Facebook Stock Price Prediction & Trading Strategy Report

Minghao Ru 001088106 Tong Yang 001302930

INFO 7390 Final Project Report

Portfolio link

https://github.com/yangtong951019/INFO7390_Final_Project

Abstract

There are four main parts in our project. In the first part, we analyze the daily returns and log daily returns of Facebook's stock price by statistical features, then we use Karman Filter & Garch models to do a regression on the returns. In the second part, we use several models includes Fama French 3 Factors, SVM bootstrap & ARIMA etc. to predict the Facebook's stock close price. We compare the prediction result to get a best model. Then we use Random Forest to combine each model to give the most accurate prediction. In the last part, we analyze Trading Strategies like MACD Signal Crossovers and Bollinger Band to give a proper investment advice for new investors.

Key word: Stock Price daily return, Close price, Prediction, Trading Strategies.

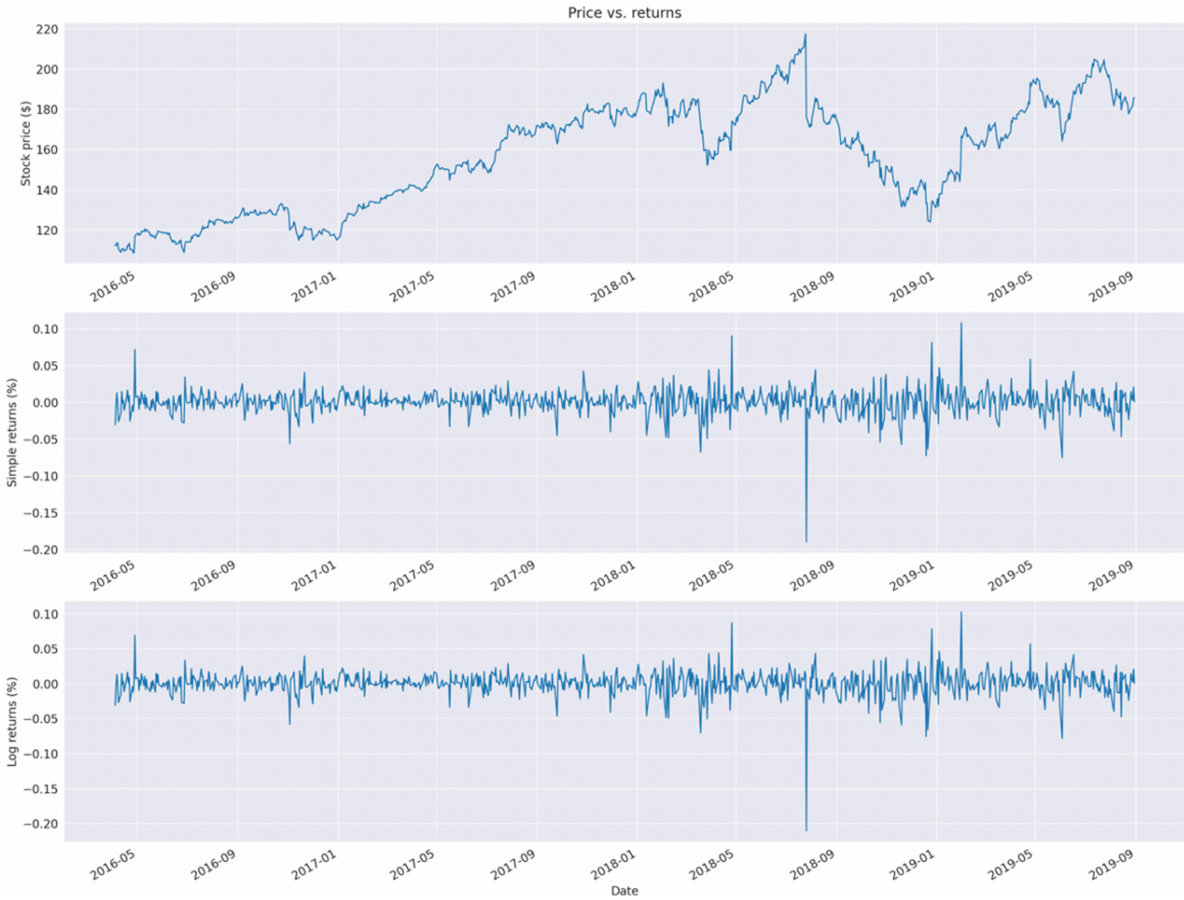
Introduction

Facebook was founded in 2004, and went public in 2012. The code for Trading on NASDAQ is 'FB'. Facebook's Market Cap is $781.80B on Dec. 17, 2020. Revenue for third quarter 2020 is $21.22B. Acquisition of Facebook includes Oculus, Instagram, Whatsapp & GIPHY etc.

Statistical Features on daily returns.

Statistical Features on (log) daily returns

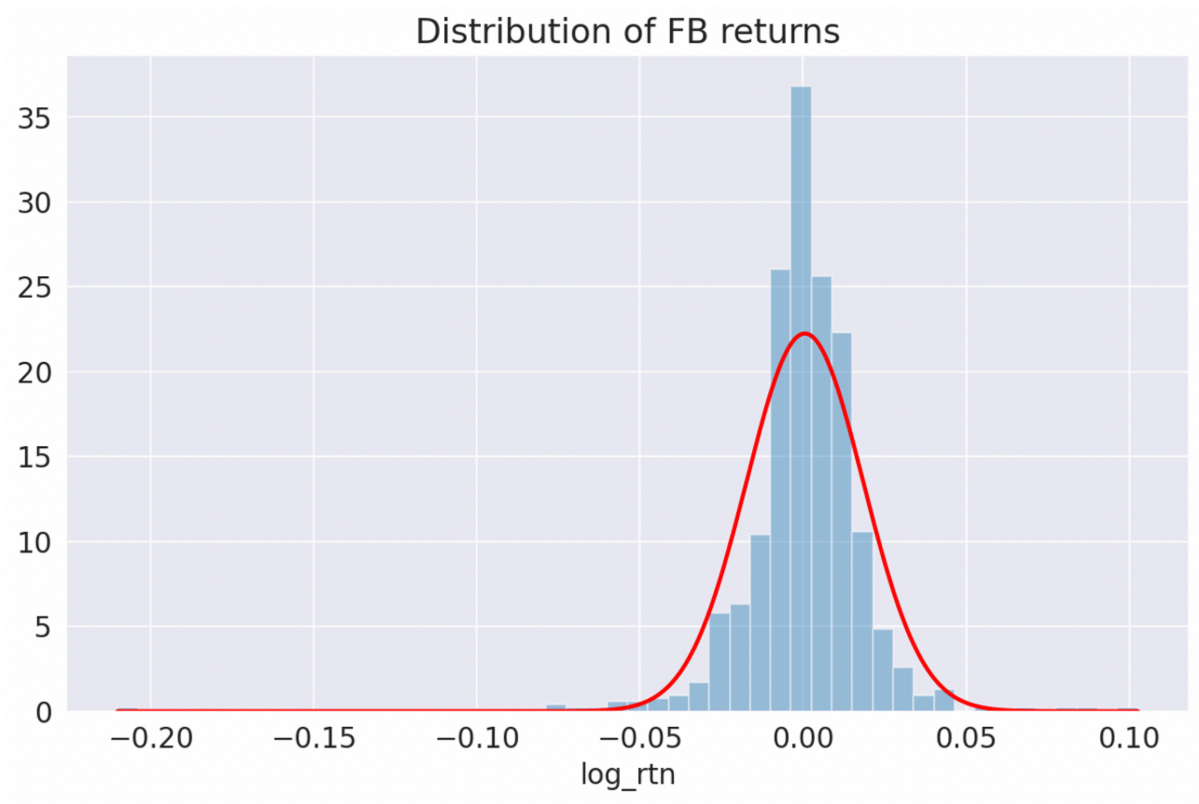
We choose to analyze daily return at the beginning, since it could make time series stationary, which is a desired property in statistical learning.



