# Abstract

The profitability of the stock market has led many researchers to study its movement with human intuition and knowledge. This paper hopes to present another way of predicting whether a stock beats the market by using machine learning algorithms trained on quarterly stock data. No qualitative data was able to be used, which put the model’s predictions at a disadvantage to financial analysts like Warren Buffet. To make up for this, only the top fifteen stock with the highest confidence of beating the market were invested in. This model predicted investments that created a 384.89% return from the last quarter of 2011 to the first quarter of 2018, which was just under three times the S&P500’s return of 130.24% and a 25.32% annual returns.

# Introduction

Even in the late 1900s, financial researchers moved in the direction of using statistical models to make predictions. In 1968, Edward Altman published his famous Z score that was able to predict whether a public company was going bankrupt using financial data from each stock. Today, financial researchers are still using statistical models to predict the movement of a stock based on momentum for day trading. Instead of following what current researchers are doing, this paper presents models based on quarterly financial statements and metrics of a stock to create a more long term trading strategy. This strategy was developed with the R programming language and the corresponding R Studio integrated development environment. There are a variety of reasons why Quantitative trading techniques like Arbitrage and High Frequency Trading were not used in this paper. These strategies are computationally expensive, very competitive, and complex. In High Frequency Trading, stocks are bought and sold in microseconds in order to catch the increase of an investment instruments by as low as a penny. In order to do this, multiple computers with high amounts of computing power and teams of professional analysts are necessary. This is also very risky because unsupervised computers are trading millions of shares of highly volatile investments. This caused many firms to trade in lower volumes and be inconsistent from year to year unlike an investor such as Warren Buffet. Another quantitative strategy is Arbitrage where a computer analyzes the same stock on multiple stock exchanges and checks if a market inefficiency has caused the price on one exchange to be different from another. However, this strategy has become almost impossible due to higher computational speeds that are correcting more of these errors between exchanges. In order for a machine learning algorithm to trade consistently and with great volume, a more long term trading approach needed to be created.

# Prediction variable

Instead of predicting the price change of a stock from quarter to quarter, the models in this paper were used to classify whether a stock did better than the S&P 500 financial index. Transforming this problem into a classification problem, made it much easier to solve. This switch also helped to scale the stock movement prediction from one quarter to the next. If the market during the training quarter drops 40% then the predictions of the next quarter will also have drops near 40% even though it is very improbable that the market would drop this much in two consecutive quarters. Therefore, a second machine learning model would have needed to be created in order to predict the movement of the entire market in order to scale the predicted movement of the stocks from quarter to quarter. Since switching to a classification problem was easier and took less work, that was used instead to scale the data.

# Model Blending

Overall, 106 machine learnings models were blended together. The positive predictions were given a weight of one and the negatives were given a weight of zero. These weights were then added up to make the final prediction. Instead of using majority vote to classify a stock as beating the market, only the stocks with the highest combined prediction weights were selected as a possible investment. This strategy aimed at increasing sensitivity at the expense of accuracy. Sensitivity, in this case, is the percent of stocks that beat the market out of the total number of stocks that were predicted to beat the market. The number of stocks chosen varied from 3 to 15 and the total returns of each are showed later in the paper. Of about 100 to 150 stocks, selecting 12 stocks produced the biggest return on investments.

# Removing Unpredictable Stocks

Another discovery that was able to boost the Sensitivity of the model was that beaten down and slow growth stocks were much harder to predict than in favor stocks and were therefore removed from the model. For the model, the beaten down stocks were classified as having a trailing year over year price increase of less than 15.28103% and in favor stocks were classified in the opposite way. This pivot point came from the median of the trailing year over year price increase of every stock from each quarter in the data set. There are two reasons why beaten down stocks are unpredictable. Even if beaten down stocks obtain good quarterly results, investors may not trust the stock until it achieves multiple quarters with good financial statistics. This number of quarters varies from each stock to the next, which makes predicting very difficult. Beaten down stocks are also less volatile than in favor stocks because beaten down stocks and slow growth stocks are value stocks and value investors are long term investors. Value investors, therefore, are more likely to hold onto an investment if the stock has a bad quarter, which could lead to misleading patterns in the data. This was a beneficial discovery since in favor stocks are usually considered growth stocks, and these stocks are more likely to have larger returns on investments than value stocks. On the contrary, growth stocks also have more risk, but the prediction model minimizes this risk while retaining the large gains of growth stocks. One bad aspect about a growth’s stock volatility is that the stock may go up huge in one quarter and plummet in the next. Due to this volatility, quarterly data was used instead of annual data. The models were originally trained with annual data and the results were subpar.



The above picture, from Yahoo Finance, exemplified the two key aspects of the prediction model stated above. The Stock, NKTR, was bought where the cross hairs point during the first quarter of 2018 and was sold at the pinnacle of the stock’s movement. The Stock went up 77.93% between these two periods. This image shows the importance of using quarterly data because the stock came back down to the price it was bought at in the quarter after the stock was sold. If annual data was used, the large gain of this stock wouldn’t have been able to be caught without the subsequent fall. This image is also important because the only time this stock was considered in our model was in the quarter we bought it in because of the large price increase right before the quarter. This tactic of only using in favor stocks prevented the model from potentially buying the stock in the previous quarters, which would have ended in a bad investment. When the models were tested without using these two strategies in several quarters, the return on investments was much closer to the market return on investment.

# Stock Data

Data Source

The data used in this paper’s models came from stockrow.com. The Income statement and the cash flow statement used trailing quarterly data and the balance sheet along with some of the stock metrics came from just the individual quarter. Stockrow didn’t provide the prices of each stock so the prices were retrieved from Yahoo Finance using their API. The training set and test set were set to be consecutive quarters. If there is any time separating the end of the training set and the start of the test set, even by one quarter, the accuracy and sensitivity would go down dramatically. In order to normalize the data, foreign companies were not included and the sectors of utilities, conglomerates, and industrial goods were also excluded.

Imputation/irregular data

Some American companies couldn’t be used because Stockrow didn’t provide complete data sets for all American stocks. Some of the stocks had many missing values, some were missing entire quarters, and some stocks were not old enough to be included. The number of stocks with too much missing information increased as the stocks decreased in their market capital. Low capital stocks, therefore, were unable to be used in the models. Other stocks had some missing values, but were exclusive to only a few features, which were kept in the data. Most of these missing data included the number of shares, retained earnings, working capital, and the financial variables in the subset of the total assets and total liabilities. These values were imputed with the MICE library in R using CART imputation. The stocks from Stockrow also had dates that were staggered by about a month or two, which meant that the quarters of each stock did not overlap. This changed the threshold of whether a stock beat the market or not because the price change of the S&P500 was calculated for each stock individually using the stocks’ own months. So a stock that beat the market could potentially have had a worse quarter than another stock that didn’t beat the market.

Clustering

Clustering was used to try to match similar quarters together and use the similar quarters as the training and test set. This had a negative effect on the results due to the small clustering size of 32 quarters and the inability to factor in the economic climate into the clustering. If statistics about the overall well-being of the economy and monthly data were used, clustering may have had a positive effect on the results, but for the models in this paper, consecutive quarters were used with quarterly data.

Preprocessing

The data from Stockrow came from six different downloadable excel files from their website that included the income sheet, balance sheet, cash flow sheet, metrics trailing, metrics quarterly, and growth files. The Growth file proved to be unimportant since it only had a few growth statistics, which were overridden when a function was called that created a growth feature for each statistic. The process of downloading these six files was done with 326 stocks and put into folders with the stocks’ ticker symbol. These files were then loaded into R and merged based on the date of the stock. This was done individually for all of the stocks and then the rows of each stock were combined together to form one data frame. During this process, a call to yahoo’s finance API allowed for the input of stock prices into the data frame.

# Training Process

First step

The first approach that was used to create a predictive model was not useful. Several financial scores like the Piotroski F-Score, Altman Z-Score, Beneish M-Score, and several other scores I created based on these were used as features, in several combination, with a random forest using three fold cross validation. The accuracy nor the sensitivity of any of the random forest models were consistently better than guessing; therefore this approach was revaluated, and was transformed into working with the financial categories (profitability, valuation, liquidity, and solvency) using Linear Discriminant Analysis (LDA). Other models and categories then stemmed from this approach.

Categories of Features

The data was separated into several different categories. The four basic financial ratios, including profitability, value, activity, liquidity, and solvency, were used. Since stocks are affected by both their present financials and future financials, several growth categories were included. The four year average growth of the stock and the predicted value of the stock in twenty years were included. The future twenty year statistic combined both the present financials and the growth of the stock by taking the sum of one and the four year average to the 20th power and multiplying that by the current statistic. This formula is the same as compound interest compounded annually. Two more categories were added to account for the present valuation and the corresponding average financial statistic from the Dow Jones Industrial Average. The former divided the statistic by the price to earnings and the latter divided the statistic by the corresponding average for the statistic from the DOW.

Machine Learning Algorithms

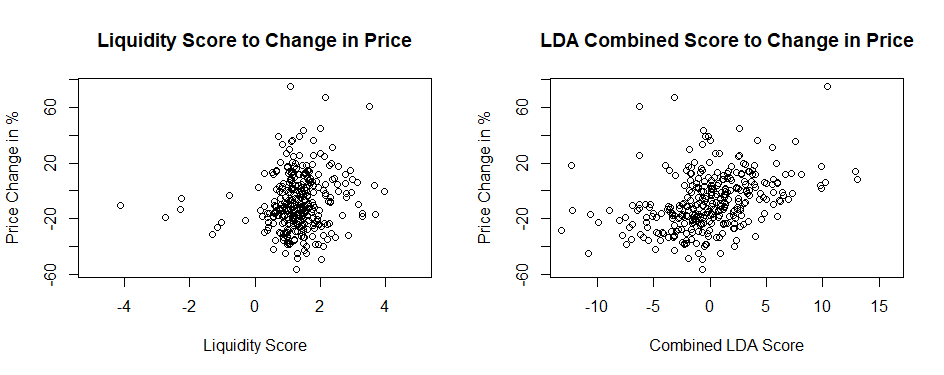
Several machine learning and dimensional reduction algorithms were then applied to these nine categories. The predictions from stepwise logistic regression, support vector machines, gradient boosted machines (GBM), and Fisher’s linear discriminant analysis (LDA) models of all nine categories were used in the final prediction. The predictions of each category were also combined with majority vote and included two other models: Multivariate Adaptive Regression Splines (MARS) and neural networks. These final predictions of each category were then added to the final model as well. Stepwise logistic regression was used instead of logistic regression because stepwise models are able to do feature selection and thus reduce the high number of dimensions (prediction features) for the model. The SVM and GBM models used three fold cross validation repeated ten times and were executed with the Caret R package. Random Forests were not used because they were more computationally expensive when compared to the GBM models. The GBM models also didn’t over fit the data because Caret runs multiple GBMs all with different tree depths and picks the best one whereas a random forest doesn’t and can therefore have too many branches. LDA predictions were used both in the final prediction of each category and LDA scores were used as features with the other machine learning algorithms. Neural networks were also used with profitability, growth, liquidity, and solvency, but they combined several different variations from each category such as statistic to PE with future twenty year statistic into the same algorithm. Principle component analysis (PCA) and the t-test was also used in combination with many different machine learning algorithms to reduce the number of dimensions of each model except for the stepwise regression models and neural networks.

Data Exploration

Usually, machine learning algorithms will be accompanied with exploratory modeling for feature selection. This was attempted, but many of the statistics offered little prediction power by themselves. It was also believed that statistics that are predictive in one quarter might not be predictive in another quarter due to the volatility of the stock market. So human feature selection was not done to avoid overfitting. Instead, PCA and the t-test were used along with stepwise regression to do feature selection.

