Predicting Stocks Closing Prices

by AvimanyuSingh

Problem | Submissions | Leaderboard | Discussions

You are given closing prices of **395** stocks for the last **5** years (**1230** days, i.e., **246** working days, *Monday* to *Friday*, per year). You are given INR **1** Crore (**1** Crore = **10** Million) and you need to invest all or some amount across a minimum of **10** stocks. Note that:

- The maximum investment in a stock is capped at INR **10** Lakhs (**1** Million = **10** Lakhs).
- The closing price on the 1230^{th} day will be considered as the trade entry point, i.e., you are only allowed to buy stocks on this day. Shares should not be bought in fractions.
- The closing price on the 1233^{rd} day will be considered as the trade exit point, i.e., the profit or loss of your portfolio will be calculated on the basis of the closing prices of the 1233^{rd} day.

Also, note that you are not provided the closing prices on the 1231^{st}, 1232^{nd} and 1233^{rd} days, so you should predict the closing prices for the 1231^{st}, 1232^{nd} and 1233^{rd} days.

Dataset

We provide the csv file (*MD5* checksum is `50260cfedc3fbd1d2c11bae1f006cc7a`): stocks_closing_prices.csv.zip containing the information of closing prices of **395** stocks for the last **5** years. The file contains the following four columns:

- `Day Sequence` : This describes the number of days in the range **1** to **1230** inclusive.
- `Weekdays` : This describes the working day, *Monday* to *Friday*, given as integers in the range **2** to **6** inclusive.
- `Stock ID` : This describes the IDs of the stocks in the range **1** to **395**.
- `Closing Price` : This describes the closing price of a stock.

Submission Details

You are required to upload the following three files:

- The output file, `predicted_prices.csv` (max allowed size is `10MB`). The file should contain the closing prices on 1231^{st}, 1232^{nd} and 1233^{rd} days for each of the **395** stocks. Moreover, it must also specify the quantity of stocks bought. If you did not invest in a particular stock then the quantity should be **0**.

  A valid output file has the following format:

```
Stock_ID,Closing_Price_1231,Closing_Price_1232,Closing_Price_1233,Quantity
1,1.1,1.2,1.3,10
2,1.2,1.2,3.3,40
3,3.1,3.0,6.5,0
4,1.2,1.2,3.3,0
5,1.2,1.2,3.3,97
.
.
.
393,2.2,2.2,2.4,100
394,1.8,1.9,3.3,0
395,1.8,1.9,3.3,0
```

  Note that:

  - The first line of the output file should contain the header, with `Stock_ID`, `Closing_Price_1231`, `Closing_Price_1232`, `Closing_Price_1233` and `Quantity` as the column names separated by a `comma`.
  - There should be exactly **396** rows including the column header. The `Stock_ID` should be in ascending order, i.e., the information of Stock ID i should be given before the information of Stock ID j if $i < j$.

- You shoud invest across a minimum of 10 stocks.
- You should not invest more than INR 10 Lakhs in any stock.
- You should not invest more than INR 1 Crore in total.

- A *PDF* file (maximum allowed size is `4MB`) providing the findings and justification on the following topics:
  - Write a few lines about training dataset quality and any errors found in the training dataset.
  - Explain the data preprocessing steps.
  - Explain and justify the model you've chosen for the prediction.
- The source code of your approach for this task. Upload a *zip* file (maximum allowed size is `5MB`) with all relevant files to reproduce your results. The submitted file must have a `README` file with a detailed description about how to run the model to predict the closing prices list for the stocks and generate the `predicted_prices.csv`. Do not forget to include links to any external libraries or packages you use for the generation of your model.

There is no limit on execution time, but the code should generate the output file: `predicted_prices.csv`.

Evaluation

We calculate the RMSE for the predicted closing prices on 1231^{st}, 1232^{nd} and 1233^{rd} days:

$$rmse_{1231} = \sqrt{\frac{1}{395} \times \sum_{i=1}^{395} (p\_1231\_actual\_closing\_price_i - p\_1231\_predicted\_closing\_price_i)^2}$$

$$rmse_{1232} = \sqrt{\frac{1}{395} \times \sum_{i=1}^{395} (p\_1232\_actual\_closing\_price_i - p\_1232\_predicted\_closing\_price_i)^2}$$

$$rmse_{1233} = \sqrt{\frac{1}{395} \times \sum_{i=1}^{395} (p\_1233\_actual\_closing\_price_i - p\_1233\_predicted\_closing\_price_i)^2}$$

We calculate the normalized RMSE:

$$nomalized\_rmse = \frac{\tanh(rmse_{1231}) + \tanh(rmse_{1232}) + \tanh(rmse_{1233})}{3}$$

We also calcuate the portfolio return on the basis of your investment:

$$portfolio\_return = \sum_{i=1}^{365} quantity_i \times (p\_1233\_actual\_closing\_price_i - p\_1230\_actual\_closing\_price_i)$$

The normalized portfolio return is calculated as:

$$nomalized\_portfolio\_return = \frac{portfolio\_return}{maximal\_portfolio\_return}$$

Here, $maximal\_portfolio\_return$ is a constant. Score of your prediction is calculated as:

$$score = nomalized\_portfolio\_return - 0.2 \times nomalized\_rmse$$

Your leaderboard score will be $10^3 \times score$.

Note

You can make any number of submissions and on the basis of the uploaded output file `predicted_prices.csv`, the checker shows a message to help you improve your model. We will consider the maximal score of all the submissions to decide the leaderboard ranking.

File Upload

Upload your files here:

predicted_prices.csv:

Documentation:

Source Code: