Team 017 Report

Introduction/Motivation

There is currently no mainstream website that provides visualizations of both fundamental/technical indicators on the same page, along with stock recommendations. Our aim for this project is to create a visualization where both fundamental data and technical indicators for a company's stock price are present. This provides better guidance for investors towards making the most lucrative financial choices. Financial professionals who are looking for a more convenient way to view both technical and fundamental indicators will be interested in this approach along with anyone looking for more informed investments.

Problem Definition

Fundamental data includes information about the underlying company and its financial situation: revenue, earnings per share, net-income, cash-flow, etc. It is used to measure a stock's intrinsic value to account for macroeconomic factors and industry trends in stock prices. Technical indicators (Credit Suisse, 2010; Bruni, 2016; Tanaka-Yamawaki, 2007; Farias Nazário, 2017; Ugur, 2014) are pattern-based signals driven by price of the stock, trading volume and crowd sentiment. These indicators allow assessment for entry and exit points in a trade based on historical data. The current literature available does not provide an insight into how technical indicators can be blended with fundamental data to make a buy prediction for a stock. Using both fundamental data and technical indicators, we attempt to determine if a particular stock will have positive short-term returns. Additionally, we could provide a list of top 10 stocks that would be advantageous to buy at their current prices.

Survey

Previous methods of stock price prediction/selection including SVMs (Stanković, 2015; Shynkevich et al., 2017; Macchiarulo, 2018), ARIMA (Adebiyi, 2015), Neural Networks (Vaiz and Ramaswamy, 2016, Shynkevich et al., 2017), Fuzzy Logic (Rodríguez-Cándido, 2021), Clustering analysis (Guo, 2020), Random Forest, Naive-Bayes (Patel, 2015), k-Nearest Neighbors (Shynkevich et al., 2017),and decision tree classification models (Al-Radaideh, 2013) have been used to predict stock direction and determine a buy/sell signal. An effective way to combine technical indicators like EMA, RSI, KDJ lines proposed by Dai, 2020 has shown to produce statistically and economically significant prediction performance.

Limitations:

Previous studies have either failed to compare different ML techniques efficiently or have not approached the problem in the most efficient way. A major discrepancy in most of these studies is variability of input window length and forecast horizon.

Additional Literature Surveyed**:**

Chong, 2008, has shown that following the RSI rule and the MACD rule outperform the buy-and-hold strategy, however, no improvement on either RSI or MACD is discussed. Cai, 2019 provides a background of ways to adjust and display stock price data. It provides a simplistic overview of some basic visualization concepts in the context of stock charting. It does not contribute to solving the fundamental problem of picking stocks to buy or sell. Friesen, 2009 provides some justification for technical indicators and explains how many technical indicators emerge in a trader's mind as evidence of confirmation bias. It identifies some relevant statistical patterns but does not generate a sound model for stock picking. Chen, 2017 is a useful resource on fundamental analysis on stock market forecasting that proposes a method for weight calculation of indicators to evaluate and select stocks. The complexity of their method puts a full implementation out of scope.

Proposed Method

<u>Overview</u>

Our financial data is sourced from the Yahoo Finance API. This data is stored in an SQLite database. We then calculate technical indicators in python utilizing pandas data-frames, which is stored with the fundamental data in the same server. Subsequently, we have used an algorithm (logistic regression) to combine the technical and fundamental signals into a probability score that is used to make a buy/sell recommendation. The database data is requested in JSON format (a REST API) via a small Flask app running on the backend. The JSON data we need is directly requested and used by D3 to provide a user-friendly interactive visualization of fundamental and technical data. Details of this visualization are in the "User Interface Layout" section below.

<u>Data</u>

Yahoo Finance provides accurate, and easily accessible real-time financial data. We have gathered previous ten years' data pertaining to the following metrics for all the companies making up the S&P 500 index. Our database size is 260MB and has approximately 1,250,000 rows.

<u>Fundamentals</u>: Share Price, Revenue, Earnings, Research and Development Costs, Income Before Tax, Price –to-Earnings ratio, Price-to-Revenue ratio, Debt-to-Equity ratio

<u>Technical Indicators</u>: These are calculated using the price data from the Yahoo Finance API

- MACD: Moving Average Convergence Divergence
  - The difference between the 26-period exponential moving average of a stock and the 12-period exponential moving average.
  - It serves as a buy or sell signal for a stock.
  - 'MACD', 'Diff_MACD' and the 'Signal' line in the visualization are a part of this indicator.
- RSI: Relative Strength Index
  - Measures the magnitude of recent changes to a company's share price which indicate whether the stock is overbought or oversold.
  - Typically, RSI values that are 70 or above indicate a stock is overbought and maybe ready for a pullback. Conversely, RSI values of 30 or below indicates a stock is oversold and maybe a good time to buy.
- EMA: Exponential Moving Average
  - A type of moving average which gives heavier weights to more recent data points.
  - Similar to other moving averages, this technical indicator is used to produce buy or sell signals based on crossovers and divergences from historical averages.
  - We have used a 9-Day EMA in our study since it is a commonly used indicator.

Approaches

Intuition/Innovation:

- The outcome of the project provides a simple and intuitive visualization that allows users to make a confident decision on buying and selling a company's stock. The visualization is made to be interactive so that the user can select the stock ticker of interest and the visualization then populates both the fundamental and technical data for the selected stock ticker. Additionally, based on the machine learning algorithm that runs in the backend, the visualization also suggests buy/sell signal for the top 10 stock tickers with highest possible returns. Our product provides a new visualization for investors that currently is not available.
- Logistic Regression has never been implemented on this combination of technical and fundamental indicators. We had initially hypothesized that this approach should provide a better indication of when one should buy, hold, or sell a stock. The hypothesis is accepted based on the results after the completion of the project.
- We also ran some simple K Nearest Neighbors and Linear SVM models to see if any other classifiers might perform better. Neither of these options were substantially better at identifying stock buying opportunities. The SVM and Logistic Regression were about 4% more accurate than the knn model. Given the interpretable output of a probability from the Logistic Regression, we chose that model.

Technical Indicators:

Using the technical indicators described above, we have obtained average return levels for each.

Logistic Regression:

- This model gives us a recommendation to buy a stock which we then used to back test.
- Factors:
  - Fundamentals
    - P/E Ratio: The ratio of a company's current share price to its earnings per share (EPS). This is used to compare companies against its historical performance and against other companies in a balanced manner.
    - P/R Ratio: A company's current share price divided by a company's revenue. This indicates whether a certain stock is overvalued or undervalued.
    - Debt/Equity Ratio: The ratio of a company's total liabilities to its equity. This indicates the debt load a company has.
  - Technical
    - MACD
    - RSI
    - Exponential MA

We have performed variable selection and tested the significance of these factors in order to see if they meaningfully help prediction or not and come to a final best model with selected factors.

User Interface Layout

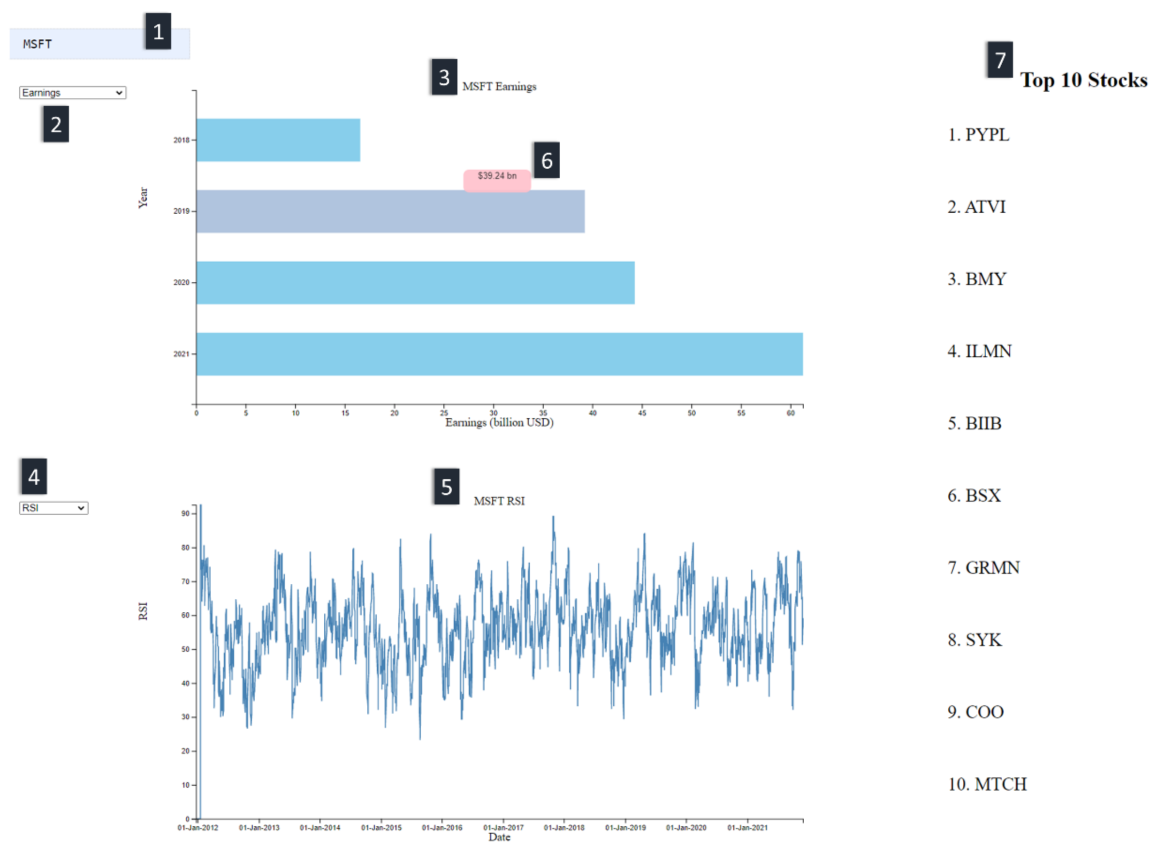

Figure 1 User Interface

- The layout is composed of the following elements:
    1. A search bar to allow users to enter their preferred stock.
    2. A dropdown menu to select the fundamental data to visualize.
    3. A bar chart showing fundamental data for a user selected stock
    4. A dropdown menu to select the technical indicator to visualize.
    5. A line chart showing technical indicators for the same user selected stock.
    6. Tooltip showing the value the bars represent for the bar chart.
    7. A section dedicated to the top 10 stocks to buy according to our logistic regression model.

Experiments and Evaluation

We have measured the prediction accuracy for the average returns of our proposed methods by backtesting those methods on a sample of 500 stocks and a 30-day time period. One key benchmark, that our methods should have preferably achieved, was to perform better than the S&P500 index. This is achieved by comparing the returns given by our logistic regression model against S&P500 for the comparable period. If the methods do not consistently beat this index, then we conclude that they would not be fit as a recommended trading strategy. The experiment answers the following questions:

- Is our predicted logistic regression model good enough to beat S&P500, over a 10-day holding period?
- Under what cases, will the model underperform and why?

Details of experiments/ observations

For our modeling parameters, we used technical indicators RSI and MACD along with fundamental data Debt-to-Equity Ratio, Price-to-Earnings Ratio, and Price-to-Revenue Ratio. This combination would give us greater confidence in the results.

First, we wanted to identify how many trading days would produce the best returns. We used logistic regression to determine the probability of a positive return on a trade. From the chart below we can see that as the number of days increases, the probability that we have a positive return also increases. This is generally true since S&P500 goes higher as the holding period increases. We did not explore beyond a 10-day holding period since that would not be considered as trading, it would rather be termed as a medium-term hold.

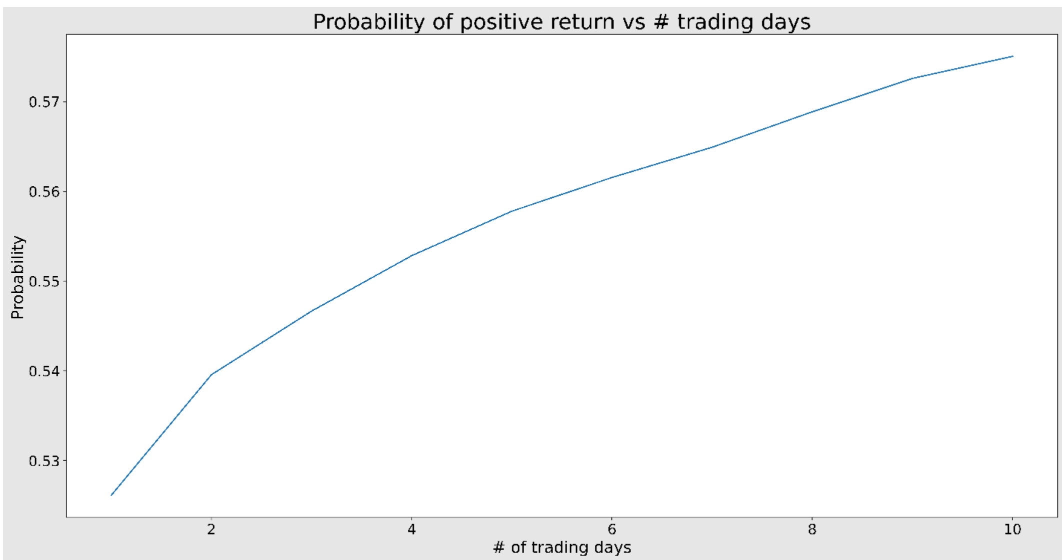

Figure 2 Probability of positive return vs # trading days

Once we settled on testing on 10-day returns, the next step was to try a few different machine learning models. Linear regression provided a very poor coefficient of determination of 0.0005. This indicated that linear regression was not going to be the right model to use for this dataset.

We ended up choosing a Logistic Regression Model to classify stocks into “buy” and “don’t buy” categories. Our model indicated that we could drop P/E ratio and diff_MACD as factors as they were not significant at the p=0.05 level. The below model was the result:

	Coef.	Std.Err.	z	P>|z|	[0.025	0.975]
RSI	0.0052	0.0001	46.4376	0.0000	0.0050	0.0054
Debt/Equity	0.0003	0.0002	2.1523	0.0314	0.0000	0.0006
Price/Revenue	0.0107	0.0012	8.9337	0.0000	0.0083	0.0130

RSI was the dominant factor, indicative of that technical indicator being far more important to our model than the other factors.

Using the logistic regression model, we generated a list of top 10 stocks to buy. They were sorted in order of highest probability of generating a positive return. See Figure 3 which compares the average return of the 10 stocks produced by the logistic regression model vs S&P500 returns. Our model outperforms S&P500 for the first 15 days, however when the market turns lower, our model underperforms. One possible explanation is that RSI is not a good indicator when the market is trending downwards. The RSI would hit a bottom very quickly which would falsely indicate to the model that it's a good time to buy. Another explanation is that S&P500 has been on an overall upward trend during the entire sample period of the last 5-10 years. This means that any model fit over that time would likely be biased towards buying stock. Therefore, it is unsurprising that our model would outperform when the index returns are positive and underperform when index returns are negative. Lastly, the S&P500 is an index of approximately 500 companies, so our small sample of 10 stocks would be expected to have a greater variance in returns, with larger swings in both directions when moves occur.

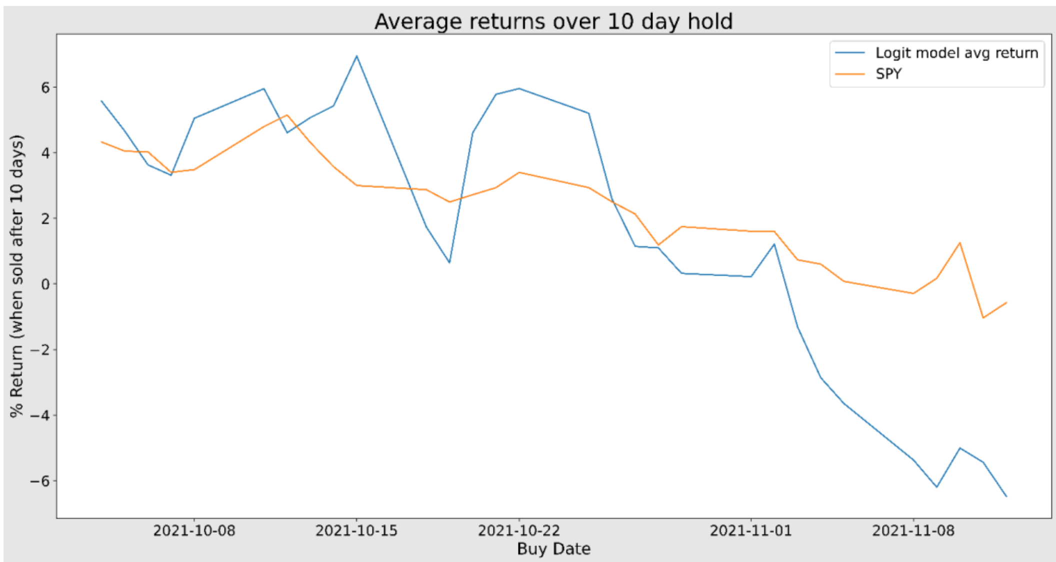

Figure 3 Comparing returns of Logit model vs SPY

Conclusion/Discussion

Overall, our logistic regression model generally outperforms the SPY benchmark on some days and underperforms on some days. One strategy to improve our model would be to detect when the market is in a downtrend. During such periods, the model can either suggest the user avoid trading or extend the hold period until an uptrend has been detected. Alternatively, we can improve the model to provide a list of stocks to short which are flashing overbought signals (only during periods of downtrend or in general). By considering both sides of a given trade, we could use the predictive power of the model to make money, even when the market is trending downward.

While we attempted to utilize a new method to identify stock opportunities, stock trading is a highly explored space with trillions of dollars at stake. If we were to identify a successful trading method, it would likely be quickly replicated, and then the opportunity would be reduced to nearly nothing (according to the efficient market hypothesis). Thus, it would be challenging to identify an easily discovered model through public data that was consistently profitable.

Group members have contributed an equal amount of work according to the plan outlined.

Citations

Adebiyi, A. A., & Adewumi, A. O. (2015). *Stock price prediction using the Arima model*. IEEE Xplore. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7046047.

Al-Radaideh, Q. A., Assaf, A. A., & Alnagi, E. (2013). Predicting Stock Prices Using Data Mining Techniques. *The International Arab Conference on Information Technology (ACIT'2013)*, *2013*. Retrieved October 13, 2021, from https://www.researchgate.net/publication/281865047_Predicting_Stock_Prices_Using_Data_Mining_Techniques.

Bruni, R. (2017). Stock market index data and indicators for day trading as a binary classification problem. *Data in Brief*, *10*, 569–575. https://doi.org/10.1016/j.dib.2016.12.044

Cai, Y., Song, Z., Sun, G., Wang, J., Guo, Z., Zuo, Y., Fan, X., Zhang, J., & Lang, L. (2019). On visualization analysis of stock data. *Journal on Big Data*, *1*(3), 135–144. https://doi.org/10.32604/jbd.2019.08274

Chen, Y.-J., Chen, Y.-M., & Lu, C. L. (2017). Enhancement of stock market forecasting using an improved fundamental analysis-based approach. *Soft Computing*, *21*(13), 3735–3757. https://doi.org/10.1007/s00500-016-2028-y.

Chong, T., & Ng, W. (2008). Technical analysis and the London stock exchange: testing the MACD and RSI rules using the FT30, Applied Economics Letters, 15:14, 1111-1114, DOI:10.1080/13504850600993598 https://doi.org/10.1080/13504850600993598

Credit Suisse. (2010). Technical Analysis - Explained. Zurich: Research & Publications SEF. Retrieved from https://www.credit-suisse.com/pwp/pb/pb_research/technical_tutorial_de.pdf

Dai, Z., Dong, X., Kang, J., Hong, L. (2020). Forecasting stock market returns: New technical indicators and two-step economic constraint method. The North American Journal of Economics and Finance, Volume 53, 2020, 101216, ISSN 1062-9408 https://doi.org/10.1016/j.najef.2020.101216.

