# Code Description

After unzipping the compressed package, you can get the working directory ‘project\_code’. Let’s say it root path ‘./’. Entering this directory, you can see a list of files. Let me introduce the role of each file and how to run this file

## Data Preparation

### Text Data Preparation

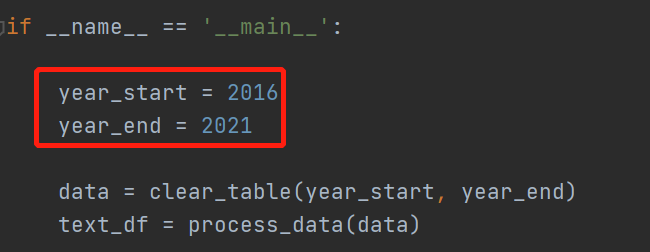
The source code ‘./bert\_classification/modeling.py’, ‘./bert\_classification/multi\_focal\_loss.py’, ‘./bert\_classification/optimization.py’, ‘./bert\_classification/tokenization .py’ are extracted from google open source bert code (<https://github.com/google-research/bert>). And I modify some parts of these code, for example, add a Softmax layer with multi-focal loss.

Firstly, download the pretrained BERT model from <https://drive.google.com/file/d/1yS3FCA-SEHsQxxslcqajcTJtVB2Tovjy/view?usp=sharing> and unzip the package you will get a folder ‘chinese\_L-12\_H-768\_A-12’ containing several files. Then move this folder to ‘./bert\_classification/data’

In ‘./bert\_classification’ is our hand-labeled training set, named ‘train\_data.xlsx’, about 3,500. We can use this to train our Bert Model: Firstly, move the folder ‘./bert\_classification’ to HKU GPU FARM, open ‘bert\_classification’ directory and execute command ‘**sh train\_bert\_model.sh**’ in the console of Linux (because the Bert model requires a GPU with high memory, we use HKU GPU FARM which is built on Linux system).

To generate a test set, firstly download the raw text data from <https://drive.google.com/file/d/1FBdfBT_KVrLOXcXei_w-2rMS1wcFY-Ey/view?usp=sharing> and unzip the package, and you will get a folder ‘舆情数据’. Move this folder to ‘./project数据’.

Then run ‘**./text\_data\_preprocess.py**’. By specifying the variable ‘year\_start = 2006’, ‘year\_end=2016’, we can generate about 10 years’ training text data in folder ‘./project数据/舆情数据/text\_data’. Then move this folder to remote folder ‘./bert\_classification’, execute command ‘python testData\_gen.py’, you can get labeled text stored in the remoter folder ‘./text\_result.xlsx’. For the back test, just specify ‘year\_start = 2016’, ‘year\_end=2021’ .



Then download training and back test ‘./text\_result.xlsx’ from remote folder, and place it in ‘./project数据/舆情数据’, rename them ‘text\_result\_ori.xlsx’ and ‘./test\_result.xlsx’ respectively.

### Factors Data Preparation

To generate factors data, firstly download the raw factors data from <https://drive.google.com/file/d/1cWdH9wBhgUGwg5heterCO2gtkSSeuruH/view?usp=sharing>, and unzip the package, and you will get a folder ‘原始数据’. Move this folder under ‘./project数据’. Then run ‘./data\_preprocess.py’, and you can get training data ‘data\_train\_date.xlsx’ and back test data ‘data\_backtest\_date.xlsx under folder ‘./project数据/因子数据’.

