Project proposal

Team members: Attila Sztreborny, Ádám Burkus, Bendegúz Birkmayer, Bojta Rácz, Roland Tuboly

Question:

Our main questions are the following:

1. How can we classify road casualties?
2. What variables explain the probability of a fatal or less serious collision?
3. Which variables decide whether a collision will result in fatality or just slight injury?

Dataset:

Reported road casualties, Great Britain, annual report 2024:
https://www.gov.uk/government/statistics/reported-road-casualties-great-britain-annual-report-2024

Three datasets: casualty, vehicle, collision

We plan to merge these three and use the most relevant variables for our research.

Baseline:

KSH research[1] and interactive map which shows the most dangerous roads in Hungary based on a dataset that includes vehicle type, casualty severity and main centers for collisions between 2019-2024. We plan to make a similar research using British data instead of Hungarian and we could compare our results in the end.

Task type: classification (logistic model, LASSO regression, decision tree)

Modeling plan:

We plan to solve a classification problem by creating multiple logistic models. Our dependent variable is going to be collision severity (1 = fatal, 0 = not fatal) and our explanatory variables are going to be age and sex of driver, road conditions, maneuver, vehicle type, among others. We are going to evaluate our models using k-fold cross validation. Apart from logistic models we will try to create a decision tree and LASSO regression.

[1] https://tinyurl.com/356e22ay