Project proposal:
Cost-sensitive machine learning methods

prof. dr Jan Mielniczuk
Institute of Computer Science PAS
miel@ipipan.waw.pl

dr Paweł Teisseyre
Institute of Computer Science, PAS
teisseyrep@ipipan.waw.pl

Description:
The aim of the project is to develop cost-sensitive feature selection methods. Nowadays, in many applications the number of features can be very large (thousands, millions), which makes fitting a classification model challenging. For this reason, most models require feature selection which can be embedded in the model or can be performed independently from the model as a pre-processing step. The significant limitation of the existing methods is that they usually assume that all features have the same cost. This assumption may be inappropriate as in some situations the acquisition of the feature values is costly. For example, in medical diagnosis, obtaining some information with naked eye is costless (e.g. sex of the patient), but each diagnostic value extracted by a clinical test is associated with its own cost. Ignoring the costs may lead to choosing features that yield a powerful model but the model cannot be used in practice as to make a prediction high cost is incurred. In such cases it is better to use a feature subset with an acceptable performance but a much lower cost. Including cost information in learning process is a challenging task, as one has to find a trade-off between quality of the feature and its cost. Aims of the research proposal include to building an open-source library with the implementations of the considered methods and performing experiments to compare the proposed methods with the existing ones.

Qualifications:

1. Major(s): computer science, mathematics
2. Experience in machine learning and statistics
3. Programming skills: R, experience with Python and Java is a plus
4. Knowledge of the English
5. Strong motivation to solve mathematical and analytical problems

Literature:

1. V. Bolon-Canedo et al., A framework for cost-based feature selection, Pattern Recognition, 2014.
2. Q. Zhou et al., Cost-sensitive feature selection using random forest: Selecting low-cost subsets of informative features, Knowledge-based systems, 2016
3. P. Teisseyre et al, Cost-sensitive classifier chains: selecting low-cost features in multi-label classification, Pattern Recognition, 2019.