LDA Unique Blending

The LDA scores were combined in several unique way that boosted the sensitivity of the model. Many of the LDA models were un-predictive by themselves, but when combined, the sensitivity was dramatically increased. LDA provides a score for each observation by multiplying a statistic by a generated coefficient and then adding it to the product of another statistic and its unique coefficient. This method is used to reduce many features into one sum of products score. These scores were then, ideally, positively correlated with the prediction feature. The normal way of predicting with an individual LDA score is by using logistic regression on it. This was used and the predictions were added to the final model. After, these scores were combined using several machine learning models including logistic regression, SVM, MARS, and neural networks by using the scores as features for each model. These scores were also preprocessed by centering the scores by subtracting the individual scores by its mean and scaling the scores by dividing the score by it standard deviation. This allowed for the addition of the scores into one score for each stock. This was then added onto the final prediction, which is the only one added that didn’t have a weight of 1 or 0. The following graphs show the predictive power of one LDA score for liquidity and the predictive power of the combined LDA score.



Using the t-test

A t-test was also used to select variables for the machine learning algorithms. This test determined whether two samples are significantly different or not. This was important because statistic that are very similar between stocks that beat the market and stocks that didn’t beat the market are not useful in prediction models. The t-test was executed with the training set between statistic of stocks that beat the market and statistic of stocks that didn’t beat the market. The t-test returned p-values with the smallest p-value indicating which statistics were significantly different. The statistics with the smallest p-values were used in the models. The number of statistics for each model ranged from 15 to 30. Statistics with p-values greater than 0.1 were generally not used unless there were less than 15 statistics with a p-value of less than 0.1. Usually a p-value of .05 is used to decide if a statistics is significantly significant, but not enough statistics had p-values that were less than .05. This t-test strategy of feature selection was used with each of the nine categories and proved to yield the most predictive models. Due to the four regular financial categories not having enough variables for the t-test to be effective, multiple variations of the statistic were used. The variations included the regular statistic, the four year average growth statistic, the statistic to the PE, and the future 20 year statistic. A t-test was run for each variation and the four smallest p-values of each were used in the prediction model. The t-test selected variables were then used with all of the algorithms that Principle Component Analysis was used with.

# Model Results

The final model was tested based on sensitivity from the last quarter of 2010 to the first quarter of 2018. The sensitivity and average price increase were recorded for each model. The average price of each model was used to calculate the total return of 100,000 invested and then reinvested in each quarter. This Process was done with the top three, six, nine, twelve, and fifteen stocks and the compared to the S&P 500 and to themselves. The following shows the sensitivity and return on investment of the highest ranked 15 stocks vs S&P500 for each quarter.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Time Periods** | **Sensitivity 15** | **Increase 15** | **Return 15** | **S&P500 Increase** | **S&P500 Return** |
| ***Totals 2011 Q4*** | 80.00% | 16.37% | $116,365.33 | 5.92% | $105,915.43 |
| ***Totals 2012*** | 60.30% | 23.34% | $143,520.22 | 14.39% | $121,161.45 |
| ***Totals 2013*** | 61.67% | 39.41% | $200,082.39 | 27.10% | $154,000.41 |
| ***Totals 2014*** | 49.29% | 12.20% | $224,491.92 | 13.63% | $174,998.30 |
| ***Totals 2015*** | 56.67% | 10.81% | $248,764.33 | -0.01% | $174,982.11 |
| ***Totals 2016*** | 41.67% | -2.18% | $243,337.24 | 9.37% | $191,384.98 |
| ***Totals 2017*** | 66.67% | 41.50% | $344,320.87 | 18.59% | $226,968.69 |
| ***Totals 2018 Q1*** | 53.33% | 8.96% | $375,178.91 | 1.44% | $230,242.44 |

The Following table shows how the number of stocks chosen from each quarter can affect the results. Twelve stocks proved to have the highest sensitivity at 58.39% and the highest return on investments at a 384.89% increase.

