Predicting severity of collisions in Seattle

Ana Cristina Bittar de Oliveira

Applied Data Science Capstone

Background

- Seattle Traffic Management Division keep records of collisions since 2004 to evaluate the severity of each accident
- The main consequences are injuries, traffic jams and the high costs related
- The evaluation of severity is crucial to prioritize measures to be taken in order to prevent or minimize impacts

Objective

- The Seattle Traffic Management Division classifies severity in four levels, and includes a ‘0’ for records without severity evaluation:
  - 3—fatality
  - 2b—serious injury
  - 2—injury
  - 1—prop damage
- The goal is to make predictions about severity of car accidents using a machine learning alghorithm.

Data

- The Seattle Open Data Portal releases weekly reports in a CSV file:
- https://opendata.arcgis.com/datasets/5b5c745e0f1f48e7a53acec63a0022ab_0.csv
- And provides metadata in a PDF file:
- https://www.seattle.gov/Documents/Departments/SDOT/GIS/Collisions_OD.pdf

Data Preparation

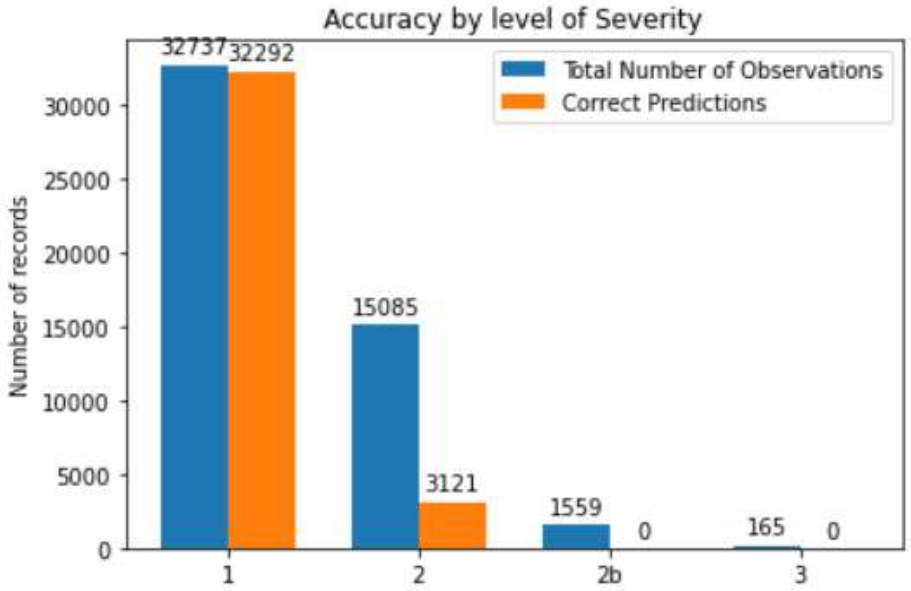
- The Data preparation was preceeded by an understanding of the dataset, leading to data cleaning and transformation

Results

- In this work Decision Tree Classifier machine learning algorithm was applied to predict severity levels of car collisions based on provided attributes.
- The available data was split into:
  - training set - 70% of the records
  - testing set - 30% of the records
- The accuracy obtained was 74%

Issues

- The model couldn’t address the highest levels of severity



Conclusion

- Decision Tree Classifier was used considering its similarity with human thinking in classifying based on simple decisions, and good computational performance
- Despite an overall accuracy of 74%, the accuracy for higher levels of security was really low, event after some balancing.