

Coursera IBM Data Science Capstone Project

Introduction

1) Problem

Car accidents are one of the leading causes of death in the world, especially among people between the ages of 20 and 65. Unfortunately, many of them are caused by high speed driving, road problems, drug or alcohol use, or weather conditions. Analyzing the factors involved in past accidents provided you can make an accurate prediction of the severity of future accidents.

These predictions could be used by emergency services, hospitals to have emergency personnel and material available to adequately attend. It can also be used by financial services such as auto and medical insurance to analyze their demand for future services and establish marketing strategies.

In addition, this serious accident situation can be notified to the nearby hospitals that can have all the equipment prepared for a severe intervention in advance.

2) Data Description

To carry out this project, the data set provided by the course has been selected, which corresponds to accidents that occurred in the city of Seattle that include automobile, bicycle and motorcycle accidents.

Only 10,000 records will be selected out of the 194,674 randomly, as some algorithms take time and are not suitable for such a large data set. severitycode will serve as a predictor field while the others will be the variables including status, addrtype, severitydesc, collisiontype, personcount, vehcount, junctiontype, sdot_coldesc, weather, roadcond, lightcond.