## MA01 Machine Learning group project instructions:

Your group will be tasked with trying to secure a client contract for a machine learning project. This involves making a detailed plan to act substantiated by research and prototyping.

In the **first phase** the project plan and accompanying research results will be presented to the client (not graded)

In the **second phase** an actual prototype of the model and application will be developed, and the final results will be presented (graded).

The project method has to be inspired by the CRISP-DM model where all steps are recognizable, but with special emphasis on steps III (Data preparation), IV (Modelling) and V (Evaluation).

**III. Data Preparation**

**Data selection:** Determine which data sets and or features within datasets will be used and document reasons for inclusion/exclusion.

**Exploratory Data Analysis (EDA):** Investigate methods useful for gaining insight into the data, explore the data using the chosen methods and report on the data properties (quality, structure, etc) supported by visualisations.

**Data cleaning:** Research different options for data cleaning. Motivate the alignment between cleaning methods and the project objectives.

***Interim presentation (concluding first phase)***

Within this data preparation step, each group chooses one specialization topic. A deep dive into the topic is performed and the results of the research on the methods relevant to the topic, as well as a detailed plan on how to apply the methods to better reach the project goal, are presented clearly and concisely in week 7.

The topics to choose from are:

* **Topic 1**: data cleaning and imputation techniques.
* **Topic 2**: dimensionality reduction techniques.
* **Topic 3**: outlier and error detection techniques.

**IV. Modelling**

**Model selection & Test design**: Determine which machine learning techniques are appropriate for the project goal and how you will determine model performance and assess robustness.

**Model realization**: a software program with your pipeline and model implementation.

**V. Evaluation**

**Model optimization**: Determine appropriate methods for fine-tuning the model and describe how the methods were applied to convincingly show that the results are robust and reliable.

**Evaluate results:** Do the models meet the business success criteria? Which one(s) should we approve for the business?

**Conclusion**: Review the work accomplished. Was anything overlooked? Were all steps properly executed? Summarize findings and correct anything if needed.