ACCIDENT SEVERITY PREDICTION

Capstone Project for the IBM Data Science Professional Certificate

Christian

Table of Contents

Introduction

The scope of this case study will be the prediction of the severity of a potential car accident, depending on various conditions, such as weather, time of the day, time of the year, road conditions, etc.

Such information might be interesting in several different scenarios. Local authorities could change the traffic flow such that the severity of incidents is kept at a minimum, especially during rush-hour or specific weather events (e.g. heavy rain or fog). Such information can also be included in future infrastructural planning. Similarly, navigation systems and especially self-driving cars could include such information in their routing, in order to ensure safety for their passengers. Lastly, first-response, such as fire-fighters or ambulance can approximate the severity of a reported accident, if more detailed information is not available (e.g. autonomous emergency call from car, but no physical person to ask questions on the phone).

About the Data

The data was collected in Seattle City, Washington, US and can be found under the following link:

Dataset: LINK TO DATASET
Description of Dataset: LINK TO DESCRIPTION

The dataset consists of 37 features, with the "SEVERITYCODE" being the label that is to be predicted. It contains the following values:

0	**Unknown**
1	**Property Damage**
2	**Injury**
2b	**Serious Injury**
3	**fatality**

The following table lists all Columns which will be integrated in the feature set for the prediction models. For further analysis and statistics, please refer to chapter "Data Analysis".

COLUMN NAME	POSSIBLE VALUES	COMMENT
X	Longitude e.g. -122.3231484	Location of accident
Y	Latitude e.g. 47.70314032	Location of accident
INCDTTM	3/27/2013 2:54:00 PM	Incident Date-Time
WEATHER	Overcast, Raining, Clear	Weather conditions
ROADCOND	Wet, Dry, Unknown	Road Conditions
LIGHTCOND	Daylight, Dark- Street lights on, Dark, Street lights off, Dusk, Dawn, unknown	Light conditions
ADDRTYPE	Block, Intersection, Alley,	Address type

Methodology

Data Analysis

Machine Learning Algorithms

Results

Discussion

Conclusion