**READ ME**

The code is in 2 parts: 1 main R file and a java file that is used for a small part of pre-processing.

First, the blank cells of columns Correct Step duration, Error step duration and Correct transaction time are filled with 0 (as these blanks hold meaning and removing them will lead to loss of information).

The first few lines of the R code are executed to remove nulls and unwanted columns that have no correlation with the prediction.

We now export this data as a csv and run it through the enclosed java codes that will convert the last 2 columns that have complex information (KC and opportunity) into numeric values.

For the java code “Preprocessing.java”, the data set alone is passed as command line argument (with only the KC column) and run in the usual way. The result data is outputted to an output file from where the column of values are copied and pasted onto the original csv file in the places of KC column.

The same is done with “Opportunity.java” with dataset (containing only column Opportunity).

Our dataset is now completely pre-processed and is again imported to R Studio.

The code is now executed sequentially (first for neural network and then ada boosting)

The results are logged in the report, and will get generated as the code is run.

The pre-processed data set is included as part of the folder (named as test\_final\_import127).