Hochschule Karlsruhe
Technik und Wirtschaft
UNIVERSITY OF APPLIED SCIENCES

Introduction to Machine Learning Challenge

Tobias Weller

07.11.2019

Task Definition

The challenge is to classify beer reviews in a team of three to five students. We consider the binary classes `positive` and `negative`. Each sample can be assigned to exactly one class. The file `train-pos.txt` contains positive reviews, the file `train-neg.txt` contains negative reviews.
The aim is to classify the evaluation dataset (`evaluation.txt`) correctly. You submit your results via email (see below). If the accuracy of your model on the test dataset is 75% or higher, you will receive a bonus of one step on your passed exam (4.0 → 3.7, 3.7 → 3.3, etc.). The submission must be handed in until 12th January 2020 23:59 (see below).
The results of the Challenge will be available in ILIAS by 14th January 2019 at latest.

Dataset

In total there are three files. The reviews are from users on beer advocate[1]. The reviews do not come from a specific beer, but consist of a mixed collection of beers. Each line in the files represents one review. The reviews are all in English.

- `train-neg.txt`: This file contains negative reviews.
- `train-pos.txt`: This file contains negative reviews.
- `evaluation.txt`: This file contains mixed reviews for which the label is not known. It will be used for evaluating the machine learning model.

Version from: Sonntag 10 November, 2019 11:31
[1] `https://www.beeradvocate.com`

Notes on Processing

- The team size is three to five students. Smaller or larger teams are not allowed. Use the ILIAS Forum if required.
- Use existing methods of machine learning for the classification.
- Do not change the order of the reviews in the file `evaluation.txt`. The evaluation is performed by comparing your predictions with the correct values.
- You may use existing tools (e.g. Weka[2]), libraries (e.g. Sklearn[3] or Keras[4]) and programming languages (Python[5] or R[6]).

Submission

The submission takes place until 12th January 2020 23:59 by email to `tobias.weller@hs-karlsruhe.de`. Submissions which take place later are excluded from the Challenge.
The following information/files must be submitted:

1. txt file with the list of classifications of the review from `evaluation.txt`.
2. Program code if used. If Weka or another Tool was used, then send a screenshot with the parameters.
3. Brief description of the approach for classification of the reviews (4 - 5 sentences).
4. Accuracy and used data split on the training dataset.
5. Full name, email address and student id of each team member, as well as a name of the team.

[2] https://www.cs.waikato.ac.nz/ ml/weka/
[3] `https://scikit-learn.org/stable/`
[4] `https://keras.io`
[5] `https://www.python.org`
[6] `https://www.r-project.org`