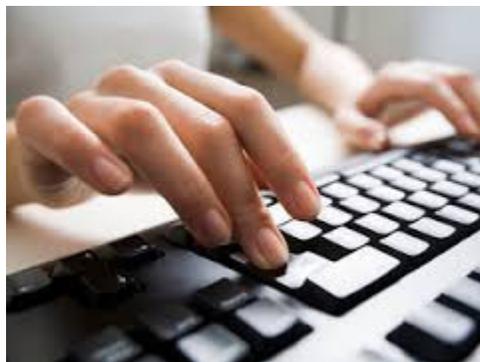


Term Project

Forensics investigation from fingerprint microbes

Recently it was discovered that DNA samples from human fingerprints are unique to individuals. Therefore, it is possible to get samples from computer keyboards and identify who is using the computer. This task provides very strong patterns and the recognition rate is high. However, the harder task is to detect which hand (left or right) the samples are gathered from. In this project, the task is to identify which hand of an individual touched the computer keyboard. This project involves assessing classification performance of clinical data gathered from DNA data on computer keyboards. The task is to perform supervised learning on the dataset and report the classification performance.



Data

There are 110 samples. Each sample contains 1956 features. Therefore each file contains a table of 110 x 1956 entries. 69 of the samples are gathered from right-hand, and 41 are from left-hand. The datasets *otu.arff*, *otu.xlsx*, and *otu.csv* are provided. They all have the same data in different formats.

Goal

With this project it is expected to have close to 90% correct classification or 0.9 ROC AUC scores. In order to achieve that you are expected to perform attribute selection (note: cross-validation in attribute selection is also required), and then go for classification with the selected attributes.

Classification Algorithms

Project Groups are free to use **ANY** classification algorithm/technology that can be found in the literature. **ANY** programming language and platforms including machine learning packages (e.g. WEKA, MATLAB, etc.) can be used. If a group programs the project, the executable build is requested. If a platform like WEKA is used, the program parameters are requested.

Performance Measures

Sensitivity, *specificity* and *AUC* is requested as the output of the program performance. If a group programs the project AUC is not necessary.

Sensitivity: = $\frac{\text{correct number of prediction of the first class}}{\text{total number of elements in the first class}}$

Specificity: = $\frac{\text{correct number of prediction of the second class}}{\text{total number of elements in the second class}}$

AUC = Area under the ROC curve.

DEADLINE IS JUNE 10TH, 2016. For all your questions and unclear points please e-mail: nalbantoglu@erciyes.edu.tr Good Luck!