From the graphs, we could see the trends of Stock price VS. daily returns. The second picture is simple return, which is defined as $R_t = (P_t - P_{t-1}) / P_{t-1} = P_t / P_{t-1} - 1$, the third one is log return, which is defined as $r_t = log(P_t / P_{t-1}) = log(P_t) - log(P_{t-1})$. We prefer log returns, since it's more stable. We could see that there was a sudden drop on both stock price and daily returns, that's because the Facebook-Cambridge Analytica data scandal that happened on 2018, which was an incident that FaceBook users' data was acquired without individual consent by Cambridge Analytica, predominantly to be used for political advertising. That scandal resulted to 24% drop of Facebooks's stock, which was $134 million at that time.

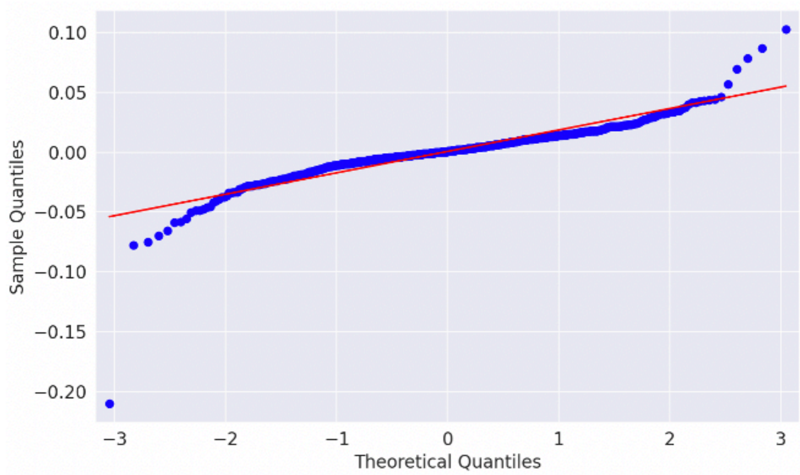
Statistical Features: Skewness and Kurtosis

When we were analyzing FB's daily returns, we found two interesting facts.



Based on the picture, we could see that the left tail is longer, and the mass of the distribution is concentrated on the right side of the distribution plot. That means large negative returns occur more often than large positives. The other fact is that the distribution has excess kurtosis, we could easily see the fat-tailed and peaked distribution from the picture, and that means large(and small) returns occur more often than expected.

