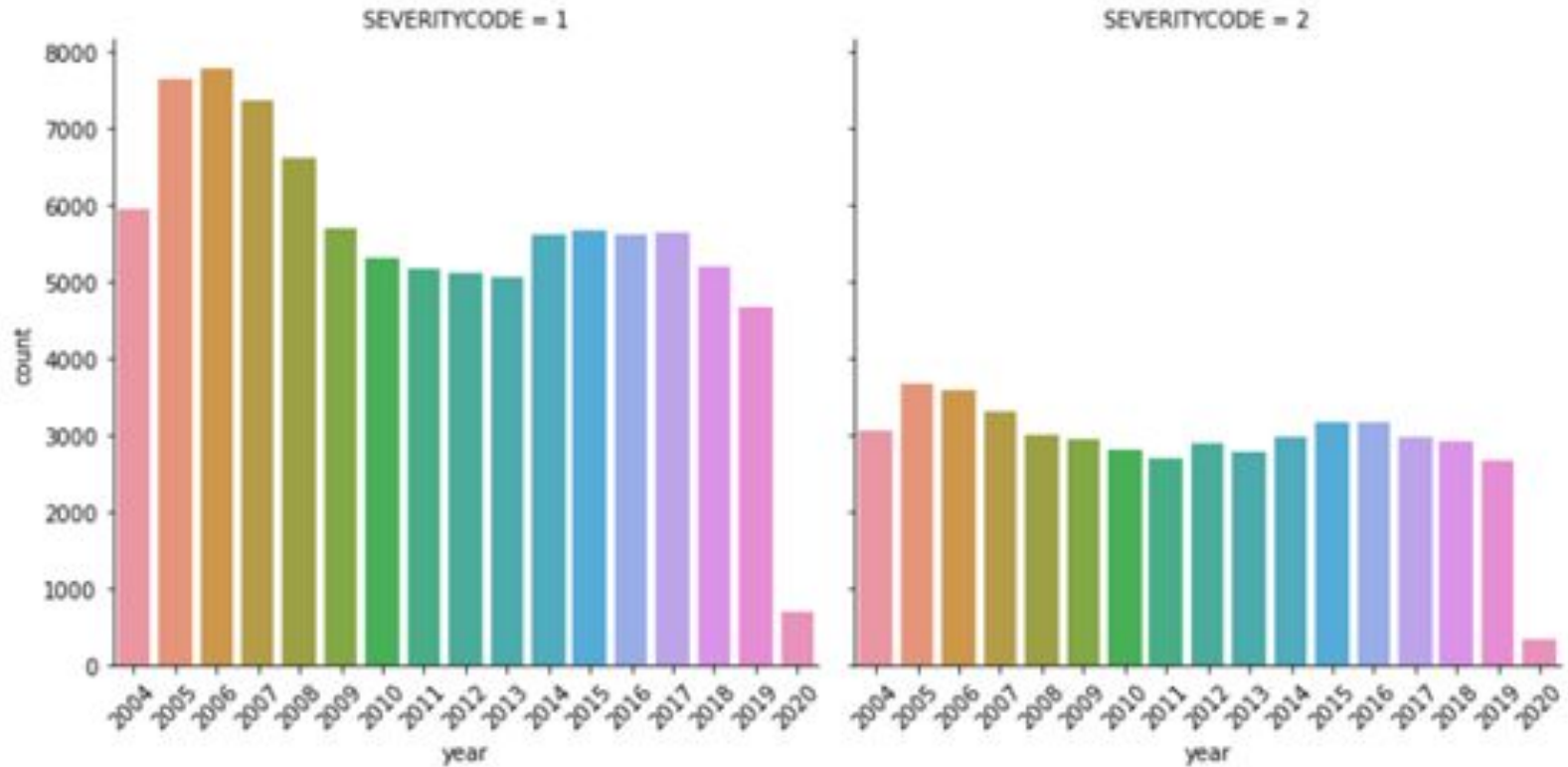
Data acquisition and cleaning

- Seattle Collisions Dataset: 194,673 car collisions from 2004-2020 with 37 attributes. This data was collected and shared by the Seattle Police Department and Accident Traffic Records Department.
- Columns containing unique identifiers, one value, redundant or missing values were removed
- Values of "unknown" or "Other" were replaced with NaN. Data coding fixed with. New date columns created
- Clean data: 143,747 car collisions and 28 variables.

