Predict the severity of an accident

Nguyen Phu Huy

September 19, 2020

# Introduction

## Background

Road traffic cause considerable economic losses to individuals, their families, and to nations as a whole. Using technology, we can predict the severity of an accident based on the weather, road condition, light condition and so on, so that you would drive more carefully or even change your travel if you are able to.

## Problem

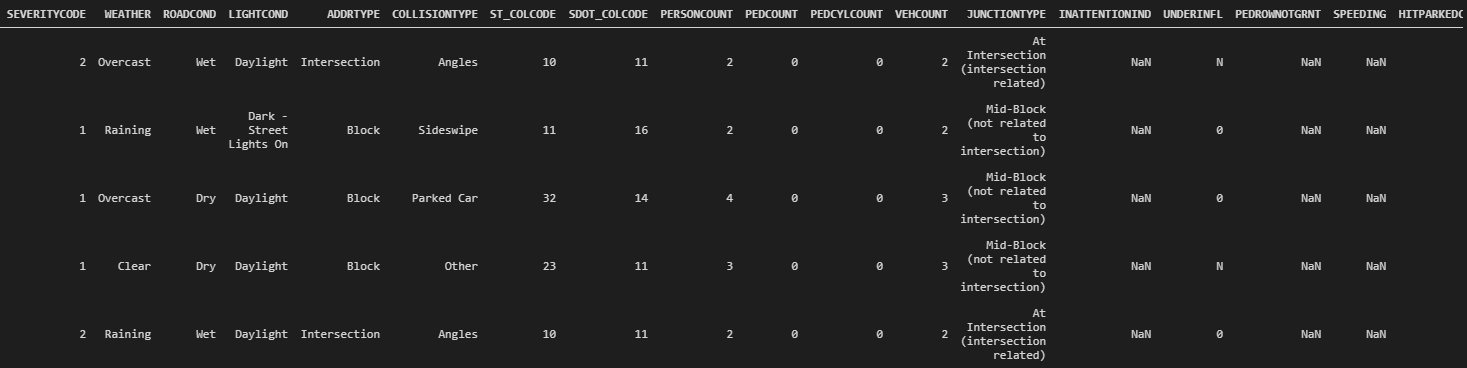
No one wants to be involved in traffic accident. If we can predict the possibility of you getting into a car accident and how severe it would be given the weather and the road conditions about, so that you would drive more carefully or even change your travel if you are able to.

For example, if it is rainy and windy, or the light condition is bad, you barely can see things clearly, you would be more likely to get into an accident.

# Data acquisition and cleaning

## Data sources

Data source is file Data-Collisions.csv downloaded from Coursera capstone course



## Data cleaning

In weather field, replace missing data with the most appeared value (Clear)

In roadcond field, replace missing data with the most appeared value (Dry)

In lightcond field, replace missing data with the most appeared value

In st\_colcode field, replace missing data with the most appeared value

In junctiontype field, replace missing data with the most appeared value

In underinfl field, replace missing data with the most appeared value, replace N with 0, replace Y with 1

In hitparkedcar field, replace N with 0, Y with 1

## Feature selection

Select the following fields to build model:

'WEATHER', 'ROADCOND', 'LIGHTCOND', 'ST\_COLCODE', 'SDOT\_COLCODE', 'PERSONCOUNT', 'PEDCOUNT', 'PEDCYLCOUNT', 'VEHCOUNT', 'JUNCTIONTYPE', 'UNDERINFL', 'HITPARKEDCAR'

Target/Predict field: SEVERITYCODE

* 3—fatality
* 2b—serious injury
* 2—injury
* 1—prop damage
* 0—unknown

# Data Analysis

We need to predict the severity of an accident based on some features like weather, road condition, light condition, the severity is classified into 5 types: fatality, serious injury, injury, prop damage, unknown. So this is a classification problem.

We will use K-Means classification to build model and make prediction

# Modeling

We divide the sample data into 80% train set and 20% test set.

After data pre-processing and cleaning, we use K-Means classification to build model

Model evaluation with k=4

Train set Accuracy: 0.7581836160731485

Test set Accuracy: 0.7449081803005009

# Conclusions

In this study, I analyzed the relationship between the severity of an accident and the weather, road condition, light condition, … I built classification models to predict how severe an accident will be. These models can be very useful in helping you to decide to drive more carefully or even change your route if you are able to.