Statistical Features: Normal Distribution

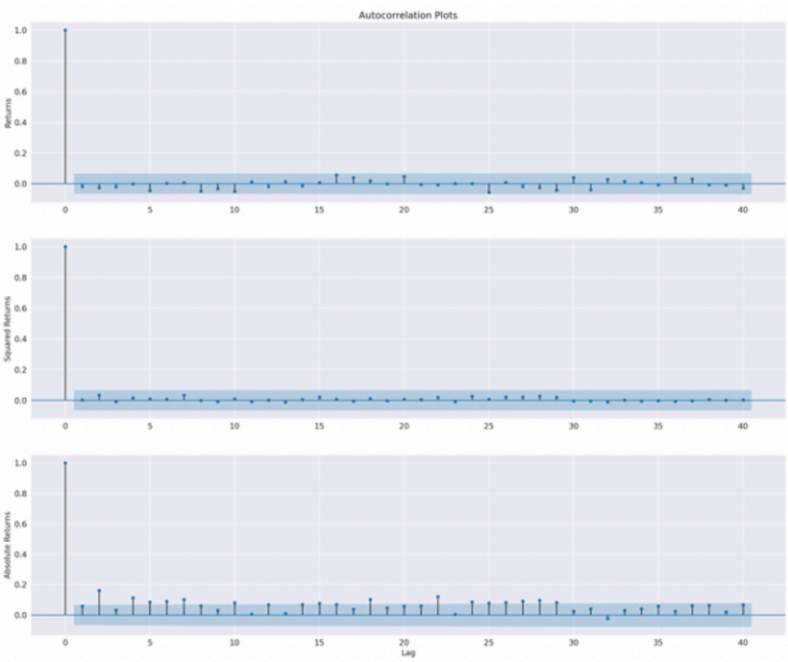


The red line represents the Standard Normal distribution, and the blue line represents returns. In the case when the returns follow the Gaussian distribution, those two lines would be aligned. However, we see that there are differences, mostly in tails. This further verifies that the returns are non-normal.

```
Range of dates: 2016-04-04 - 2019-08-30
Number of observations: 860
Mean: 0.0005
Median: 0.0006
Min: -0.2102
Max: 0.1027
Standard Deviation: 0.0179
Skewness: -1.8682
Kurtosis: 25.3077
Jarque-Bera statistic: 23170.83 with p-value: 0.00
```

Follow data are Jarque-Bera Normality Test, which confirms our suspicions, with p-value small enough to reject the null hypothesis stating that the data follow normal distribution.

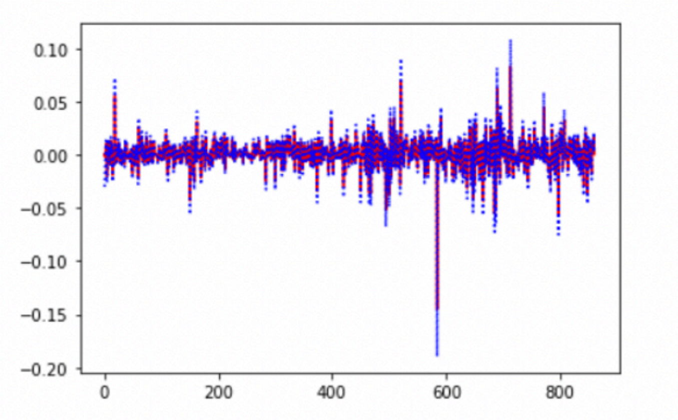
Statistical Feature: Autocorrelation



Above is the autocorrelation plot, the blue area indicates the 95% confidence interval, points outside of it are statistically significant. We see that for log returns, there are only a few significant points hence no significant autocorrelation. Small and slowly decreasing autocorrelation is easier to observe for squared returns than in case of absolute returns.

Kalman Filter

The Kalman filter is an optimal estimation algorithm that uses noisy observations of a system over time to estimate the parameter of system and predict future observations. The Kalman filter and state space models are strongly connected, and it combines information sources to give better estimates. The application of Kalman filter are Navigation and control systems(Spaceships, Robots, touchpads), Trading strategies and econometrics.




```
Optimization terminated successfully.
         Current function value: 4.371765
         Iterations: 2
         Function evaluations: 42
         Gradient evaluations: 7
```

RMSE: 0.00426517197744397

Above is our regression of daily returns using Kalman Filer, which fits really well with 0.004265 RMSE.

Generalized Autoregressive Conditional Heteroskedasticity(Garch)

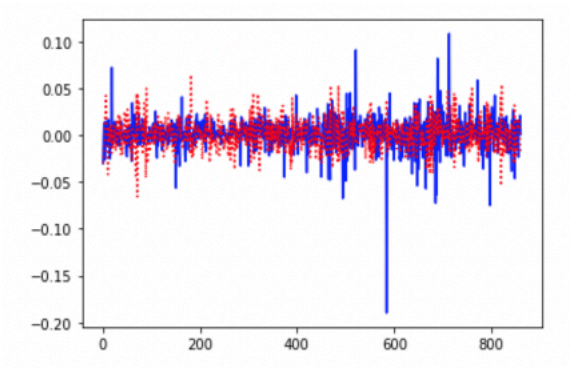
Garch could deal with burst of Volatility, which in our case, is the drop of stock prices due to 2018 scandal. The basic idea of Garch is that the Time Series today is based on random error, time series yesterday and the volatility yesterday.

$$r_t = \mu + \sigma_t \varepsilon_t$$

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i (r_{t-i} - \mu)^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2$$

Parameters: [np.mean(Y),0.0001, 0.2, 0.5]

RMSE: 0.0243994528256253



We could see that Garch gives less accurate results than Kalman filter, which has 0.02439 RMSE.

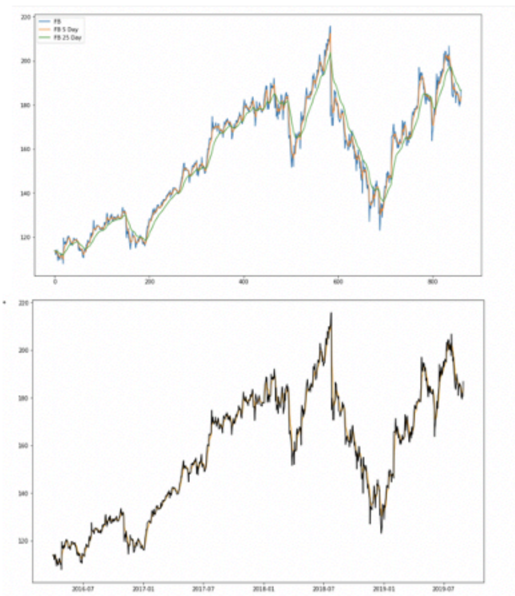
Supervised learning Models for Open Pricing

Exponential Moving Average(EMA)

Moving averages that are calculated by assigning greater weights to more recent prices and proportionately lower weights to prices farther in the past and then averaging those together.

EMA using **pandas.df.ewm**

Relation between EMA and Open values



We could see that the EMA prediction results fit really well with real open values, which gives RMSE 5.05.

Fama French 3 Factor Model

The Fama French 3-Factor model is an asset pricing model that expands on the capital asset pricing model by adding size risk and value risk factors to the market risk factors.

The model has three factors: size of firms, book-to-market values and excess return on the market. For this model, we used a combination of ADS Index and Fama/French 3 Factors[Daily] with FB data.

Below is the regression graph of FF3.

$$R_{it} - R_{ft} = \alpha_{it} + \beta_1(R_{Mt} - R_{ft}) + \beta_2 SMB_t + \beta_3 HML_t + \epsilon_{it}$$

Support Vector Machine(SVM)

Support Vector Machine is a machine learning technique used in recent studies to forecast stock prices. Support vector Machines are one of the best binary classifiers. They create a decision boundary such that most points in one category fall on one side of the boundary while most point in the other category fall on the other side of the boundary.

Radial Basis Function (RBF): $G(x_j, x_k) = \exp(-\frac{1}{2\sigma^2}||x_j - x_k||^2)$

Moving Average Convergence Divergence(MACD)

MADC line is one of the most important features in finance. It has two key elements, which are MACD line and signal line. The formula for MACD is 12 period EMA – 26 period EMA. Signal line is nine-day EMA of the MACD. It is time to sell the stock when MACD falls below the signal line, and it is time to buy the stock when MACD rises above the signal line.

Random Forest

Random forest consists a large number of individual decision tress that operate as an ensemble. A large number of relatively uncorrelated models(trees) operating as a committee will outperform any of the individual constituent models.

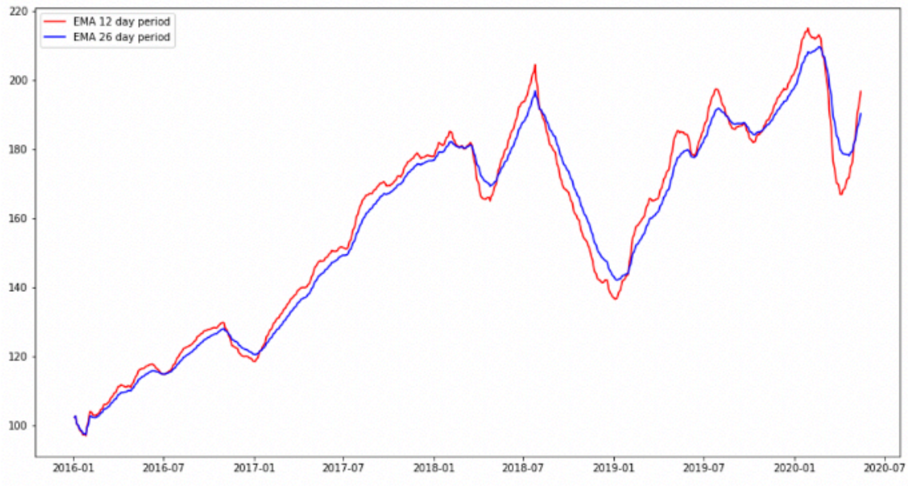
AR_RMSE: 1.9076290240331346
EMA_RMSE: 5.050240539306411
MACD_RMSE: 5.1444098494264665
SVM_RMSE: 7.280697329212049

