

# Computational Statistics

## Tree-based methods

The dataset `Hdma` in the R package `Ecdat` contains a cross-section of 2381 mortgage applications. The goal of this exercise is to predict the decision of the bank (variable `deny`) based on the other variables.

1. Read the data. Make several plots to explore the relation between each individual predictor and the response variable.
2. Split the data into a training set and a test set.
3. Build a decision tree for this problem (use package `tree`). Evaluate its performance on the test set.
4. Prune the tree using cross-validation. Evaluate the performance of the pruned tree on the test set.
5. Apply the bagging and random forests techniques on this dataset (use package `randomForest`). Evaluate the performance of the obtained classifiers on the test set.
6. Compare the above results with those obtained by other classification methods (logistic regression, neural networks, etc.).