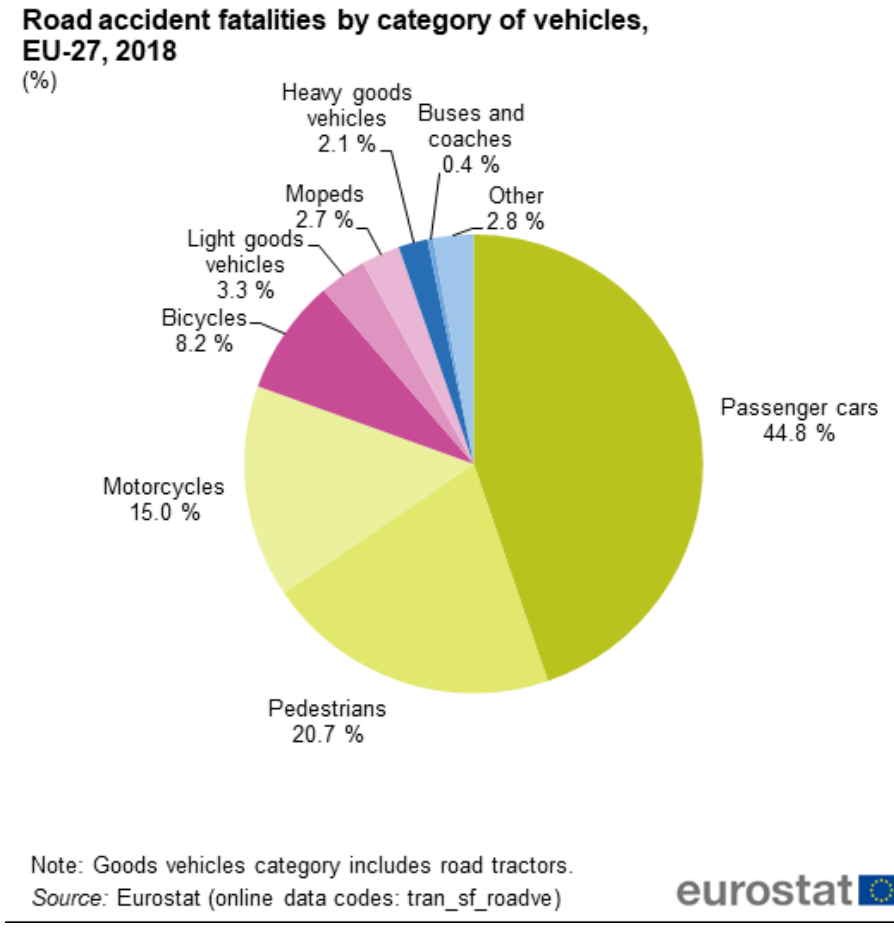
Predicting the Car accident severity

PROBLEM DESCRIPTION

- road traffic crashes are a leading cause of death in the world and the leading cause of non-natural death or healthy citizens for all age groups, according to CDC
- road traffic crash death rate is over three times higher in low-income countries than in high-income countries, according to WHO
- according to Eurostat statistic "Road accident fatalities by vehicles" car drivers and passengers represented the largest category of road traffic deaths in the EU in 2018, with 44.8% of all road traffic fatalities



BUSINESS UNDERSTANDING

- **Objective**

develop the navigation app that could warn the car drivers about the possibility of getting into a car accident on the chosen route and how severe it would be

- **The target audience**

car drivers with smartphones

- **Stakeholders**

governments of the countries that are interested in the reducing of car accidents on the roads

- **Question**

can we predict for given route the severity of possible car accident in real time for any region?

