Capstone Project - Car Accident Severity

Ida Kusumawati

INTRODUCTION

Problem Description and Discussion

Traffic accidents are bad events to experience. Frequent accidents cause many losses for many people. In addition to claiming lives, accidents also cause material losses because these events can result in traffic jams or damage to public infrastructure that have been the target of collisions so that national development is hampered. The Seattle government, through the Seattle Department of Transportation, has released data on accidents that have occurred in Seattle since 2004 and it is always updated every week. The uploading of the data shows how concerned the Seattle government is about this topic. Reporting from Colburn Law, it is known that almost every day, there are collisions in Seattle. In 2016, collisions severity occurred as follows: 58% properly damage only, 27% injury collisions, and 1% serious injury collisions. It is also known that the "common" type of collisions occurred were car collisions, with a total of 3,644 incidents in 2016. Of course this number is not a low number. Departing from this background, this project was created. Analysis and prediction of car accident severity from various variables in the dataset using machine learning.

Data Description

In order to fulfill my grade on this IBM Data Science capstone project, I used an example of data uploaded by the Seattle government through the Seattle Department of Transportation in .csv format. The dataset is available in both categorical and numerical data. The dataset has been downloaded via the Coursera course. In addition to documents in .csv format, there is also metadata. Metadata contains all the information related to that data. There is a code to categorize an accident severity. Based on what is written in the metadata, it can be seen that this dataset was recorded from 2004. The dataset contains variables related to collisions in Seattle. Some of the variables are weather, road conditions, location, type of collisions, and etc. The existing variables will be linked and analyzed to determine what factors affect the car accident severity. After analyzing it, a car accident severity prediction program using Machine Learning will be made.