

# Read Me

This is a short explanation about the different directories and their contents.

## ***Data* Directory**

All data used in our processes are stored here. There are two sub directories explained in the following:

***CSV* Directory:** All gathered data is stored here in a .csv format. The different CSV files (see our project report in 'Section 2 *Gathering, Structure and Size of the Data Set*' for more information on the content) are numbered in the order they were gathered and used in pre-processing. The file *FINAL\_PREPROCESSED\_DATA.csv* is the result of the pre-processing process (cf. '*1.PreProcessingData*' process described below).

***RapidMiner-Format* Directory:** In this directory are six sub directories each containing the data used for data mining (i.e., the data after pre-processing). As described in our project report (in 'Section 3 *Pre-Processing*'), we have six data sets with different number of classes and each with and without aggregation. The data in the subdirectories are in a RapidMiner format (the result of the *Store*-operator).

## ***RapidMiner* Directory**

All RapidMiner processes are stored here. There are two sub directories explained in the following:

***1.PreProcessing* Directory:** The pre-processing processes are stored here and roughly explained in the following. For more information, please read the comments in the respective process.

- *1.PreProcessingData*: This process has four CSV files as input (data gathered at the different resources). These are joined together and pre-processing methods are applied. The result is one final pre-processed data set stored as CSV file.
- *2.DiscretizeAndAggregate*: This process has the final pre-processed CSV file as input. The data is discretized in a different number of classes and the aggregation process described in the project report (in 'Section 3 *Pre-Processing*') is applied.

***2.Data Mining Directory***: The classifiers processes on which we focused are stored in this directory. Each process retrieves the six different data sets and applies the respective classification algorithm. For further information, please read the comments in each process.