Apple is a great company. They build outstanding hard wares and soft wares. Many of us enjoy the convenience of cooperation of iPhones, iPads, Apple Watches and Macs. Then a group called “apple fan” merged. However, some apple fans love apple so much that they even want to buy apple stocks to prove that their love is TRUE love. Just like this guy. I believe he must be a perfect husband if he has been married.

This guy wants to buy stocks but only apple stocks. However, he loves his money too. The last thing he wants to see is seeing apple get defeated by other tech companies just like Google, SUMSANG or HUAWEI. Nevertheless, losing money comes just behind. It is the second last thing he wants to see. So, we are trying to build a model to invest apple and make a profit, or to say, minimize his loss.

Now, let’s introduce yijie to tell you some process before conducting the expriments.

…

It seems that this guy is eager to do something terrible. But for sake of the safety of his wallet, we must do the experiments to prove the efficiency of our model.

Now, let us conduct the first experiment. This experiment is to find out if social media sentimental data has an influence on the result of the prediction, and if so, we are going to find out the winner language model. The base data set is pure stock price data as mentioned before. We also generated EWMA data as different columns as YEJ said. Then, we use two different sentimental analysis models, VADER and FinBERT to evaluate the sentimental score of each tweet. VADER is a traditional machine learning method; it is widely used today because it is really fast and resource-saving. As to FinBERT, it is a pretrained large model. It is much slower than VADER, and it costs a lot of computing resources. However, the outcome of FinBERT is very marvel. Before the experiment, we have guessed that the winner will be FinBERT. Let’s look at the experiment results. From the error table, FinBERT didn’t fail our expectation. It is the sort of the art! It has the best performance in all three indexes. Now, let’s bring our winner to the next level.

In the next level, we use the dataset of base combined with FinBERT sentimental scores. In this experiment, we are going to find out which network works best in stock price prediction. The candidates are RNN, GRU, LSTM and MLT. RNN is simple, yet useful. It is fast but has a few limitations. GRU and LSTM can remember earlier data, GRU has less gate than LSTM. It combines long-term gate and short-term gate as a new memory gate. We also give MLP a chance because it has a large number of parameters. We want to see if it makes a miracle. We must explain why we didn’t use the models in the right side of the table. For bidirectional models, our reason is that our mission is a financial prediction task, and it is obvious that we cannot retrieve data from the future, otherwise we would not working so hard on this model even at the risk of losing our hairs. And another fantastic model we gave up is CNN. It is usually useful in matrix related tasks instead of the time series tasks.

We also had a guess before the competition, we believe that the champion will come from LSTM or GRU.

Here, you can see the loss and predictions of different models. The blue line is the reality. To our surprise, MLP really did make a miracle! It defeated GRU, LSTM and RNN and killed the game!

To be more direct, let’s look at the error table. MLP shows much better performance on all three indexes! We don’t know the reason, too, but we have to respect the result. A game is a game and a word is a word, MLP is going to the final.

In the last competition, we use MLP to predict on the test data. Note that this is the data the model has never seen before. We trained the model with training dataset and dev dataset.

In our simulation on quant trading, our strategy is using the model to predict the close price of the next day, if the predicted price is higher than the actual close price of the previous day, we will buy the stock at the last moment, aka 3:59 am. Otherwise, we will sell the stock.

The result of the experiment is very astonishing. Our model not only didn’t lose money, but also profited much more than the stock itself! Now, we can give the model to the crazy apple fan!

The presentation today consists of four parts. We will introduce an overview on our research, and then, we will talk about how data is prepared, including data collection and preprocessing. Later, we will do some feature engineering to help building the model. After that, is the most exciting part: building a good model to help that guy!

I think the stock market is a chaos system, so the price data alone is far from enough to do the prediction. So, we got many comments from social media to help the prediction. However, it is just an assumption. We need to prove the data is helpful so that we can continue our experiment. So, the first experiment we are going to do is using pure stock price data, which will be called as base in this pre later and compare it with data with sentimental scores. We will use a simple network in the first-round prediction competition, because we only need to prove that sentimental data is useful. After having the champion of sentimental analysis models, we are going to do experiments on prediction networks. Winner winner, chicken dinner. The winner of the second competition will be brought to the final experiment: simulated quantitative trading. We will see whether our model is useful or maybe just a piece of shit.