German Research Center for Artificial Intelligence
UNIVERSITÄT DES SAARLANDES
Chair in Business Administration
esp. Information and Service Systems
Univ.-Prof. Dr.-Ing. Wolfgang Maaß

Task Description

Dataset:
A clinical experiment was performed where the blood samples of 30 participants were collected. The dataset consists of the haematological profile of each participant. The haematological profile consists of the following attributes:

- HB: Haemoglobin
- RBC: Red blood cell (erythrocyte) count
- RET%: Reticulocytes percentage
- RDW-SD: Red cell distribution width (standard deviation)
- RDW-CW: Red Blood Cell Distribution Width - Coefficient of Variation
- OFF-HR: OFF-HR Score
- LFR: Low Flourescence Reticulocytes percentage
- IRF: Immature reticulocyte fraction
- MFR: Medium Flourescence Reticulocytes percentage
- HFR: High Flourescence Reticulocytes percentage

The dataset consists of 821 profiles from 30 participants. For each participant (represented by 'ParticipantID'), there are about 8-35 profiles in the dataset.

Task:
Develop an algorithm which can detect whether the profiles are from the same or different participants. E.g. for testing, if we give 3 profiles to the algorithm, it should report '0' if all the profiles are from the same participant and report '1' if any one of the profile is from a different participant. Remember, the testing set should not contain the 'ParticipantID'.

- Please feel free to perform multivariate analysis by using any of the statistical methods.
- You can use any of the machine learning/deep learning algorithm for solving this task.
- You can also use any domain knowledge specific to blood samples.
- In the end, you should be able to report appropriate metric which could be used to assess the performance of your trained algorithm.
- It is fine even if you don't get the best results. We are interested in knowing your creativity and approach to attack the problem.

Deadline: **Monday, 24.05.2021, 17:00 CEST**

Submission:

Your submission should be in the form of zip file, which includes:

- Code repository (written in python)
- 1-page document for the explanation of your approach
- Recent CV
- Current transcript of Saarland University