Assignment – Machine Learning for Prediction of Train Delays

The Swedish Transport Administration (Trafikverket) registers all train deviations from scheduled arrival times. In Sweden, a train is said to be punctual if it arrives at the measuring station within 5 minutes of the planned arrival time. However, every delay over 3 minutes is recorded as a delay in the system. These are the delays that are included in the analysis.

In this assignment, you will work on a csv/xls file. This file consists of different trains operating across various routes in Sweden. The delays in these trains due to railway vehicles are recorded in the data as 'registered delays'. This assignment includes data pre-processing, data visualisation, feature engineering, machine learning and decision support.

This assignment will be considered complete when a demonstration has been shown to the teacher and the assignment report has been submitted to Canvas.

Here is the detailed break-up of the task list.

Tasks to pass this assignment with a grade 3: Feature engineering

Use the provided data file in the assignment. Go through the steps of data pre-processing, feature engineering, initial data visualisation (geopandas can be used), and visualise the analysis of the data (latitude and longitude, train number, route, place etc.). Theorise which parameters will be important for model development. Justify your decisions.

Additional tasks for grade 4: Prediction

Prediction—Perform this task as predicting a continuous variable, e.g., delays, based on multiple features. Train and test suitable models. Which model worked best for you, and why was it so?

Additional tasks for grade 5: Decision Support

Describe and develop a decision support system utilising the results of the analysis performed in the previous steps. What and how information should be presented to the domain experts to maximise the utilisation of your results? Justify your decisions.