

Predicting Car Accident Severity

By: Jonathan Klausner

1. Introduction

1.1 Background

Car accidents are very common on the road and happen every day across the country. In fact, Motor vehicle accidents are one of the leading causes of accidental death in the United States. Some accidents can be as small as lightly bumping into a parked car to six car pile-up on a major highway resulting in any amount of deaths and injuries. Insurance needs to be able to determine the severity of each accident to determine the amount of coverage they are offering. Outside forces such as weather, road and light conditions can help to predict how bad an accident could be as well as the frequency. People that speed or drive under the influence are more likely to cause a more severe accident than those who follow all traffic laws. Overall, public safety could benefit by this knowledge in order to encourage smarter decision making while on the road to avoid such accidents.

1.2 Business Problem

The desired result of this project is to be able to accurately predict the severity of a car accident based on a number of aspects, circumstances and features. Various features that could be useful in this project include road conditions, car and pedestrian count

1.3 Interest

This problem could be of interest to anyone planning long car trips during uncertain weather or times as well as emergency response agencies that need to determine how much support must be sent to an accident sight. Anyone who drives a car would and should have an interest in accident severity statistics for safety reasons.

2. Data

The data being used in this project would be the shared data set from the Seattle Department of Transportation (SDOT). This data set contains information that is updated weekly and includes all types of collisions. Collisions will display at the intersection or mid-block of a segment. Timeframe: 2004 to Present. The data includes features such as counts of pedestrians, cars, etc. Useful features also include road and weather conditions, speed of the car, whether the drivers were under the influence and severity descriptions. The selected features will have to be tested for viability and connection to the selected dependent variable which will be SEVERITYCODE in this case. Variables such as INTKEY are not necessary to the actual detail of the accidents and

also have a ton of missing inputs. Thus, variables like that will be omitted. Using python and its various packages, the data will be broken down and simplified in order to have clean results for predicting severity levels of accidents. The process will be broken down in the methodology section.