Data Analytics

Classification:

IF checking\_status = <0 AND credit\_history = existing paid AND purpose = radio/tv AND employment = >=7 THE good (6.0/1.0)

6/1 means that out of 6 cases, one is incorrectly classified.

1. **Introduction**

Briefly introduce the aims of this coursework.

1. **Data Preparation**
   1. **Data Cleaning**

Describe the way the data is cleaned by the use of OpenRefine.

**2.2. Data conversion**

Describe how the cleaned data is converted to data sets, which can be analysed by

algorithms.

1. **Data Analytics**
   1. **Classification**

What it is

Pruning

What the j48

Rules (5-7)

Accuracy of algorithm

Confusion matrix

* 1. **Regression**
  2. **Association**

What it is

Where it is used

What is apriori

Parameters changed (a rule that has a confidence of over 0.9)

Rules (6)

* 1. **Clustering**

What it is

Algorithm simplekmeans

Table with clusters

Size of the clusters

How many cases fall into that

Visualisation:

