

## Term Assignment (40%) - 2017

Choose data sets from the UCI Machine Learning Repository.

<http://archive.ics.uci.edu/ml/>. You will need at least two data sets, one for regression and one for classification. You will carry out three analyses.

For each data set you should

- Explore the data (producing tables and graphs)
- If necessary set aside some data for testing purposes (define training set and testing set)
- Create a prediction model
- Apply the model to the test data.
- Evaluate the results.

You should apply three data mining techniques.

- Regression
- Decision Trees
- kNN

Write a report. The report should document your three analyses as follows

- Give an overview of the data sets.
- Data exploration (tables and graphs)
- Definition of training and testing sets
- Model generation & information on the model.
- Predictions for the test data
- Evaluation of the model(s).

Submit a single zipped R project (folder).

Include

- data sets in a “data” folder
- R scripts
- report as a PDF file.