Wavelet Machinle Learning Challenge

Task

Your task is to write a classifier that takes as input the database given in CSV format containing biometric features from 10 people and demonstrate the capability of the classifier in classifying people based on their biodata. The idea is to have a model based on a user biodata that can later detect if the new biodata is from the user or not. So for this dataset you have to train 10 classifiers for each user and test the performance of each one against your test data from the same user and the data from other users. Your goal is to maximize the average accuracy of these classifiers.

Dataset

The `interview_biodb.csv.zip` file contains a csv file `interview_biodb.csv`.

File structure

The file has 26 columns and 234263 rows. Each row correspond to a single heart beat from each person.

22 of the columns are biometric features of those beats, called `feat1` through `feat22`. The first 4 columns give additional information on where each beat comes from:

- `_user_id`: user id, to be used as labels. Each user has a number of beat datasets that each belong to a calendar date.
- `_calendar_date`: date of each dataset of beats. Each of these contains a number of captures, identified by `_cap_seq`.
- `_cap_seq`: id of each capture in each date. They are numbered in order they are measured each date.
- `_date_time`: exact moment when each beat in a capture is being measured.

Output

The classifier code and any output results that demonstrates the capability of the classifier.

Language/Library

You may use any language or libraries you wish. Our recommendation is to use Python and Scikit-learn toolkit (or other machine learning libraries written in Python); however, using other language will not count against you if you are more comfortable in another environment.

Additional Notes

Please let us know if the requirements are unclear.