Car accident severity

IBM Applied Data Science Capstone Project

Jan Luca Neduchal

Table of contents

Introduction/Business Problem

- data given by Coursera
- effort to reduce the frequency of car collisions
- predict severity of car accidents given the current weather, road and visibility conditions
- alert drivers when the current conditions are bad
- analysing the impact weather conditions have on car accidents

Data

Data requirements

- large amounts of data
- numerical data
- data needs to be trained and tested for our results to be as accurate as possible

Data

Data description

- Data-Collisions.csv
- target variable == SEVERITYCODE
- SEVERITYCODE is used to measure the severity of a car accident from 0 to 4 within the dataset
- to weigh the severity of car accidents we also use WEATHER, ROADCOND and LIGHTCOND

- before we can analyse the data, we need to balance it

Methodology

- after analysing and cleaning our data we can use it for our machine learning models
- KNN: predicting the severity code of an outcome by finding the most similar data point within k distance
- Decision Tree: displaying all possible outcomes to analyse the consequences of a decision
- Logistic Regression: predicting one of our two severity codes (1 or 2)
- after that we test the accuracy of our models

Results

Machine Learning Model	KNN	Decision Tree	Logistic Regression
Jaccard Similarity Index	0.564001947698565	0.566436570904820 6	0.526021825680978 4
F1-Score	0.540177530897430 8	0.545059793738944 4	0.511602093963383
LogLoss	—	—	0.684953538319888 7
Variable	k	max_depth	C
Most accurate amount of variable	25	7	6

Discussion

- beginning: change of data type from ‚object‘ to ‚int8‘ for our algorithms
- after that: sample down majority class until values of both classes (1 and 2) match
- then: analyze and clean data to prepare it for the ML models
- followed by: testing accuracy of our models using the Jaccard Similarity Index, F1-Score and LogLoss

Conclusion

- based on historical data
- relation between weather conditions and frequency of car accidents
- impact on whether or not driving could result in a car accident leading to property damages (class 1) and/or injuries (class 2)