

Deployment of Project

The project is run end-to-end by running the `Project_Execution_Script.sh` included in the submission folder.

Pre-installed Programs

The following must be installed before running the shell script:

1. R
2. Kaggle API
3. Java and all necessary sub-components to run the jar files.
4. Hadoop

Running the Script

To run the project, call the `Project_Execution_Script.sh` in a Linux Terminal. The following must be done beforehand:

1. Ensure Hadoop is already running and that the HDFS can be accessed using the commands “`hdfs dfs`”. If other commands are used on the host computer to access the HDFS, then replace the “`hdfs dfs`” with the alternative commands in the script.
2. Ensure that R packages can be installed using the “`install.packages`” command in R.
3. Run the `Project_Execution_Script.sh` script in the Linux terminal ensuring it is deployed from the same directory the script is saved in.