Can we predict a car accident's severity?

Carisa Li

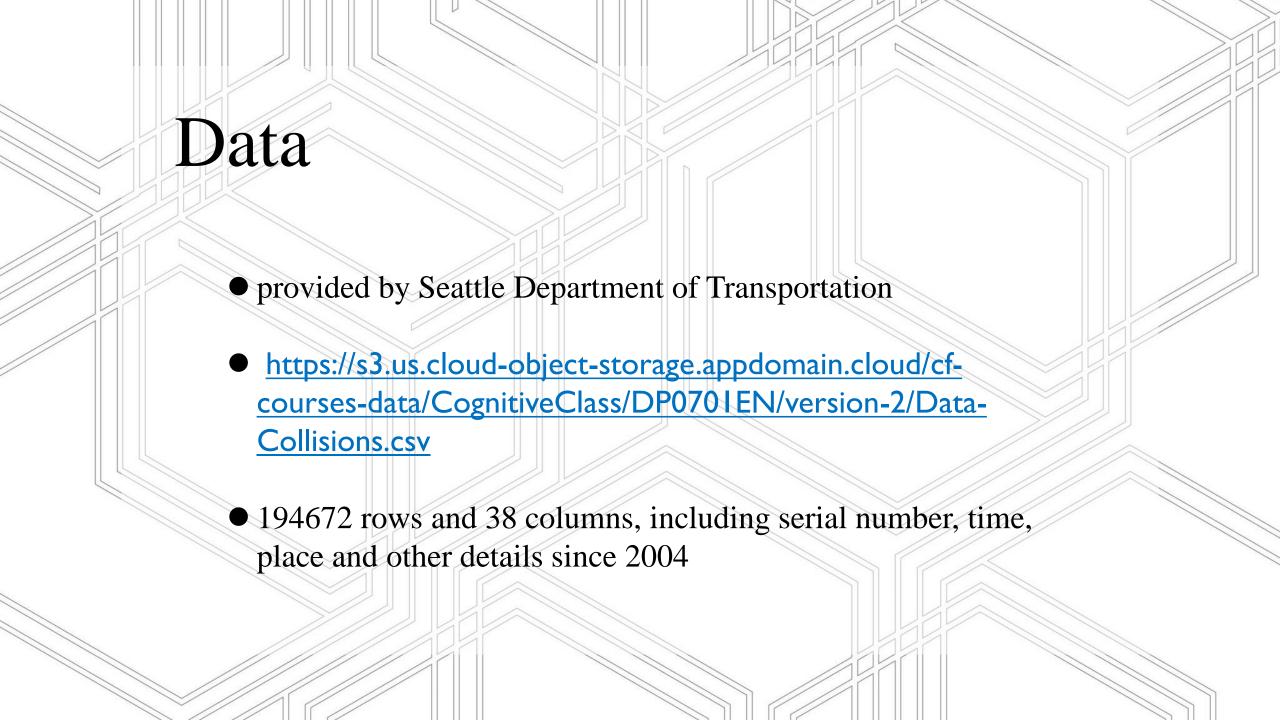November 18, 2020

Introduction

- using the dataset which is about the collisions and try to predict the severity of a car accident

- help traffic police and hospital to decide next action, be prepared and allocate resources

- the injured can get better treatment and these car accidents can have less impact on others

Data

- provided by Seattle Department of Transportation

- https://s3.us.cloud-object-storage.appdomain.cloud/cf-courses-data/CognitiveClass/DP0701EN/version-2/Data-Collisions.csv

- 194672 rows and 38 columns, including serial number, time, place and other details since 2004

Pre-processing

- For some condition in WEATHER (weather), LIGHTCOND (light condition) and ROADCOND (road condition) are too rare, I combine them into 0 (NaN, unknown, other) to simplify model.

- replace yes/no with 1/0 and use one hot encoding let all columns become int or float type.

Model

- Trained by 70% of data, test by other 30% of data

- provided by Decision Tree with max_depth = 8

- Jaccard index = 0.757838

- F1-score = 0.847295

Suggestion

- train staff member to ask suitable questions, get the necessary information, and predict severity through the model

- record new data in detail

- retrain new model on the regular basis