Random forest RMSE:
1.6211889048736239

We could easily compare the accuracy between individual model and Random Forest. The Random Forest gives the best prediction accuracy with 1.6212 RMSE.

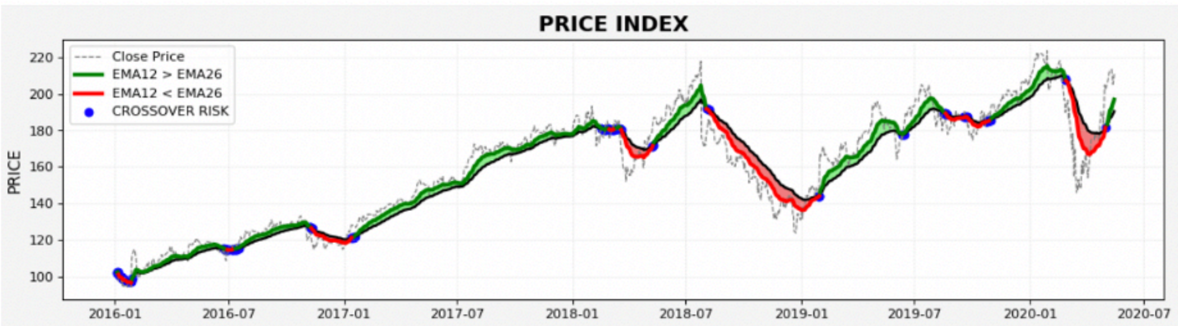
Trading Strategies

EMA and MACD Signal Crossovers.



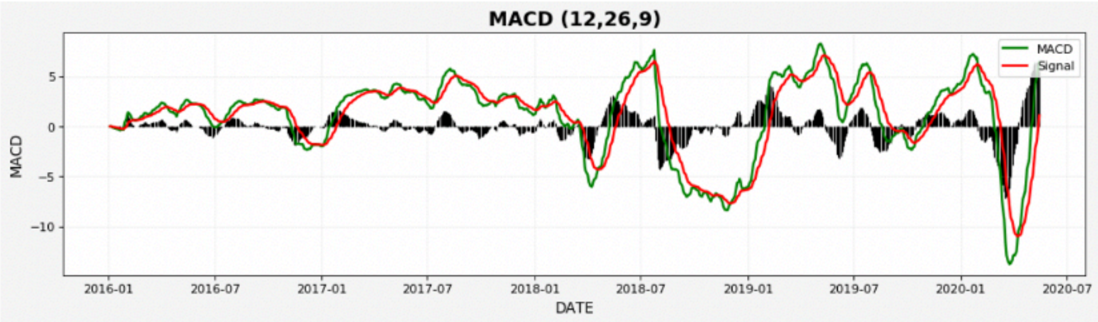
Above is the graph of EMA 12 day period and EMA 26 day period, so we could observe the MACD line from its by formula 12 period EMA – 26 period EMA.

Trading Strategies-Crossover



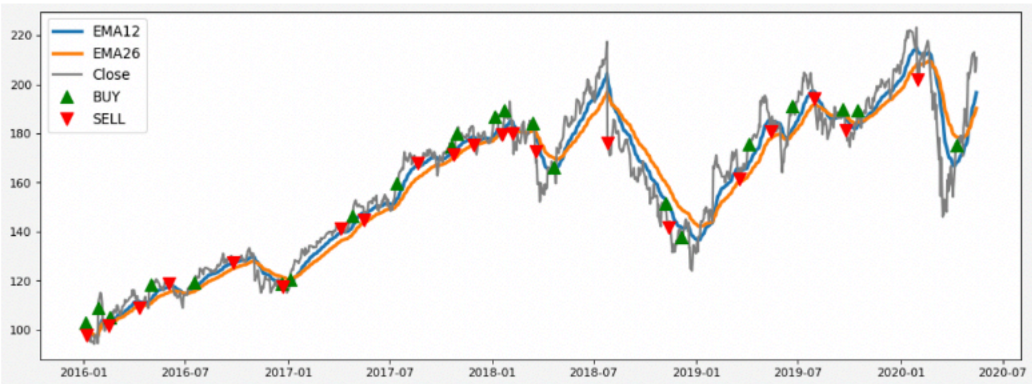
We could observe when EMA12>EMA26 and EMA12<EMA26, and we could get the crossover risk from that graph.

Buy Sell & Hold



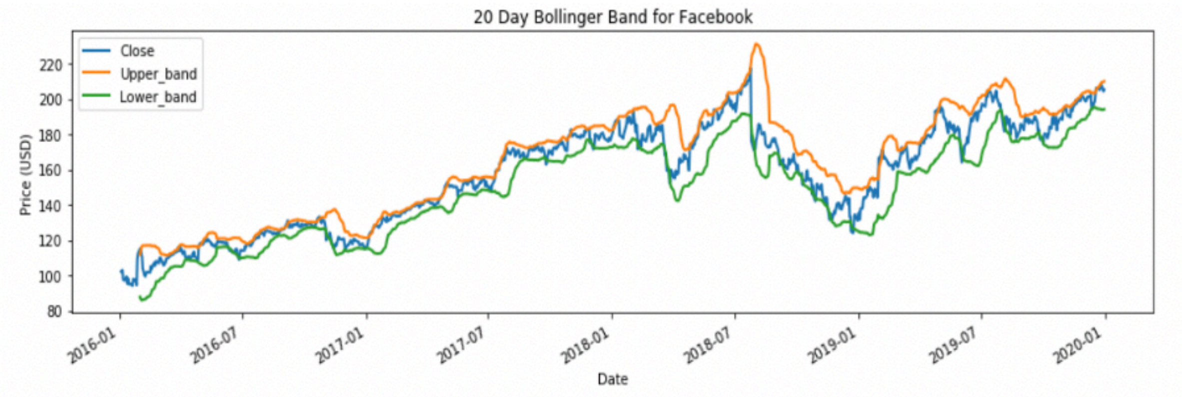
We could observe the MACD line and the Signal line from the graph, which could let investor know when MAC line lies beyond the signal line, and when falls behind the signal line.

Buy and Sell



From that graph, we could see the buy and sell signal, so the investor could have a brief idea on when to buy/sell the stock. And our strategy will based on that signal, the results is in the next part.

Bollinger Band



Bollinger band is also an important feature. The indicator is made up of 3 lines – a middle band and two outer ones. The middle band is moving average, usually with the period of 20. Usually the outer bands are set 2 standard deviations above and below the middle band. The use of Bollinger Bands varies widely among traders. Some traders buy when price touches the lower Bollinger Band and exit when price touches the moving average in the center of the bands. Other traders buy when price breaks above the upper Bollinger Band or sell when price falls below the lower Bollinger Band. Moreover, the use of Bollinger Bands is not confined to stock traders; options traders, most notably implied volatility traders, often sell options when Bollinger Bands are historically far apart or buy options when the Bollinger Bands are historically close together, in both instances, expecting volatility to revert towards the average historical volatility level for the stock.

When the bands lie close together, a period of low volatility is indicated. Conversely, as the bands expand, an increase in price action/market volatility is indicated. When the bands have only a slight slope and track approximately parallel for an extended time, the price will generally be found to oscillate between the bands as though in a channel.

Traders are often inclined to use Bollinger Bands with other indicators to confirm price action. In particular, the use of oscillator-like Bollinger Bands will often be coupled with a non-oscillator indicator-like chart patterns or a trendline. If these indicators confirm the recommendation of the Bollinger Bands, the trader will have greater conviction that the bands are predicting correct price action in relation to market volatility.

