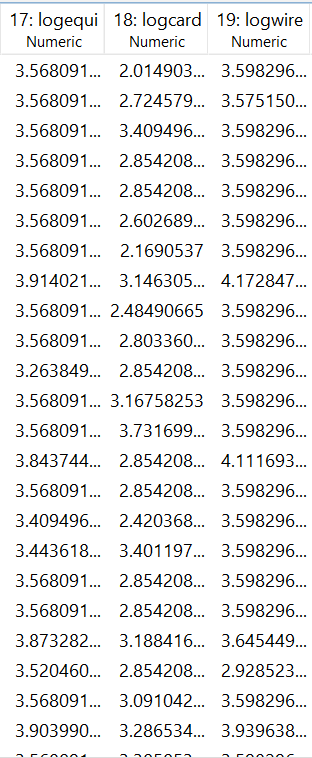
**21L-6269 Eesha Tariq Data Mining Lab BDS-6A2 Lab-03**

|  |  |  |  |
| --- | --- | --- | --- |
| **SNo** | **Preprocessing Performed** | **Result** | **Reasoning** |
| 1 | -R 2,3 | Removed ‘temperature’ and ‘humidity’ | Both the attributed just removed have little to no relation with the target variable (play). In other words, the change in temperature and/or humidity will not impact the chances of a person playing. |
| 2 | ReplaceMissingValues | Replaced missing values in ‘temperature’ and ‘humidity’ | Replaced the missing values with means of the respective columns, as the distribution of each column is approximately normal. In other words, the values are concentrated around the mean of the data. |
| 3 | ReplaceMissingValues (small\_telco\_labOne.csv) | Replaced missing values in ‘logequi’, ‘logcard’ and ‘logwire’ | Replaced the missing values with means of the respective columns, as the distribution of each column is approximately normal. (Fig 3.1) |
| 4 | -R first-last -O 3.0 -E 6.0 | Columns with outliers identified.  (Removed extra attributes generated where no outlier is detected) | Attributes that have outliers (values outside the interquartile range) have been identified.    detectionPerAttribute – True shows the outliers and extremevalues for each attribute |
| 5 | -S 0.0 -C 23 -L last | Outliers dealt with | Outliers from above removed where the ‘outlier’ column in true/yes. |
| 6 | -S 1.0 -T 0.0 | Normalized | Numeric values normalized (changed to fit a certain range to ease calculations) from the range [0.0 – 1.0]. Ranges can be changed. Applied [-1.0, 1.0] with scale 2 as well! |

A screenshot of a computer

Description automatically generated

**Fig 3.1**