Modeling of CO Emissions from Cars

Christine Al-Thifary | Elisabeth Grasia Putri | Kin Man Lam

Master of ICT – Master of Data Science

Supervisor: Dr. Rosalind Wang

Carbon Monoxide (CO) is one of the most dangerous and poisoning gas emissions All vehicles testing has concerned to limit this gas emission. Therefore, we would like to analyse this data to predict it. The project steps consist of data pre-processing, detailed methods analysis, and a comparison between each of the methods. In data pre-processing, we cleaned the dataset, such as get rid missing values, combining data, simple statistical analysis (calculating mean and standard deviation for each feature), and plotting. We use several methods of analysis such as Neural Network, Support Vector Machine(SVM), Naive Bayes, and Lasso Regression. We did regression in Neural Network and Lasso Regression; also, classification on SVM and Naive Bayes which combined Gaussian and Categorical Naïve Bayes. The project is successfully completed. As a result, the highest accuracy score is given by SVM (87%). It is followed by Neural network (85%), Naïve Bayes (74%), and Lasso regression (37%). The feedback regarding this result is quite satisfying, we got 36 out of 40 points. In conclusion, using SVM is the best choice to predict CO emissions. For our team members personally, this project leads us to practice more by looking at the real problem dataset and applying each method in detail.