

Predicting the severity of Car Accidents in Catalonia

Marc Bara

August 25th, 2020

Introduction / Problem

Road accidents are a major world economic and social problem of loss of lives and properties in many countries around the world. Key reports indicate the number of fatalities from road accidents per year of about 1.3 million and 50 million injuries [1] or an average of 3000 deaths/day and 30,000 injuries/day. Furthermore, its consequences have an impact on economic and social conditions in terms of health care costs of injuries and disabilities. The World Health Organization (WHO) [2] estimated the economic costs derived from road accidents reached 518 billion USD per year in high income countries and 65 billion USD per year in medium and low income countries.

It is true that the safety of vehicles has been increasing in the latest years/decades with the use of new materials and technology in vehicle manufacturing, and the awareness of this global problem is evident. Apart from this trend, there is an increasing need for including in this equation also road conditions. According to "Vehicle Safety 2018" by the European Commission, "Increasingly, vehicle systems which can integrate vehicle and road network interventions (integrated systems) are being pursued."

Indeed, there is a special interest in tackling this problem from all possible points of view. One of them is not only to try to reduce the number of total accidents, but trying to reduce the severity of the accidents that do occur as well. From this perspective, we may think of a possible machine learning system, operated by local/regional administration, that could enable decision making in near-real time. Depending on local road conditions, traffic data, weather conditions, type of roads, day of the week, etc., the system will try to predict the severity of the accident; severity meaning possible fatalities involved. Depending on the maturity of this system, it could be also embedded on board the vehicle dashboard to support driver decisions (as avoiding high-risk situations or conditions in route planning). Based on existing databases, we propose to create a machine learning algorithm to warn us about the probability of fatalities happening, in case of accident.

Data Source

Given the problem to tackle, we will be using relevant data with all possible descriptors that help our algorithm to succeed. In particular, we will focus on the area of Catalonia, in Spain, including Barcelona and its region.

The data set to be used for algorithm training and testing is here:

<https://analisi.transparenciacatalunya.cat/Transport/Accidents-de-tr-nsit-amb-morts-o-ferits-greus-a-Ca/rmgc-ncpb>

It consists of a structured data set with information on traffic accidents with deaths or serious injuries that have occurred in Catalonia since 2010. Among the data fields we find road, light and weather conditions when the accident occurred, number of fatalities or injured people, type of road, time/day of the week, etc. Quantitatively speaking, it

records more than 16,000 accidents from 2010 to 2018, and for each of them we find 58 fields.

References

- [1] W.H. Organization
Global status report on road safety: time for action
World Health Organization (2009)
- [2] M.M. Peden, W.H. Organization
World report on road traffic injury prevention
World Health Organization (2004)