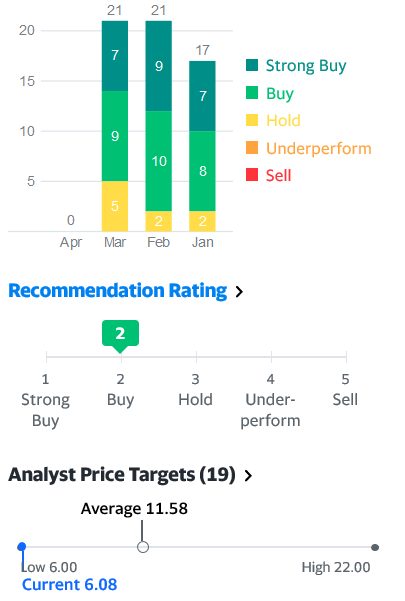
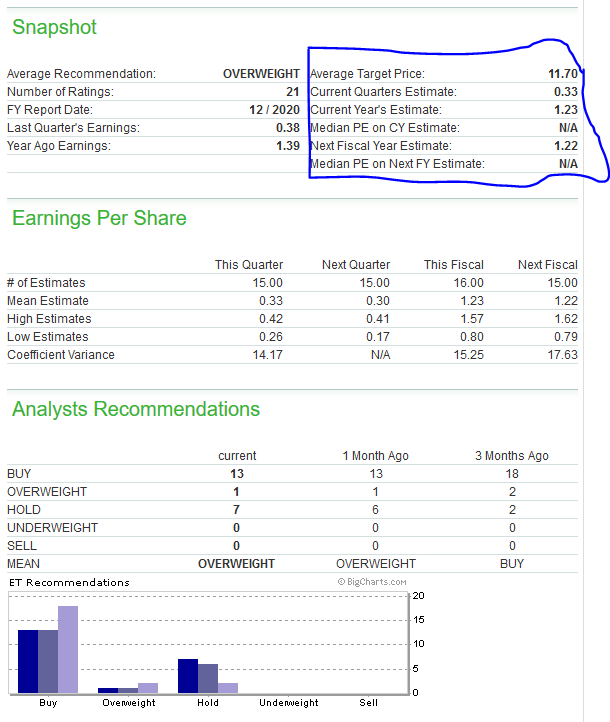
**Project: Stock price forecasting using LSTM/ARIMA and investment decesion**

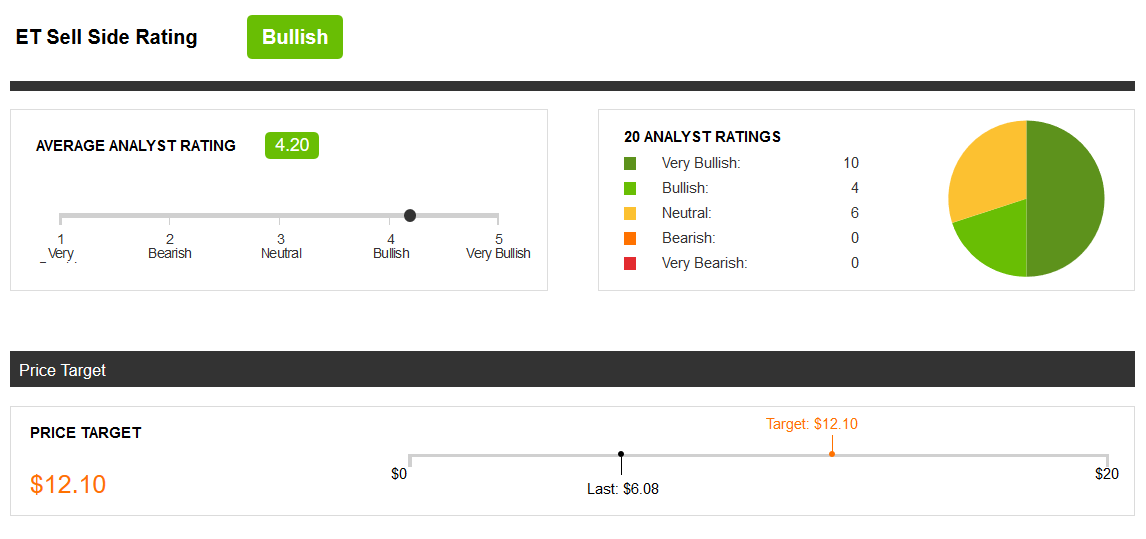
**Problem:** I have been investing in the stock market for a few months starting from last August. The recent fluctuations in the market have made new investors like me vulnerable. I have invested using the forecasts and 1-year estimates collected from CNN, Yahoo, Market Watch, seeking alpha, NBC websites. For example, see the figure below for the stock price of energy transfer taken from CNN finance.



