

MICROSOFT STOCK PRICE FORECASTING

Predictive Modeling based on Contextual Factors

Mar. 2019, by Yuzi Liu

OVERVIEW

- The project objective is to forecast the stock price of MSFT at the end of 2019 based on contextual factors
- Conducted Feature engineering to determine 10+ variables, created cycle replication
- Performed data cleaning and data integrated with ETL tools; Prepared data for modeling via Alteryx
- developed an ensemble forecast model based on 7 different time series models (Random Forest, Stepwise Regression, etc.)
- Built What-if interactive dashboard for MSFT stock monthly price on Tableau to evaluate different scenarios

Techniques and Tools used: Alteryx, Tableau



ABOUT MSFT

Founded in 1975, Microsoft Corporation is the leading provider of software, services and solutions to businesses and consumers worldwide.

- Time of IPO: 3/13/1986
- Time became the DOW 30 component: November 1999



EXPLORE FACTORS

There are external factors and internal factors which will influence MSFT stock price. **Internal factors** in question are a factor that comes from within the company in the form of company performance as well as the fundamentals of the company. **External factors** such as fluctuations in inflation rates, interest rates and the exchange rate that determines the transaction in the capital market so it is important for investors to consider matters that could affect the stock price as it will ultimately have an impact on yield that will be obtained by the investor.

I have determined the following variables for MSFT price prediction:

- Internal: Revenue of Microsoft and its competitors (Apple, IBM, etc.)
- External: S&P 500 index, NASDAQ Index, DOW 30 Index, Real GDP growth rate, Interest Rate (Federal Funds Rate), Consumer Sentiment Index

In addition to that, I create two independent variables to capture both the **Market Cycle** and the **Economic (Business) Cycle**.

MODEL BUILDING PROCESS

After conducting some research, I chose to use two datasets; one is from 1/31/2001 to 2/28/2019; the other is from 3/31/1986 to 2/28/2019. There are reasons for using these time ranges:

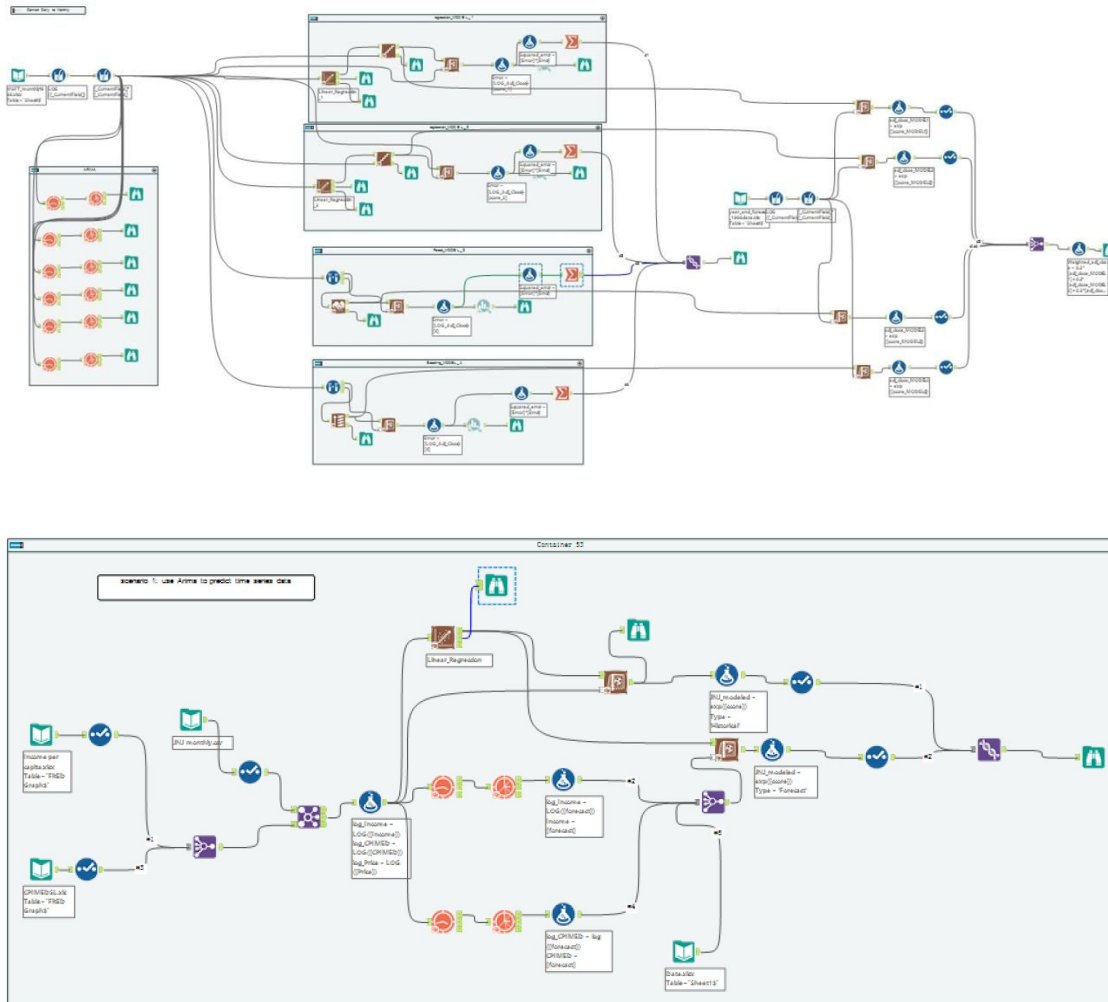
- a. The market structure has undertaken significant changes since Microsoft's IPO on 3/13/1986.
- b. Microsoft has become a component of DOW 30 since November 1999.

Noticing that the company announces last quarter's revenue in the first month of the new quarter, I move the data back one quarter to match the information. As to the real GDP growth rate, it is published at the last month of each quarter, so it's appropriate to use previously published rate for the first two months, newly published rate for the last month (of each quarter). By using the interpolation method, I was able to convert the quarterly data into monthly.

Knowing that the stock market tops out before business cycle peaks, and bottoms out before cycle troughs, I create two independent variables to capture both the market cycle and the economic(business) cycle. I replicated Market cycle by looking through the change of S&P 500 historical graph, assigning "Bull" to 1, "Bear" to -1, and the smooth change to 0. The economic cycle is replicated by assigning 1 to expansion (from through to peak), - 1 to contraction (from peak to through). I got the information from the National Bureau of Economic Research.

I first did some association analysis to see the correlation among variables. The models I used including **Linear Regression with Stepwise selection, Boosted regression, Random forest, ARIMA time series**. Among all these models, linear regression has the most normal distribution shape of errors; Random Forest has the lowest MSE (Mean Squared Error). All the models are robust while explaining more than 97% variation of Log adjusted close price. (Linear regression: Adjusted R-squared 97.7% with MAPE=1.8%, MSE=0.006)

ALTERYX WORKFLOW



Dependent variable:

- MSFT stock price (monthly)

Independent variables with some variable transformation (Log, Quadratic):

- Microsoft Revenue (quarterly)
- Apple Revenue (quarterly)
- IBM Revenue (quarterly)
- Real GDP per capita (quarterly)
- S&P 500 Index (monthly)
- NASDAQ Index (monthly)

- Consumer Sentiment Index (monthly)
- Market Cycle (binary)
- Economic Cycle (binary)

Results and Insight:

In order to predict the stock price for December 31, 2019, I need to predict all the independent variables as well. For some of the indicator independent variables, I chose to use the prediction from the SMEs (Subject matter expertise) --GDP growth rate and interest rate prediction from Wall Street Journal Economic Survey, S&P 500 Year End prediction from the CNBC Market Strategist Survey. I used year-end Nasdaq and DOW30 index forecast from the Financial Forecast Center (recommended by the American Economic Association). As for the revenues and cycles, I used ARIMA time series models for prediction.

I end up getting seven models from 2 Alteryx workflows. I first weighted my 2001 models (2 linear regression, 1 forest) and 1986 models (2 linear regression, 1 forest, 1 boosted regression) based on their performance on MSE. Then I give weight to the weighted results of 2001 and 1986 dataset on an 80%:20% ratio also based on overall MSE performance. The final results expect the 12/31/2019 stock price for Microsoft to be \$121.69, a 20.14% increase compared to \$101.29 at 12/31/2018.