### Merge factors data with text data

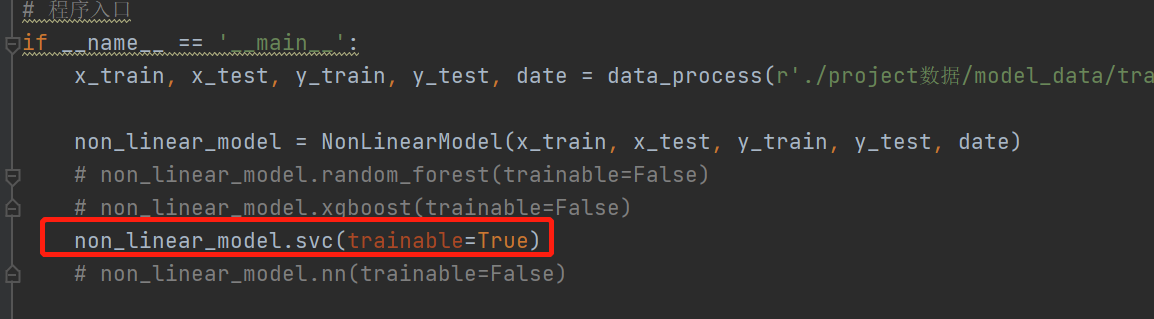
To merge factors with text factor, run ‘./merge\_factor\_with\_text.py’, then you can get the final training data ‘train\_data.parquet.gzip’ and back test data ‘backtest\_data.parquet.gzip’ under ‘./project数据/model\_data’.

### Factors Selection

To select the factors, just run ‘./factor\_selection.py’, and the result ‘factorFilter\_result.xlsx’ will be saved under folder ‘./project数据/因子数据’.

## Model Training

Run ‘./non\_time\_factor\_model.py’, you can train Random Forest, Xgboost, Pca+Svm, Neural Network classification model using data ‘train\_data.parquet.gzip’ generated in above steps. And the trained model will be saved under ‘./project数据/model/’. In detail, if you want to train and save the a specific model, you can set the parameter ‘trainable=True’ like the below picture shows.

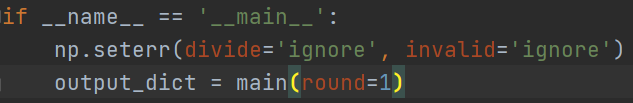


Similarly, run ‘. /time\_factor\_model.py’, you can train the ‘RNN’, ‘LSTM’, ‘CNN’, ‘CNN+LSTM+ATTENTION’ classification model.

## Optimization

Run ‘./Optimization.py’ setting the parameter ‘round=1’ (img1) in main function, you can get a json file ‘output\_code.json’ showing which stocks to buy every month within back test period. Then upload ‘output\_code.json’ together with ‘./joint\_quant1.py’ to JointQuant (<https://www.joinquant.com/>). In detail, place joint\_quant1.py under ‘策略研究->策略列表’ (img2) and place output\_code.json under ‘策略研究->研究环境’ (img3). Then run ‘joint\_quant1.py’ on joint quant (img4), the program will select top 10 stocks according to ROE every month, and saved in the file ‘stock\_choosen\_10.json’. Then download it and move it to root path ‘./’.

