Capstone Proposal

Stock Target Price Forecast and Risk Estimate

GETTY

Background

The stock market has high potential and requirements to apply machine learning since it has a huge volume of data. Currently Wall Street has applied machine learning to predict stock prices or make related market trade decisions. However, there are many problems with directly forecasting stock prices. One is the difficulty of collecting all related data to build a forecast model because the stock price is related to many direct or indirect information most of which are hard to digitize. Besides, unpredictable futures make stock prices predict hard to close to true prices.

Statement

This project aims to build a machine learning model that could choose the best stock which has a high probability to gain profit in the future. To avoid unpredictable stock prices, this project only predicts if the stock could gain promising profit rather than give a specific future price. Promising profit is defined as increasing 50% of the original stock prices in a two-year period. To add more information, we also forecast the stock price in the near future as a middle step of the whole project. After training the forecasting model, we use the 1,6,12

month forecast result as three inputs for the promising profit stock predict model, with other company information together to train this classification model.

Data and Input

The data used include S&P 500 companies (there are 505 companies in total) stocks, including daily trading price as well as company information. Data was downloaded according to Yahoo Finance APIs `yfinance`.
To fit the forecasting model, time series price data will be tailored into a 3 years period, two year of training and one year of testing. Since most stock has more than 3 years of trading history, there are more than 2000 training time series data
As we want to choose the most promising stock now, we need to leave the company stock information in 2021 alone. To fit the classification model we tailor stock information into a 2 year period, for example, 2018-2020. We use company information in 2018 as well as the target price we trained in the previous model to predict if in 2020 stock price will increase 50%. We don't want to consider the earlier time stock, which may not represent the current stock market situation. So, the earliest time we trace back is 2010, so there are 5 periods of data for each company (if the company started trading before 2010). There are more than 2000 pieces of data for training and testing, after a 75/25 split, there are more than 1500 data points for training (not sure if 2000 is enough for training and testing this model, if not, I will add other stocks).

Solution Statement

In this project, Two models will be built, one is the forecast model, which will forecast the price of stocks in the future one year, by using time series data of stock market close price. The other model is a classification model. to classify if a stock will gain promising profit in 2 years. The classification model will select the 1,6,12 month price of the previous forecast model as input features, with other features including information about the company itself. These features may include: sector, revenue, earnings, dividends, Investment Banking's Perspective, 52 weeks high, 52 weeks low, etc

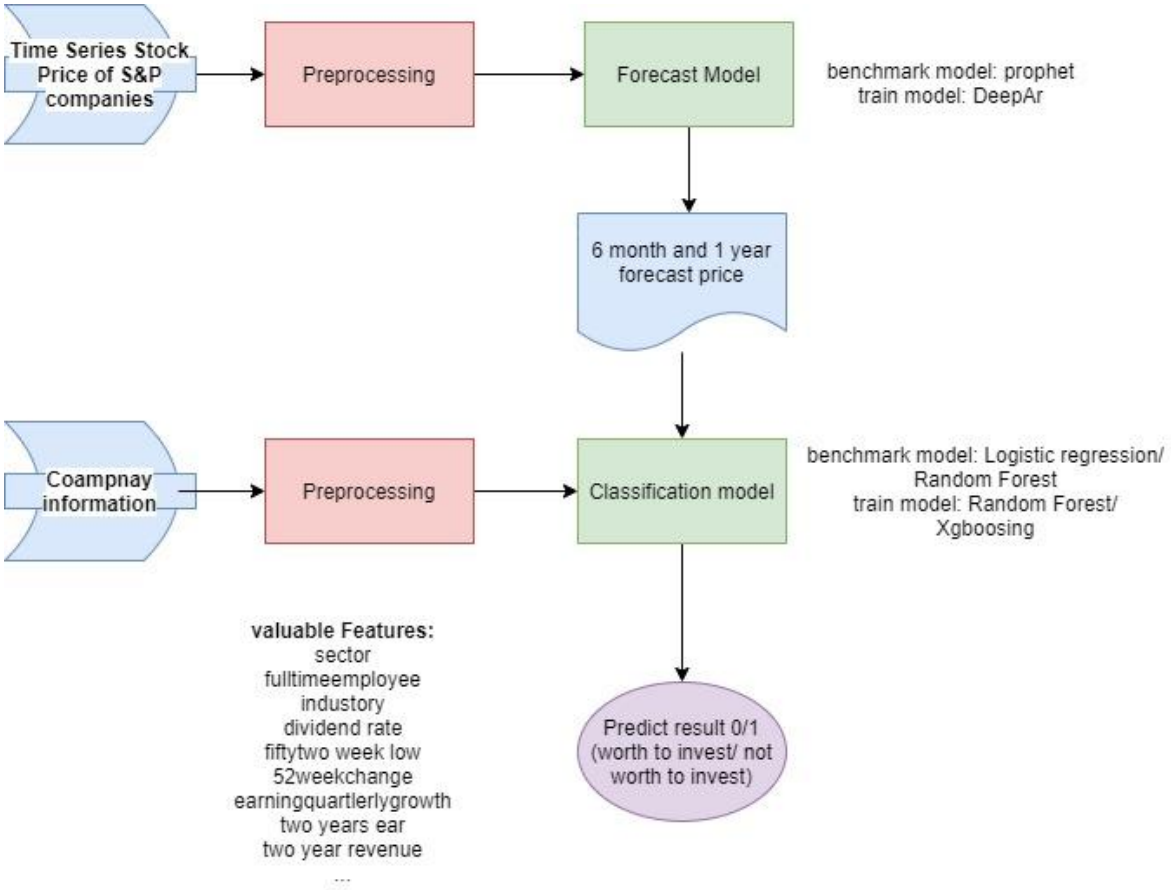
Benchmark model

To create a benchmark model, a simple forecast model (using `prophet`) and a simple classification model (logistic regression) will be built as the benchmark models and improvement will be made based on these benchmark models.

Evaluation Metric

Precision will be used as the evaluate metric. Because for this project, what we really care is that the stocks we choose are stocks that will truly gain profit in the future, and we don't need to consider the performance of the stocks that aren't chosen. So, to evaluate the model performance, calculate the precision will be enough

Project Design



Resource

- Yahoo Finance API: https://github.com/ranaroussi/yfinance
- Blog: Stock Price Prediction – Machine Learning Project in Python: https://data-flair.training/blogs/stock-price-prediction-machine-learning-project-in-python/
-