DATA ACQUISITION

- **Data source**

dataset SDOT Traffic Management Division (2004-nowadays)

with weekly update frequency

includes many attributes that describe all the circumstances of the accident

- **Dataset**

in total, 194,673 rows and 38 features in the raw dataset

- **Data Preprocessing**

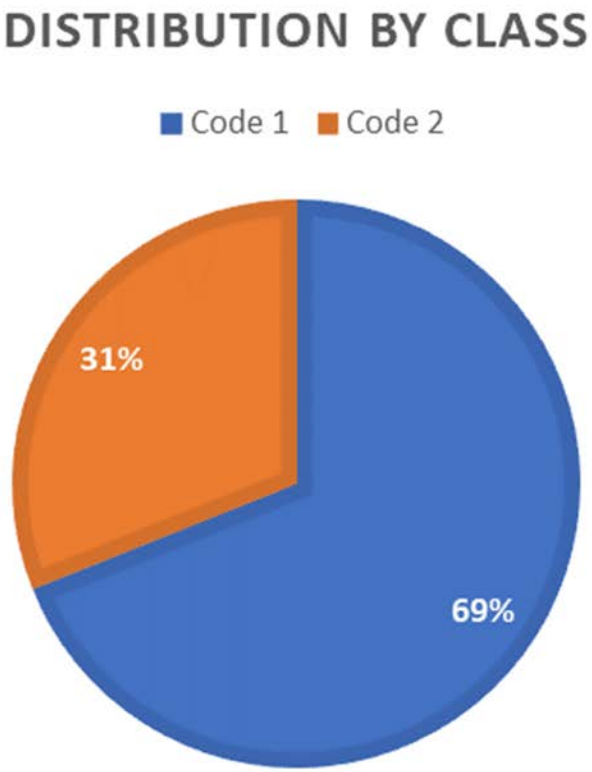
duplicate, highly similar or unnecessary features were dropped