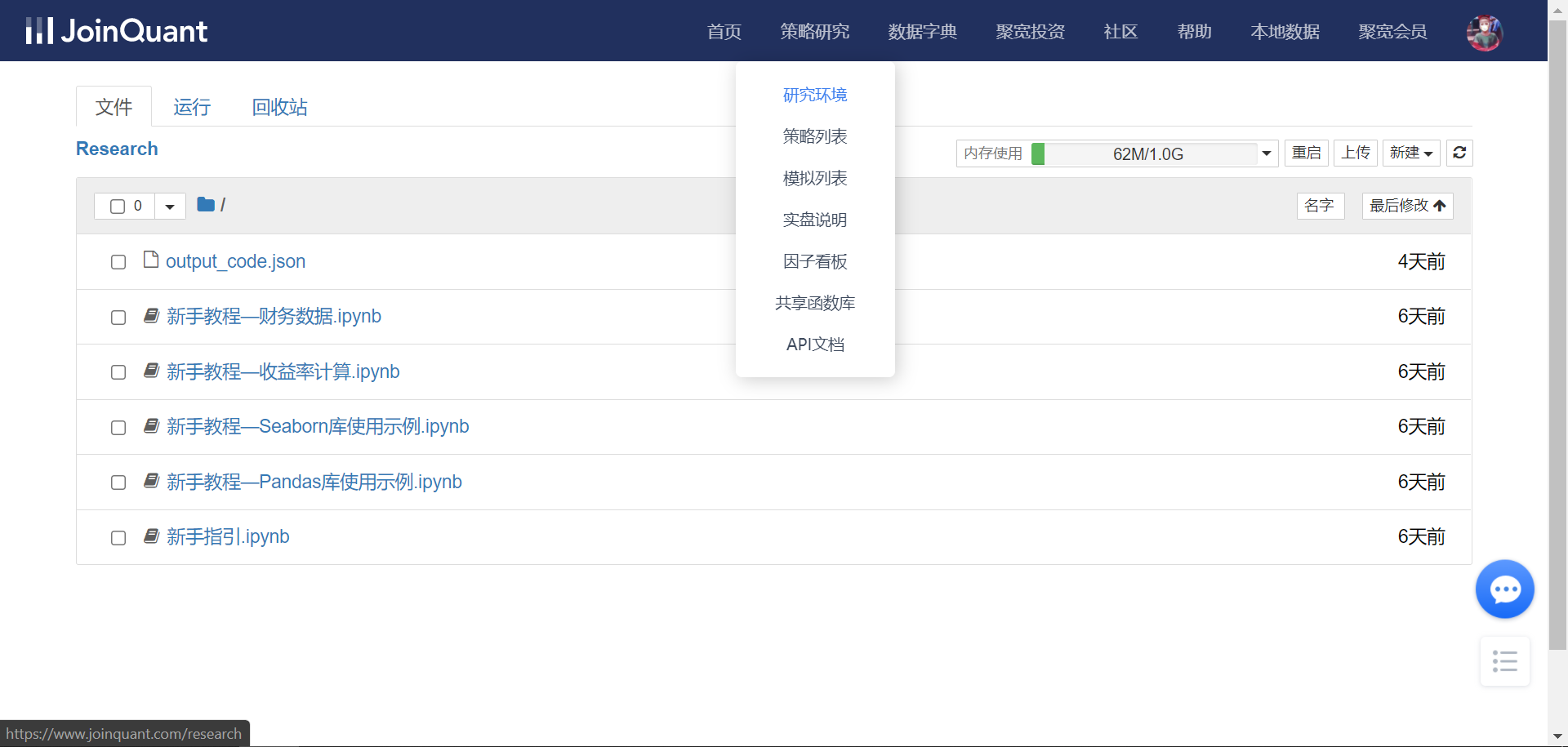
Then, Run ‘./Optimization.py’ setting the parameter ‘round=2’ (img5) in main function, you can get a json file ‘weight\_code.json’ showing the internal weights within stock portfolio. Similarly, upload ‘weight\_code.json’ and ‘joint\_quant2.py’ to JointQuant (img6, img7) and run ‘joint\_quant12.py’ on joint quant (img8), the program will return the final yield curve. You can allocate different weight between stock and bond as img9 shows. By doing so, you can get different yield curve.



