Predicting standardized absolute returns using rolling-sample textual modelling

***DataCleansing-public.py***

- The data cleansing process of the public data.

***DataCleansing-private.py***

- The data cleansing process of the private data.

***RiskModelling.py***

- Shows how we transfer the HSI data to risk using GARCH modelling.

***rolling\_LDA.py***

- Contains a class of the rolling LDA process as stated on the paper.

- The formation of the customised topic scores are also embedded in this class.

***rolling\_analysis.py***

- Contains a class to do some preliminary analysis on the rolling window data.

- Most of the functions are not made into use for the outcome of the result discussed on the paper.

***regression.py***

- Contains a class that utilise the result from rolling LDA to perform a regression-based prediction on the risk which is drawn from RiskModelling.py

- Also include the generation of report to tell which threshold should be optimal to use in that particular dataset.

***result\_analysis.py***

- Analyze the result produced in regression.py with the help of graphs like histogram, heat map, confusion matrix, and word cloud.

- The related analysis are discussed in the paper to show how our proposed method is helpful on the prediction of market volatility.

***sample\_run\_demo.ipynb***

- A full demo to show how we use the rolling\_LDA.py, rolling\_analysis.py, regression.py, and result\_analysis.py to perform prediction and related analysis.

- This version is our original version to perform the analysis. The codes will be duplicated with rolling\_LDA.py, rolling\_analysis.py, regression.py, and result\_analysis.py. Those .py files are trying to make each step clearer. Since our proposed method involves quite a number of visualisation, so we will encourage running jupyter notebook in order to get a better understanding of the entire flow.