

Predicting Car Accident Severity

By: Prabhleen Singh

Table of Contents

1.Introduction & Business Understanding <ul style="list-style-type: none">• Description of Problem• Background	3
2.Data Understanding <ul style="list-style-type: none">• Description & Source of Data• Use of the Data	4

Introduction & Business Understanding

Description of Problem

To minimize the number of car crashes in a city, an algorithm has to be developed to predict the severity of an accident given the current conditions of weather, road and visibility.

The main goal of this report is to shed light on why accidents happen, how they can be predicted based on environmental factors and how a model can be created that can alert drivers to be careful or postpone their road trip. This could be a business proposition for automobile sector as security/caretaking systems like these could be a game-changer.

Another target audience could be the local health institutes, police, insurance companies etc. The use of this model will allow them to be aware of all circumstances and provide services to aid the victims.

Background

In most cases, carelessness while driving, drug use, drunk driving, speeding, and many other crimes are the primary cause of injuries that can be prevented by stringent laws enforced.

In addition to the above causes, weather, visibility or road conditions are major uncontrollable variables that can be prevented by uncovering data-hidden trends and advising local government, police and drivers on highways, targeted highways and alarm systems.

Data Understanding

Description & Data Source

The data is obtained from the example data set in the course. It is Type of Collisions data from Seattle, WA, provided by the Seattle Police Department and recorded by Traffic Records in a timeframe of 2004 to present.

The data consists of 37 independent variables and 194, 673 rows. The dependent variable and also the predictor that we will use for our model is "SEVERITYCODE" that contains numbers encoding different levels of severity caused by accident from 0 to 4.

Severity codes are as follows:

- 0: Little to no Probability (Clear Condition)
- 1: Very Low Probability – Chance of Property Damage
- 2: Low Probability – Chance of Injury
- 3: Mild Probability – Chance of Serious Injury
- 4: High Probability – Chance of Fatality

Data Set Summary

<i>Data Set Basics</i>	
Title	Collisions—All Years
Abstract	All collisions provided by SPD and recorded by Traffic Records.
Description	This includes all types of collisions. Collisions will display at the intersection or mid-block of a segment. Timeframe: 2004 to Present.
Supplemental Information	
Update Frequency	Weekly
Keyword(s)	SDOT, Seattle, Transportation, Accidents, Bicycle, Car, Collisions, Pedestrian, Traffic, Vehicle

The data is unbalanced in some attributes and existence of null values is there in many records. The Data needs to be preprocessed, cleaned & balanced before analytics.

Use of Data

In our use case, we can see that Severity is impacted by several attributes, the most important ones being “WEATHER”, “ROADCOND” and “LIGHTCOND”.

These features have the highest influence over the accuracy of the predictions.