Guidelines for Reproducibility of a Journal Article Entitled "Using Machine Learning to Detect Liquidity Risk in Banks" By Rweyemamu Barongo and Jimmy Mbelwa

16th November, 2023

Code Organisation:

The study has three parts implemented in both Python files and Jupyter Notebook files as follows:

- Part 1: Data validation, cleaning, exploration, and transformation to required dataset (Implemented in **Prog_I_Building_Dataset_From_Collected_Data.py** or **Prog_I_Building_Dataset_From_Collected_Data.ipynb**)

- Part 2: Factors and features identification (Implemented in **Prog_II_Factor_Analysis_and_Feature_Selection.py** or **Prog_II_Factor_Analysis_and_Feature_Selection.ipynb**)

- Part 3: Model building and tests (Implemented in **Prog_III_Model_Building_and_Evaluation.py** or **Prog_III_Model_Building_and_Evaluation.ipynb**)

How to Run codes:

Jupyter notebook (.ipynb) can be run Prog_IXXX, followed by Prog_2XXX, followed by Prog 3XXX.

Python files can be run using a reproducible run button that executes a run.sh file

Data Organisation:

Source data are in the **data/original** folder, static data are in **data/static**, and results and models will be stored in **results**. The remaining folders are for passing intermediate data. The output of Prog_I will be stored in folder **data/final**, and the output of Prog_II will be stored in **data/model_input**.

Inquiries

Please send inquiries to rbarongo@gmail.com