# Data Mining Project To-Do List

October 3, 2018

# 1 Assignment 1 - Data Understanding

| Task Index | Task | Person | Output | Remarks | Status | Deadline |
|---|---|---|---|---|---|---|
| 1 | Determine the quality of the data wrt:<br>- Syntactic Accuracy<br>- Semantic Accuracy | Maddalena | - Analysis of the syntactic accuracy of all attributes in the data with plots and a textual explanation of the most interesting results<br><br>- Analysis of the semantic accuracy of all attributes in the data with plots and a textual explanation of the results | | To-Do | 8/10/2018 |
| 2 | Find outliers via boxplots | Gemma | - Boxplot of the attributes in the dataset to identify outliers<br>Textual explanation of the results obtained via the boxplots | | To-Do | 8/10/2018 |
| 3 | Discover new or confirm expected dependencies or correlations between attributes. | Daniele | - Scatter matrix and scatter plot of the attributes in the dataset and textual explanation of the results obtained via the scatter plots | | To-Do | 8/10/2018 |
| 4 | - Detect and examine missing values, possibly hidden by default values.<br>- Check specific application dependent assumptions | Riccardo | - Analysis to identify missing values in the dataset, with an explanation of the identified missing values and eventually plots thereof<br>- Plot of the distribution of attributes in the dataset<br>[QQPlot for identifying the normal distribution of attributes] | | To-Do | 8/10/2018 |
| 5 | BONUS: Identify possible data normalization to be applied to the data | | Data properly normalized and ready to be fed to data mining algorithms | | To-Do | |