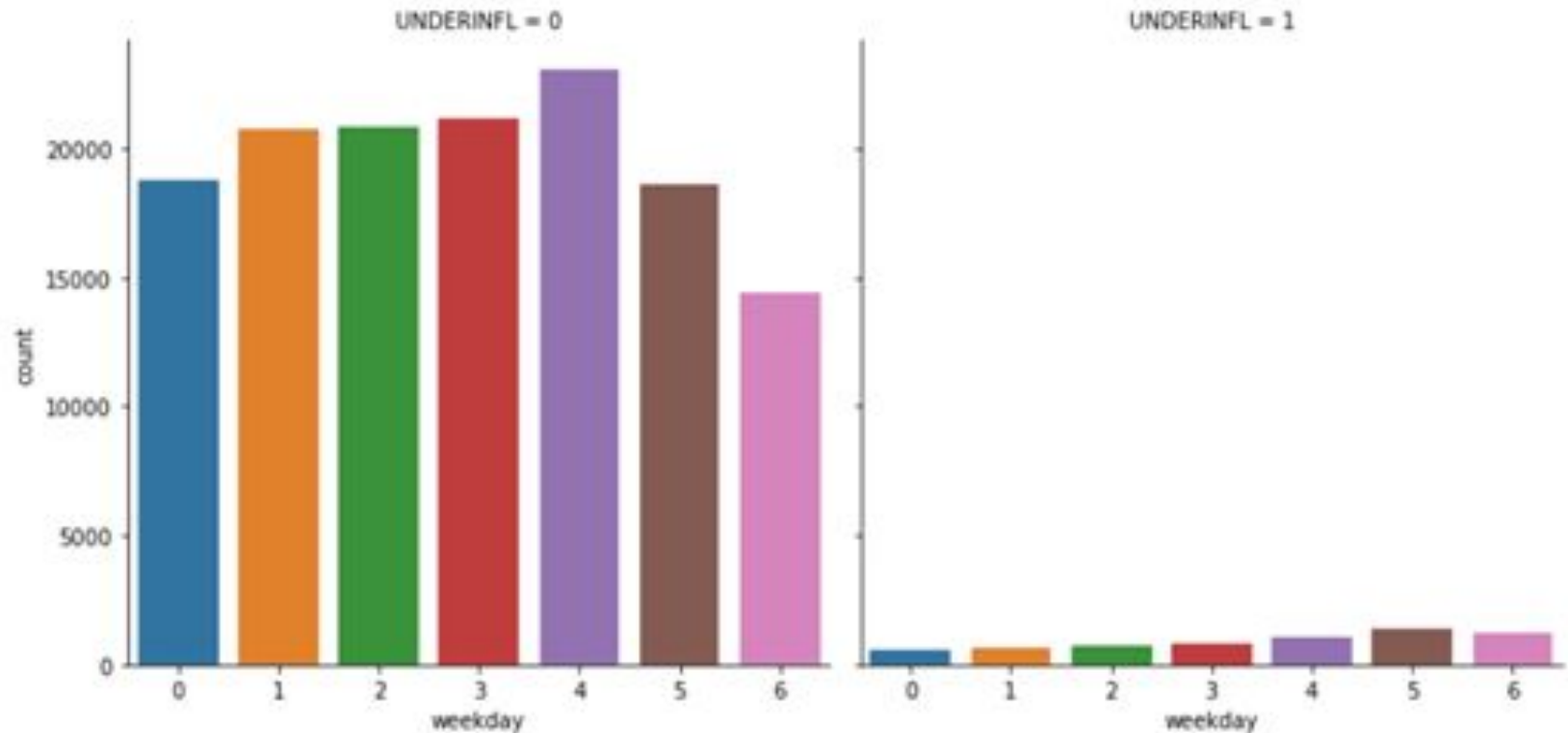
Collision Severity by Collision Type



Collision Severity by Year



Collisions Under the Influence of Alcohol or Drugs by Day of the Week

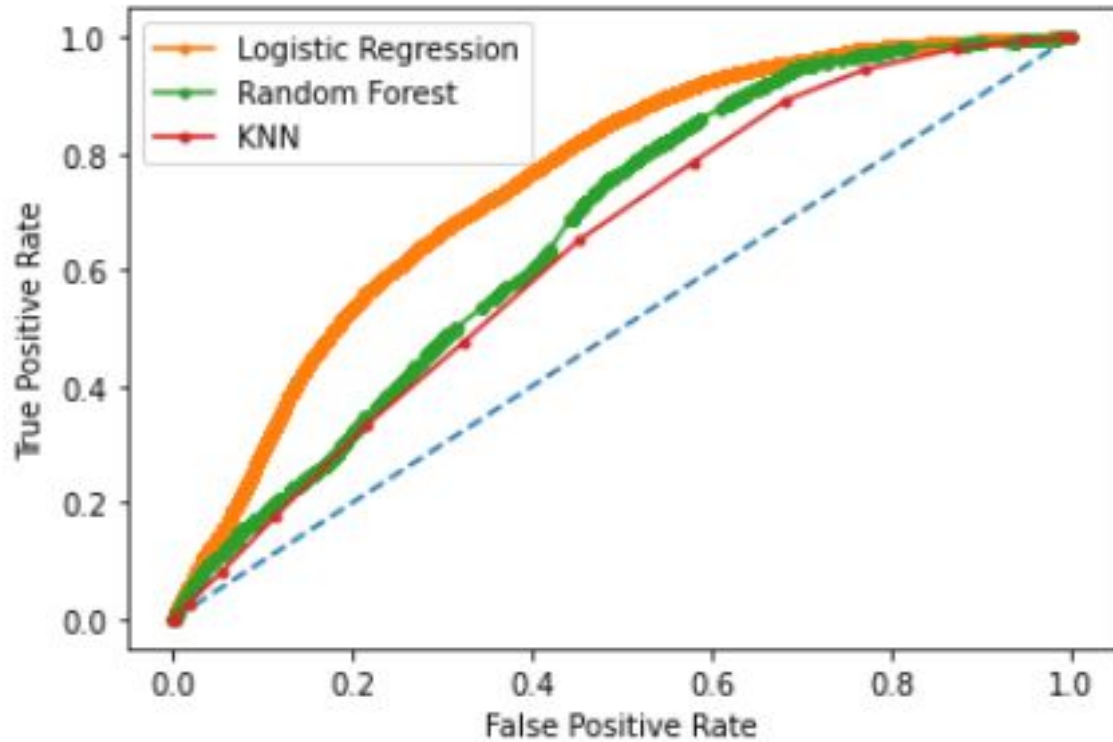


Classification Models

	Precision	Recall	F1-Score	AUC
K-Nearest Neighbor	0.70	0.71	0.70	0.64
Random Forest	0.74	0.73	0.69	0.63
Logistic Regression	0.73	0.73	0.70	0.64

All models have similar results, the logistic regression was best overall.

Classification Models



Conclusion

- Analyzed car collision data and the connection of different factors to collision severity.
- Collisions involving pedestrians are the most fatal, that weather and road and light conditions have little effect, and that driving offenses such as DUIs and speeding are dangerous
- These models can help future decision-making for drivers, insurance companies or local authorities to prevent the next fatal collision from taking place.