The figure shows the last 12 months of the stock price for this company (current price is $6.07) and the one-year median forecast of $12. The high and low represents the confidence interval. This forecast is made by stock market analysts.  Based on this forecast, analysts think that there would be a 97.4% increase in this stock price in the next 12 months. Similar forecasts are available on other websites as well.  The estimates change in every quarter due to earnings reports or due to any special situations like the current coronavirus fluctuations or any internal company issues.

Similar stock prediction from other sources are also available. For example, see the yahoo estimate (left), market watch estimate (right), seeking alpha (below).



So, the CNN estimate is $12, yahoo estimate is $11.58 and the Marketwatch estimate is $11.70 and seeking alphas is $12.10. The average of estimates in 11.84. The current price of the stock is $6.1. This indicates that if I invest in this stock there is a strong chance of making a 94% return on investment in one year!

The problem is:

1. The underlying rationale for these predictions is not known.
2. There are thousands of stocks available. It’s not possible to manually look for potential stocks with high yield estimates.

For example, I used the above intuitions to invest in two oil stocks. And currently at 40% down on my investments. Why? Because my profile is not diversified enough. For example, due to coronavirus effect, retail and oil were hurt severely compared to tech stocks.

In this project, I aim to develop a simple LSTM based model for predicting the upward/downward trend for a stock using historical stock price data and predict future stock price. I would like to identify one or two best stocks per category (retail, tech, oil, consumer, financial, etc.) from the forecast. Then I would tabulate the results and compare it with the analyst estimates from various other sources. The prediction can also be validated using analyst estimates.

Finally, the investment decision can be made based on a combination of the return of investment and the diversity of the portfolio. This approach is similar to the ensemble technique used in a random forest.

User’s/client: Financial institutes, stock investors. But the primary user would be new/small investors like me who want to make a data-driven decision using a simple approach rather than going through the numbers and balance sheets used by the financial analysts which can be daunting.

**Dataset**: I used yahoo stock API to scrap stock price for the stocks.

- Used pandas datareader API to scrap stock prices data from ‘yahoo’.  
- Data was then manipulated to create the following DataFrame where:

- Rows represent time series of date.

- Columns represent a hierarchical table of stock tickers and prices info.

I scrapped the stocks with following tickers:

MSFT: Microsoft

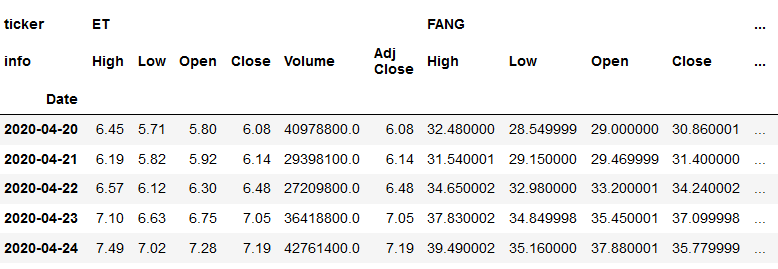
ET: Energy transfers

SPG: Simon properties group

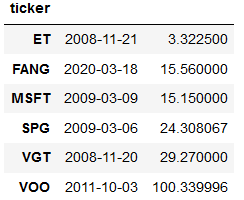
FANG: Diamondback energy

VGT: Vanguard technology ETF

VOO: Vanguard S&P500 ETF



**Data analysis:**

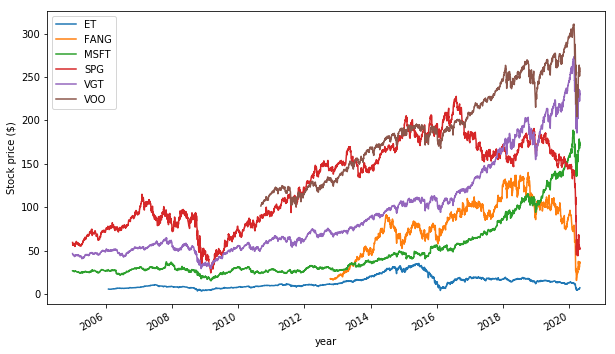
First, lets see how the stocks are performing currently compared to their all-time high and lows.

All time low

|  |  |  |
| --- | --- | --- |
|  |  |  |
| Current price | All time low price | All the high price |

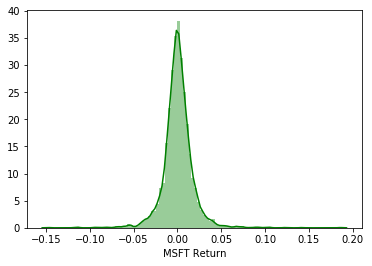
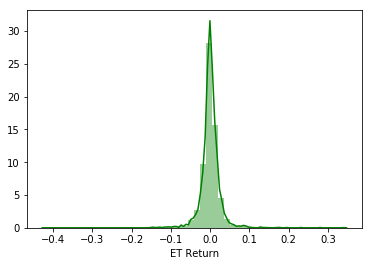
In the table above, we show the current, all the high and all-time low-price side by side for the stocks of our analysis. SPG, FANG, ET are close to their all-time low. Again, MSFT,VOO, VGT are very close to their all-time high.

This can be further visualized in the chart below where I have plotted the time series of the stock prices.

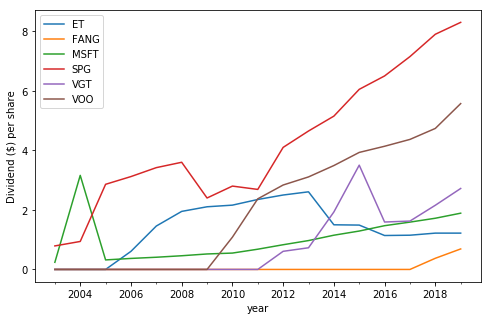


It can be seen that Microsoft, VGT and VOO prices grew considerably in the last 15 years. FANG, and ET stocks declined considerably recently especially due to lack of oil demand.

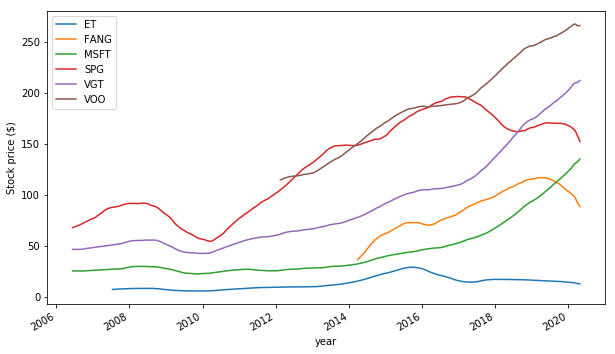
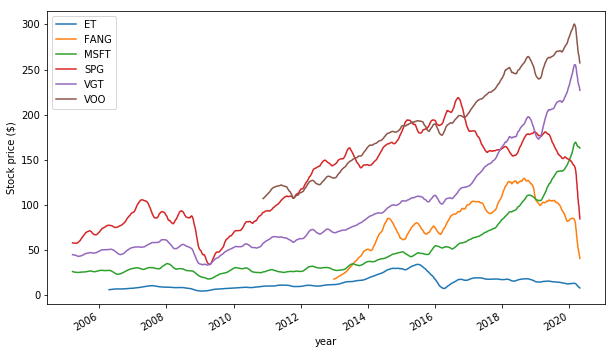
Next, I checked the distribution plot of the standard deviation of the return (calculated using the percentage change) indicates that ET stock is more fluctuations in prices compared to MSFT.



Next, I check the yearly paid dividends by those stocks. SPG, MSFT, VOO are very stable dividend payers. While ET pays dividend at a decent ratio, but the dividend is not growing.



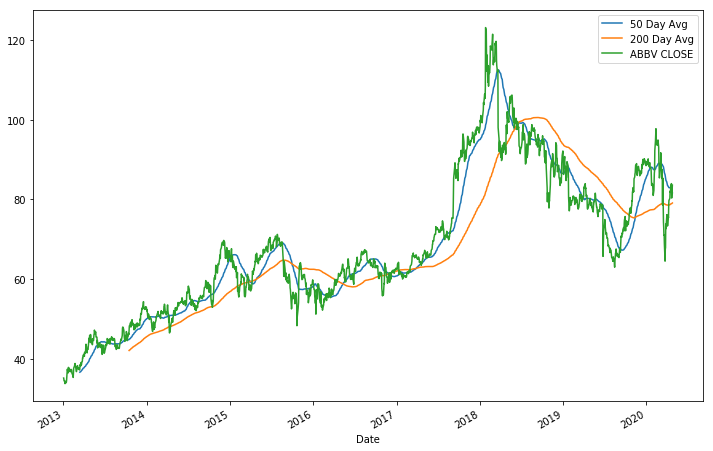
Next, we plot the stock price using the 50 day and 365 day rolling average. Time series of the 200/365 day rolling average gives a high-level idea about the stock type. For example, MSFT is a still growing (growth stock). Energy transfer pays dividends to the shareholders. It’s a dividend stock.



365 day rolling average

50 day rolling average

Next, we check the golden cross behavior i.e., When 50day moving average cuts the 200day moving average, it is called the golden cross. 50day moving average cuts the 200day moving average and the slop is positive, it’s a bullish signal (buy). 50day moving average cuts the 200day moving average and the slop is positive, it’s a bearish signal (sell). The figure above for the ticker ‘ABBV’ (a pharmaceutical stock) validates the above-mentioned fact



Next, we check the Correlation between stock price. Correlation between stocks can be used to identify similar stocks. For example, VGT (technology index and VOO (S&P 500 index are heavily correlated)). To enhance diversity of the portfolio, avoid investing all of your investments in the same bucket