|  |  |  |  |
| --- | --- | --- | --- |
| **Number of Stocks** | **Sensitivity** | **Increase** | **Return** |
| ***Three stocks*** | 56.41% | 255.57% | $355,573.97 |
| ***Six Stocks*** | 56.41% | 287.55% | $387,551.09 |
| ***Nine Stocks*** | 58.39% | 345.24% | $445,239.60 |
| ***Twelve Stocks*** | 58.01% | 384.89% | $484,890.70 |
| ***Fifth Teen Stocks*** | 56.86% | 275.18% | $375,178.91 |
| ***S&P500*** | ---------- | 130.24% | $230,242.44 |

No matter what number of stocks were used, the models always beat the S&P500 by a considerable amount. This may seem surprising due to the low sensitivity, but the stocks that didn’t beat the market were usually close to zero and therefore didn’t hurt the very high price increases of the stocks that did beat the market. This was because only the stocks with the best financials, according to the models, were used.

The models in this paper also had highs in lows. During 2012, 2013, 2017, and 2018, our models well outperformed the market, but in 2014 and 2016 our models didn’t pick stocks that did better than the market. The most probable explanation for this is that only in favor stocks were used. These stocks are more overvalued than beaten down stocks and were therefore more likely to respond to stock market corrections and bear markets more than beaten down stocks. It also seemed that the bear market had some quarters where the market did very good and the market did very badly. This variation of bear market is less expressed in bull market where the stock market only goes up. This variation could possibly have created too much change between the training set and the test set, which meant that the training set wouldn’t be a good representation of the test set.

The following table shows the fifth teen stocks’ returns vs Warren Buffet’s returns.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Time Periods** | **Sensitivity 15** | **Increase 15** | **Return 15** | **Buffet Increase** | **Buffet Return** |
| ***Totals 2011 Q4*** | 80.00% | 16.37% | $116,365.33 | 16.37% | $116,365.33 |
| ***Totals 2012*** | 60.30% | 23.34% | $143,520.22 | 17.00% | $136,147.44 |
| ***Totals 2013*** | 61.67% | 39.41% | $200,082.39 | 33.00% | $181,076.10 |
| ***Totals 2014*** | 49.29% | 12.20% | $224,491.92 | 27.00% | $229,966.64 |
| ***Totals 2015*** | 56.67% | 10.81% | $248,764.33 | -12.00% | $202,370.64 |
| ***Totals 2016*** | 41.67% | -2.18% | $243,337.24 | 23.00% | $248,915.89 |
| ***Totals 2017*** | 66.67% | 41.50% | $344,320.87 | 22.00% | $303,677.39 |
| ***Totals 2018 Q1*** | 53.33% | 8.96% | $375,178.91 | -1.57% | $298,909.65 |

Not only were the models in this paper able to beat the S&P500, but they were also able to beat Warren Buffet’s returns. The 2011 quarter 4 returns of Buffet’s portfolio was unable to be found so the return from the fifth teen stocks by our models were used as Buffet’s return for that quarter. However, it is important to note that Buffet handled much more money than the models in this paper did. This huge volume of money could not be used with this way of trading because stocks don’t offer up enough shares for Buffet to buy with his money in a given quarter. Instead, Buffet must take several years building up his positions in companies and must buy more stocks that he may not feel perfectly comfortable with. Before Buffet handled so much money, he averaged returns around 20% annually. This 20% increase is very similar to the increase in this paper at 20.93% for all fifth teen stocks and 25.32% for twelve stocks.

# Considerations:

The sensitivity of the models increased when smaller time periods of the sample data were used. This was the case in the transition from annual to quarterly data. This meant that monthly data may have worked better than the quarterly data, but Stockrow didn’t provide monthly data. This was also believed to be the case when determining whether to use different model for each financial sector. However, this change decreased the sensitivity of the model because of a decrease in training data for each sector. Many of the machine learning algorithms over fit the data and GBMs didn’t even have enough observations to use. A larger data set was unable to be attained because Stockrow didn’t have a complete data set for each stock and only stocks with in favor stocks were used. When the trailing annual year over year price change threshold was lowered to obtain more samples with the individual sectors, the sensitivity became even worse. If monthly data was run for each individual sector, the return on investments presented of this paper may be able to be increased.

The returns showed by this paper may not be 100% accurate. When a stock reports good earnings, then usually the stock will go up a couple of percentage points immediately before the stocks were bought, which would adulterate the returns showed by the model. However, the same concept applies to stocks that don’t beat the market. Stocks that beat the market could also go down before they were able to be bought, which would increase the returns. Because of this, it seemed that this situation would both hurt the models and help them, which at the very worst would decrease the returns by a couple of percentages, but might actually increase the effectiveness of the models. Another consideration that these models do not take into account is qualitative data. The prediction results and the investors intuition and knowledge could be used together to increase the returns of the investments.

The returns showed in this paper would have also been decreased by taxes. Since short term investments include stocks held for less than a year, the stocks in the investment models would be taxes dramatically more because short term investments are taxed as ordinary income. Long term stocks would only be taxed at a rate of between 15% and 20%.

As noted early, Stockrow’s data was riddled with missing data and missing quarters. If this data was corrected, the predictive power of the models would probably increase dramatically. Missing value were imputed with CART imputation, but this process can sometimes be ineffective and create patterns in the data that shouldn’t be here. The missing data points were minimized to a couple of variables from Stockrow, but these variables were used extensively with creating the variables from the financial categories: Profit, value, Liquidity, solvency, and activity. This is why it was believed that perfect imputation of these variables would result in a much more predictive model.

# Model Features and Stocks

The above variables were used to create the growth, statistic to PE, future, and statistic to market features.

Profit:

Gross Margin, ROA, ROE, Profit Margin, Free Cash Flow Margin, Return on Capital Employed, Operating Profit Margin, Return on Operating Cash Flow, Return on Retained Earnings, Cash Flow Return on Investment, Return on Debt, Return on Average Equity, Cash Return on Assets, Return on Average Assets, Return on Research Capital, Cash Return On Capital Invested

Value:

PE, PCF, PS, PB, PEG, PEGY, EPS, Cash EPS, Net Asset Value per Share, Times Preferred Dividends Earned

Activity:

Asset Turnover, Merchandise Inventory Ratio, Working Capital Turnover Ratio, Fixed Asset Turnover Ratio, Account Receivables Turnover, Account Payables Turnover, Inventory Turnover, Average Days of Payables, Average Days of Receivables, Inventory Turnover, Cash Turnover, Days Sales Outstanding, Days Payable Outstanding, Days Inventory Outstanding, Operating Cash Flow to Sales, Days Working Capital, Days Cash on Hand, Sales to Administrative Expenses, Investment Turnover, Sales to Equity, Inventory to Sales, Sales to Operating Income, Goodwill to Assets, Free Cash Flow to Sales

Solvency:

Cash Flow to Debt, Debt Ratio, Equity Ratio, Working Capital to Debt, Current Cash Debt Coverage, Interest Coverage, Asset Coverage, Interest Expense to Debt, Capitalization Ratio, Debt to EBITDA, Long-term debt ratio, Net Debt to EBITDA, Cash Flow Coverage, Financial Leverage Index, Non-Current Asset to Net Worth, Long-term debt to Equity Ratio, Fixed-Assets to Net Worth Ratio, Total Debt To Total Assets, Debt to Equity Ratio, Current Ratio

Liquidity:

Cash Ratio, Quick Ratio, Cash to Working Capital, Inventory to Working Capital, Sales to Current Assets, Sales to Working Capital, Net Working Capital Ratio, Acid Test, Cash Flow Coverage, Current Ratio

Continuous Variables:

Weighted Average Shares, Long-term debt, Total Assets, Revenue, Net Income, Cost of Revenue, Gross Profit, Research and Development (R&D) Expenses, Selling, General and Administrative (SG&A) Expenses, Operating Expenses, EBITDA, EBIT, Cash and cash equivalents, Investments Current, Cash and short-term investments, Inventory, Current Assets, Goodwill and Intangible Assets, Assets Non-Current, Current Liabilities, Liabilities Non-Current, Total Debt, Total Liabilities, Shareholders Equity, Financing Cash Flow, Investing Cash Flow, Operating Cash Flow, Average Receivables, Average Payables

Stocks:

A, AAL, AAPL, ABC, ABMD, ABT, ACN, ADI, ADM, ADP, ADS, ADSK, AET, AFG, AFL, AGN, AIG, AIV, ALGN, ALK, ALKS, ALL, ALV, ALXN, AMD, AMG, AMP, AMT, AMTD, AMZN, ANTM, AON, APD, ARCC, ARW, ATHN, ATVI, AVB, AXP, AZO, AZPN, BA, BAC, BAX, BBT, BDX, BEN, BG, BIIB, BIO, BK, BLK, BLL, BMY, BOKF, BR, BRO, BSX, BWA, BX, BXP, C, CA, CACC, CAG, CAT, CB, CBS, CBSH, CCI, CCL, CE, CELG, CFR, CFX, CHD, CHRW, CI, CL, CLX, CMCSA CME, CMG, CAN, COF, COLM, COO, COST, CPB, CPRT, CSCO, CSL, CSX, CTL, CTSH, CVS, CVX, DAL, DE, DHR, DISCA DISH, DPS, DRI, DVA, DWDP, DXCM, EA, EBAY, ECL, EFX, EGN, ELS, EMN, EMR, EPD, EQR, ESRX, EW, EWBC, EXAS, EXPE, EXR, F, FAST, FDS, FDX, FFIV, FIS, FISV, FITB, FLEX, FLIR, FNF, GD, GE, GGG, GILD, GIS, GLW, GPC, GS, GWW, HAL, HAS, HCP, HD, HEI, HES, HFC, HIG, HLF, HOG, HP, HRL, HST, HSY, HUBB, HUM, HUN, IBM, ICE, INGR, INTC, INTU, IP, IPG, IPGP, IR, IRM, ISRG, IT, JAZZ, JBHT, JBLU, JCI, JNPR, JPM, JWN, KAR, KEY, KIM, KMB, KMX, KO, LAMR, LB, LEN, LKQ, LLL, LLY, LMT, LNC, LOW, LPT, LUV, MA, MAC, MAR, MCD, MCK, MDLZ, MET, MHK, MKL, MMM, MNST, MO, MOH, MRK, MRO, MRVL, MSFT, MSI, MTB, MTN, MU, NFLX, NKE, NKTR, NLY, NNN, NOC, NSC, NTAP, NUE, NVDA, NWL, OHI, OMC, ORCL, ORLY, OXY, PAYX, PCAR, PEP, PFE, PG, PGR, PH, PII, PKG, PLD, PM, PNC, POOL, PPG, PRU, PSA, QCOM, REG, REGN, RF, RGA, RHI, RHT, RJF, RMD, ROL, ROP, RPM, RSG, RTN, SBAC, SCCO, SCHW, SCI, SEE, SEIC, SHW, SIRI, SJM, SLB, SLG, SNPS, SNV, SPG, SRPT, STI, STT, STZ, SUI, SYK, TEL, TGT, TIF, TJX, TMK, TMUS, TR, TRV, TXN, TYL

UAL, UDR, UHAL, ULTA, ULTI, UNH, UNM, UNP, UPS, URI, UTX, V, VLO, VRTX, VTR, VZ, WAT, WBA, WBC, WCN, WFC, WHR, WLK, WM, WMB, WMT, WPC, WST, WU, XOM, XPO, XRX

# References:

“NKTR Interactive Stock Chart | Nektar Therapeutics Stock.” Yahoo! Finance, Yahoo!, finance.yahoo.com/chart/NKTR.

“Strategies And Secrets Of High Frequency Trading (HFT) Firms.” Investopedia, Investopedia, 14 Aug. 2017, [www.investopedia.com/articles/active-trading/092114/strategies-and-secrets-high-frequency-trading-hft-firms.asp](http://www.investopedia.com/articles/active-trading/092114/strategies-and-secrets-high-frequency-trading-hft-firms.asp).

“Arbitrage.” Investopedia, Investopedia, 22 Apr. 2018, www.investopedia.com/terms/a/arbitrage.asp.