rows with missing values were also dropped

cleaned data contains 4 features

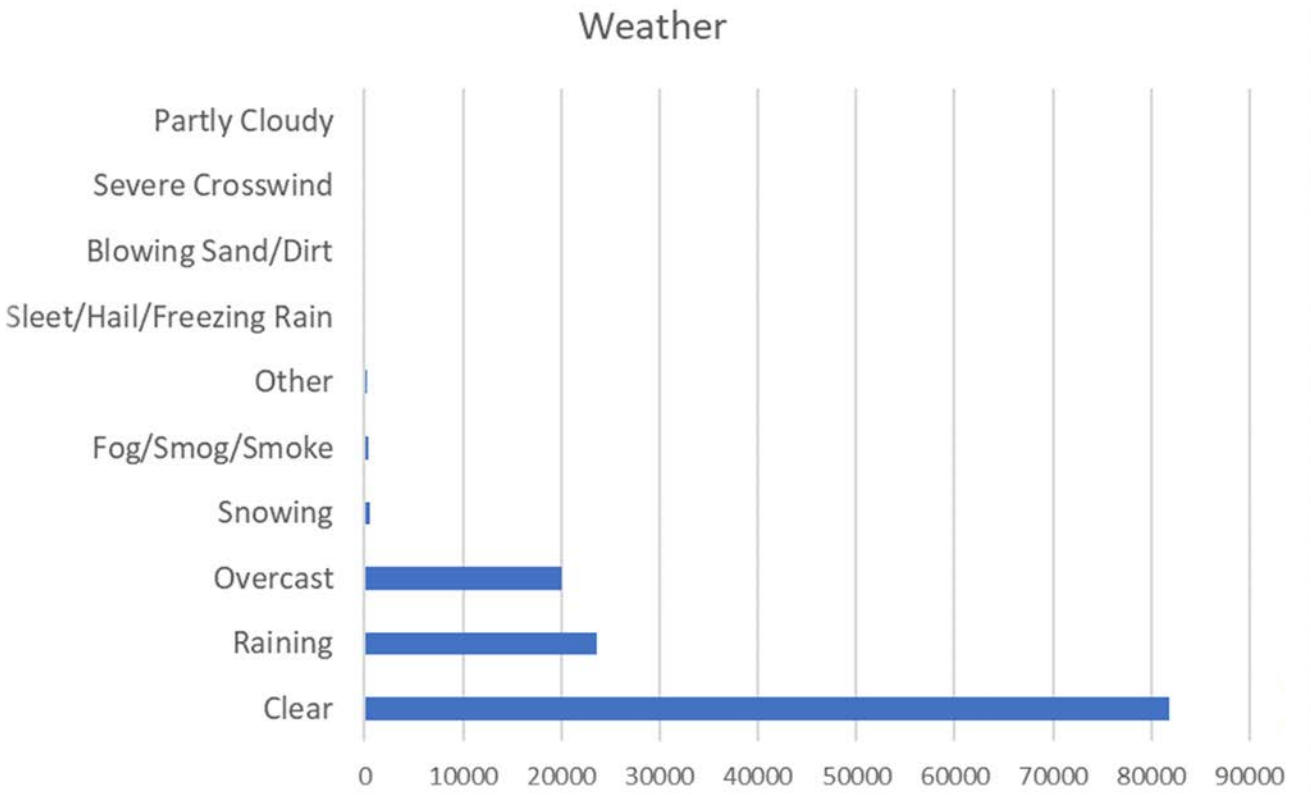
Seattle Department of Transportation

DISTRIBUTION BY SEVERITYCODE



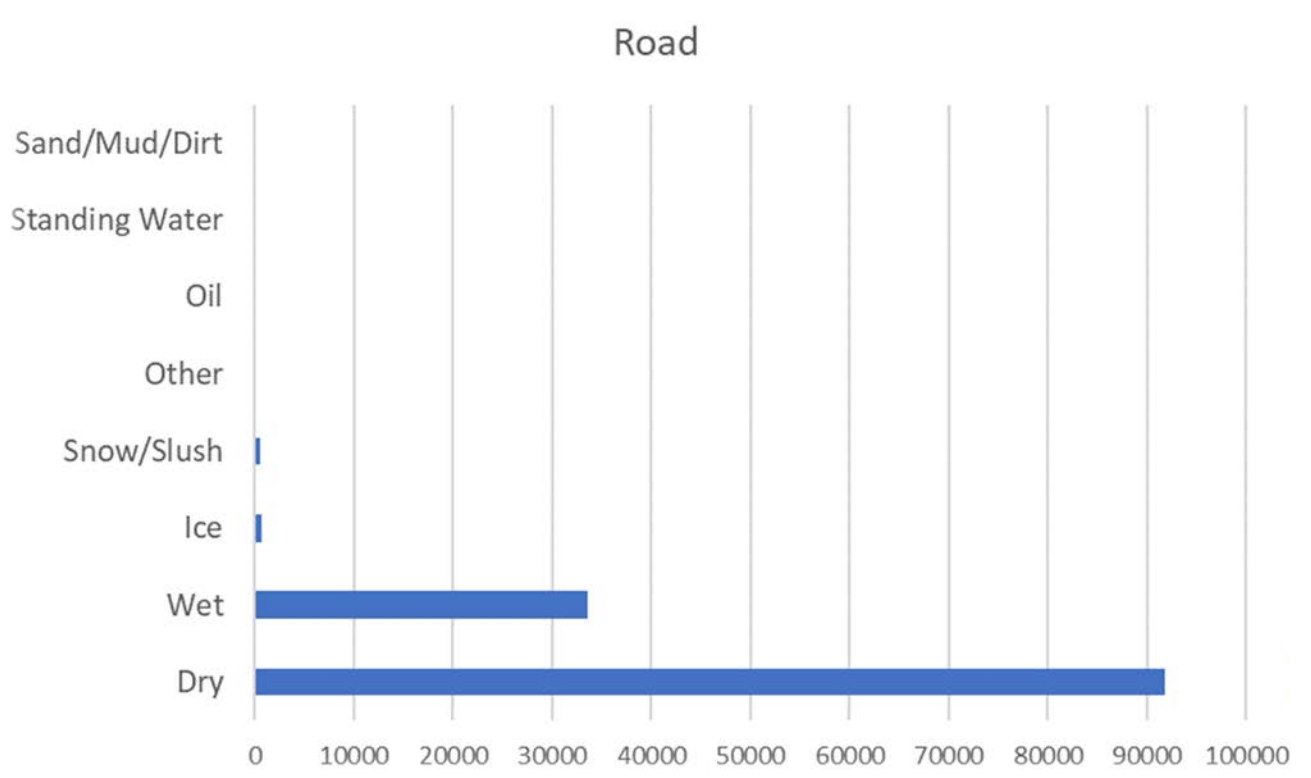
- 90851 cases of property damage only and 40925 cases of injury collisions
- distribution is uneven, which can negatively affect the model
- no fatalities accidents left after pre-processing the data