The individual forecasts are as follows:

2001 dataset:

Record #	Date	adj_close_MODEL1	adj_close_MODEL2	adj_close_MODEL3	Weighted_adj_close
1	2019-12-31	135.430094	148.128434	103.822002	126.596359

Record #	Mean_Squared_error
1	0.005707
2	0.005787
3	0.004923

1986 dataset:

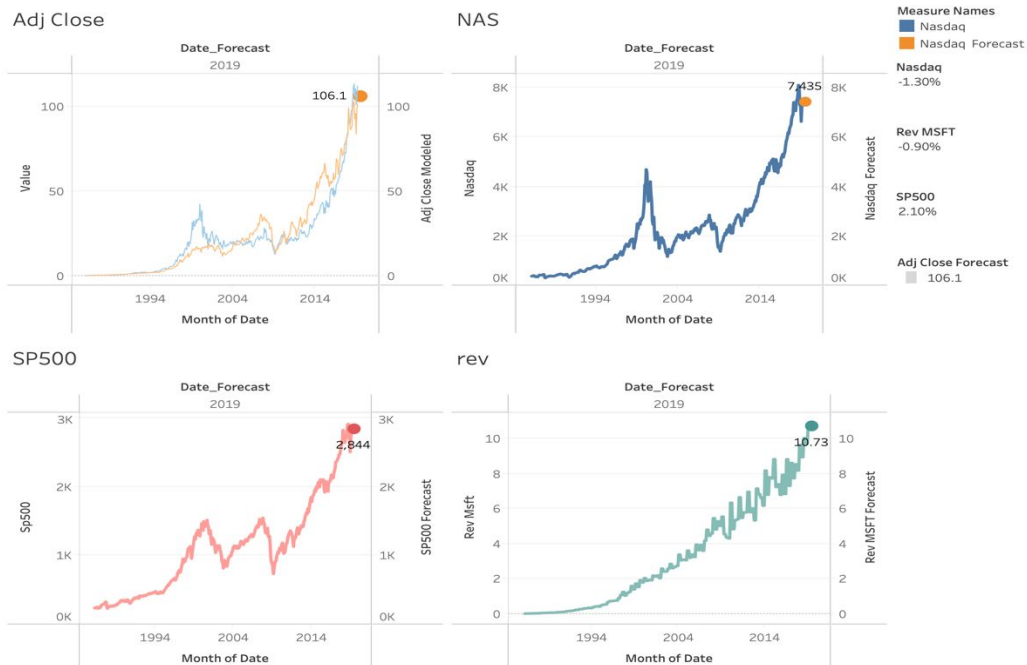
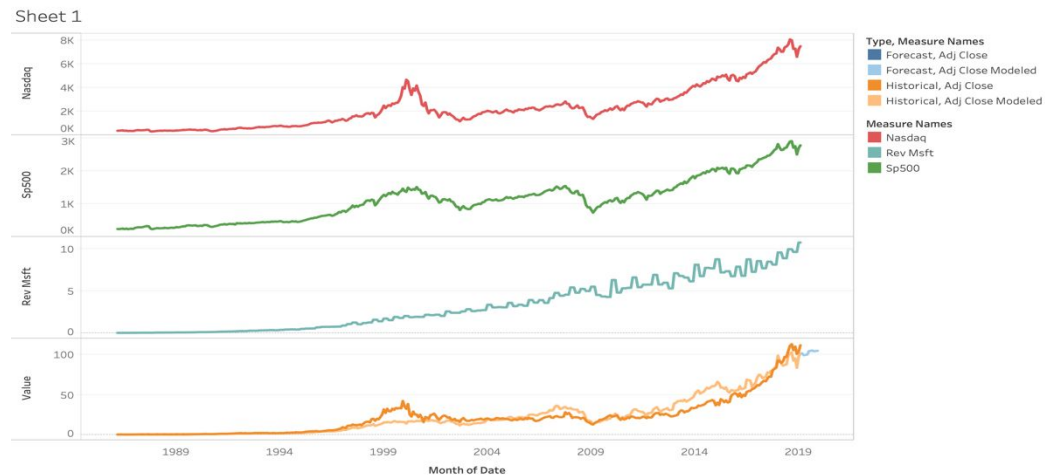
Record #	Date	adj_close_MODEL1	adj_close_MODEL2	adj_close_MODEL3	adj_close_MODEL4	Weighted_adj_close
1	2019-12-31	114.573473	108.887599	108.518775	82.745512	102.071501

Record #	Mean_Squared_error
1	0.02248
2	0.023255
3	0.014202
4	0.007905

TABLEAU VISUALIZATION

After modeling, I planted what-if model result into Tableau, then we can build a What-if Dashboard, on which we can change our predictors and forecast date to simulate the price change.

Below are sample Tableau dashboards for some of the variables:



The Alteryx Workflow can be found [here](#).