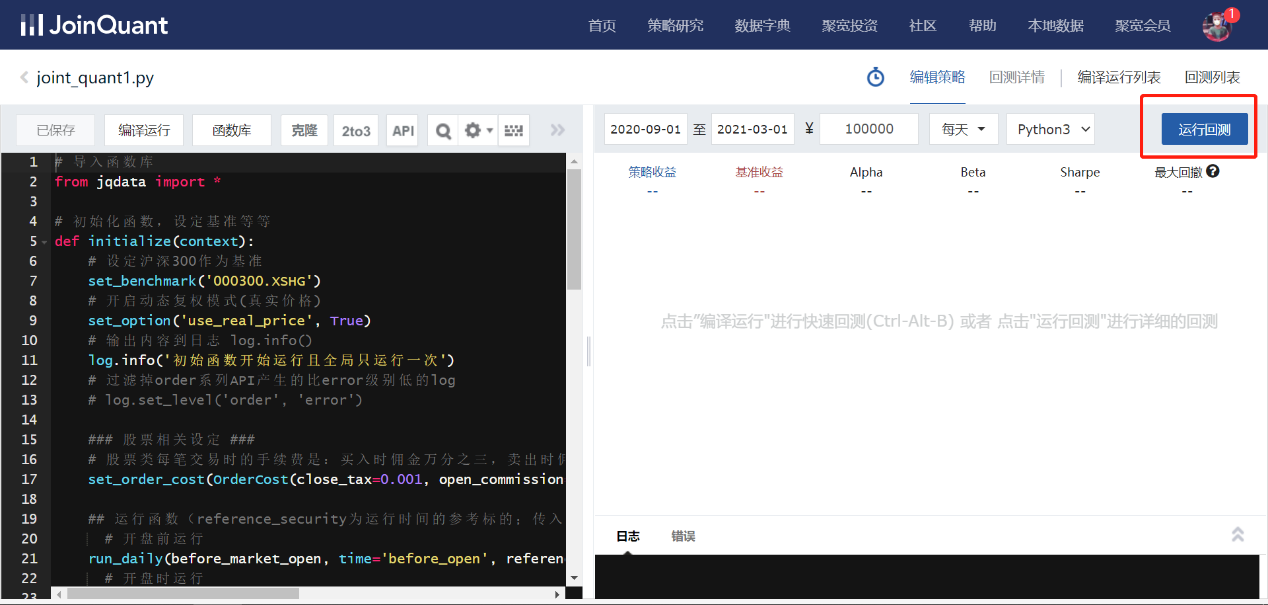
*(img1)*



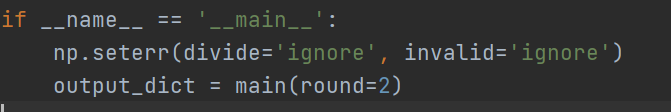
*(img2)*



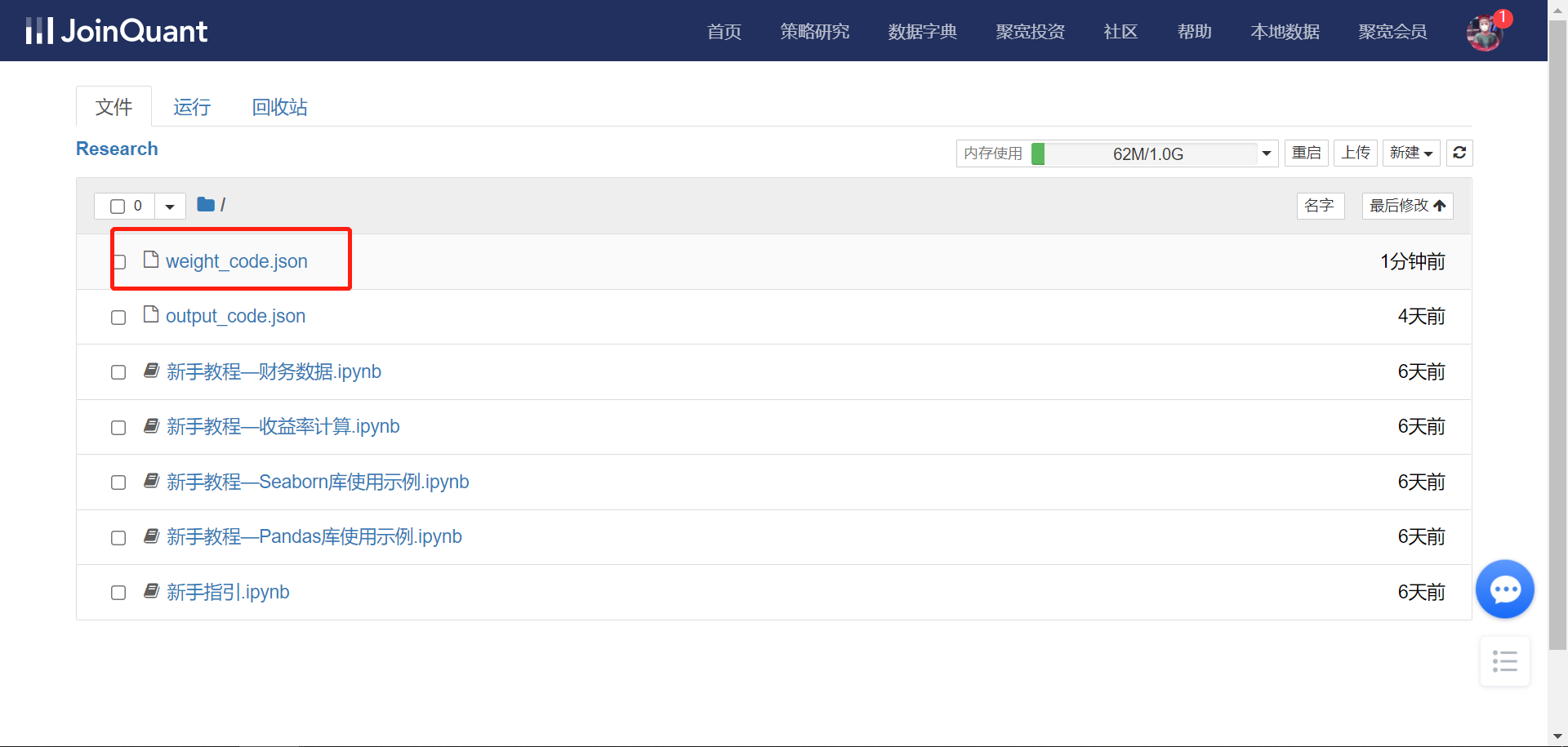
*(img3)*



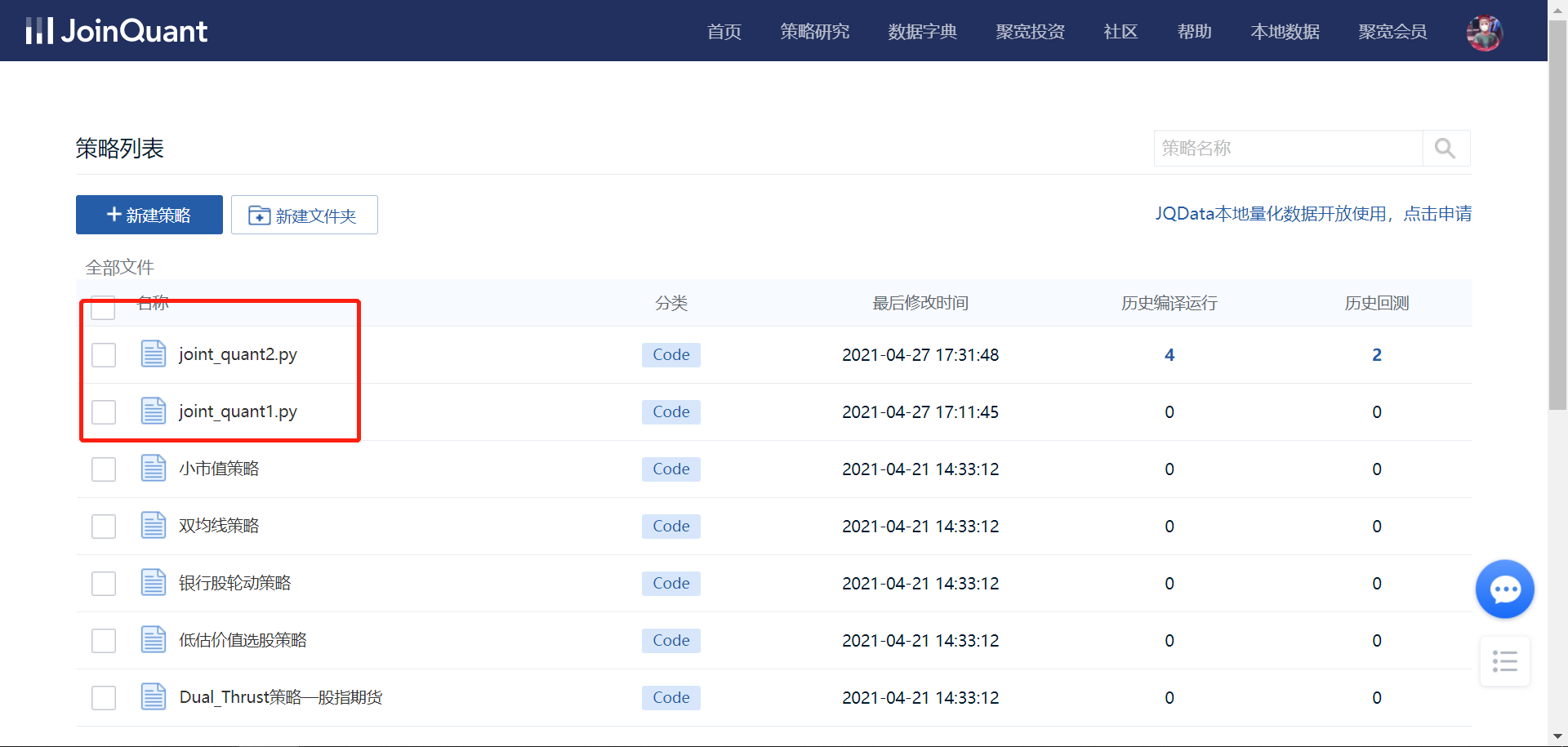
*(img4)*



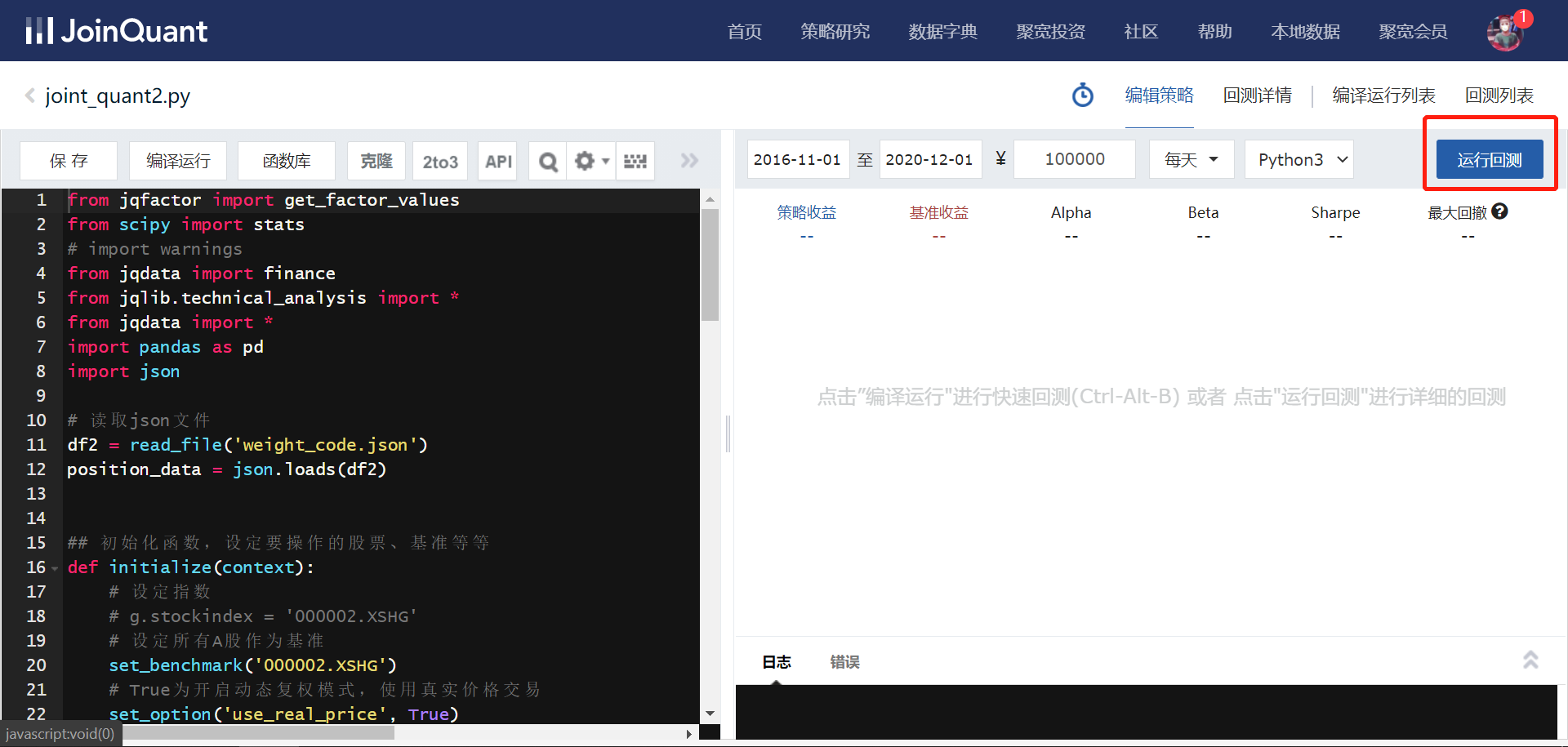
*(img5)*



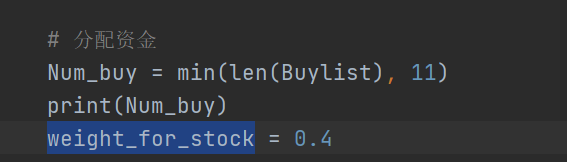
*(img6)*



*(img7)*



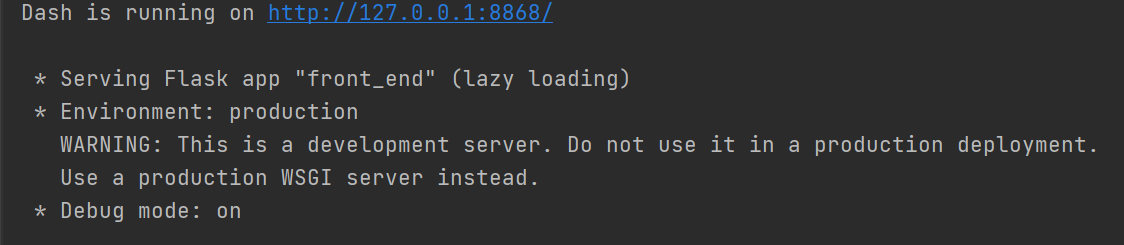
*(img8)*