ACCIDENTS BY EACH TYPE OF WEATHER CONDITIONS



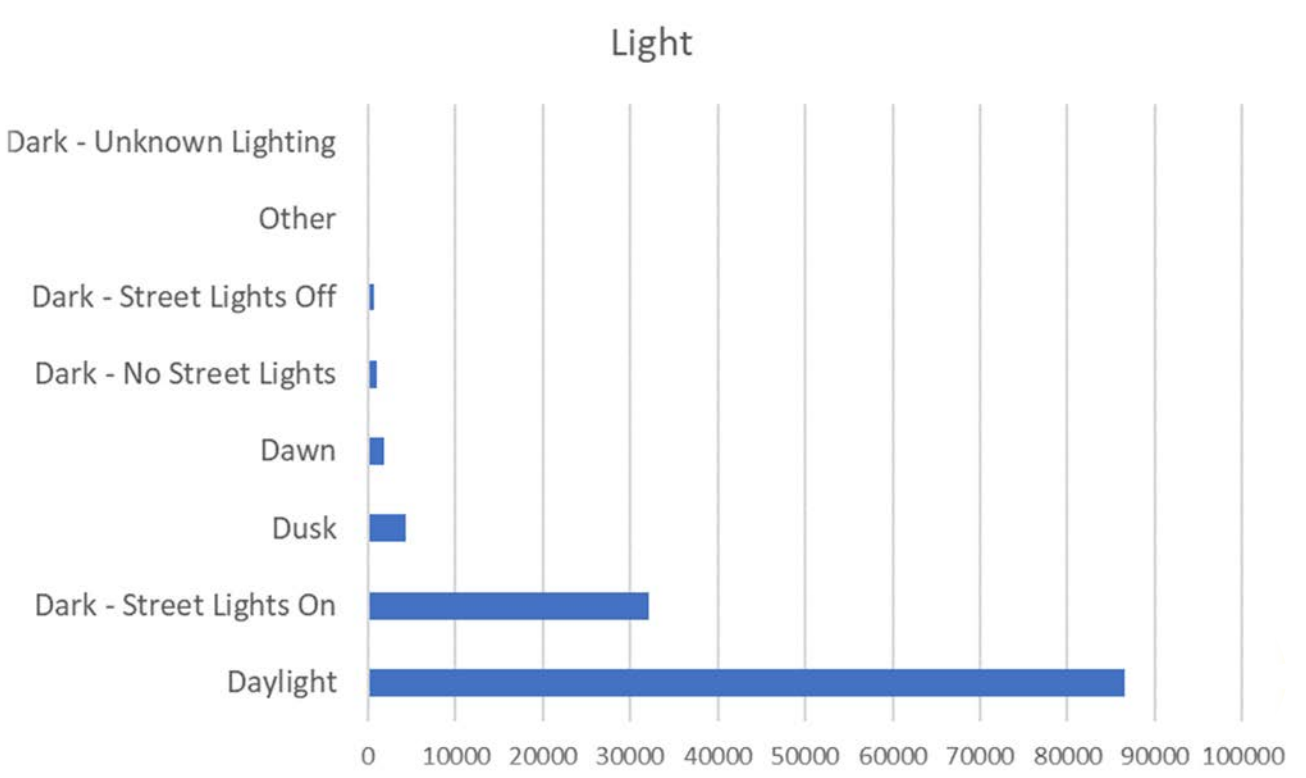
- most accidents happened in clear weather
- many accidents occurred in overcast and during rain
- other weather conditions are unlikely

