IBM Data Science Specialisation

Capstone Project

Title: Predicting automobile crash severity in New Zealand

Fernando Castellanos - September 2020

Problem description and background

According to the World Health Organization, road traffic injuries caused an estimated 1.35 million deaths worldwide in the year 2016 (*Global status report on road safety 2018, WHO*). The report indicates that more than half of all the deaths are among vulnerable users: cyclists, pedestrians or motorcyclists. Additionally, road traffic injuries are currently the leading cause of death for children and young adults. These statistics provide a regrettable picture: even with all the safety devices installed in modern cars (such as seat belts, airbags, anti-lock brakes, shatter-resistant glass and head restraints), fatalities due to automobile crashes still occurring in high numbers. Hence, it is important to have tools to help reduce the number of accidents and their related fatalities.

Thanks to the advancement of data collection and data analysis, it is possible now to better understand how automobile crashes happen and how to predict their outcome (severity). Practical use of such a predictive tool would be to enable the transport, security and emergency local agencies to analyse incoming reports of accidents and being able to determine on-line the severity of such a crash. This can be used to dispatch the adequate emergency response and activate any traffic controls.

Estimating the severity of an automobile crash is difficult and its accuracy will depend on the data available and the type of model used to define the problem. Nowadays, transport agencies around the world accumulate a huge variety of traffic information, and in most instances, they collect data of all reported automobile crashes. These datasets have usually well-defined features describing aspects such as weather, location, light and road conditions at the time of the crash and its severity. Making use of these datasets will give us better chances to construct a usable model.

My goal is to predict the severity of a given automobile crash sample. All files for the project will be available in my personal Github profile here.