Farias Nazário, R. T., e Silva, J. L., Sobreiro, V. A., & Kimura, H. (2017). A literature review of technical analysis on stock markets. *The Quarterly Review of Economics and Finance*, *66*, 115–126. https://doi.org/10.1016/j.qref.2017.01.014

Fernando, J. (2021, November 3). *Moving average convergence divergence (MACD)*. Investopedia. https://www.investopedia.com/terms/m/macd.asp.

Fernando, J. (2021, November 3). *Price-to-earnings (P/E) ratio*. Investopedia. https://www.investopedia.com/terms/p/price-earningsratio.asp.

Fernando, J. (2021, November 3). *Relative strength index (RSI)*. Investopedia.https://www.investopedia.com/terms/r/rsi.asp.

Friesen, G. C., Weller, P. A., & Dunham, L. M. (2009). Price trends and patterns in technical analysis: A theoretical and empirical examination. *Journal of Banking & Finance*, *33*(6), 1089–1100. https://doi.org/10.1016/j.jbankfin.2008.12.010

Guo, Y. (2020). Stock trading based on principal component analysis and clustering analysis. *IOP Conference Series: Materials Science and Engineering*, *740*, 012129. https://doi.org/10.1088/1757-899X/740/1/012129

Hargrave, M. (2021, November 3). *Price-to-sales (P/S) ratio*. Investopedia. https://www.investopedia.com/terms/p/price-to-salesratio.asp.

Macchiarulo, A. (2018). *Predicting And Beating The Stock Market With Machine Learning And Technical Analysis*. icommercecentral.com. https://www.icommercecentral.com/open-access/predicting-and-beating-the-stock-market-with-machine-learning-and-technical-analysis.pdf.

Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2015). Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques. *Expert Systems with Applications*, *42*(1), 259–268. https://doi.org/10.1016/j.eswa.2014.07.040

Rodríguez-Cándido, N. P., Espin-Andrade, R. A., Solares, E., & Pedrycz, W. (2021). A compensatory fuzzy logic model in technical trading. *Axioms*, *10*(1), 36. https://doi.org/10.3390/axioms10010036

Shynkevich, Yauheniya & Mcginnity, T.M. & Coleman, Sonya & Belatreche, Ammar & Li, Yuhua. (2017). Forecasting Price Movements using Technical Indicators: Investigating the Impact of Varying Input Window Length. *Neurocomputing. 10.1016/j.neucom.2016.11.095.*

Stanković, J., Marković, I., & Stojanović, M. (2015). Investment strategy optimization using technical analysis and predictive modeling in emerging markets. *Procedia Economics and Finance*, *19*, 51–62. https://doi.org/10.1016/S2212-5671(15)00007-6

Tanaka-Yamawaki, M., & Tokuoka, S. (2007). Adaptive use of technical indicators for the prediction of intra-day stock prices. *Physica A: Statistical Mechanics and Its Applications*, *383*(1), 125–133. https://doi.org/10.1016/j.physa.2007.04.126

Ugur, S. & Murat, A. (2014). TN-RSI: Trend-Normalized RSI Indicator for Stock Trading Systems with Evolutionary Computation. *Procedia Computer Science, vol. 36, no. C, Elsevier B.V, 2014, pp. 240–45, doi:10.1016/j.procs.2014.09.086.* https://doi.org/10.1016/j.procs.2014.09.086

Vaiz, S. J., & Ramaswami, M. (2016). A Study on Technical Indicators in Stock Price Movement Prediction Using Decision Tree Algorithms. *American Journal of Engineering Research (AJER)*, *5*(12), 207–212.e-ISSN: 2320-0847 p-ISSN : 2320-0936 https://www.ajer.org/papers/v5(12)/Z05120207212.pdf