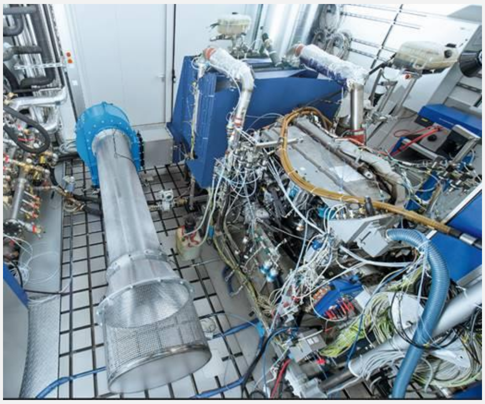
DATA SCIENCE CHALLENGE 2020

TASK ANNOUNCEMENT

CONTEXT

Combustion engines are high tech machines equipped with a wide array of sensors. These sensors have to fulfil a plethora of requirements regarding geometry, power consumption, reliability or maximum cost. Simply put: under certain conditions, having adequate estimated values and avoiding the need to install extra sensors beats having exact measurements. However, since computations need to be done with hardware available within the machine (i.e.: engine control unit in the car), performance plays an essential role.

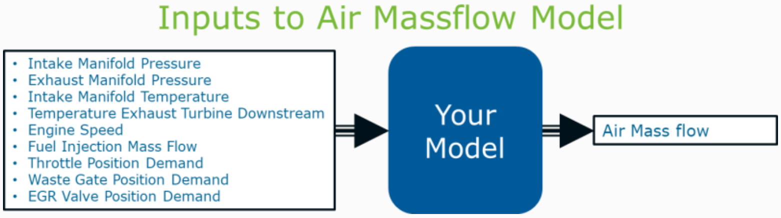


AVL List GmbH is the world's largest independent company for development, simulation and testing technology of powertrains (hybrid, combustion engines, transmission, electric drive, batteries and software) for passenger cars, trucks and large engines.

For this year's data science challenge, time series data of vehicle road trips are provided. The task is to estimate Fresh Air Mass (MfAir, fresh air mass to engine) as computationally inexpensive as possible within a certain margin of error. This information is pivotal for determining the adequate fuel mass to the engine. The estimation can be achieved by analyzing the measurements of several available sensors, e.g. engine speed, fuel injection mass flow, exhaust manifold pressure etc.

TIME FRAME & PROCESS

Your challenge consists of estimating MfAir in eval.csv by applying a model you trained with the training dataset train.csv. Since we will only evaluate the "speed" of your model.predict() call, there are some rules that need to be followed:



1. Your solution needs to be developed in an instance of the **docker container jupyter/tensorflow-notebook**
    a. Sign up for docker hub here, docker image includes everything mentioned here, here is a tutorial if needed
    b. Only solutions built on the docker container in Python are valid. Installing extra libraries is not allowed
2. Submission of solution and constraints
    a. Main rule: By running your notebook in our own instance of the docker container, we need to be able to **get predicted results for eval.csv by just hitting "Run all".** Your model.predict() - call (or equivalent) needs to be clearly identifiable
    b. For your solution to be considered "valid", **only 5% of all estimated values for eval.csv are allowed to have an absolute error of more than 3 standard deviations of MfAir in eval.csv**. We will provide a method for evaluation.
    c. Submit ONE Jupyter (IPython) notebook including your complete solution (see 2a)
    d. If your model took a long time to train (>5m) or was trained on dedicated high performance hardware, please also provide the model as a file. Your model.predict() call should then be preceded by reading the model from the file in the same directory as the notebook (see 2a)
    e. Your solutions will solely be ranked on the time it takes to predict MfAir for eval.csv in our docker instance

Until February 2, 2020 you can submit your final solution as described before via e-mail (dsc@know-center.at). Results will be ranked based on the time it takes to run model.predict(). After verifying the results, we will announce the top 5 and invite them to the AI-Know conference on February 11, 2020, where the winner will be made public!

Best of luck to you and happy number-crunching!

KNOW-CENTER GMBH
Research Center for Data-Driven Business & Big Data Analytics
Inffeldgasse 13/6, 8010 Graz, Austria

Funded by the Austrian COMET Program (Competence Centers for Excellent Technologies), we thank our funding bodies: