8-2020

We Can Use Machine Learning to Determine Which Financial Ratios are Best for Investors

Collin Butterfield
Utah State University

We Can Use Machine Learning to Determine Which Financial Ratios are Best for Investors

by

Collin Butterfield

A thesis submitted in partial fulfillment
of the requirements for the degree

of

MASTER OF SCIENCE

in

Financial Economics

Approved:

Dr. Tyler Brough, Ph.D., Finance
Major Professor

Paul Fjeldsted, MBA, Management
Committee Member

Dr. Pedram Jahangiry, PhD, Economics
Committee Member

UTAH STATE UNIVERSITY
Logan, Utah
2020

ABSTRACT

We Can Use Machine Learning to Determine

Which Financial Ratios are Best for Investors

by

Collin Butterfield, Master of Science

Utah State University, 2020

Major Professor: Dr. Tyler Brough
Department: Economics and Finance

This study develops and tests the hypothesis that the machine learning algorithm, Random Forests, can be used to systematically pick financial ratios that would be best for indicating market trends and be used subsequently to perform comparable analysis to speculate whether a firm is over- or under-valued. Results show that financial ratio selection differs depending on the market sector to which a firm pertains. We examine the 11 financial sectors representing the key areas of the economy. We also look at four possible trading strategies that an investor could have: month-long, quarter-long, semi-annual, and annual to capture differing trading horizons.

Keywords: Financial Analysis, Comparable Analysis, Machine Learning, Random Forests, Market Sector, Investor Strategy, Feature Selection

ACKNOWLEDGMENTS

I would like to thank Dr. Tyler Brough for making available to me his time and knowledge. I would especially like to thank my committee members, Dr. Pedram Jahangiry, CFA and Paul Fjeldsted, CFA for their support and assistance throughout the entire process.

I give special thanks to my wife Michelle, family, friends, and colleagues for their encouragement, moral support, and patience as I worked my way from the initial proposal writing to this final document. I could not have done it without all of you.

Collin Butterfield

CONTENTS

1. **Introduction**:

When it comes to financial and comparable analysis, a question that has plagued many financial analysts is, “Which financial ratios should be used to evaluate the price of the firm?” Ratio analysis started with the current ratio as a method to gauge credit worthiness (Beaver, 1966). And since then there have been an abundance of new and interesting financial ratios, all claiming to be the best at helping to value a publicly traded company. From personal experience there is no correct ratio to use, it is left to the opinion of the analyst to which they then must defend why they chose that financial ratio leading to the respective valuation.

In this study I show that there is a systematic way to determine which financial ratios are the best to use in financial analysis. I will be using the machine learning algorithm called Random Forests to classify whether the market is in an upward trend or a downward trend and then extract the features importance list from the algorithm. The feature importance list will indicate which financial ratios are doing a better job at predicting the trend in the market. For this study I will look at the 11 financial sectors that represent the key areas of the economy and break them down into four different investing horizons – we can also think of these as investing strategies. The four investing horizons are: monthly, quarterly, semi-annually, and annually. I will then compare the different strategies across market sectors, looking for similarities and differences.

2. **Process**:

The reason I am using Random Forests is because of its unique capacity to record and track which features are doing the best job, on average, at reducing the error rate across all decision trees formed. The feature space will be made up of industry standard financial ratios. The Random Forest classifier is an ensemble method that decorrelates all the decision trees used during bootstrap aggregation thus keeping the bias of the model low while also lowering the overall variance. All our Random Forest classifiers' hyper-parameters were tuned using a grid search using the following criterion:

- Max Depth: 1, 10, 100
- Criterion: Gini, Entropy
- Max Features: Auto, Square root

All data sets were split into a training set (80%) and a test set (20%); and then were trained using the training set and then predicted onto the test set. Since we are not interested in making this a trading strategy, we do not need to focus on trying to increase accuracy. Our goal is to use the Random Forest classifier's feature importance to learn which ratios, form the feature space, do the best job at producing the associated accuracy.

To use the Random Forest classifier, we need a target variable that relates to the question proposed. We create a binary target variable that represents a 'buy' or 'sell' signal in the market, 1 for 'buy' and 0 for 'sell' respectively. We use the 11 SPDR ETF funds' prices as the dataset to build our target variables. The very nature of the funds is to track the respective sector using a basket of the most popular companies within that sector. We use the ETF funds as a proxy for any company within the respective sector. For example, let's assume a new technology company, like Apple, recently IPO'd and analysts are busy trying to determine if its stock price is over or undervalued. We would use the financial ratios determined by our results derived from the SPDR ETF fund XLK. The 'buy' or 'sell' signal is created by taking the moving average of the fund's prices. We know that when the price crosses above the moving average, the market is trending upward, creating a ‘buy’ signal. If the price crosses below the moving average, then the market is trending downward, indicating a ‘sell’ signal. Thus, for each sector we have four target variables to accompany the four investment horizons.

- Monthly: $Buy_{20} = if\left(price_{adj} > MA(20)_{price}\ , 1, 0\right)$
- Quarterly: $Buy_{60} = if\left(price_{adj} > MA(60)_{price}\ , 1, 0\right)$
- Semi-Annually: $Buy_{100} = if\left(price_{adj} > MA(100)_{price}\ , 1, 0\right)$
- Annually: $Buy_{200} = if\left(price_{adj} > MA(200)_{price}\ , 1, 0\right)$

The feature space will consist of seven popular financial ratios used for financial and comparable analysis. The ratios are: Price/Earnings (P/E), Price/Book Value (P/Bk), Price/Sales (P/Sales), Enterprise Value/ Trailing 12-month EBITDA (EV/T12EBITDA), Enterprise Value/Trailing 12-month Sales (EV/T12Sales), Price/Earning-to-Growth (PEG), Price/Free Cash Flow (P/FCF). We will also include the number of trades made per day (Volume) and the last price from the day before (lag_price). Volume will be added to give more information to the model to learn from. The reason we include lag_price is because many econometric literature state that yesterday’s price is the best predictor of today’s price, if this is the case we will be able to use this idea to gauge how well our model is doing, meaning we should see lag_price higher up in the feature importance list.

The data for the feature space comes from taking a weighted average from three companies of the top holdings in each individual sector, aiming for a higher representation of the holdings – ideally around 50%. For example, looking at the technology sector ETF, XLK, we used Microsoft (MSFT), Apple (AAPL), and Visa (V), representing 47.11% of all the holdings in the XLK ETF fund. Then each ratio would be calculated using a weighted average based off market capitalization. Below is an example of how the P/E ratio was calculated for XLK. The market caps of MSFT, AAPL, and V are $1,347B, $1,254B, and $359.9B respectively. Thus, w_1, w_2, and w_3 are .4549, .4235, and .1216 respectively.

$$\left\{\frac{P}{E}\right\}_{XLK} = \left\{\frac{P}{E}\right\}_{MSFT} * w_1 + \left\{\frac{P}{E}\right\}_{AAPL} * w_2 + \left\{\frac{P}{E}\right\}_{V} * w_3$$

The following was done for all the financial ratios. Since the target variables were created using moving averages, we took the moving averages of all the ratios according to the respective investment horizon to smooth out the noise of the day to day.

3. Results:

As predicted, there is a difference to which financial ratios are to be used dependent on the market sector. In this section I will discuss each sector individually according to investment horizons.

Keep in mind that we cannot use lag_price nor volume in financial analysis to formulate a prediction and/or valuation. Both were used to aid the model in learning and to gauge whether our model is following current econometric theory. It is also important to note that we cannot use the associated values of the feature importance list to determine the proportionate strength of one ratio over the other. We can only determine that one does a better job than the other (for this reason the associated numbers are not shown in figures, only in the appendix).

Energy:

For the energy sector, the target variable was created using the Energy Select Sector SPDR ETF (XLE) that tracks companies that generate revenue tied to crude oil, natural gas and other commodities. Typically, this ETF is used for short-term buy and hold strategies; it is commonly used to transfer risk exposures (Research, 2020).

The feature space was created using the market cap weighted average (fig. 1) financial ratios of Chevron Corporation (CVX), Exxon Mobil Corporation (XOM), and Kinder Morgan, Inc. (KMI), representing 50.9% of all the holdings in the ETF. Both

Company	Market Cap ($ Billions)	Weight
CVX	170.98	49.6%
XOM	143.05	41.5%
KMI	30.47	8.8%

Figure 1

XOM and KMI are missing more than 25% of the data associated with P/FCF. For this reason, we decided to take out P/FCF as a feature seeing that two of the tree companies that are being used as an estimate for the entire energy sector are missing data for it. Then the four investment horizon moving averages were created for each feature in the feature space (MA(20), MA(60), MA(100), MA(200)).

In the monthly horizon we see that the PEG ratio is the feature doing the best at reducing the error rate in the model. Followed by the P/Bk and P/E ratios (Fig. 2). It would make sense that if the PEG ratio is the best ratio for predicting a 'buy' signal, that the P/E ratio would not be far behind seeing that the PEG ratio is formed from the P/E ratio. We also see that lag_price is the second-best predictor of a 'buy' signal in the market. This is a good sign that our model is working in our favor since we know that lag_price should be the best predictor of today's price.

At the quarterly level we see a bit of a change. P/Bk falls out of the top 3 and EV/EBITDA becomes the third best ratio at reducing the error rate of the model – in other words, it is the third best at determining an upward trend in the market. However, we do see lag_price jump up to the best predictor. This is inline with our theory, again giving confidence that the model is working appropriately.

Looking at the semi-annual investment horizon, we see that EV/EBITDA is now the best ratio (Fig.2). It would seem that as the investment horizon gets longer, the more important EV/EBITDA becomes. In the monthly model EV/EBITDA was the fourth best predictor. Take note that lag_price is still the best predictor of the market signal.

The last investment horizon is the annual strategy. Here we see a similar story as the semi-annual strategy, except P/Bk is not back in the second spot (Fig. 2). However, it is interesting to note that lag_price dropped to the sixth best predictor. This could possibly be due to the fact that the smoothing done by the 200-day moving average lost (rolled out) some of its tacit knowledge that is held within prices.

To summarize the energy sector, an analyst should be looking at the PEG and/or P/E ratios when performing analysis in the monthly or quarterly investment horizon. But if the analyst wants to extend that horizon, they should be looking at the EV/EBITDA and/or P/E ratios. For a more detailed information about the Energy Sector's features space results, see appendix 1).

Sector: Energy (MA feature space)				
Top 3 Ratios	**Monthly (20)**	**Quarterly (60)**	**Semi-Annually (100)**	**Annually (200)**
1st	PEG	PEG	EV/EBITDA	EV/EBITDA
2nd	P/Bk	P/E	PEG	P/Bk
3rd	P/E	EV/EBITDA	P/E	P/E

Figure 2

Materials:

For the materials sector, the target variable was created using the Materials Select Sector SPDR ETF (XLB) that tracks companies that are engaged in the extraction or production of natural resources as a way to generate revenue. Typically, this ETF is used for both long-term and short-term buy and hold strategies; it is commonly used to balance exposure to the US equity market (Research, 2020).

The feature space was created using the market cap weighted average (fig. 3) financial ratios of Air Products and Chemicals, Inc. (APD), Ecolab Inc. (ECL), and The Sherwin-Williams Company (SHW), representing 23.3% of all the holdings in the ETF.

Company	Market Cap ($ Billions)	Weight
APD	47.84	32.7%
ECL	52.07	35.6%
SHW	46.46	31.7%

Figure 3

Only APD was missing data; the PEG ratio was missing less than 1%, so we filled it in with the median of the associated year. Then the four investment horizon moving averages were created for each feature in the feature space (MA(20), MA(60), MA(100), MA(200)).

From our monthly results, we see that the financial ratios that are doing the best to reduce the error rate of the model are: First P/Bk, Second PEG, and Third P/FCF (Fig. 4). This suggests that in the short-term, analysts should be evaluating based off company book value and equity. We also see that lag_price is the best predictor. This is a good sign that the model is following theory.

In the quarterly model we see a change in the order. P/FCF becomes the best ratio to look at when doing financial analysis. P/FCF is followed by the PEG ratio and P/Bk ratio. This suggests that as the investment horizon increases, analysts should be looking closer into the equity of the firm. Lag_price remains at still as the best indicator, causing no alarm.

The semi-annual tells a similar story to the monthly and quarterly results. We see in figure 11 that the best ratios for determining a market trend upward are: P/Bk, P/FCF, and PEG. It is good to see that over different investment horizons the same ratios are showing up. This gives strong evidence that no matter the investment horizon length, analysts can rely on the P/FCF and P/Bk ratios as a tool for analysis.

The annual investment model is the exact same story as the semi-annual model. This shows strong evidence that P/Bk and P/FCF are great ratios for determining a 'buy' signal. One interesting note; lag_price fell in rank, only from first to second, this does raise the same concern as in the energy sector results. Are we losing the knowledge that price carries by smoothing it out too much?

To summarize the materials sector, analysts should be looking at a firm's book value and equity to know whether a company over or undervalued. The length of the investment horizon does not matter. The PEG ratio is also a good tool to use. For additional information on the materials feature space results, see appendix 2.

Sector: Materials (MA feature space)				
Top 3 Ratios	**Monthly (20)**	**Quarterly (60)**	**Semi-Annually (100)**	**Annually (200)**
1st	P/Bk	P/FCF	P/Bk	P/Bk
2nd	PEG	PEG	P/FCF	P/FCF
3rd	P/FCF	P/Bk	PEG	PEG

Figure 4

Industrials

For the industrials sector, the target variable was created using the Industrials Select Sector SPDR ETF (XLI) that tracks companies that include transportation, providers of commercial and professional services, and manufacturers of capital goods to generate revenue. Typically, this ETF is used for long-term buy and hold strategies; it is commonly used because of its high liquidity (Research, 2020).

The feature space was created using the market cap weighted average (fig. 5) financial ratios of Honeywell International Inc. (HON), Union Pacific Corporation (UNP), and The Boeing Company (BA), representing 14.9% of all the holdings in the

Company	Market Cap ($ Billions)	Weight
HON	95.85	34.9%
UNP	105.92	38.6%
BA	72.77	26.5%

Figure 5

ETF. This is the lowest representation of holdings used in this analysis. In future research more data will be used to increase overall percent holdings. Only BA was missing data, we filled them in with the median of the associated year. Missing data for year 2020 was filled by the median of 2019. Then the four investment horizon moving averages were created for each feature in the feature space (MA(20), MA(60), MA(100), MA(200)).

According to our model, the monthly strategy's best ratios are: P/Bk, EV/EBITDA, and PEG (fig. 6). Lag_price is the best predictor; this is in line with the econometric theory. According to the data, there is no real difference between EV/EBITDA and the PEG ratio, EV/EBITDA is only fractionally better.

The quarterly strategy tells a different story. The best ratios are: P/FCF, PEG, and P/E. Only the PEG ratio is similar. This suggests that the industrials sector is more sensitive to the investment horizon. Lag_price is still the leading feature for determining a market trend.

For the semi-annual model, the results seem very similar to that of the quarterly model. The ratios an analyst should be using for a mid-term investment horizon are: PEG, P/FCF, and P/E. The similarity between the quarterly and semi-annually models suggest that analysts should be looking at firm's cash flows and earnings.

The annual model introduces the P/Bk value back into the feature importance. For a year+ investment horizon, the best ratios that determine a 'buy' signal are: P/FCF, P/Bk, and P/E. This is the first annual model that has lag_price as the leading feature. This supports our econometric theory and the idea that prices hold the tacit knowledge of the market.

To summarize the industrials sector, for most strategies, analysts should be using a combination of the P/FCF, P/E and PEG ratios. However, it is important to point out that in the short-run, analysts should be looking more at book and enterprise value. For a more in-depth analysis of the industrials sector, see appendix 3.

Sector: Industrials (MA feature space)				
Top 3 Ratios	**Monthly (20)**	**Quarterly (60)**	**Semi-Annually (100)**	**Annually (200)**
1st	P/Bk	P/FCF	PEG	P/FCF
2nd	EV/EBITDA	PEG	P/FCF	P/Bk
3rd	PEG	P/E	P/E	P/E

Figure 6

Consumer Discretionary:

For the consumer discretionary sector, the target variable was created using the Industrials Select Sector SPDR ETF (XLY). Typically, this ETF is used as a rotation strategy for when the market is in a state of recovery (Research, 2020).

The feature space was created using the market cap weighted average (fig. 7) financial ratios of Amazon.com, Inc. (AMZN), The Home Depot, Inc. (HD), and McDonald's Corporation (MCD), representing 44.7% of all the holdings in the ETF. Both

Company	Market Cap ($ Billions)	Weight
AMZN	1200.00	76.7%
HD	227.91	14.6%
MCD	136.82	8.7%

Figure 7

HD and MCD had missing data for P/Bk, 11% and 34% respectively. I decided to remove them from the analysis. Then the four investment horizon moving averages were created for each feature in the feature space (MA(20), MA(60), MA(100), MA(200)).

The monthly model determined that the best ratios are: PEG, P/FCF, and P/E. Similar to the energy sector, it is not a surprise that P/E and PEG would show up together. Lag_price was the second most important predictor.

In the quarterly model P/FCF soars to the top of the list (app. 4) as the best predictor of a market trend. This is followed by P/E and EV/EBITDA (fig. 8). This suggests that a company's cash flows are important to understanding overall company health.

For the semi-annual model we see that the best ratios are: P/FCF, P/E, and PEG. We can see a pattern developing for the short- and mid-term investment horizons. An analyst should be researching a company's earnings and cash flows.

The annual model is completely different compared to all the rest. The best ratios are: PEG, P/Sales, and EV/Sales. Lag_price drops to the fifth best indicator, suggesting that the smoothing has rolled out some of the powerful predictive knowledge of price. For a long-term strategy, a company's revenue should be looked at carefully.

In summary, the consumer discretionary sector in the short- and mid-term investment horizons can be best analyzed using the P/FCF and P/E ratios. But if an analyst is thinking about pitching a long-term buy or hold strategy, they should use the PEG and P/Sales ratios. For a more in-depth analysis of the consumer discretionary sector, see appendix 4.

Sector: Consumer Discretionary (MA feature space)				
Top 3 Ratios	**Monthly (20)**	**Quarterly (60)**	**Semi-Annually (100)**	**Annually (200)**
1st	PEG	P/FCF	P/FCF	PEG
2nd	P/FCF	P/E	P/E	P/Sales
3rd	P/E	EV/EBITDA	PEG	EV/Sales

Figure 8

Consumer Staples:

For the consumer staples sector, the target variable was created using the Consumer Staples Select Sector SPDR ETF (XLP). Typically, this ETF is used when investors want a tilt exposure towards the firms of the market that do well if the economy is in a downturn (Research, 2020).

The feature space was created using the market cap weighted average (fig. 9) financial ratios of The Procter & Gamble Company (PG), The Coca-Cola Company (KO), and PepsiCo, Inc. (PEP), representing 36.7% of all the holdings in the ETF. There was not missing data; the four investment horizon moving averages were created for each feature in the feature space (MA(20), MA(60), MA(100), MA(200)).

Company	Market Cap ($ Billions)	Weight
PG	292.94	43.0%
KO	200.66	29.4%
PEP	187.80	27.6%

Figure 9

In all four investment horizons, lag_price is the best predictor for reducing the overall error rate, on average, for the model. This is strong support that our model is performing well according to econometric theory. For the monthly model the best ratios are: PEG, P/FCF, and P/Bk.

For the quarterly model we see that the best ratios for an analyst to use are: P/FCF, PEG, and EV/Sales. Like the monthly model, a company within this sector in the short- to mid-term investment horizon can be analyzed using cash flows and earnings (fig. 10).

The semi-annual random forest classifier determined that the best financial ratios are: PEG, P/Bk, and P/Sales. And for the annual model the best predictors are: EV/EBITDA, P/Bk, and P/Sales. It seems that as the investment horizon increases, the more an analyst should look at the enterprise and book values, and even investigate the firm's revenues. For a more in-depth analysis of the consumer staples sector, see appendix 5.

Sector: Consumer Staples (MA feature space)				
Top 3 Ratios	**Monthly (20)**	**Quarterly (60)**	**Semi-Annually (100)**	**Annually (200)**
1st	PEG	P/FCF	PEG	EV/EBITDA
2nd	P/FCF	PEG	P/Bk	P/Bk
3rd	P/Bk	EV/Sales	P/Sales	P/Sales

Figure 10

Health Care:

For the health care sector, the target variable was created using the Health Care Select Sector SPDR ETF (XLV). Typically, this ETF is used in long-term strategies and for when investors want to tilt their exposure towards lower risk industries (Research, 2020).

The feature space was created using the market cap weighted average (fig. 11) financial ratios of Johnson & Johnson (JNJ), UnitedHealth Group Incorporated (UNH), and Merck & Co., Inc. (MRK), representing 23.8% of all the holdings in the ETF. Only MRK had missing data for the PEG ratio (< 1%), so we filled it with the corresponding year's median. The four investment horizon moving averages were created for each feature in the feature space (MA(20), MA(60), MA(100), MA(200)).

Company	Market Cap ($ Billions)	Weight
JNJ	409.98	46.3%
UNH	270.33	30.6%
MRK	204.23	23.1%

Figure 11

Just like in the consumer staples sector, the lag_price is near the top of the feature importance list for all models. Lag_price should be the best predictor of today's price according econometric theory, and this supports that theory.

Our monthly model determined that the best ratios are: PEG, P/FCF, and P/E; suggesting that a company's earnings and cash flows can be used to analyze and even compare stock price. A pattern to note; when PEG or P/E show up in the top 3 ratios, the other is usually not too far behind. This is most likely do to the fact that the PEG captures a lot of the P/E ratio's knowledge.

In figure 12 we can see that for the quarterly model's best features are: P/FCF, P/Sales, and EV/EBITDA. And if we look at the semi-annual model, the best ratios are: EV/EBITDA, P/Sales, and P/Bk. This suggests that for a mid-term strategy, an analyst should be using a firm's revenue and enterprise value to determine a buy or hold strategy.

In the long-term, the best ratios are: EV/EBITDA, P/FCF, and P/E. We can see that as investment horizon increases the more ideal the EV/EBITDA ratio becomes. Interestingly, the long and short runs have more in common with each other than the mid-term strategies.

In summary, investment horizon plays more of a role when it comes to determining which financial ratio to use. In the monthly and annual models, it is best to use a firm's enterprise values, cash flows and earning to determine a price. But in the mid-term, an analyst should incorporate the revenue of the company. For a more in-depth analysis of the health care sector, see appendix 6.

Sector: Health Care (MA feature space)				
Top 3 Ratios	**Monthly (20)**	**Quarterly (60)**	**Semi-Annually (100)**	**Annually (200)**
1st	PEG	P/FCF	EV/EBITDA	EV/EBITDA
2nd	P/FCF	P/Sales	P/Sales	P/FCF
3rd	P/E	EV/EBITDA	P/Bk	P/E

Figure 12

Financials:

For the financials sector, the target variable was created using the Financials Select Sector SPDR ETF (XLF). This ETF is exposed to US policy changes but pays a dividend that is useful in times when the economy is down (Research, 2020).

The feature space was created using the market cap weighted average (fig. 13) financial ratios of JPMorgan Chase & Co. (JPM), Bank of America Corporation (BAC), and Wells Fargo & Company (WFC), representing 23.2% of all the holdings in the ETF. There is not data for EV/EBITDA nor EV/Sales, so we dropped them both from this sector analysis. The four investment horizon moving averages were created for each feature in the feature space (MA(20), MA(60), MA(100), MA(200)).

Company	Market Cap ($ Billions)	Weight
JPM	288.76	47.5%
BAC	202.79	33.3%
WFC	116.69	19.2%

Figure 13

Even though we had to drop both EV/Sales and EV/EBITDA, that did not affect the model in any extreme way. The lag_price feature was still the best predictor in all four investment strategies.

For the monthly model, we see that the best ratios for determining a market trend are: P/FCF, PEG, and P/Sales. We end up seeing P/FCF in all the models, this suggests that a financial firm or institution's health can be seen via their cash flows (fig. 14).

In the quarterly model P/Sales jumps to the top of the ratio list, followed by P/Bk and P/FCF. For the financial sector, in the short- and mid-term investment horizon, it appears that revenue is a factor that comes up often in determining a 'buy' signal.

In the semi-annual model, the best ratios are: P/Bk, P/Sales, and P/FCF. Although P/FCF is in all the models, it is important to note that in the mid-term, the company's book value and sales are doing a better job at cluing us in on the real value of the firm.

Our annual random forest classifier determines that the best ratios are: PEG, P/FCF, and P/Bk. Interestingly, like the health care sector, there is this dumbbell effect. The short and long run both have PEG and P/FCF being ratios one and two as the best predictors, unlike the mid-term.

In summary, the financials sector can be valued by looking at a firm's cash flows. For a more in-depth analysis of the financials sector, see appendix 7.

Sector: Financials (MA feature space)				
Top 3 Ratios	**Monthly (20)**	**Quarterly (60)**	**Semi-Annually (100)**	**Annually (200)**
1st	P/FCF	P/Sales	P/Bk	PEG
2nd	PEG	P/Bk	P/Sales	P/FCF
3rd	P/Sales	P/FCF	P/FCF	P/Bk

Figure 14

Information Technology:

For the information technology sector, the target variable was created using the Information Technology Select Sector SPDR ETF (XLK). Typically, this ETF fund splits its assets between the technology and communication services investing across all technology sectors (Research, 2020).

The feature space was created using the market cap weighted average (fig. 15) financial ratios of Microsoft Corporation (MSFT), Apple Inc. (AAPL), and Visa Inc. (V), representing 47.1% of all the holdings in the ETF. Only Visa had missing data; we filled them with the corresponding year's median. The four investment horizon moving averages were created for each feature in the feature space (MA(20), MA(60), MA(100), MA(200)).

Company	Market Cap ($ Billions)	Weight
MSFT	1347.00	45.5%
AAPL	1254.00	42.4%
V	359.91	12.2%

Figure 15

For the information technology sector, all the models are basically the same except the monthly model. The short-term model determines that the best ratios are: PEG, P/FCF, and P/E. Like the other models, P/FCF is important to determing a 'buy' signal. I have noticed a trend that in the short-term, P/E and PEG show up a lot more and then slowly disappear when the investment horizon increase.

All other random forest clssifiers, quarterly, semi-annally, and annually, determine that the best ratios are: P/FCF, P/Bk, and EV/EBITDA – in no particular order (fig. 16) . An analyst should use the P/FCF ratio when performing analysis, and then should compare to the enterprise and book value estimations. For a more in-depth analysis of the information technology sector, see appendix 8.

Sector: Information Technology (MA feature space)				
Top 3 Ratios	Monthly (20)	Quarterly (60)	Semi-Annually (100)	Annually (200)
1st	PEG	P/FCF	P/FCF	EV/EBITDA
2nd	P/FCF	EV/EBITDA	P/Bk	P/FCF
3rd	P/E	P/Bk	EV/EBITDA	P/Bk

Figure 16

Utilities:

For the utilities sector, the target variable was created using the Utilities Select Sector SPDR ETF (XLU). This sector typically captures the part of the market with high distribution yields, yet relatively low volatility (Research, 2020).

The feature space was created using the market cap weighted average (fig. 19) financial ratios NextEra Energy, Inc. (NEE), Dominion Energy, Inc. (D), and Duke Energy Corporation (DUK), representing 30.3% of all the holdings in the ETF. All were missing a significant amount of data from P/FCF, so we removed that feature from this sector analysis. The 4 investment horizon moving averages were created for each feature in the feature space (MA(20), MA(60), MA(100), MA(200)).

Company	Market Cap ($ Billions)	Weight
NEE	119.23	47.9%
D	65.80	26.5%
DUK	63.68	25.6%

Figure 19

Lag_price is the best indicator for all four investment horizons in the utilities sector. This is brought up every time since this is one of the only ways to gauge how my models are doing compared to previous econometric research.

In the monthly classifier we see that the best ratios are: EV/EBITDA, EV/Sales, and PEG. In the quarterly model EV/EBITDA drops slightly in importance, but it picks up P/E. We can see from figure 20 that for both the monthly and quarterly investment horizons, earnings and enterprise value should be considered for valuation.

For the semi-annual and annual classifiers, the ratios switch up a bit. The P/E ratio is still important, but now we have the P/Bk in the long-term and P/Sales in both semi- and annual models. This suggests that when the investment horizon increases then the book value and especially sales become more important for valuation. For a more in=depth analysis of the utilities sector, see appendix 9.

Sector: Utilities (MA feature space)				
Top 3 Ratios	**Monthly (20)**	**Quarterly (60)**	**Semi-Annually (100)**	**Annually (200)**
1st	EV/EBITDA	P/E	P/E	P/Bk
2nd	EV/Sales	EV/EBITDA	PEG	P/E
3rd	PEG	PEG	P/Sales	P/Sales

Figure 20

<u>Telecommunication Services</u>

For the telecommunication services sector, the target variable was created using the Telecommunication Services Select Sector SPDR ETF (XLC). This sector is very new – 2018. However, because of that reason we do not have much data to work with – only 427 max. The feature space was created using the market cap weighted average (fig. 17) financial ratios Facebook, Inc. (FB), Alphabet Inc. (GOOGL), and Netflix, Inc.

(NFLX), representing 35.8% of all the holdings in the ETF. Only NFLX had missing data; we decided to forward fill the missing data for NFLX even though we were missing

Company	Market Cap ($ Billions)	Weight
FB	534.52	33.5%
GOOGL	874.36	54.8%
NFLX	185.33	11.6%

Figure 17

70% of P/FCF. But since the existence of FB and GOOGL, it appears that P/FCF is important. This fact along with the fact that NFLX only makes up 11% of the weighting gave me justification to fill in the missing data (But I have considered to do the analysis without NFLX). The four investment horizon moving averages were created for each feature in the feature space (MA(20), MA(60), MA(100), MA(200)).

For the monthly model, the best ratios are: PEG, P/Sales, and P/FCF. Lag_price is the best indicator – a good sign. If we look at the quarterly and semi-annual results as well, we see that P/Sales and EV/Sales are both very good predictors of our target variable. This suggests that in the short- and mid-term investment horizon lengths, analysts should be using the firm's earnings and revenue to calculate a price.

For the annual model, the results differ. First thing to point out is that lag price drops to the bottom of the list. We are unclear to why this happened,however, this could be simply due to the fact that we have less than 500 observations and we got rid of half of them with the moving average 200 The best ratios are: P/Bk, EV/EBITDA, and P/FCF. Analysts should be using the book and enterprise value of a firm in the long-term. For a more in-depth analysis, see appendix 10.

Sector: Telecommunication Services (MA feature space)				
Top 3 Ratios	Monthly (20)	Quarterly (60)	Semi-Annually (100)	Annually (200)
1st	PEG	P/Sales	P/Sales	P/Bk
2nd	P/Sales	EV/Sales	EV/Sales	EV/EBITDA
3rd	P/FCF	PEG	EV/EBITDA	P/FCF

Figure 18

Real Estate:

For the Real Estate sector, the target variable was created using the Real Estate Select Sector SPDR ETF (XLRE). This sector is very new – 2015.

The feature space was created using the market cap weighted average (fig. 21) financial ratios American Tower Corporation (REIT) (AMT), Prologis, Inc. (PLD), Crown Castle International Corp. (REIT) (CCI), representing 35.2% of all the holdings in the ETF. Because this is such a new ETF there is a lot of missing data. We dropped P/FCF and P/E. And then filled in the rest with a backward fill. With the amount of data that was filled, out of the 11 analyses, this one is the most incorrect and should be taken lightly since more data is needed. The four investment horizon moving averages were created for each feature in the feature space (MA(20), MA(60), MA(100), MA(200)).

Company	Market Cap ($ Billions)	Weight
AMT	110.91	44.9%
PLD	68.02	27.5%
CCI	68.03	27.5%

Figure 21

Since we were missing a lot of data and that fact that we had to drop two of the ratios from the feature space, I was expecting to see lag_price be the best predictor. That was the case for all four classifiers.

From the results seen in figure 22 we can see that in the short- and mid-term that sales and book value are factors in determining the value of a firm. But as the investment horizon transitions to the long-term, earnings and book value become more important for valuation.

In summary, the real estate sector is still new and with as much missing data there is in the data set, we are not as confident that this analysis is correct. But from what we do have, an analyst should focus on using sales and book value in the short- and mid-term strategies but make a slight adjustment to the earnings of the firm in the long-term.

Sector: Real Estate (MA feature space)				
Top 3 Ratios	**Monthly (20)**	**Quarterly (60)**	**Semi-Annually (100)**	**Annually (200)**
1st	P/Sales	PEG	PEG	PEG
2nd	P/Bk	P/Bk	P/Sales	P/Bk
3rd	EV/Sales	EV/Sales	P/Bk	EV/EBITDA

Figure 22

4. Special Note:

One item we have not addressed is that fact that this is time series data. All the features in the feature space and the target variable itself are not independent and identically distributed (idd). The random forest classifier has no awareness of time and takes all data given to it as independent and identically distributed. Since the random forest cannot extrapolate, it will have a hard time predicting values outside the range of the training set (TILGNER, 2019). There are ways to get around this by data pre- and post-processing such as differencing, time delay embedding, and feature engineering. This area will be explored in further research, our model right now is a preliminary study to see if there appeared to be any patterns or accuracy success. This was one of the reasons for the lag_price feature. This was the main gauge to see if our model was still performing in a similar way had it been given idd data.

Another item to note, is that the data only goes as far back as 2010. Since 2010 the market has had incredible momentum with great positive returns. None of our models incorporate a recessionary period of sorts. Future analysis would use 2007-2009 data.

5. Conclusion:

In this study we test the hypothesis that the machine learning algorithm, random forests, can be used to systematically pick financial ratios that would be best for indicating market trends; and be used subsequently to perform comparable analysis to speculate whether a firm is over- or under-valued. The motivation underlying this analysis has come from the age-old question that plagues many financial analysts of which financial ratio should they use to know if a company is under- or over-valued? We obtain the data for 33 different companies and create the target variable using the 11 SPDR ETF funds to create a model for the 11 different market sectors. This model is meant to be a proxy for every company in the associated market sector.

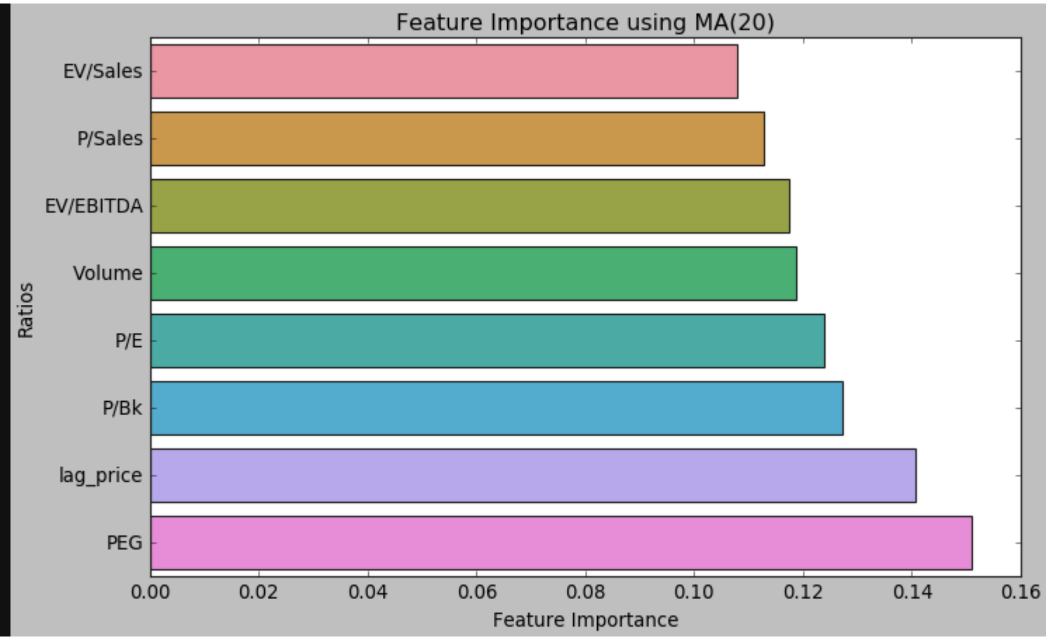
We also look at varying investment horizons that an analyst may be interested in. As proxy for the short-, mid- and long-term horizons we use a 20-, 60-, 100- and 200-day moving average to create the target variables and to smooth out the feature space, getting rid of any noise that be lingering in the data.

We find that the best financial ratios to use are dependent on the sector and investment horizon. There are many patterns in all the different sectors including, but not limited to, P/E and PEG seem to be closely related and P/Bk and EV/EBITDA should be used in the long-term. These models will be continued to be tested and further researched since there are many implications for financial analysts, portfolio managers, and even day-traders, giving them a better way to value and a better access the tacit knowledge that prices hold.

Appendix 1: Energy Sector Feature Importance, feature space has been rolled

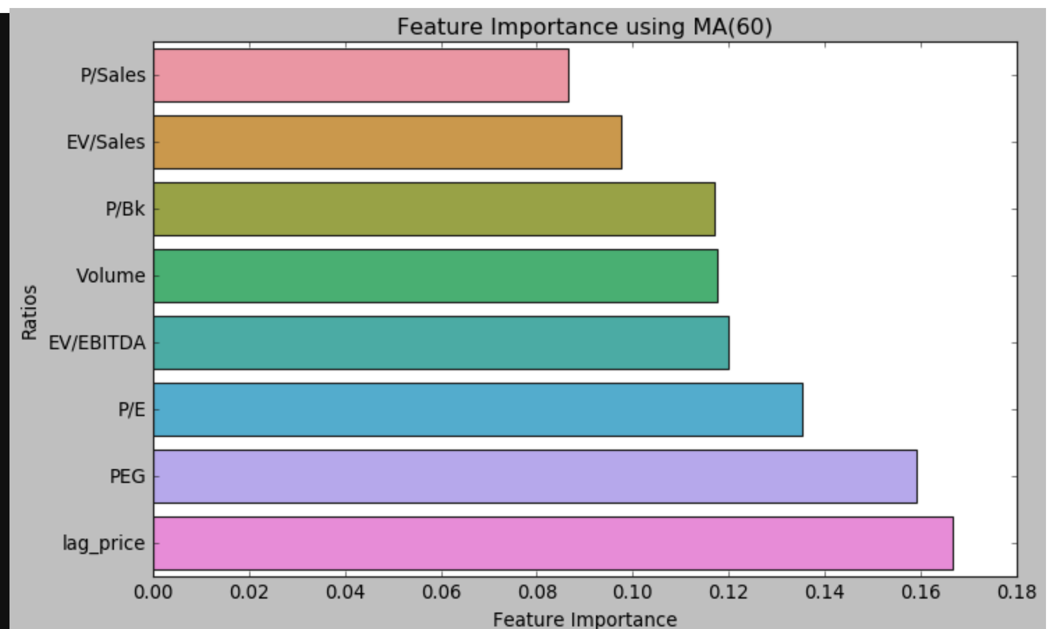
```
Accuracy = 0.84
Precision = 0.84
Recall = 0.85
----------------------------------------
The mean accuracy is:  0.867
The std. dev. of accuracy is:  0.012
```

Features	Feature_importance
EV/Sales	0.10799
P/Sales	0.11275
EV/EBITDA	0.11747
Volume	0.11878
P/E	0.12386
P/Bk	0.12737
lag_price	0.14071
PEG	0.15106

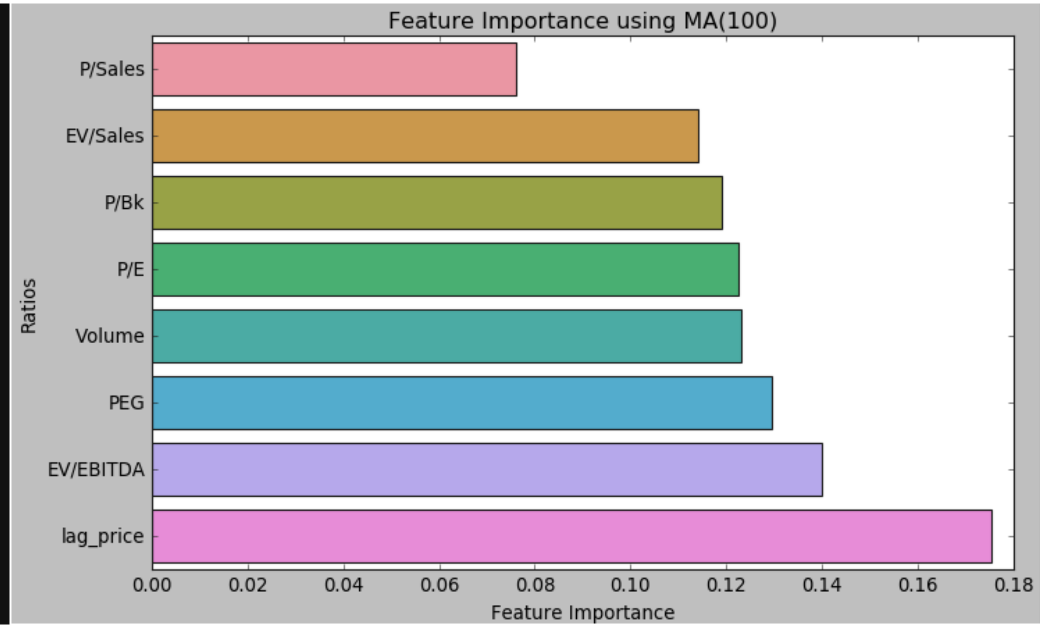



```
Accuracy = 0.96
Precision = 0.95
Recall = 0.98
----------------------------------------
The mean accuracy is:  0.932
The std. dev. of accuracy is:  0.01
```

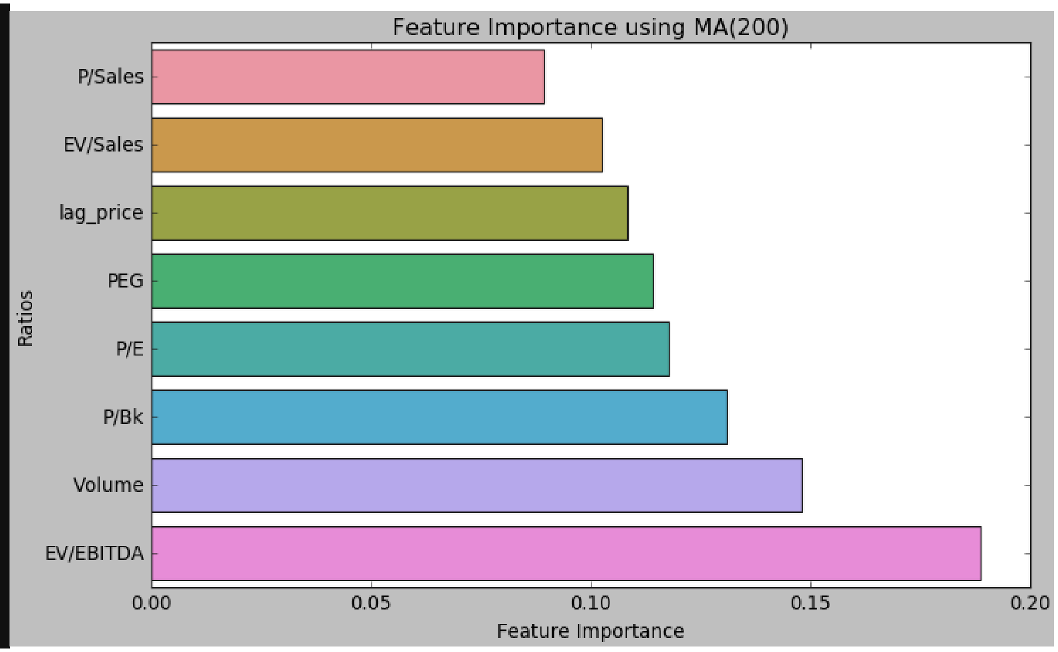
Features	Feature_importance
P/Sales	0.08649
EV/Sales	0.09766
P/Bk	0.11715
Volume	0.11774
EV/EBITDA	0.11988
P/E	0.13545
PEG	0.15902
lag_price	0.16661



```
Accuracy = 0.96
Precision = 0.96
Recall = 0.96
-----------------------------------------
The mean accuracy is:  0.945
The std. dev. of accuracy is:  0.008


+------------+----------------------+
| Features   |   Feature_importance |
|------------+----------------------|
| P/Sales    |               0.0761 |
| EV/Sales   |              0.11419 |
| P/Bk       |              0.11909 |
| P/E        |              0.12245 |
| Volume     |              0.12315 |
| PEG        |              0.12963 |
| EV/EBITDA  |              0.13991 |
| lag_price  |              0.17548 |
+------------+----------------------+
```

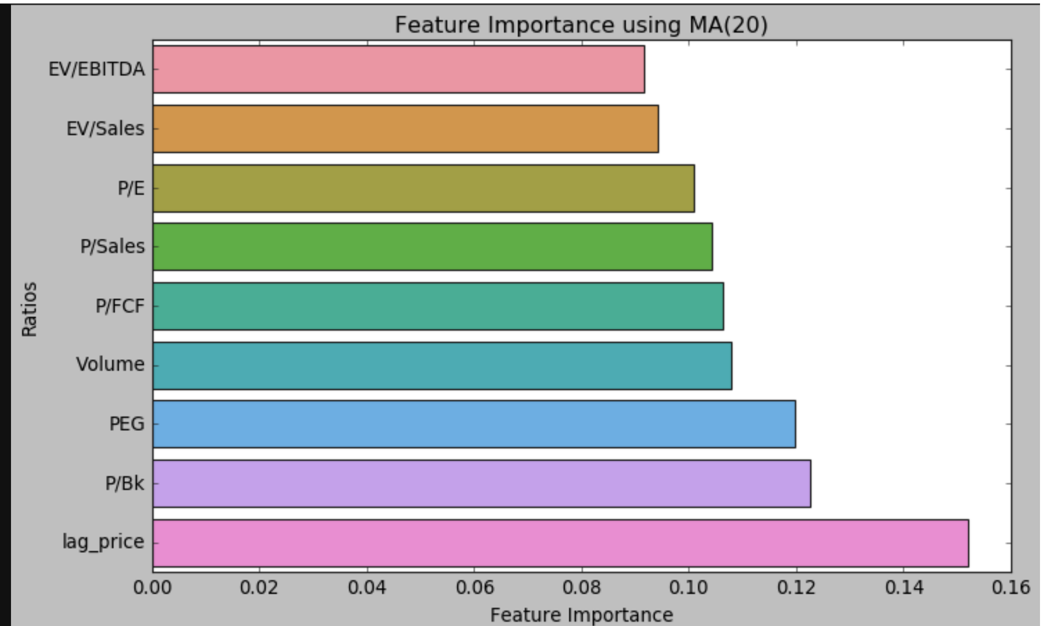



```
Accuracy = 0.97
Precision = 0.97
Recall = 0.97
-----------------------------------------
The mean accuracy is:  0.965
The std. dev. of accuracy is:  0.009


+------------+----------------------+
| Features   |   Feature_importance |
|------------+----------------------|
| P/Sales    |              0.08929 |
| EV/Sales   |              0.10256 |
| lag_price  |              0.10845 |
| PEG        |              0.11408 |
| P/E        |              0.11782 |
| P/Bk       |              0.13094 |
| Volume     |              0.14807 |
| EV/EBITDA  |               0.1888 |
+------------+----------------------+
```

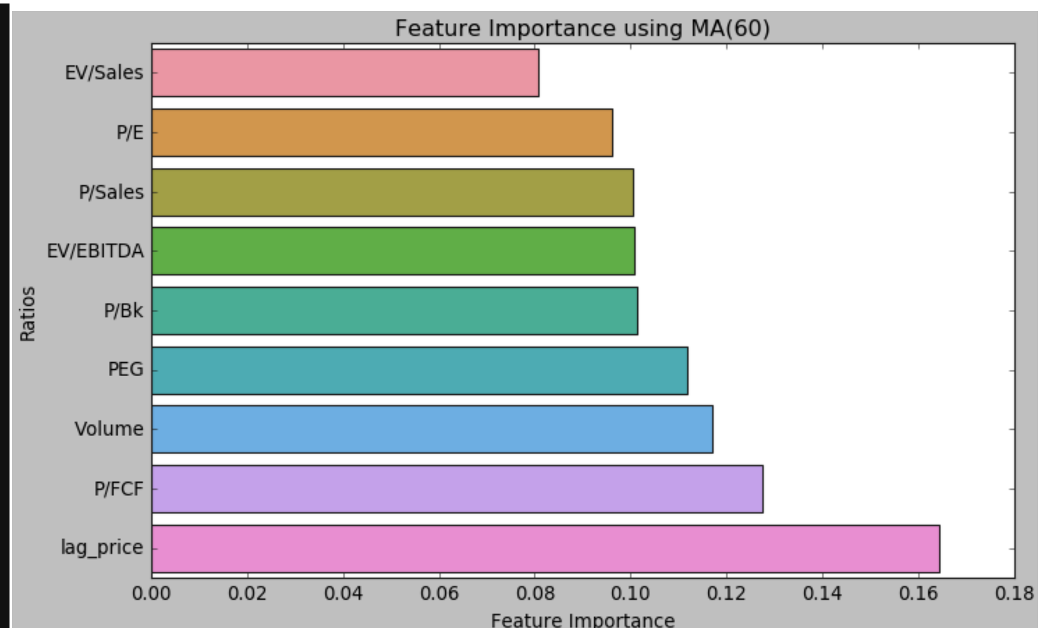


Appendix 2: Materials Sector Feature Importance, feature space has been rolled

```
Accuracy = 0.83
Precision = 0.86
Recall = 0.87
-----------------------------------------
The mean accuracy is:  0.856
The std. dev. of accuracy is:  0.016


+------------+---------------------+
| Features   |  Feature_importance |
|------------+---------------------|
| EV/EBITDA  |             0.09169 |
| EV/Sales   |             0.09427 |
| P/E        |             0.10095 |
| P/Sales    |             0.10439 |
| P/FCF      |             0.10635 |
| Volume     |             0.10796 |
| PEG        |             0.11977 |
| P/Bk       |             0.12252 |
| lag_price  |             0.1521  |
+------------+---------------------+
```



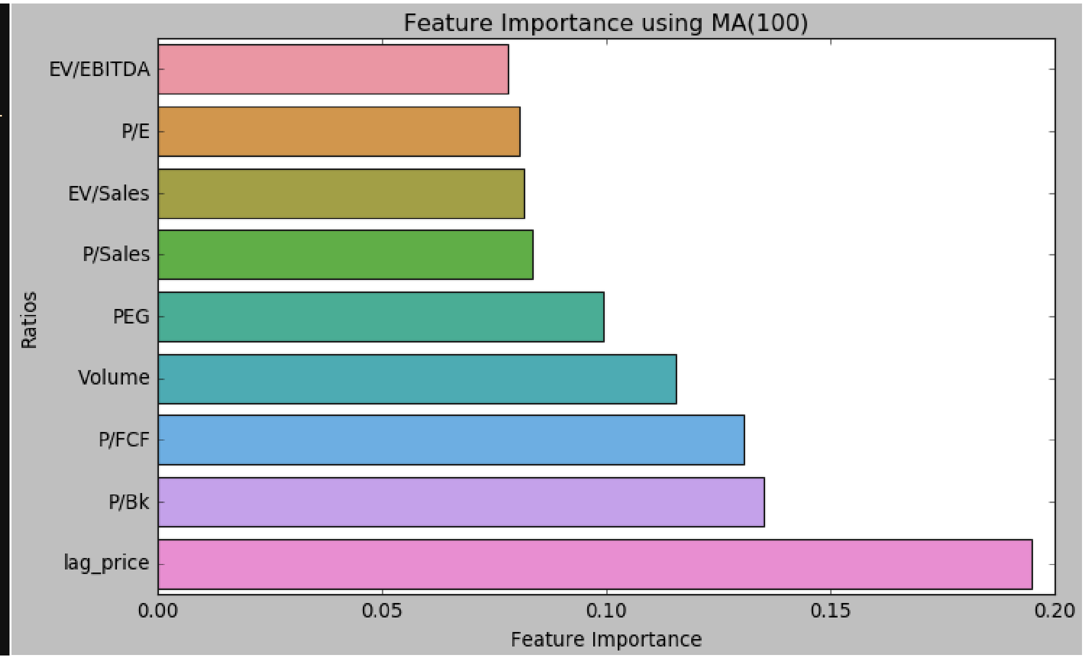

```
Accuracy = 0.94
Precision = 0.93
Recall = 0.97
-----------------------------------------
The mean accuracy is:  0.916
The std. dev. of accuracy is:  0.021


+------------+---------------------+
| Features   |  Feature_importance |
|------------+---------------------|
| EV/Sales   |             0.08061 |
| P/E        |             0.09625 |
| P/Sales    |             0.10049 |
| EV/EBITDA  |             0.10076 |
| P/Bk       |             0.1013  |
| PEG        |             0.11172 |
| Volume     |             0.11697 |
| P/FCF      |             0.12763 |
| lag_price  |             0.16427 |
+------------+---------------------+
```



Accuracy = 0.92
Precision = 0.95
Recall = 0.94

The mean accuracy is: 0.93
The std. dev. of accuracy is: 0.015

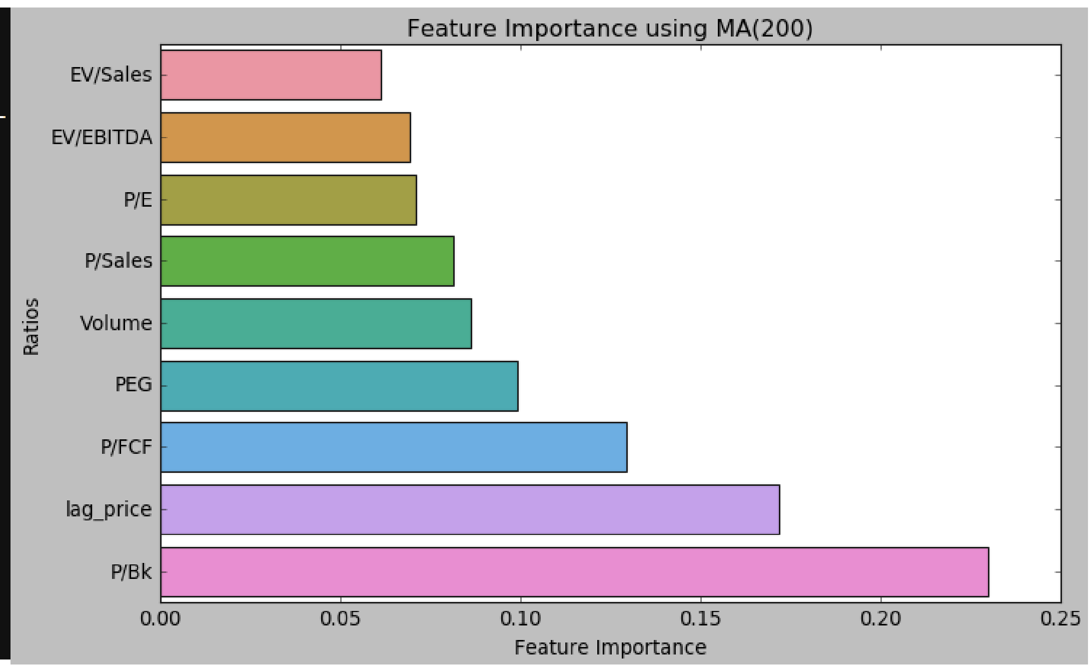
Features	Feature_importance
EV/EBITDA	0.07815
P/E	0.08065
EV/Sales	0.08178
P/Sales	0.08346
PEG	0.09942
Volume	0.1156
P/FCF	0.13069
P/Bk	0.13523
lag_price	0.19502



Accuracy = 0.96
Precision = 0.98
Recall = 0.97

The mean accuracy is: 0.952
The std. dev. of accuracy is: 0.01

Features	Feature_importance
EV/Sales	0.06151
EV/EBITDA	0.06923
P/E	0.07115
P/Sales	0.08141
Volume	0.08616
PEG	0.09908
P/FCF	0.12955
lag_price	0.17201
P/Bk	0.2299

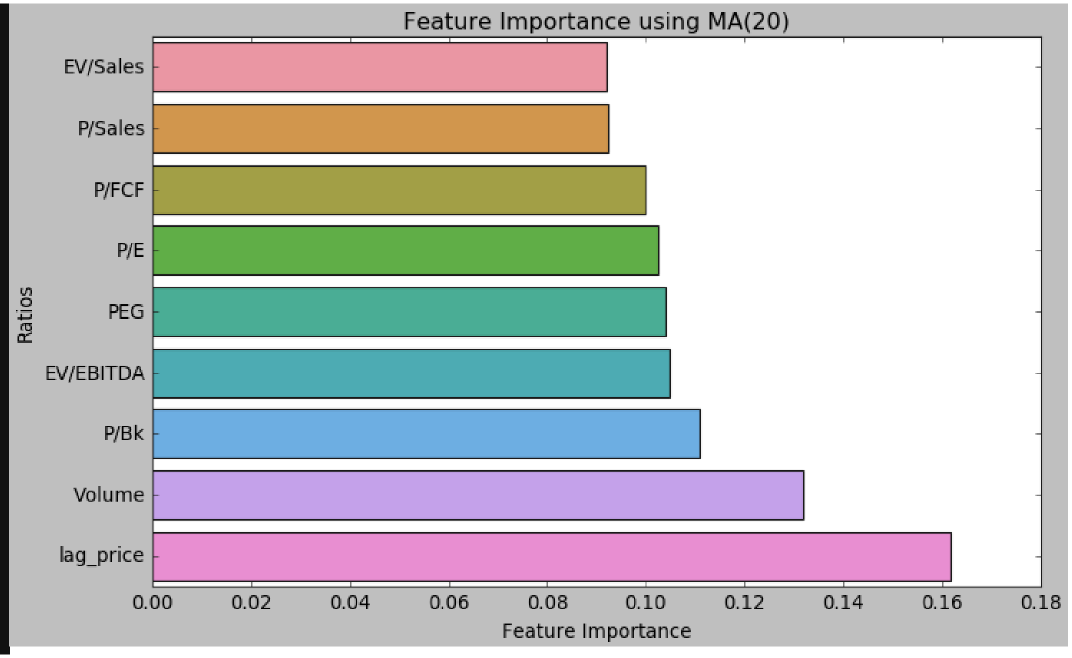


Appendix 3: Industrials Sector Feature Importance, feature space has been rolled

Accuracy = 0.86
Precision = 0.89
Recall = 0.88

The mean accuracy is: 0.848
The std. dev. of accuracy is: 0.007

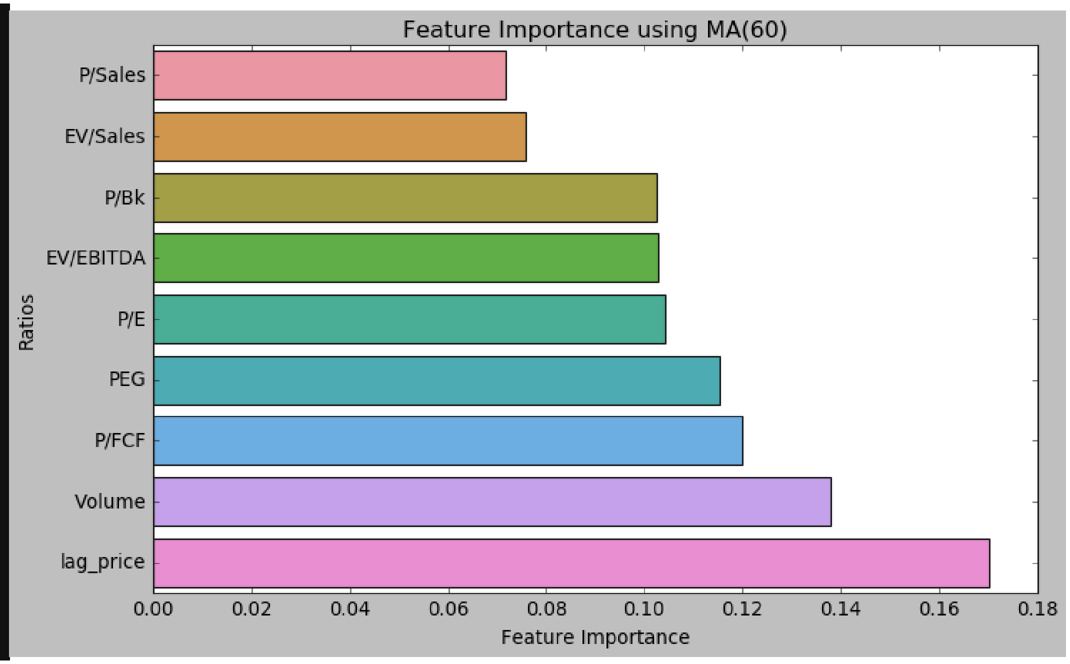
Features	Feature_importance
EV/Sales	0.09208
P/Sales	0.09248
P/FCF	0.09978
P/E	0.10253
PEG	0.10403
EV/EBITDA	0.1047
P/Bk	0.11092
Volume	0.13179
lag_price	0.1617



Accuracy = 0.93
Precision = 0.94
Recall = 0.97

The mean accuracy is: 0.92
The std. dev. of accuracy is: 0.013

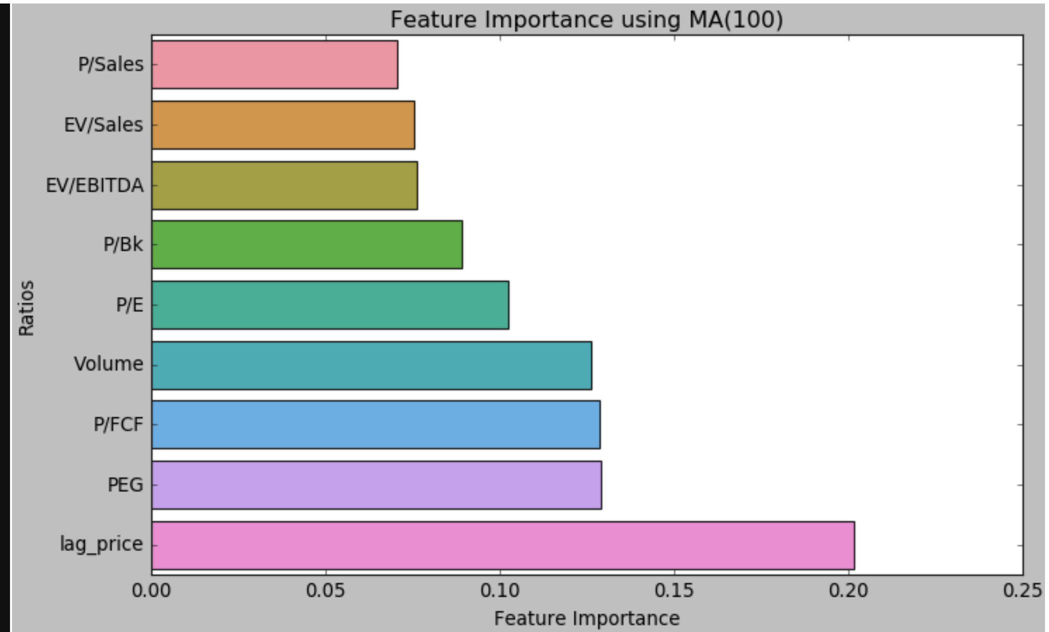
Features	Feature_importance
P/Sales	0.07162
EV/Sales	0.07572
P/Bk	0.10242
EV/EBITDA	0.10291
P/E	0.10414
PEG	0.11517
P/FCF	0.12005
Volume	0.13789
lag_price	0.17008



Accuracy = 0.94
Precision = 0.96
Recall = 0.96

The mean accuracy is: 0.96
The std. dev. of accuracy is: 0.005

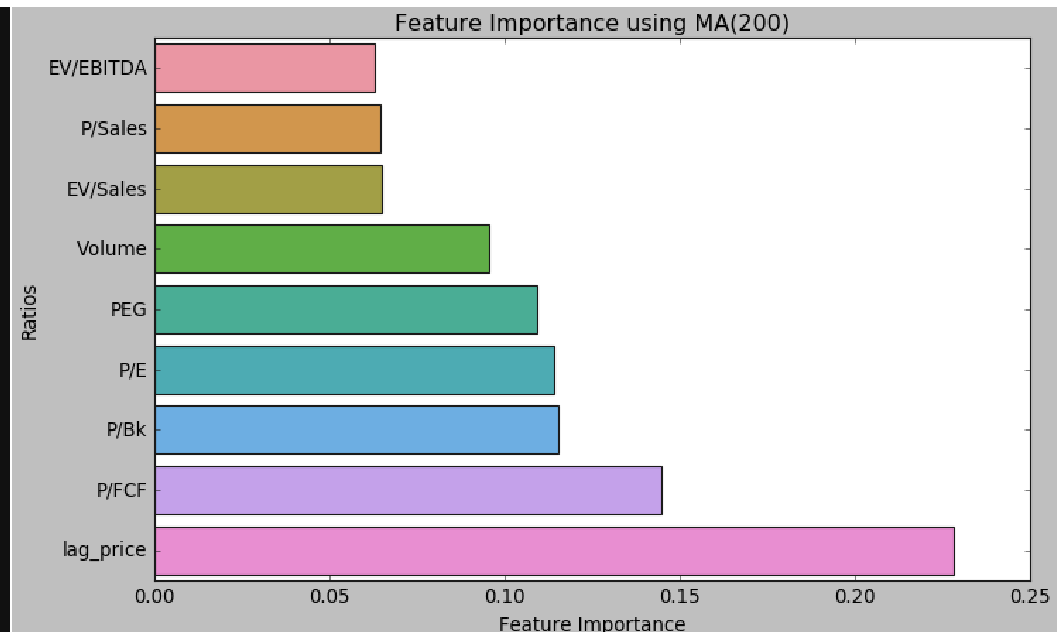
Features	Feature_importance
P/Sales	0.07044
EV/Sales	0.07547
EV/EBITDA	0.07637
P/Bk	0.08907
P/E	0.10267
Volume	0.1264
P/FCF	0.12882
PEG	0.12915
lag_price	0.20161



Accuracy = 0.96
Precision = 0.98
Recall = 0.98

The mean accuracy is: 0.976
The std. dev. of accuracy is: 0.012

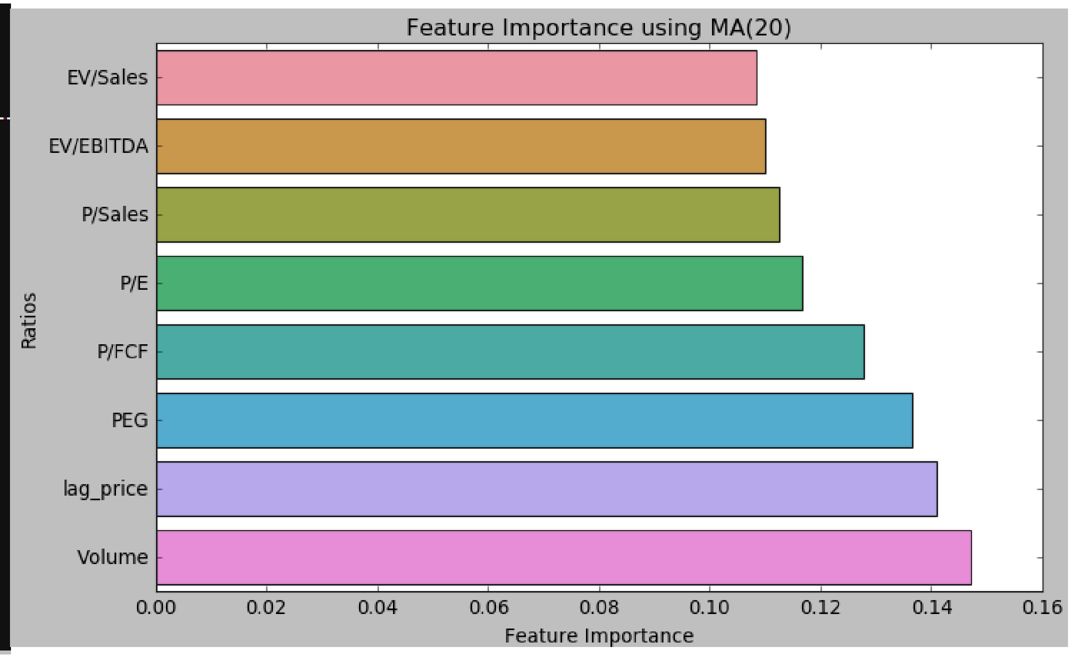
Features	Feature_importance
EV/EBITDA	0.06286
P/Sales	0.0646
EV/Sales	0.06502
Volume	0.0955
PEG	0.10952
P/E	0.11398
P/Bk	0.11544
P/FCF	0.14494
lag_price	0.22814



Appendix 4: Consumer Discretionary Sector Feature Importance, feature space has been rolled

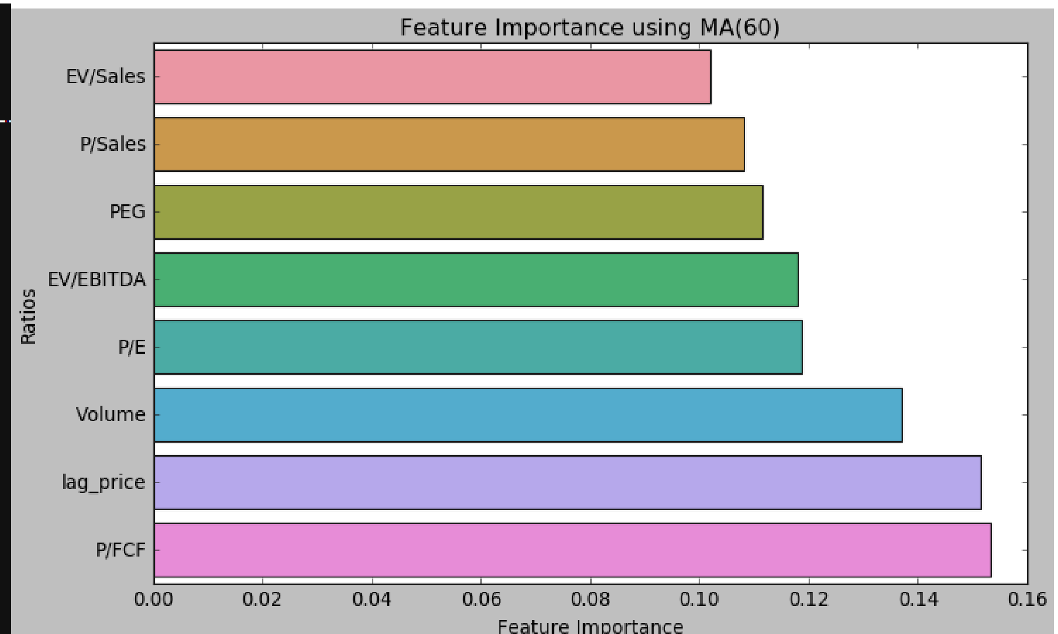
```
Accuracy = 0.87
Precision = 0.89
Recall = 0.91
------------------------------------------
The mean accuracy is:  0.854
The std. dev. of accuracy is:  0.023
```

Features	Feature_importance
EV/Sales	0.10853
EV/EBITDA	0.10991
P/Sales	0.11264
P/E	0.11664
P/FCF	0.12776
PEG	0.13653
lag_price	0.14088
Volume	0.14713




```
Accuracy = 0.92
Precision = 0.92
Recall = 0.97
------------------------------------------
The mean accuracy is:  0.931
The std. dev. of accuracy is:  0.004
```

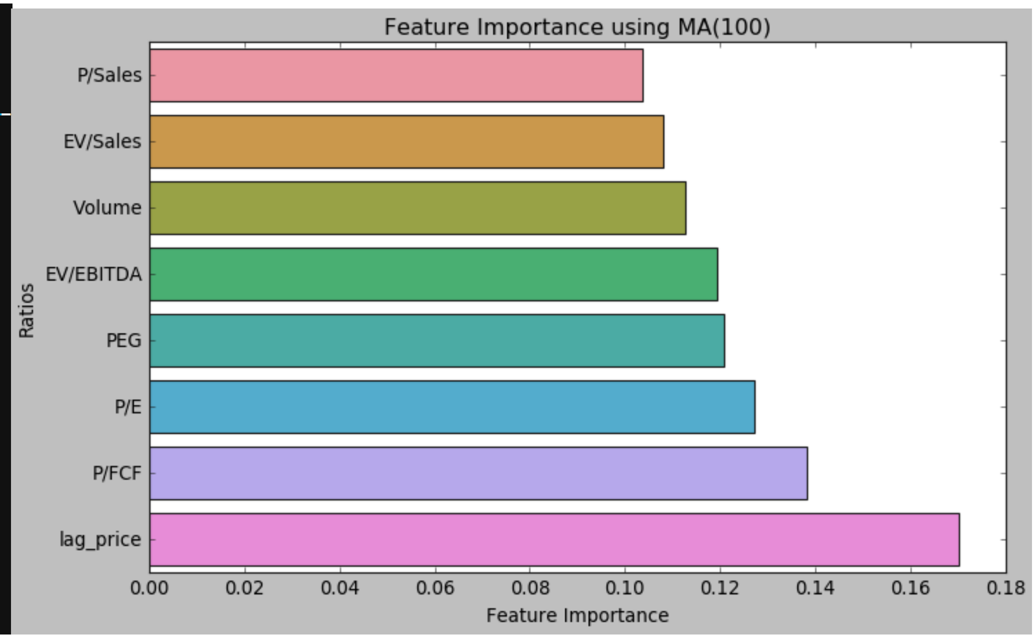
Features	Feature_importance
EV/Sales	0.10191
P/Sales	0.10806
PEG	0.11141
EV/EBITDA	0.11799
P/E	0.11863
Volume	0.13707
lag_price	0.15151
P/FCF	0.15341



Accuracy = 0.95
Precision = 0.96
Recall = 0.97

The mean accuracy is: 0.948
The std. dev. of accuracy is: 0.006

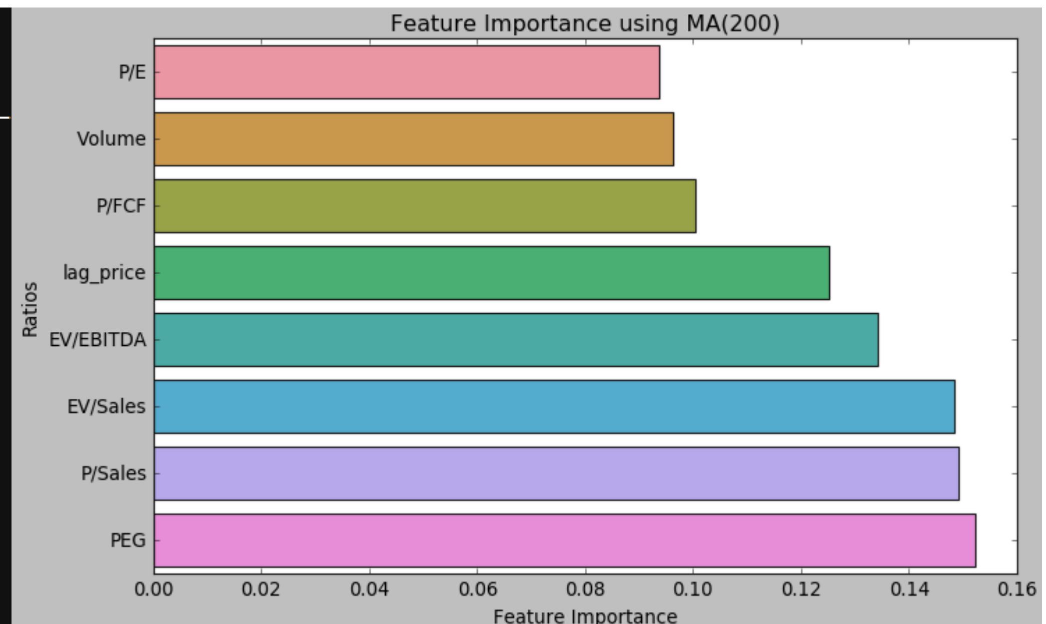
Features	Feature_importance
P/Sales	0.10377
EV/Sales	0.10798
Volume	0.11264
EV/EBITDA	0.11935
PEG	0.12089
P/E	0.12711
P/FCF	0.13809
lag_price	0.17017



Accuracy = 0.97
Precision = 0.98
Recall = 0.98

The mean accuracy is: 0.978
The std. dev. of accuracy is: 0.008

Features	Feature_importance
P/E	0.09382
Volume	0.0964
P/FCF	0.10047
lag_price	0.1252
EV/EBITDA	0.13425
EV/Sales	0.14848
P/Sales	0.1492
PEG	0.1522

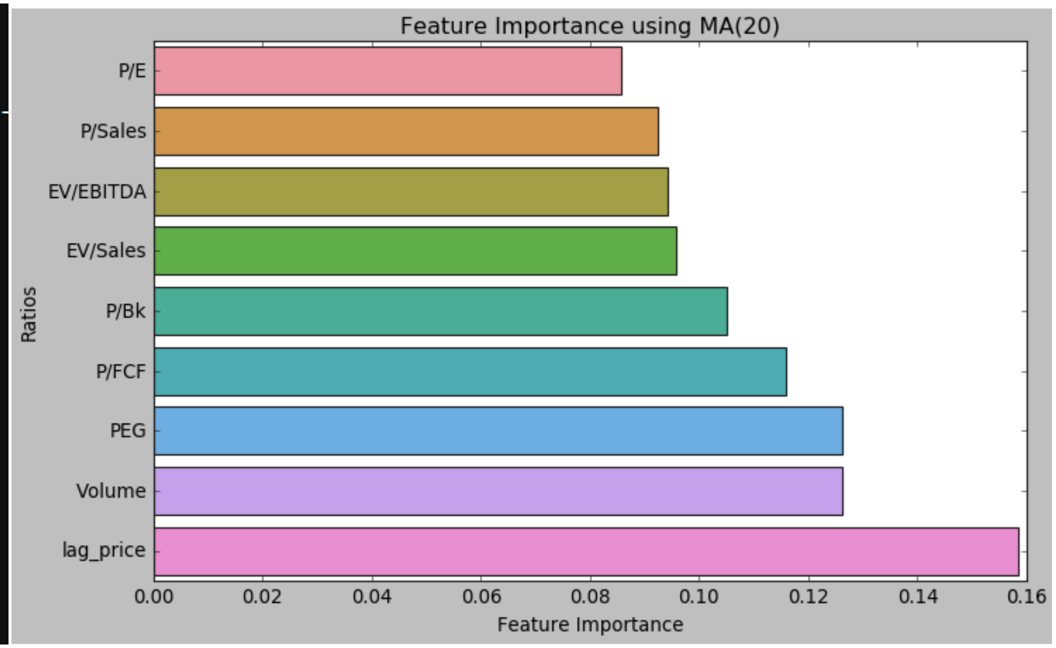


Appendix 5: Consumer Staples Sector Feature Importance, feature space has been rolled

Accuracy = 0.87
Precision = 0.89
Recall = 0.92

The mean accuracy is: 0.846
The std. dev. of accuracy is: 0.008

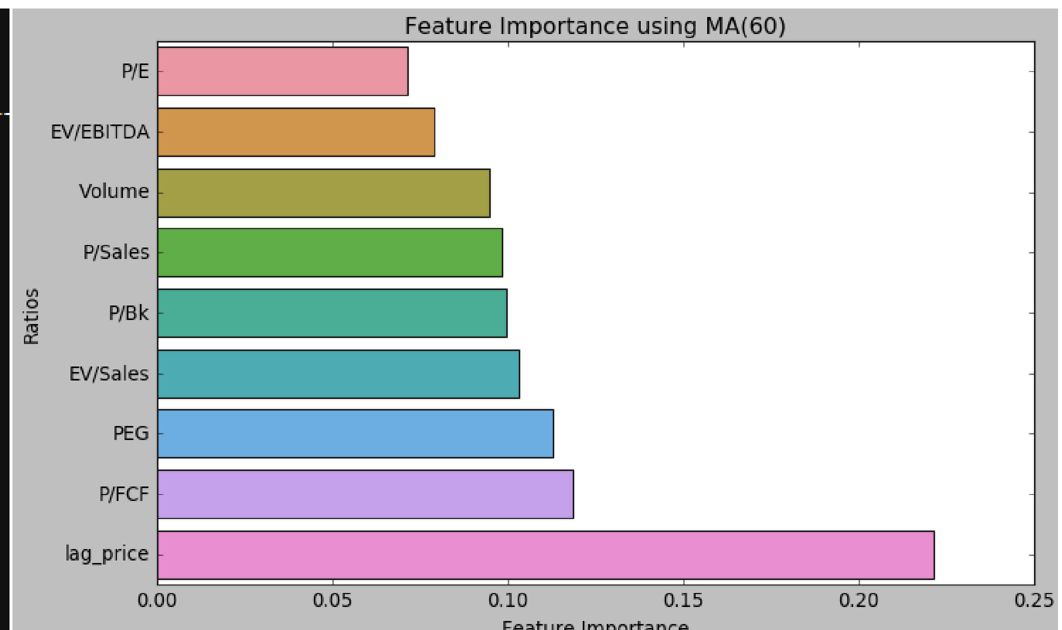
Features	Feature_importance
P/E	0.08568
P/Sales	0.09251
EV/EBITDA	0.09416
EV/Sales	0.09586
P/Bk	0.105
P/FCF	0.11596
PEG	0.12612
Volume	0.12615
lag_price	0.15856



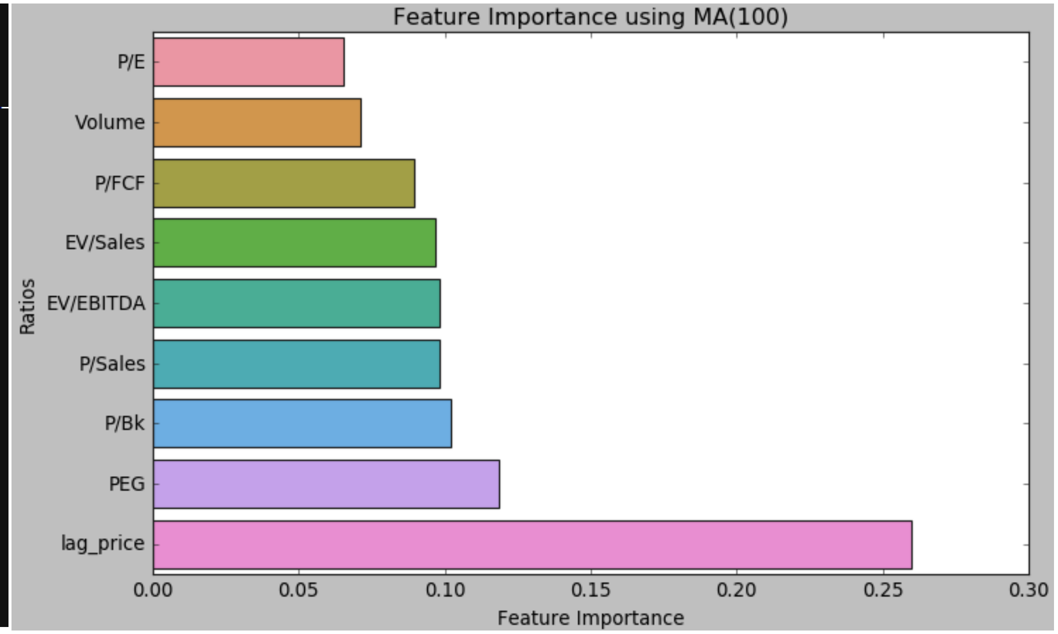
Accuracy = 0.93
Precision = 0.95
Recall = 0.96

The mean accuracy is: 0.927
The std. dev. of accuracy is: 0.006

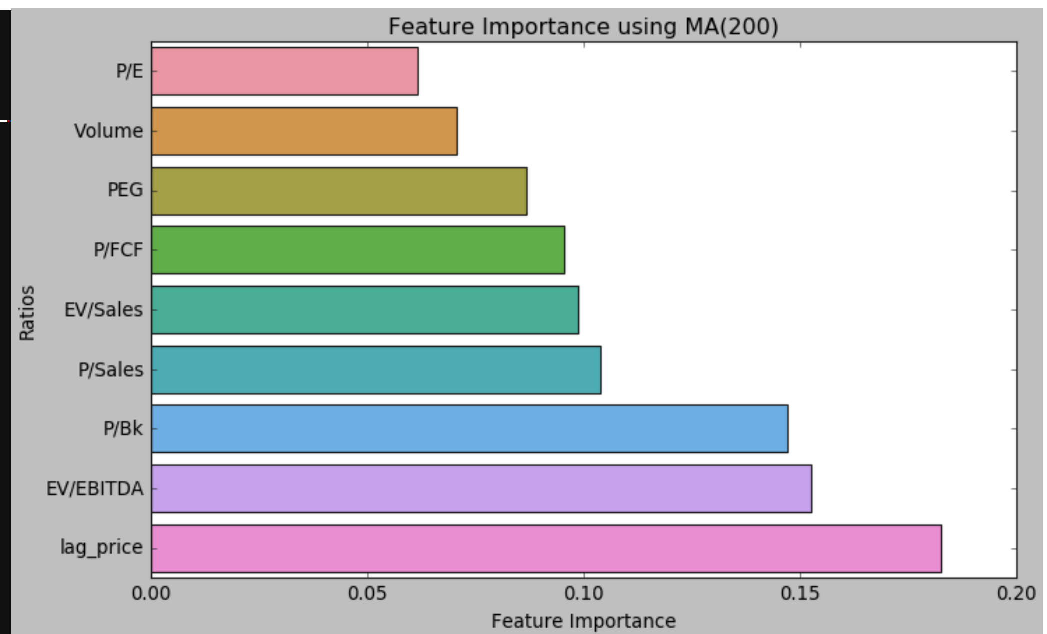
Features	Feature_importance
P/E	0.0715
EV/EBITDA	0.07906
Volume	0.09478
P/Sales	0.0986
P/Bk	0.09974
EV/Sales	0.10313
PEG	0.11295
P/FCF	0.11879
lag_price	0.22147



```
Accuracy = 0.95
Precision = 0.96
Recall = 0.98
-----------------------------------------
The mean accuracy is:  0.947
The std. dev. of accuracy is:  0.013


+------------+----------------------+
| Features   |   Feature_importance |
|------------+----------------------|
| P/E        |               0.0656 |
| Volume     |              0.07116 |
| P/FCF      |              0.08939 |
| EV/Sales   |              0.09662 |
| EV/EBITDA  |              0.09831 |
| P/Sales    |               0.0985 |
| P/Bk       |              0.10193 |
| PEG        |              0.11866 |
| lag_price  |              0.25985 |
+------------+----------------------+
```

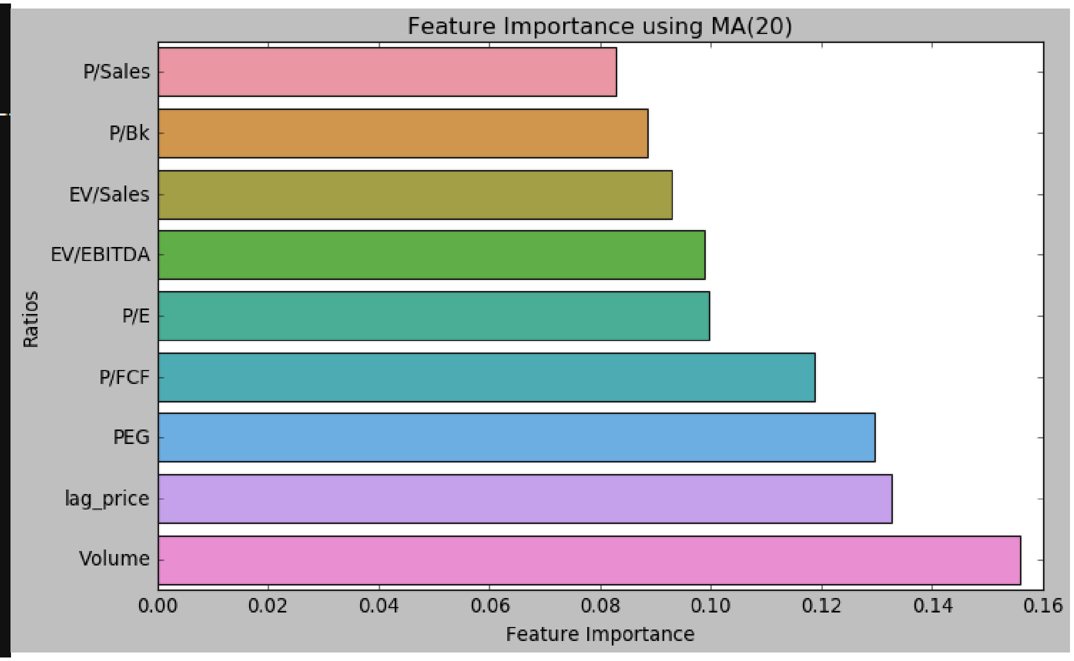



```
Accuracy = 0.97
Precision = 0.98
Recall = 0.99
-----------------------------------------
The mean accuracy is:  0.965
The std. dev. of accuracy is:  0.008


+------------+----------------------+
| Features   |   Feature_importance |
|------------+----------------------|
| P/E        |              0.06176 |
| Volume     |               0.0708 |
| PEG        |              0.08669 |
| P/FCF      |              0.09558 |
| EV/Sales   |              0.09876 |
| P/Sales    |              0.10392 |
| P/Bk       |              0.14729 |
| EV/EBITDA  |              0.15255 |
| lag_price  |              0.18265 |
+------------+----------------------+
```

