1 Welcome to the IntroML-2019 competition!

In the *IntroML-2019 competition* your goal is to use the AVP data set[1] for building & training a model-pipeline that generalizes as well as possible. If your model-pipeline performs at least as good as the one from the tutors, you will get 10 bonus points. Your model-pipeline will be tested on a secret test set where the results will be summarized by the **averaged**[2] **F1-score**.

1.1 Automatic evaluation

Because of the huge number of participants we developed a system for automatically running and evaluating the submission. Because of this system it is necessary[3] to follow some **strict guidelines** as explained in the next subsection. Further details about the system's hardware can be found in Section 1.1.3. Finally, **do not miss the deadlines as listed in Section 1.1.2**

1.1.1 How to submit

1. List all used packages (incl. their version number) in a file called `REQUIREMENTS.txt` - you can use `pip freeze` for listing all installed packages (incl. their version numbers).

2. Put your model-pipeline (code for training and prediction) into the provided file `submission.py`. You can do anything you like in the file but **you must not change the signature of the function** `my_submission` - the function `my_submission` is the main entry point and will be called by our system for evaluating your model-pipeline. Inside this function you must create and fit your model-pipeline, and return the predicted class labels of the testing samples - training samples as well as the test set (without labels) are provided as an argument to the function.

Submit the files `REQUIREMENTS.txt` and `submission.py` by **mail** to aartelt@techfak.uni-bielefeld.de - the subject of the mail must be *IntroML competition 2019*. Please list all group members in your mail and indicate if you do not want your names to be listed on the final ranking published in the LernraumPlus.

1.1.2 Deadlines

1. *First intermediate evaluation:*[4] Thursday, 19.12.2019, 11:59pm

2. *Second intermediate evaluation:*[5] Sunday, 12.01.2020, 11:59pm

3. **Final deadline:** Sunday, 19.01.2020, 11:59pm

1.1.3 Hardware

You submission will be executed in a fresh, fully patched, virtual machine running a 64 bit *Xubuntu* 18.04 - the machine has a *single* Intel(R) Xeon(R) W-2135 CPU core and 4 GB of working memory (*no GPU available*).

Your submission will be executed using *Python 3.6* - all packages (if necessary) as listed in your `REQUIREMENTS.txt` file will be installed before running your submission. There will

[1] *André's vacation pictures* data set - available in the learning space as `avp_dataset.npz`

[2] you have to set `average='macro'` when using the `sklearn.metrics.f1_score` function

[3] Otherwise your submission will be rejected!

[4] This deadline is voluntary! Your model-pipeline will be evaluated on a subset of the secret test set - you can use this score an indicator of the performance of your model-pipeline

[5] see the previous footnote

be a **time-out after 5 minutes** - your submission will be rejected automatically if it is interrupted by the time-out!

Warning

There will be serious consequences if you try to "attack"[a] our system (e.g. by injecting malware). Attacking our systems will be considered as a criminal offence:

1. You will be excluded from the course.
2. The incident will be reported to the dean and the student office - which might result in an exmatriculation (you have to leave university for ever).
3. In case of a cyber attack, we will report this incident to the police - which will result in a prosecution by the prosecutor's office[b].

Do not think that we will not detect your attack because you are that smart - we will!

[a] including fraud attempts
[b] recall that the university is an authority.