

The background of the image is a close-up, high-angle shot of a severely damaged car windshield. The glass is shattered into numerous sharp, jagged fragments, with a prominent spiderweb pattern of cracks radiating from a central point. The shards of glass are translucent and reflect light, creating a complex, fractured texture. In the center of the image, there is a dark blue rectangular text box with a thin white border. Inside this box, the text "PREDICTING SEVERITY OF CAR ACCIDENTS BASED ON WEATHER" is written in a clean, white, sans-serif font, all in uppercase letters. The text is centered both horizontally and vertically within the box.

PREDICTING SEVERITY OF CAR ACCIDENTS BASED ON WEATHER

BUSINESS UNDERSTANDING

- Governments, public health organisations, insurances, car manufacturers are potential interested parties for a model which predicts accidents based on weather conditions
- It can be used by governments or car manufacturers in order to warn car drivers.
- European Union wants to have zero deaths out of traffic accidents by 2050
- With this model, at least the number of accidents due to bad weather conditions could be reduced.
- The target is to predict the severity of an accident based on weather, road and light conditions



DATA ACQUISITION AND CLEANING

- Data, which was provided by coursera was used: <https://s3.us.cloud-object-storage.appdomain.cloud/cf-courses-data/CognitiveClass/DP0701EN/version-2/Data-Collisions.csv>
- Empty rows were dropped
- Categorical values transformed to a numerical value
- Due to **imbalanced target variable**, sample sized will be limited to approx 120,000
- Training data= 80 %, Testing data = 20%



KNN

- The first model is a K-nearest neighbours ML model. Different Ks between 1-500 were conducted (due to performance not at once, not documented in the notebook)
- Since the accuracy score for a rather low k (e.g. 7) was quite low with approx, the range for possible ks was increased. The best k was found at 91 with an Accuracy score of 0,55. Because of the high ressources neede for this operations, it needs to be evaluated if this is the correct model type.
- As stated in the table, F1 score and also the Jaccard Similarity Score are rather low.
- **The built model is not good for prediction**

	precision	recall	f1-score	support
1	0.50	0.92	0.65	11457
2	0.52	0.09	0.15	11375
avg / total	0.51	0.51	0.40	22832

Jaccard Similarity Score: 50.58 %
F1-Score: [0.65146882 0.15107968]
Train Accuracy: 50.55078621172967 %



SVM (SUPPORT VECTOR MACHINES)

- For this model the model was set to RBF which leads to a Jaquard score of 0.55.
- Like KNN, also this model doesn't seem to be a good prediction model



CONCLUSION AND FUTURE DIRECTIONS

- Features like road condition, light condition, weather do not influence the accident rate as we expected.
- It appears that the accident occurred the most during daylight, when the weather was clear, and road condition was dry.
- Therefore, other, not wheather-related attributes need to be evaluated.

