Data Mining and Decision Systems

08338

Assessed Coursework

Data Mining of Legacy Data

Stage 2. Data Cleaning

Student Number: STUDENTID

PDF Due: 2pm 16 November 2015 via E-Bridge

Thursday, 24 September 2015

# Data Cleaning

Data cleaning of the given data ( **BaseData-All** ) is described in Table 1 below, based on the worksheet **DataCleaning**.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Attribute Type** | **Value Type** | **Issue** | **Number of Records** | **Base Repair** | **Alternative Repairs** | |
|  |  |  |  |  |  |  |
| All | All | Duplicated Record |  |  |  |  |
| All | Nominal | Homonym (case) |  |  |  |  |
|  |  |  |  |  |  |  |
| Input | Nominal | Missing Value |  |  |  |  |
| Input | Numeric | Missing Value |  |  |  |  |
| Input | Nominal | Unknown Value |  |  |  |  |
| Input | Numeric | Unknown Value |  |  |  |  |
| Input | Nominal | Null Value |  |  |  |  |
| Input | Numeric | Null Value |  |  |  |  |
|  |  |  |  |  |  |  |
| Target (Risk) | Nominal | Missing Value |  |  |  |  |
| Target (Risk) | Nominal | Unknown Value |  |  |  |  |
| Target (Risk) | Nominal | Null Value |  |  |  |  |

*Table1. Summary Table of Data Cleaning of the given ACW data*

Now add some words describing the issues with this data and table, for instance alternative means to obtain “clean” records.