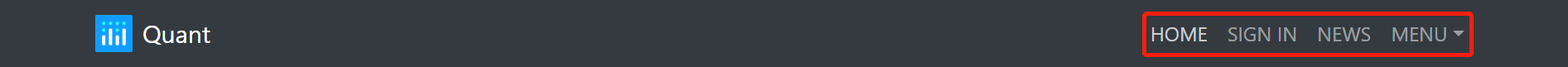
*(img9)*

## User Interface

To see the user interface, just run ‘./front\_end.py’, and the program will return a URL like this:

 Enter this URL in browser, you will see the webpage.

The web page is made up of 3 parts. You can switch to these 3 parts by clicking ‘HOME, SIGN IN, NEWS’ respectively. The ‘MENU’ part will be developed in the future after the project.



The first part is ‘Home Page’. You can switch to this page by clicking ‘HOME’. And you will see the yield curve and some parameters of our strategy.

The second part is ‘Questionnaire’. You can switch to this page by clicking ‘SIGN IN’. And you will see a questionnaire to investigate investors’ risk appetite.

The third part is ‘NEWS’. You can switch to this page by clicking ‘NEWS’. You can choose the amount of news you want to see, by clicking ‘Confirm’ button, the program will extract latest financial news and shows their titles on the page like below:



If you are interested in one news. You can click it and the webpage will be redirected to a specific webpage which contains full content of this news.