Examination Term Paper: Dataset Elaboration

In this document, you will find the information regarding preparation of your *examination term paper* for the lecture "Authentication". The term paper is about 8-10 pages, however, more pages are welcome.

1. Consider at least one dataset of your choice for binary classification task (or several datasets if you wish, you might also consider the datasets discussed in our practical sessions). You will find many examples for datasets on https://archive.ics.uci.edu/ml/index.php. Have a look, e.g. on *Most Popular Data Sets* or datasets, which are categorised for *Classification Task*. Further dataset examples can be also found, e.g., on https://www.kaggle.com/datasets
2. Analyse and compare performances of at least two state-of-the-art classifiers (e.g. SVM, LDA, k-NN, naïve Bayes, classifier based on Mahalanobis distance) for the features in the original form (you might test also the classifier results for normalised features).
3. Analyse the classifier performances within established Feature Extraction or/and Feature Selection methods (e.g. FDR based or PCA based). Describe if considered Feature Extraction or/and Feature Selection methods are able to, e.g., increase accuracies, reduce complexity or execution time, or not.
4. Analyse the classifier performances within advanced methods (e.g. combinatorial or margin-based refinement, multiple classifiers). Describe if considered advanced methods are able to increase accuracies, reduce complexity or execution time, or not.

Important rules you have to know and to follow:

1. Consider MATLAB as programming language within its classification and other libraries.
2. At least one of advanced methods from lectures have to be tested and analysed in the term paper.
3. Additional advanced methods (not discussed in lectures) are welcome.

An example for the possible (such in bold style are necessary) topics in the *table of contents* of your term paper is following:

1. **Introduction** (e.g. historical information and general description of the application and corresponding dataset: Why is the dataset especially interesting? What kind of problems can it solve?)
2. Task solution by state-of-the-art classification methods.
3. Increasing classification performance: Feature Extraction or/and Feature Selection.
4. Advanced classification methods.
5. Reduction complexity and execution time.
6. **References**.

Please, take into account, that all information you will found out about the dataset of your choice is important and belongs to your term paper! There are datasets, where considered state-of-the-art classification methods might be best performing. In that case, Feature Extraction or/and Feature Selection as well as advanced classification methods will fail. That kind of information belongs to your term paper as well.