Day Trade Result using Random Forest

Metrics	Values
Amount	10000
Shares	300
Number of days	607
Total profit	217595.97702026367
Total loss	107078.98864746094
Profit%	21.75959770202637
Net profit	110516.98837280273
Profit factor	2.0321071366919687
Profit days	375
Loss days	231
Winning rate	0.6177924217462932
Average Net Profit Per Trade	182.0708210425086
Average Daily Return	449.8710129261017
Daily Return STD	312.5259972445435

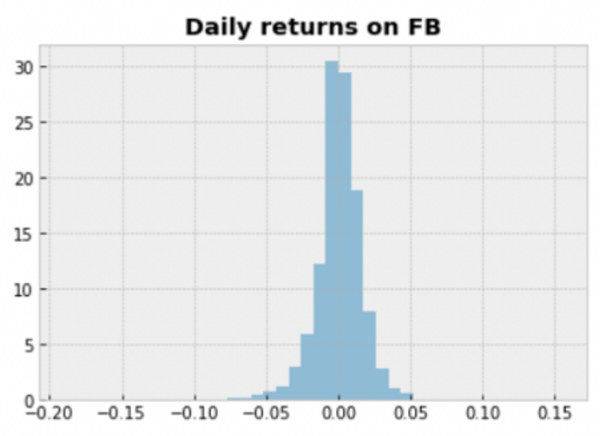
We could see that day trade strategy by random forest gives a really well result. We purchase 300 shares of stock by $10000, the investment period is 607 days, we buy and sell in each one of 607 days, then we could get net $110516 net profit with profit% 21.7595.

Long Short using Random Forest

Metrics	Values
Amount	10000
Shares	300
Number of days	607
Total profit	111552.00576782227
Total loss	53141.97692871094
Profit%	11.155200576782226
Net profit	58410.02883911133
Profit factor	2.099131651001004
Profit days	181
Loss days	181
Winning rate	0.2981878088962109
Average Net Profit Per Trade	96.22739512209445
Average Daily Return	222.61243791329233
Daily Return STD	166.5753606066162

We could see that long short trading strategy performs not as well as day trade. We could only get 58410 net profit, with profit %11.155.

Value at Risk using Historical Bootstrap



We calculate the value at risk to see the investment risks. From the graph, we could see that the 0.05 empirical quantile of daily returns is at -0.025. That means for 95% confidence, our daily loss will not exceed 2.5%.If we have a $1M investment, our one-day 5% VaR is 0.025*$1M = $25k.

Conclusion

In a nut shell, the analysis of daily returns could let us know that Facebook's daily return is not so stable, and the large(small) returns occur more often than we expected. Since the Facebook-Cambridge Analytica scandal, it is hard to predict the sudden drop of the stock prices in 2018. We tried Garch model, which could do the prediction based on the burst of volatility. However, the model performs not so well on daily returns. For the model prediction part, we try several models includes EMA, Fama French 3 Factors, SVM bootstrap and autocorrelation model, and each of them gives relative accurate results. However, the random forest combines each model and give the most accurate result with only 1.6211 model RMSE. If a new investor follows our strategy, he/she could possibly get $110516 net profit on day trade net profit and $58410 on long short net profit with $10000 investment money and 300 shares.

References

https://jingyan.baidu.com/article/1974b289560c8af4b1f77430.html

https://blog.csdn.net/a934079371/article/details/105479990

https://blog.csdn.net/weixin_34326179/article/details/89580456

https://www.kaggle.com/utkukubilay/time-series-analysis-in-python

https://www.kaggle.com/hasitparmar/stock-price-using-linear-regression-and-lstm-98

https://www.kaggle.com/roshanpk/notebook40b36ebaf8

https://www.kaggle.com/honeysingh/intro-to-recurrent-neural-networks-lstm-gru

https://www.kaggle.com/subbhashit/stock-time-series-using-gru

https://www.kaggle.com/pbizil/random-forest-regression-for-time-series-predict

License