

CAR ACCIDENT SEVERITY

BUSINESS UNDERSTANDING (PROBLEM STATEMENT)

According to the Global status report on road safety, conducted by World Health organization in 2013, based on the information gathered on road safety from 182 countries, accounting for almost 99% of the world's population, the total number of road traffic deaths are unacceptably high at 1.24 million per year.

Every day, many accidents occur with various degree of severity based on different external and internal factors. Therefore, it is important to define the factors leading the severity of accidents and to generate a model which predicts the severity type of accidents best. Based on these developed models, it is possible to help the authorities take some actions to reduce the risks.

The purpose of this study is to try to understand the factors that contribute to the severity factor of a vehicle collision in the city of Seattle by using the data from the city of Seattle recorded between the years 2004 and 2020. In order to predict the severity of the collision based on various factors such as road and wet conditions in dataset, different machine learning algorithms were developed. Based on the understanding the impact of different factors on the collision outcome, it is hoped to provide meaningful insights into how to prevent such collisions so that the drivers can be alerted in advance. After Exploratory Data Analysis and Data Cleaning, predictive models were developed and evaluated by different metrics

DATA

The data for this project was provided by IBM's Applied Data Science Capstone Project. This dataset includes many features to build a robust predictive machine learning algorithm. This data contains all types of collisions which occurred between the years of 2004 and 2020 in the city of Seattle. It has a total of 194673 records and 38 different attributes.

METHODOLOGY

After careful exploratory data analysis, some variables were selected that researcher think, drive the severity level of accidents. For this undertaking, the following variables were selected to build a robust machine learning model.

	SEVERITYCODE	WEATHER	ROADCOND	LIGHTCOND	COLLISIONTYPE	INATTENTIONIND	UNDERINFL	SPEEDING
0	2	Overcast	Wet	Daylight	Angles	NaN	N	NaN
1	1	Raining	Wet	Dark - Street Lights On	Sideswipe	NaN	0	NaN
2	1	Overcast	Dry	Daylight	Parked Car	NaN	0	NaN
3	1	Clear	Dry	Daylight	Other	NaN	N	NaN
4	2	Raining	Wet	Daylight	Angles	NaN	0	NaN

The target variable of the study is the SEVERITY CODE, which is used to measure the severity level of an accident.