Files Description:

1. Data_processing.ipynb , Data_processing_2.ipynb – these two files contains pre-processing of data. The script generates 2 files final1.csv, processedFeatures.csv respectively.

   Note: The data is collected from multiple sources. Every stock and feature have their own seperate csv files. The data is aggregated into final1.csv and processedFeatures.csv which contains only required data. These files are generated using the above notebooks. As we are not providing all the data sources, It is not required to run the above files. We have attached final1.csv and processedFeatures.csv in the data folder of zip file.

2. Stock Prediction.ipynb – The code containing the designing, training and evaluation of the model. It has three parts:

   PART - I Loding the data

   In this part we load the aggregated data and process the data into the format that fits the model input shape.

   Part - II: Desing and Training the Model

   In this part the model is designed and trained for 35 epochs. The model is saved at the end of the training. (“stockmodel.h5”)

   Part - III: Evaluation of the Model

   In this part the trained model is loaded back into the jupyter notebook and evaluated over the test data

   Note: You can skip part 2 and directly continue to part 3 if you do not have enough compute power to train the model. we already provided “stockmodel.h5” file, which is trained under GPU for 35 epochs. The trained model is loaded in the part III

3. stockmodel.h5 - Trained model
4. data/final1.csv – The processed data that contains stock values of over 410 companies for the last 5 years.
5. data/processedFeatures.csv- The file contains data on important features that affect the economy over the last 5 years.

Required Python Dependencies:

1. Keras
2. Matplotlib
3. Pandas
4. Numpy
5. Scikit-learn
6. Jupyter notebook
7. Glob

Software Requirements:

1. An IDE (VS CODE or Google Colab preferable)
2. Python 3.6+
3. Anaconda or pip