ACCIDENTS BY EACH TYPE OF ROAD CONDITIONS



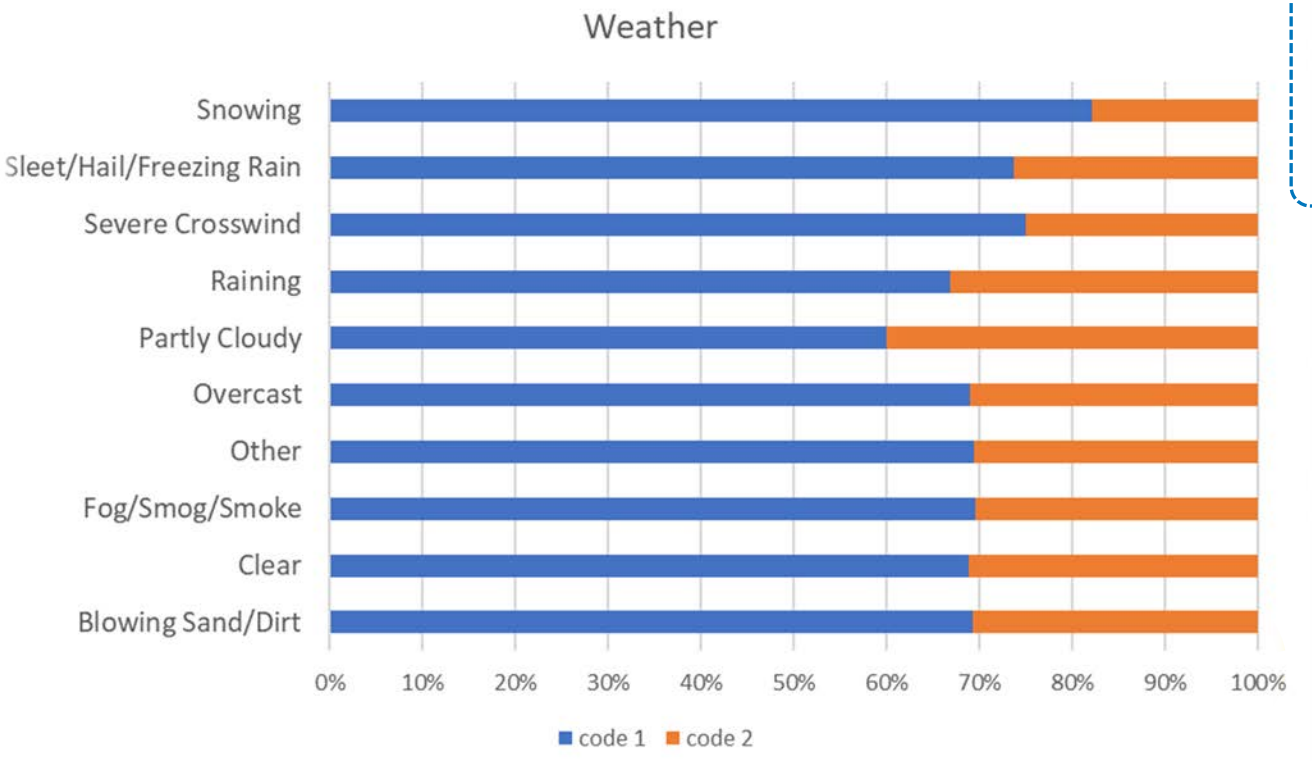
- most accidents occurred on a dry road
- many accidents also happened on a wet road
- other road conditions are unlikely

