Preventing accidents

BETTINA HARTLICH, 25 AUGUST 2020

102 people

... are killed in a car accident per day!*

The Problem

- Every killed or injured human in an car accident is one too much.

- The circumstances when an accident happens might be very divers but:

  How can accidents be prevented?

- I want to look into the different conditions of accidents like weather or road conditions but also look into all other possible conditions that could predict a higher probability that an accident could happen

The Data

- The Seattle SDOT Traffic Management Division and Traffic Records Group have gathered all collisions provided by SPD and recorded by Traffic Records. This includes all types of collisions in Seattle.

- The data gives the details per accident e.g.:

    - Severity of collisions: Tells us about the extent of damage - property damage, injury or fatality

    - Collision type (head on, involved pedestrians or cyclists)

    - Affected persons (if cyclists, pedestrians or vehicles where involved)

    - Address: alleys, blocks or intersections

    - What is known about the driver's situation when the accident happened (attentive/ unattentive, under influence or too speedy)

    - Weather, road and light conditions (rain, dry or wet road etc.)

- Overall there are 194673 accidents recorded.

The Methodology

- The data was cleaned and missing data was filled by using functions like .isnull() or .fillna

- Data was also transformed into better readable values e.g.: „Yes/No" into „0/1".

- Looking at the correlation helped to determine which attributes could help to predict an accident to happen

- Logistic Regression, Support Vector Machine, Decision Tree and Random Forest Classifier where used

- The different sores where performed and the Decision Tree method looks to be the best predictor.

The Results

- The ones with a little higher correlation (corr > 0.2) are accidents at intersections or accistents that involved parked cars, pedestrians or cyclists.

- The results of the diffrend Machine Learning methods are the following:

Model	Accuracy	F1 Score	Precision	Recall
Logistic Regression	0.751637	0.845567	0.749243	0.970311
Decision Tree	0.752023	0.84704	0.74594	0.979841
Random Forest	0.751946	0.847004	0.745864	0.979878
Support Vector Machin	0.75061	0.84761	0.741155	0.989774

- The best method to predict the severity of an accident is the Decision Tree method.

Conclusion

- When I started this projects, I assumpted that bad weather would be a big influence of probability of an accident. It does not look like it when looking at this big data set of Seattle.

- Other factors are so much more important to look at, when you want to prevent accidents from happening.If you as a driver approach a intersection, parked cars, pedestrians or cyclists: be more careful.

- Also the ones responsible for designing the intersections, pedestrian crossings, cycle lanes and parking lots should be mindful in design them. They could improve signals at intersection for the different participants of traffic (cars, pedestrians and cyclist) so that it is more clear who goes or drives where.