Predicting Stock Prices report

Dataset

The dataset is provided by HackerRank.

The dataset contains the price information of 395 stocks from day 1 to day 1230. The task is to predict the stock price in day 1231, 1232 and 1233, and decide how many stocks should be bought to maximize the profit.

The dataset is provided in good quality. No errors were found.

Preprocessing

There is not much preprocessing task.

Each stock is processed independently, so the only task need to be done is to divide the dataset into 395 individual datasets.

Prediction

The task of prediction is done by using ARIMA model implemented in *forecast* package[1] in R.

For each stock, we first use the data from day 1 to day 1295 as training data and the price information from day 1296 to day 1300 as testing data. There are two predictions are performed for each stock, one is done by *ets* and one is done by *arima.* The RMSE value in the testing dataset is used to decide which prediction will be used afterwards. RMSE is calculated by using the package *Metrics*[2]*.*

The true prediction is done by using all information from 1230 days as the training data.

Quantity Buying

The task of decide the quantity for each stock is simple.

First we calculate the relative difference, i.e. how many percent the price increase from day 1230 (actual) to day 1233 (predicted).
Then we rank the stock by the relative increase of price.

---

[1] https://cran.r-project.org/web/packages/forecast/
[2] https://cran.r-project.org/web/packages/Metrics/

Then, from high to low, we spend as much money as possible to buy the stock, until no money left.