ACCIDENTS BY EACH TYPE OF LIGHT CONDITIONS



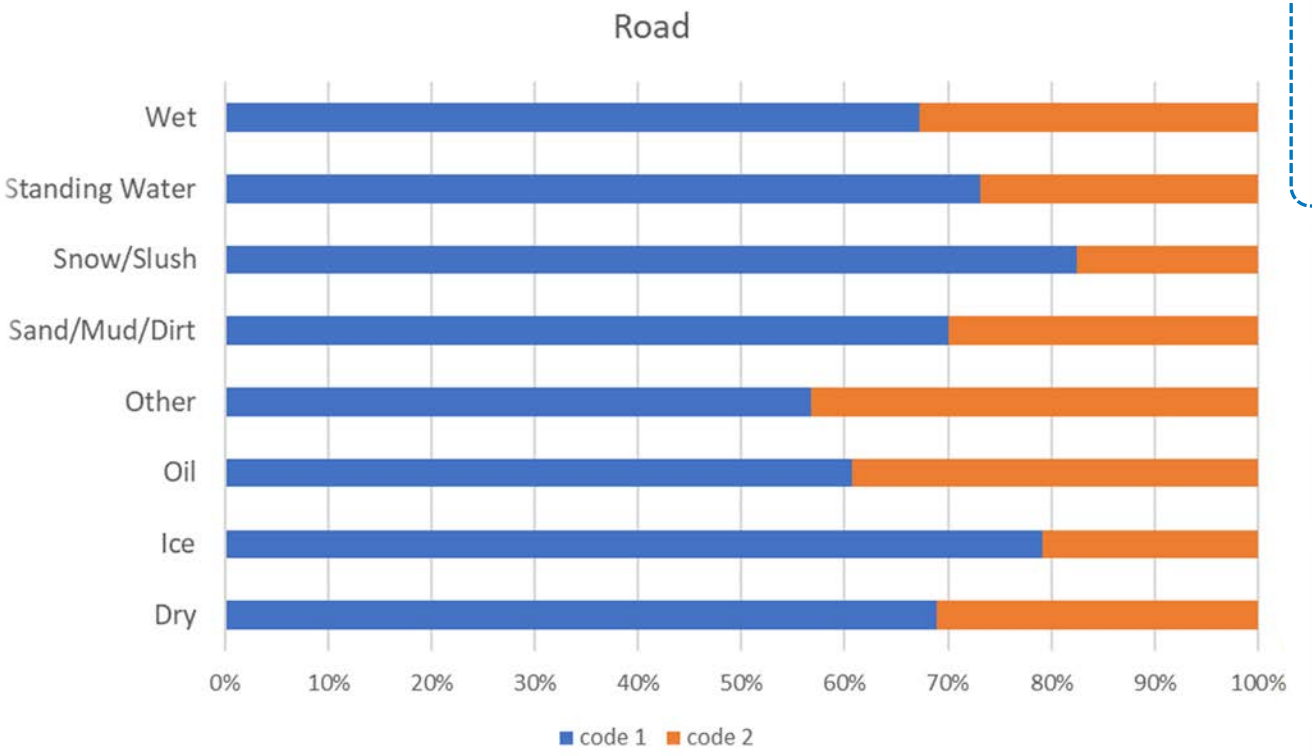
- most accidents occurred in daylight
- many accidents also happened in the dark with the lights on
- other light conditions are unlikely

PERCENTAGE BY EACH TYPE OF WEATHER CONDITIONS



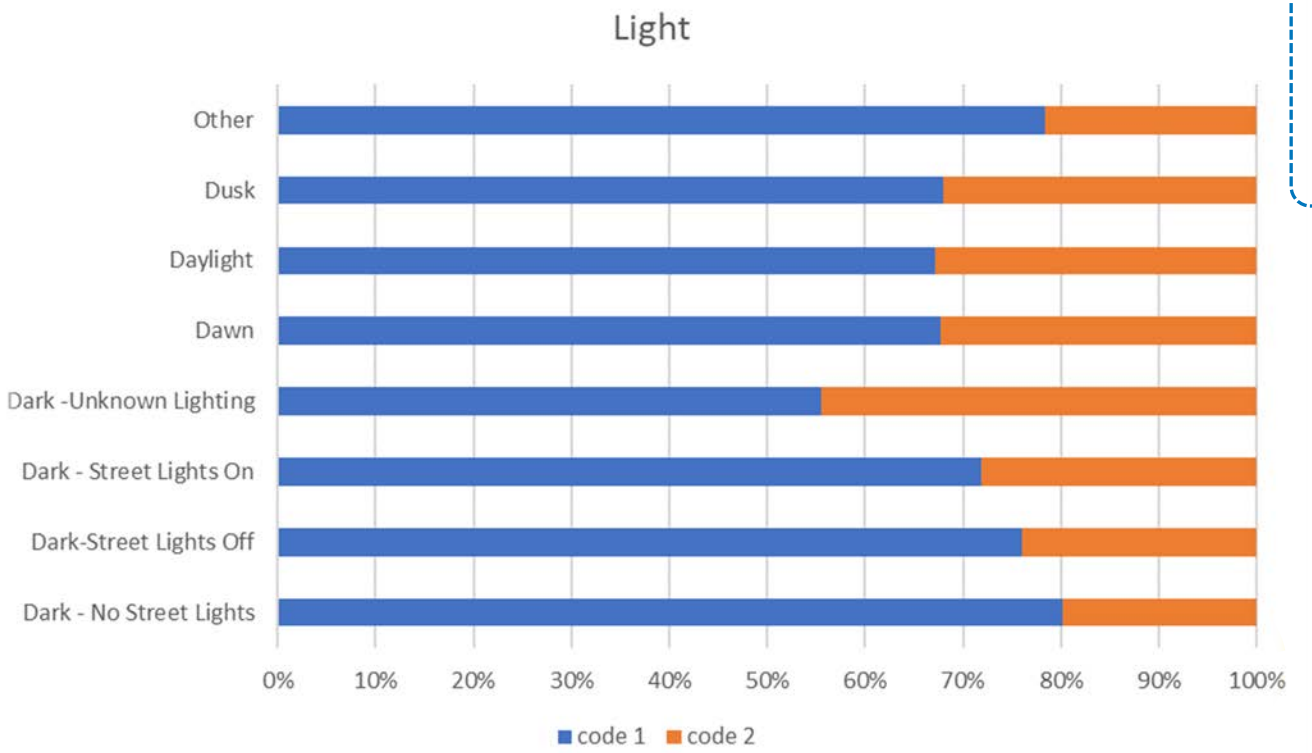
- distribution of the data for the most frequent weather conditions (Clear, Raining and Overcast) is almost perfectly correspond to the general distribution

PERCENTAGE BY EACH TYPE OF ROAD CONDITIONS



- distribution of the data for the most common conditions (Dry and Wet) is very close to the general distribution

PERCENTAGE BY EACH TYPE OF LIGHT CONDITIONS



- distribution of the data for the most common conditions (Daylight and Dark-Street Lights On) corresponds to the distribution of data across classes

MODEL EVALUATION RESULTS

Algorithm	Jaccard	F1-score	Log Loss
KNN	0.689777	0.565658	NaN
Decision Tree	0.690991	0.564857	NaN
SVM	0.690869	0.565135	NaN
Logistic Regression	0.691021	0.564816	0.615403

➢ among the individual models, the Logistic Regression model performed the best (~69.1% accuracy), though the differences between models were very small.

CONCLUSION AND FUTURE DIRECTIONS

➢ **Results:**

-The accuracy of the models does not actually differ from the general distribution of data across classes.

-This confirmed fears that the severity of road accidents does not actually correlate with the parameters chosen.

➢ **Ideas for future:**

- look for other factors that could be obtained in real time and that would have an impact on the severity of accidents

- fatal cases must be considered without fail

- take into account the type of car body, because this can significantly affect the consequences of an accident