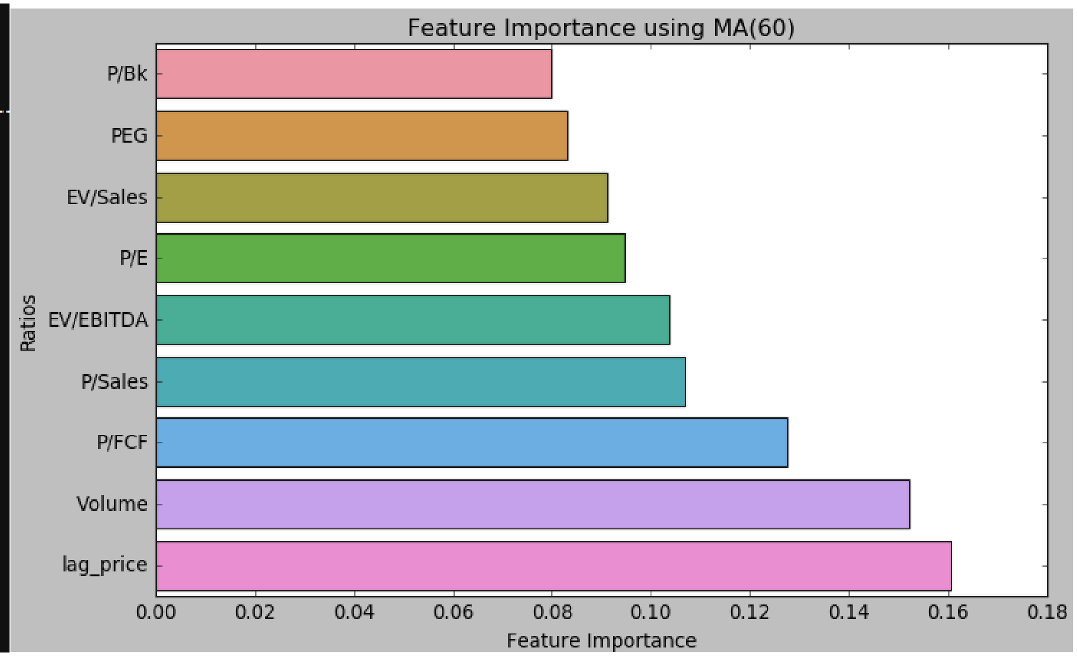


Appendix 6: Health Care Sector Feature Importance, feature space has been rolled

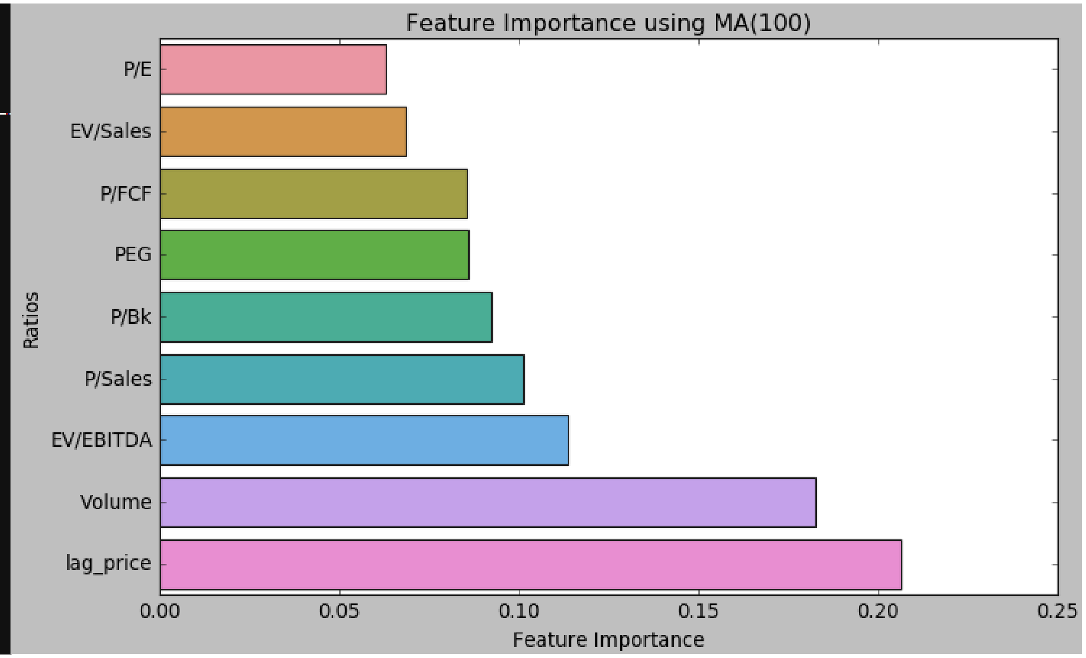
```
Accuracy = 0.88
Precision = 0.89
Recall = 0.92
----------------------------------------
The mean accuracy is:  0.85
The std. dev. of accuracy is:  0.018


+------------+----------------------+
| Features   |   Feature_importance |
|------------+----------------------|
| P/Sales    |              0.08298 |
| P/Bk       |              0.0885  |
| EV/Sales   |              0.09306 |
| EV/EBITDA  |              0.09875 |
| P/E        |              0.09975 |
| P/FCF      |              0.11865 |
| PEG        |              0.12965 |
| lag_price  |              0.13274 |
| Volume     |              0.15592 |
+------------+----------------------+
```



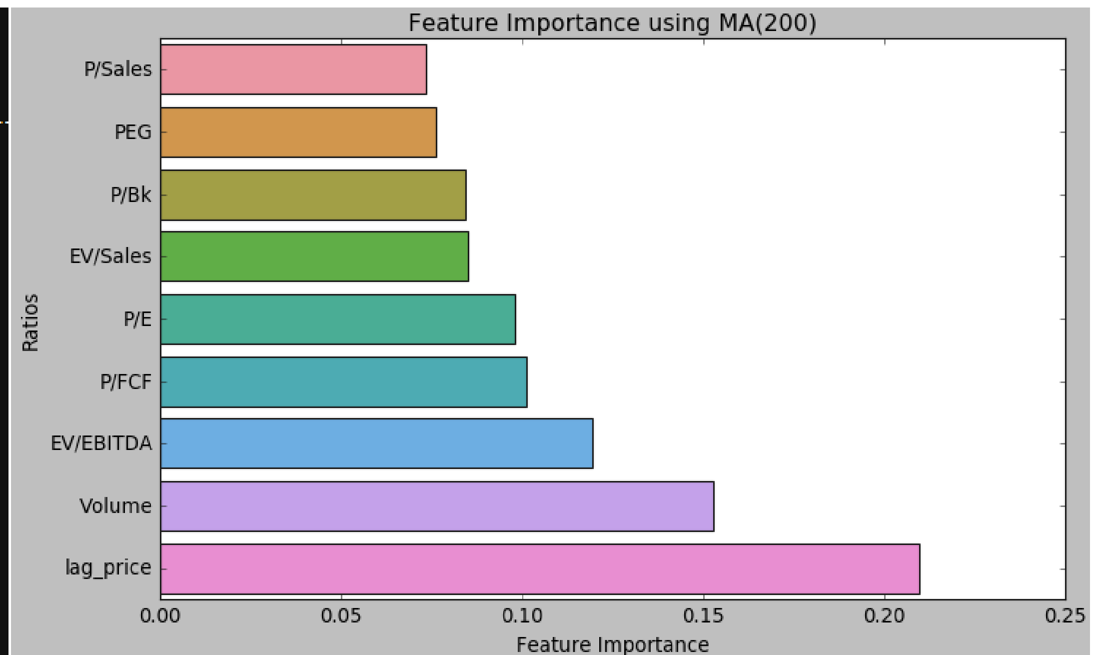

```
Accuracy = 0.94
Precision = 0.96
Recall = 0.97
----------------------------------------
The mean accuracy is:  0.929
The std. dev. of accuracy is:  0.009


+------------+----------------------+
| Features   |   Feature_importance |
|------------+----------------------|
| P/Bk       |              0.08001 |
| PEG        |              0.08297 |
| EV/Sales   |              0.09114 |
| P/E        |              0.09481 |
| EV/EBITDA  |              0.10378 |
| P/Sales    |              0.10674 |
| P/FCF      |              0.1276  |
| Volume     |              0.15222 |
| lag_price  |              0.16073 |
+------------+----------------------+
```



```
Accuracy = 0.95
Precision = 0.97
Recall = 0.96
---------------------------------------
The mean accuracy is:  0.954
The std. dev. of accuracy is:  0.007


+------------+---------------------+
| Features   |  Feature_importance |
|------------+---------------------|
| P/E        |             0.06295 |
| EV/Sales   |             0.06851 |
| P/FCF      |             0.08568 |
| PEG        |             0.08606 |
| P/Bk       |             0.09239 |
| P/Sales    |             0.10131 |
| EV/EBITDA  |             0.11374 |
| Volume     |             0.18283 |
| lag_price  |             0.20654 |
+------------+---------------------+
```




```
Accuracy = 0.95
Precision = 0.96
Recall = 0.98
---------------------------------------
The mean accuracy is:  0.955
The std. dev. of accuracy is:  0.006


+------------+---------------------+
| Features   |  Feature_importance |
|------------+---------------------|
| P/Sales    |             0.07344 |
| PEG        |             0.07606 |
| P/Bk       |             0.08433 |
| EV/Sales   |             0.08496 |
| P/E        |             0.09805 |
| P/FCF      |             0.10109 |
| EV/EBITDA  |             0.11935 |
| Volume     |             0.15304 |
| lag_price  |             0.20968 |
+------------+---------------------+
```

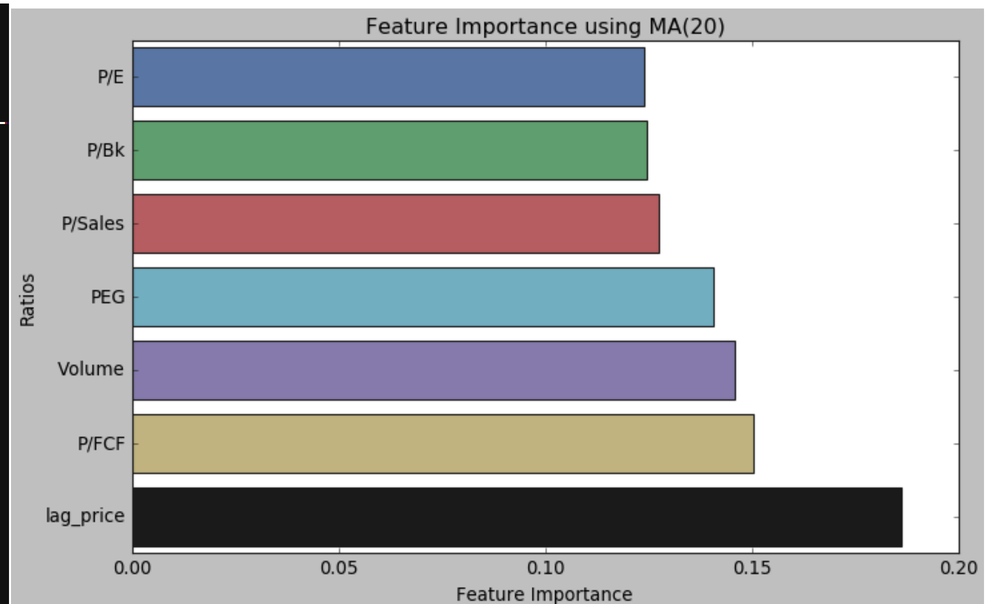


Appendix 7: Financials Sector Feature Importance, feature space has been rolled

Accuracy = 0.84
Precision = 0.88
Recall = 0.87

The mean accuracy is: 0.863
The std. dev. of accuracy is: 0.009

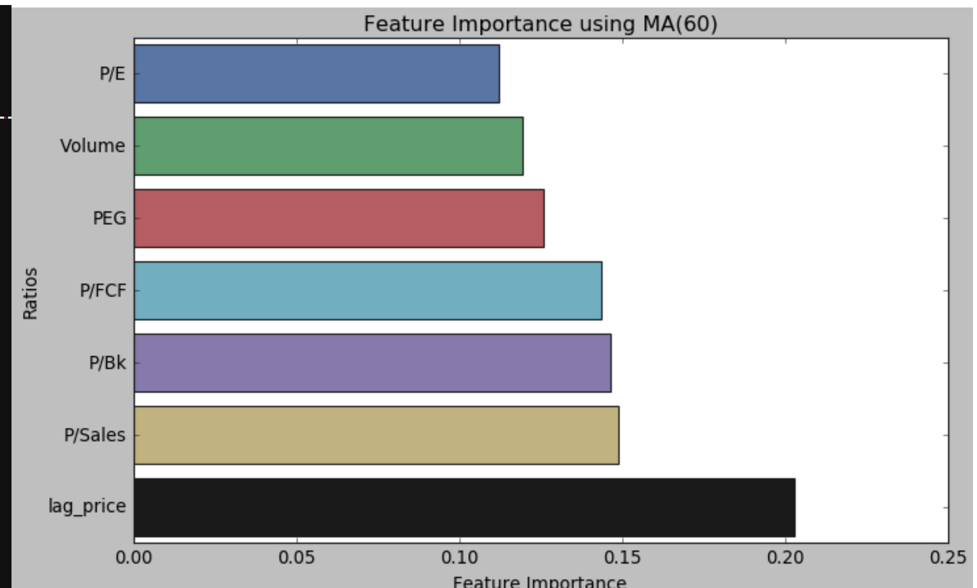
Features	Feature_importance
P/E	0.12407
P/Bk	0.12458
P/Sales	0.12771
PEG	0.14085
Volume	0.14593
P/FCF	0.15053
lag_price	0.18631



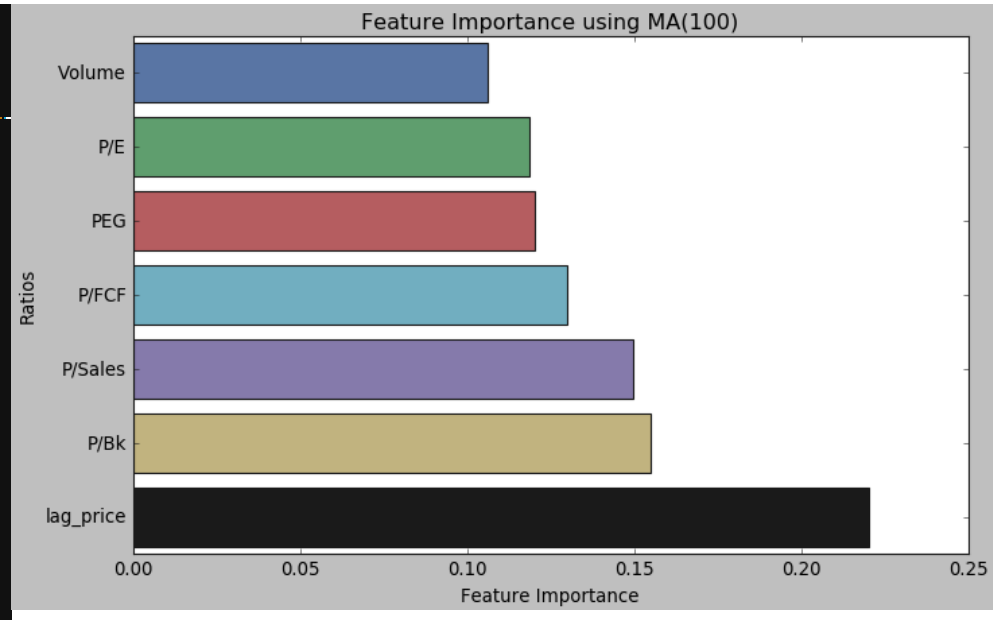
Accuracy = 0.93
Precision = 0.94
Recall = 0.96

The mean accuracy is: 0.93
The std. dev. of accuracy is: 0.007

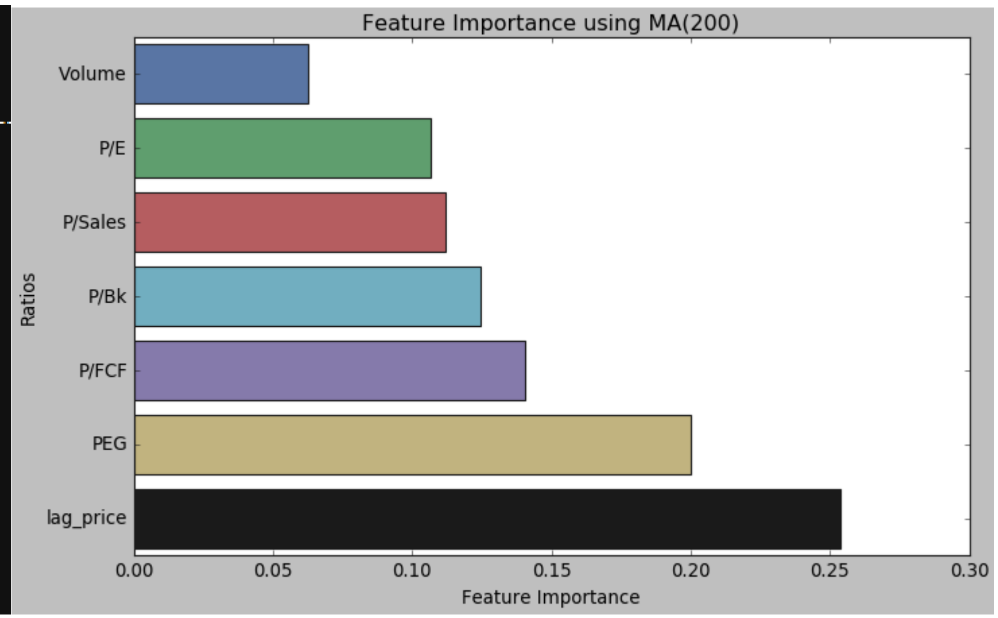
Features	Feature_importance
P/E	0.11228
Volume	0.11968
PEG	0.12582
P/FCF	0.14371
P/Bk	0.14676
P/Sales	0.14891
lag_price	0.20284



```
Accuracy = 0.95
Precision = 0.96
Recall = 0.98
---------------------------------------
The mean accuracy is:  0.93
The std. dev. of accuracy is:  0.006


+------------+----------------------+
| Features   |   Feature_importance |
|------------+----------------------|
| Volume     |              0.10632 |
| P/E        |              0.11856 |
| PEG        |              0.12025 |
| P/FCF      |              0.12994 |
| P/Sales    |              0.14973 |
| P/Bk       |              0.15498 |
| lag_price  |              0.22022 |
+------------+----------------------+
```

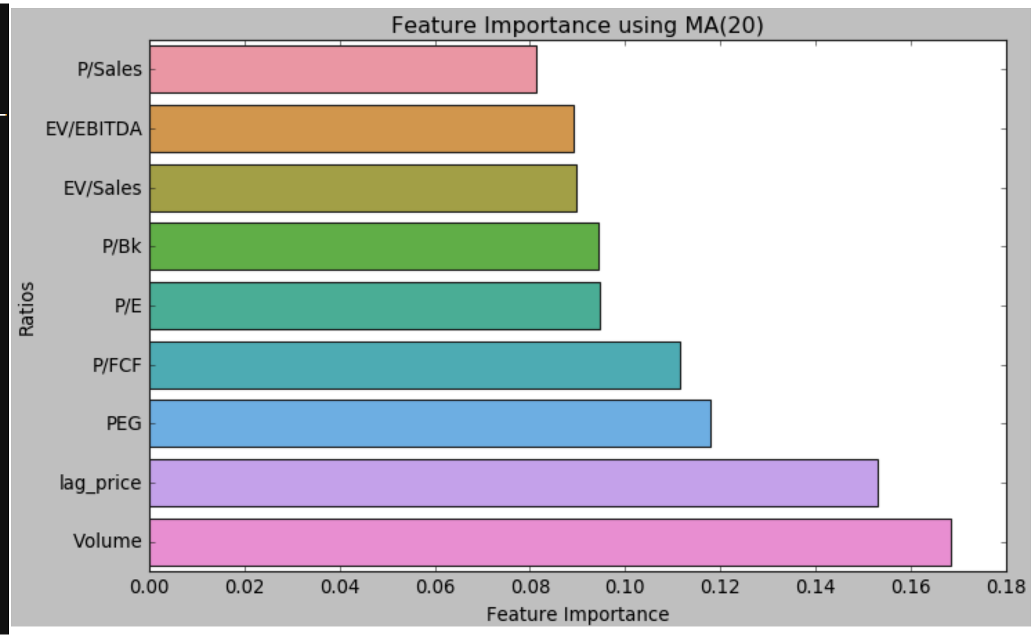



```
Accuracy = 0.97
Precision = 0.98
Recall = 0.98
---------------------------------------
The mean accuracy is:  0.972
The std. dev. of accuracy is:  0.006


+------------+----------------------+
| Features   |   Feature_importance |
|------------+----------------------|
| Volume     |              0.06269 |
| P/E        |              0.10664 |
| P/Sales    |              0.11188 |
| P/Bk       |              0.12463 |
| P/FCF      |              0.14065 |
| PEG        |              0.19985 |
| lag_price  |              0.25366 |
+------------+----------------------+
```

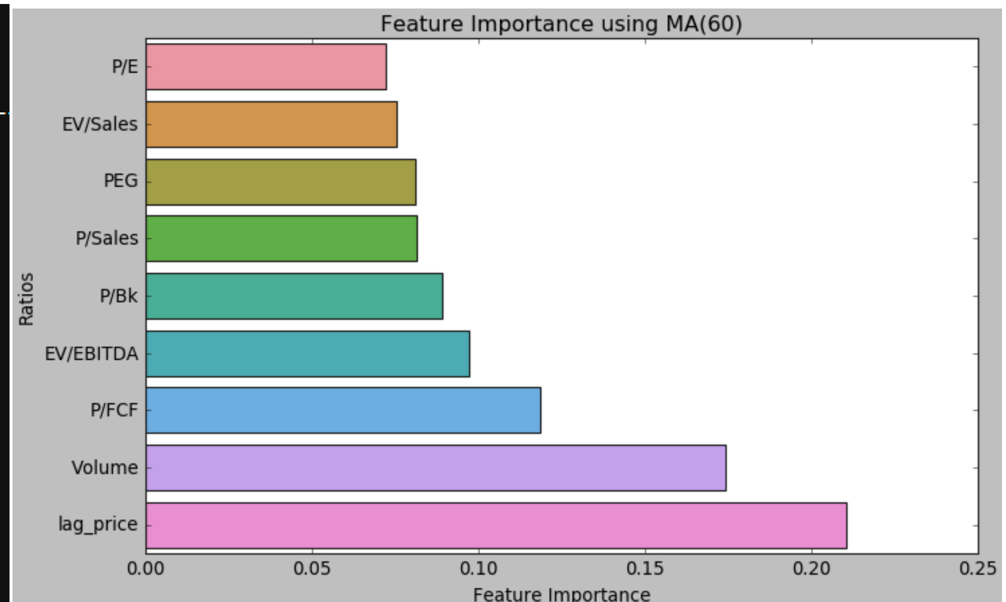


Appendix 8: Information Technology Sector Feature Importance, feature space has been rolled

```
Accuracy = 0.86
Precision = 0.88
Recall = 0.92
---------------------------------------
The mean accuracy is:  0.864
The std. dev. of accuracy is:  0.006


+------------+----------------------+
| Features   |   Feature_importance |
|------------+----------------------|
| P/Sales    |              0.08121 |
| EV/EBITDA  |              0.08927 |
| EV/Sales   |              0.08969 |
| P/Bk       |              0.09445 |
| P/E        |              0.09457 |
| P/FCF      |              0.11157 |
| PEG        |              0.11781 |
| lag_price  |              0.15292 |
| Volume     |              0.16852 |
+------------+----------------------+
```



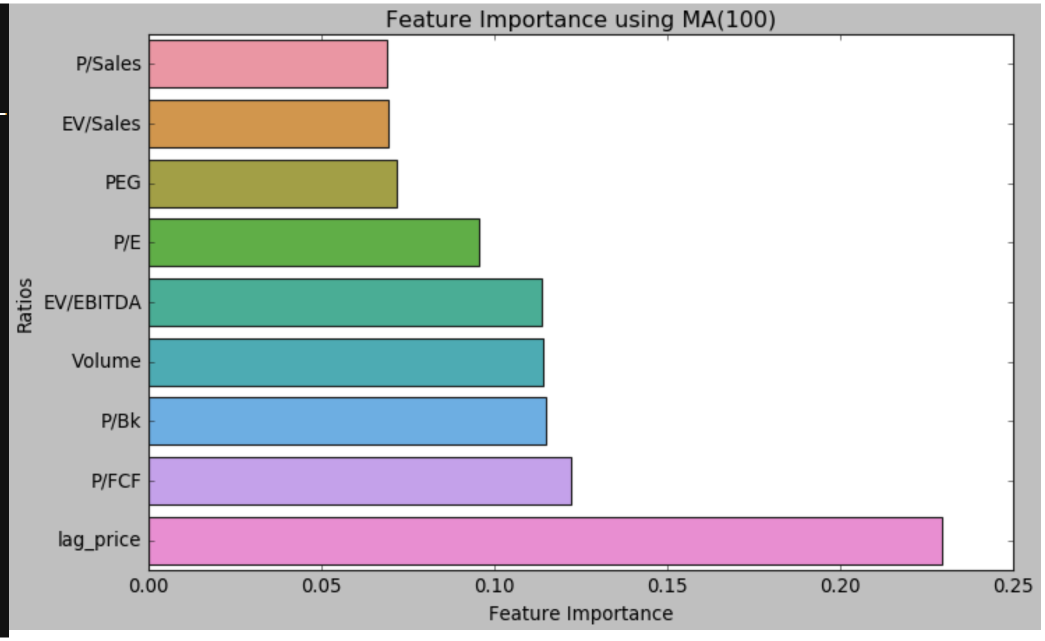

```
Accuracy = 0.91
Precision = 0.91
Recall = 0.97
---------------------------------------
The mean accuracy is:  0.924
The std. dev. of accuracy is:  0.015


+------------+----------------------+
| Features   |   Feature_importance |
|------------+----------------------|
| P/E        |              0.07206 |
| EV/Sales   |              0.07552 |
| PEG        |              0.08107 |
| P/Sales    |              0.08169 |
| P/Bk       |              0.08899 |
| EV/EBITDA  |              0.09713 |
| P/FCF      |              0.11864 |
| Volume     |              0.17424 |
| lag_price  |              0.21065 |
+------------+----------------------+
```



Accuracy = 0.95
Precision = 0.96
Recall = 0.97

The mean accuracy is: 0.951
The std. dev. of accuracy is: 0.009

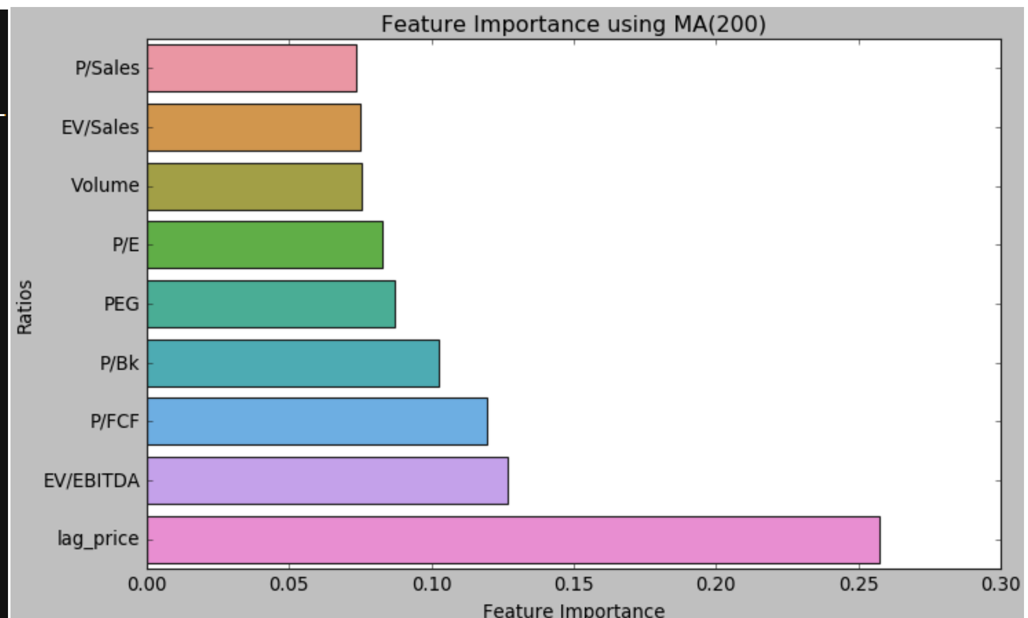
Features	Feature_importance
P/Sales	0.06896
EV/Sales	0.06946
PEG	0.07174
P/E	0.09546
EV/EBITDA	0.11359
Volume	0.11433
P/Bk	0.11499
P/FCF	0.12205
lag_price	0.22942



Accuracy = 0.97
Precision = 0.97
Recall = 1.0

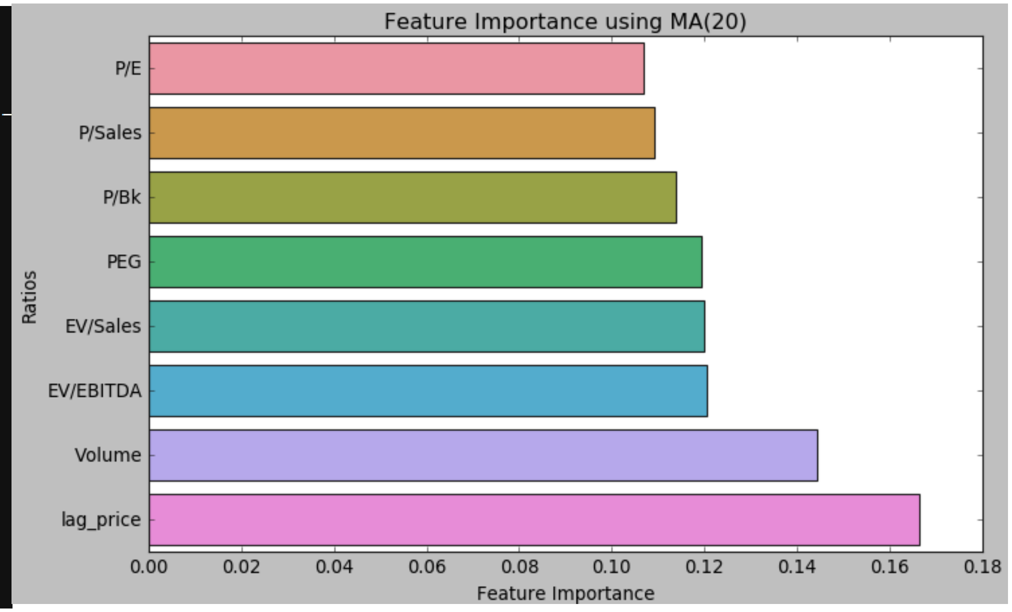
The mean accuracy is: 0.976
The std. dev. of accuracy is: 0.01

Features	Feature_importance
P/Sales	0.07353
EV/Sales	0.07497
Volume	0.07557
P/E	0.08277
PEG	0.08701
P/Bk	0.10258
P/FCF	0.11939
EV/EBITDA	0.12666
lag_price	0.25752

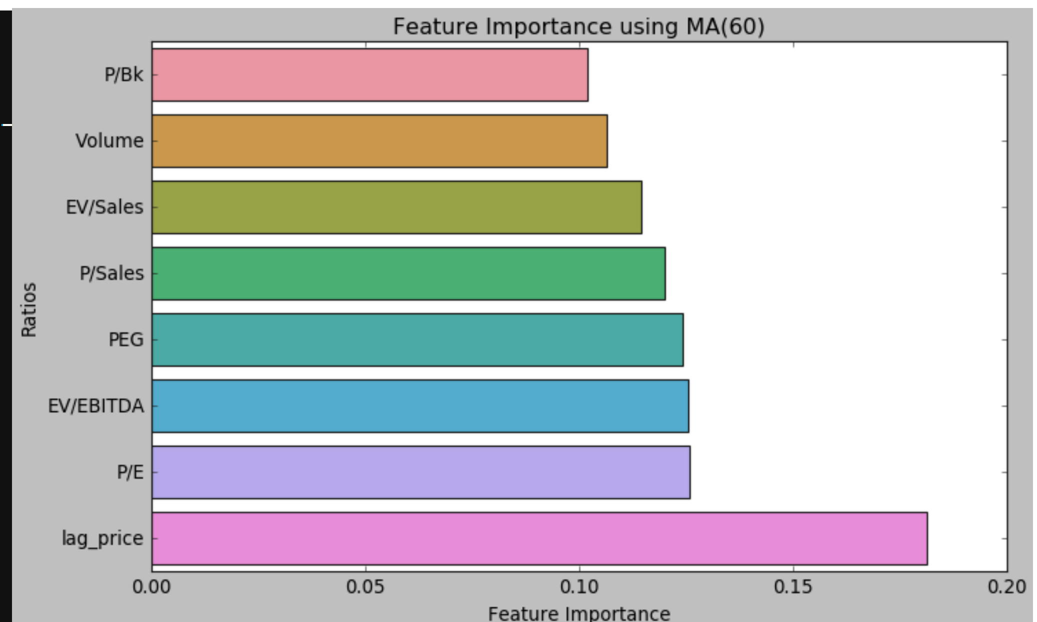


Appendix 9: Utilities Sector Feature Importance, feature space has been rolled

```
Accuracy = 0.87
Precision = 0.89
Recall = 0.91
----------------------------------------
The mean accuracy is:  0.831
The std. dev. of accuracy is:  0.013


+------------+----------------------+
| Features   |   Feature_importance |
|------------+----------------------|
| P/E        |              0.10692 |
| P/Sales    |              0.10911 |
| P/Bk       |              0.11371 |
| PEG        |              0.11925 |
| EV/Sales   |              0.12005 |
| EV/EBITDA  |              0.12039 |
| Volume     |              0.1442  |
| lag_price  |              0.16637 |
+------------+----------------------+
```



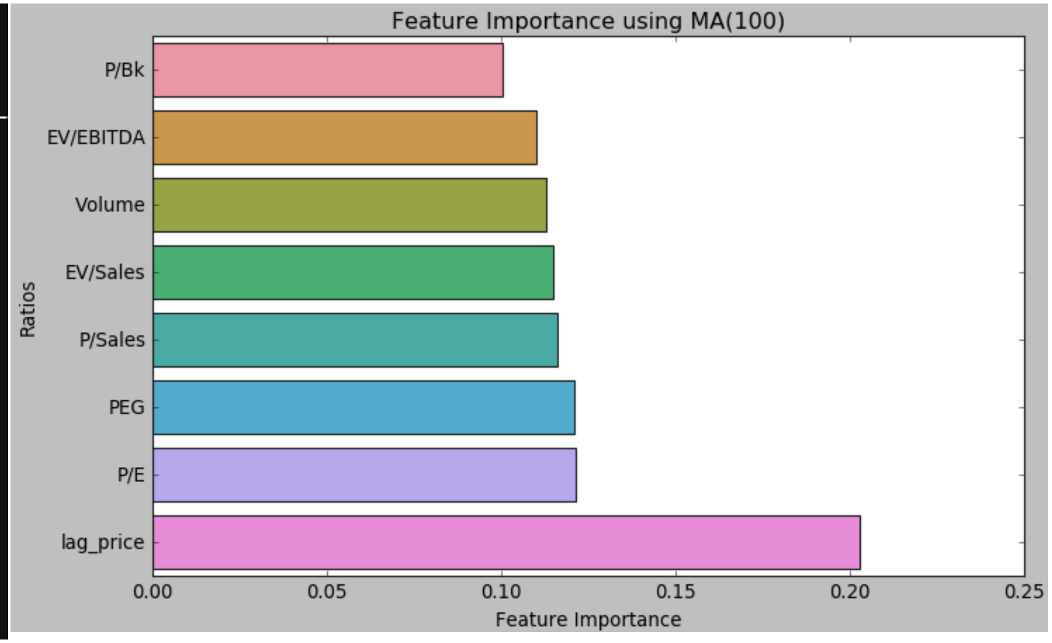

```
Accuracy = 0.93
Precision = 0.95
Recall = 0.96
----------------------------------------
The mean accuracy is:  0.921
The std. dev. of accuracy is:  0.012


+------------+----------------------+
| Features   |   Feature_importance |
|------------+----------------------|
| P/Bk       |              0.10205 |
| Volume     |              0.10644 |
| EV/Sales   |              0.11457 |
| P/Sales    |              0.11991 |
| PEG        |              0.12433 |
| EV/EBITDA  |              0.12543 |
| P/E        |              0.12595 |
| lag_price  |              0.18133 |
+------------+----------------------+
```



