Solution Design Document

Pairs Trading strategy with the application of Deep Learning

Himanshu Kumar

Overview

This project aims to enhance the pairs trading using deep learning techniques to predict cointegration of asset pairs in future periods. Pairs trading is a market-neutral strategy that profits from the price deviations between two historically co-integrated assets. By introducing a deep-LSTM architecture, our solution seeks to improve the accuracy of selecting pairs that are likely to remain cointegrated

Solution Architecture

This hosted Code repository contains 4 major folders namely:
1. Fetch_data_via_api - (Fill the API via Alpha Vantage and pull & store data)
2. Stationarity and Coint - (Research, Feature engineering and storing data for model training)
3. Train via LSTM - (Contains code to train the LSTM model)
4. Pairs Trading Models - (Uses the data & Models to actually trade and analysis)

How to reproduce the results :

Simpler way : download the folder 🖼 4. Pairs Trading Models upload in gdrive and run via colab else -

   Step 1 : Download the Pairs Trading Models folder (4) to your local PC
   Step 2 : Upload this folder in your google drive
   Step 3 : In utils folder you will need to put data, you can use either of the methods -
   ➢ Download data from this link OR
   ➢ Run the script by providing your API from Alpha Vantage in folder Fetch_data_via_api
   Step 4 : Run all the Cells in the Pairs Trading.ipynb file in the uploaded folder, You will need to Authorise Gdrive access once prompted to access the Model and Data stored

Detailed functionalities of each folder and files

1. Fetch_data_via_api : This folder contains 2 files helper.py and main.py, helper contains all the helper functions reqd. To fetch and preprocess the data, whereas to fetch data you need to obtain free API from https://www.alphavantage.co/   and run the main.py to get all the data.

2. Stationary and Coint : This folder aims at preparing dataset for the Neural network training in an appropriate format. Pair_Selection_Process.ipynb file is the heart for this folder and upon execution you will create a dataset named LSTM_NN_dataset.h5, you can find this dataset here.

3. Train via LSTM : This folder contains Build_model.ipynb file which actually contains the neural network architecture, it trains Hedge ratio predictor model and Classifier model on the LSTM_NN_dataset.h5 dataset. You can find these trained models stored in the folder named Model.

4. Pairs Trading Models : This folder contains 2 subfolders - Models (contain different trained models) and Utils (contains data necessary), Pairs Trading.ipynb script contains the complete infrastructure built to trade and compare all the models across different time periods and metrics.

How to run Locally vs Colab : The provided code base is structured in a way that if you run on colab there should not be any issues if you put necessary data in your drive. To run locally you will need to change the way how data is being accessed and will need to run requirements.txt rest everything remains the same.