Accuracy = 0.93
Precision = 0.96
Recall = 0.95

The mean accuracy is: 0.946
The std. dev. of accuracy is: 0.01

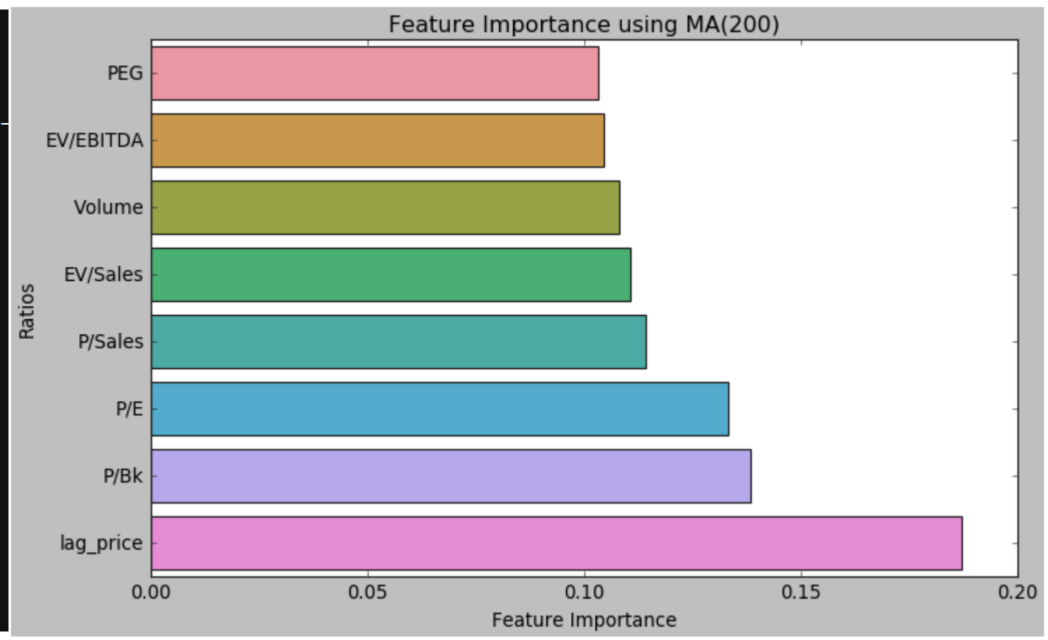
Features	Feature_importance
P/Bk	0.10025
EV/EBITDA	0.11009
Volume	0.11298
EV/Sales	0.11502
P/Sales	0.11618
PEG	0.12119
P/E	0.12146
lag_price	0.20283



Accuracy = 0.97
Precision = 0.98
Recall = 0.99

The mean accuracy is: 0.957
The std. dev. of accuracy is: 0.01

Features	Feature_importance
PEG	0.10333
EV/EBITDA	0.1045
Volume	0.10818
EV/Sales	0.1106
P/Sales	0.11424
P/E	0.1334
P/Bk	0.13849
lag_price	0.18727

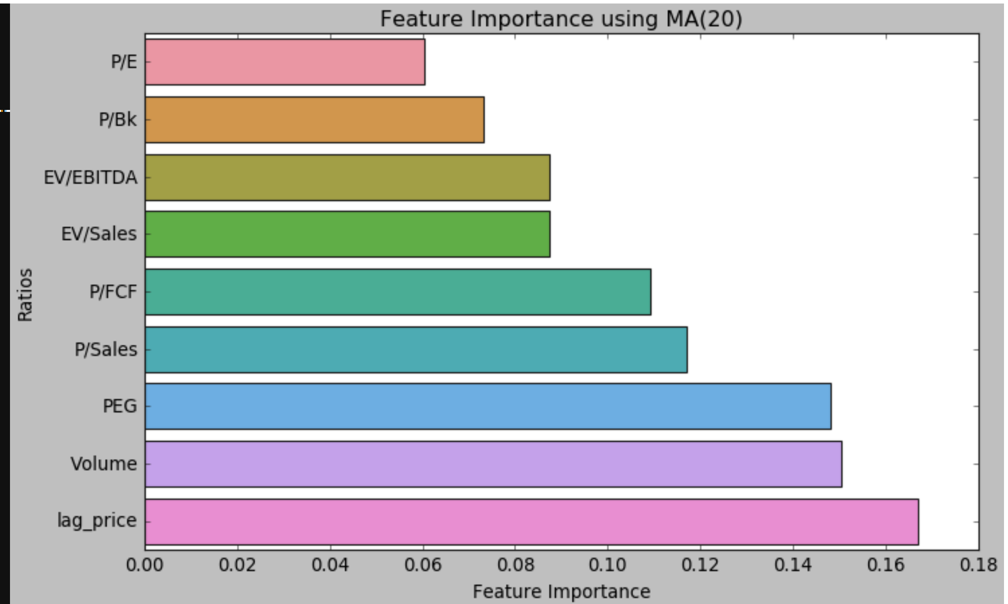


Appendix 10: Telecommunication Sector Feature Importance, feature space has been rolled

Accuracy = 0.87
Precision = 0.89
Recall = 0.9

The mean accuracy is: 0.895
The std. dev. of accuracy is: 0.025

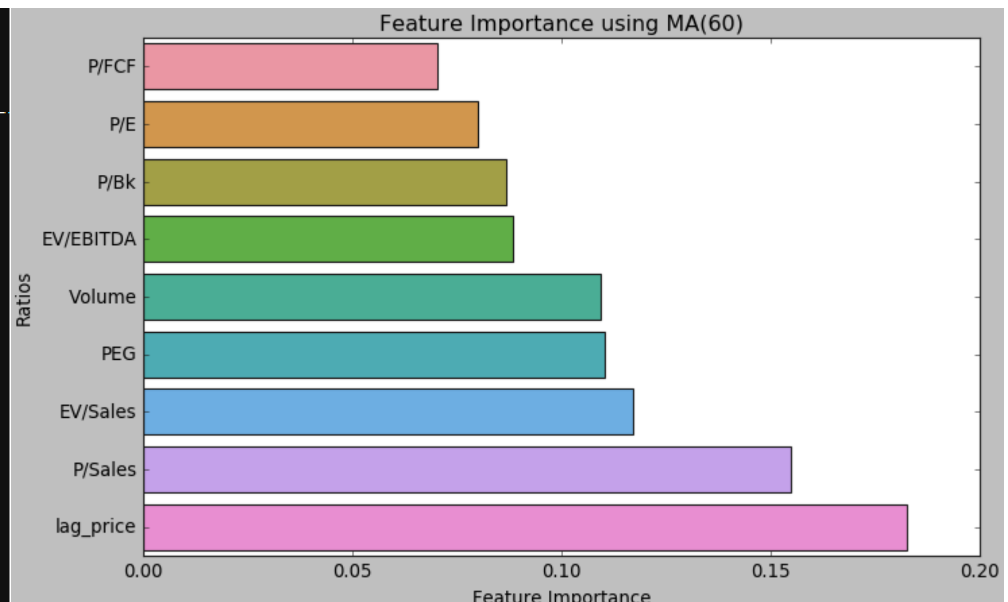
Features	Feature_importance
P/E	0.06047
P/Bk	0.07307
EV/EBITDA	0.08731
EV/Sales	0.08741
P/FCF	0.10927
P/Sales	0.11706
PEG	0.14796
Volume	0.15051
lag_price	0.16694



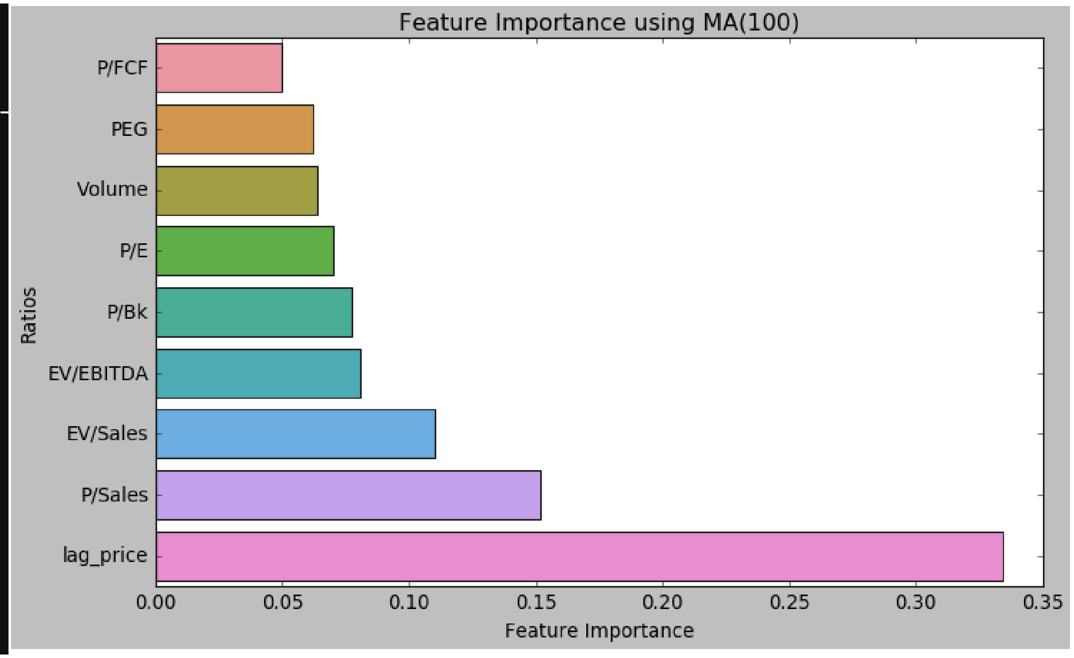
Accuracy = 0.95
Precision = 1.0
Recall = 0.92

The mean accuracy is: 0.929
The std. dev. of accuracy is: 0.039

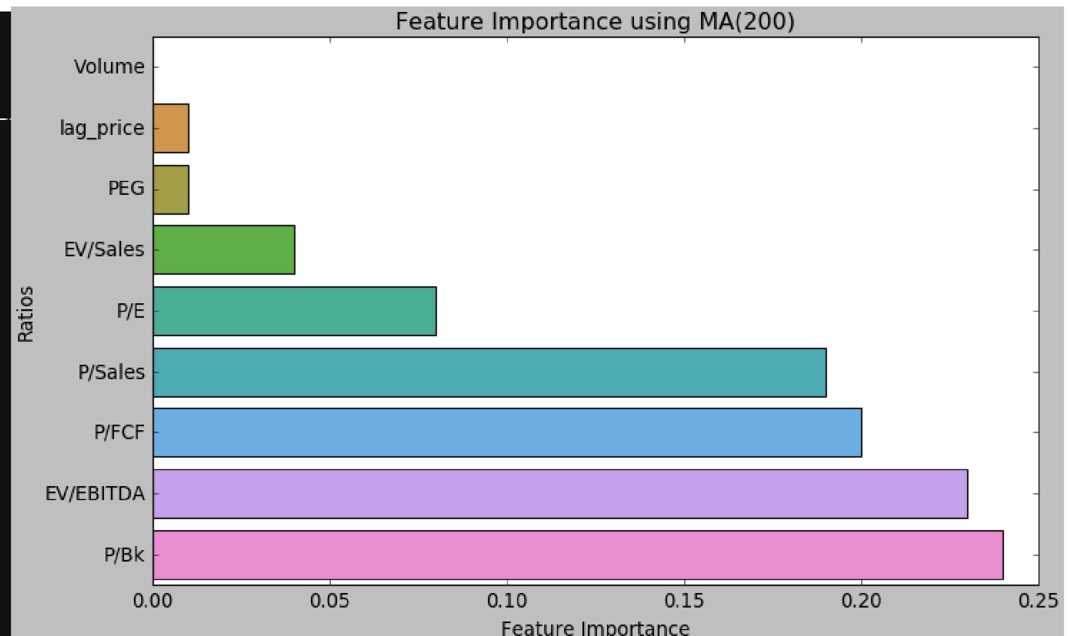
Features	Feature_importance
P/FCF	0.07028
P/E	0.0802
P/Bk	0.08671
EV/EBITDA	0.08846
Volume	0.10931
PEG	0.1105
EV/Sales	0.11702
P/Sales	0.15502
lag_price	0.18251



```
Accuracy = 0.96
Precision = 0.94
Recall = 1.0
-----------------------------------------
The mean accuracy is:  0.96
The std. dev. of accuracy is:  0.021


+------------+---------------------+
| Features   |  Feature_importance |
|------------+---------------------|
| P/FCF      |             0.04976 |
| PEG        |             0.06242 |
| Volume     |             0.06365 |
| P/E        |             0.07033 |
| P/Bk       |             0.07737 |
| EV/EBITDA  |             0.08065 |
| EV/Sales   |             0.10997 |
| P/Sales    |             0.15188 |
| lag_price  |             0.33398 |
+------------+---------------------+
```




```
Accuracy = 0.96
Precision = 1.0
Recall = 0.96
-----------------------------------------
The mean accuracy is:  0.985
The std. dev. of accuracy is:  0.012


+------------+---------------------+
| Features   |  Feature_importance |
|------------+---------------------|
| Volume     |              0      |
| lag_price  |              0.01   |
| PEG        |              0.01   |
| EV/Sales   |              0.04   |
| P/E        |              0.08   |
| P/Sales    |              0.19   |
| P/FCF      |              0.2    |
| EV/EBITDA  |              0.23   |
| P/Bk       |              0.24   |
+------------+---------------------+
```

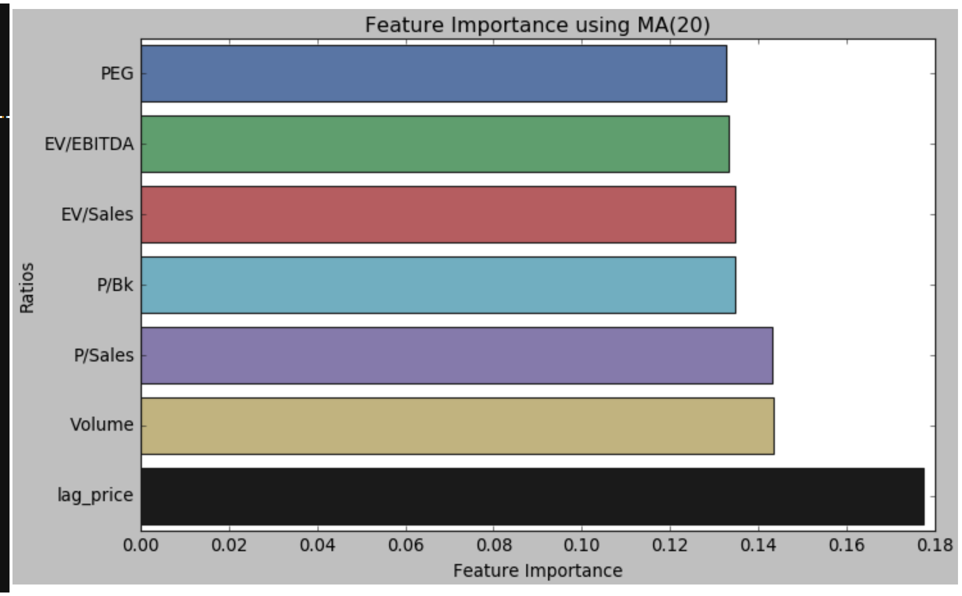


Appendix 11: Real Estate Sector Feature Importance, feature space has been rolled

Accuracy = 0.86
Precision = 0.88
Recall = 0.87

The mean accuracy is: 0.861
The std. dev. of accuracy is: 0.016

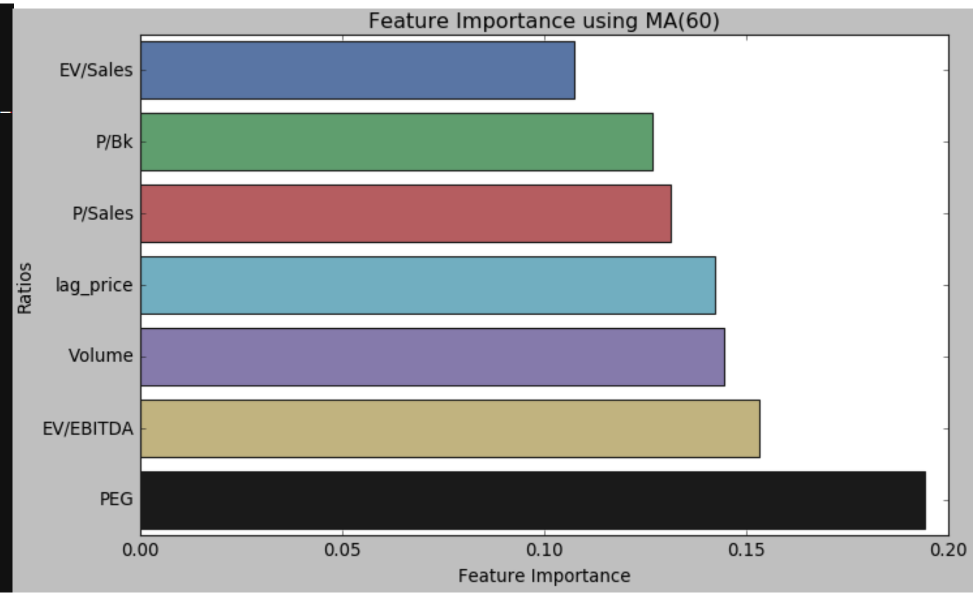
Features	Feature_importance
PEG	0.13277
EV/EBITDA	0.1333
EV/Sales	0.13487
P/Bk	0.13489
P/Sales	0.14309
Volume	0.14353
lag_price	0.17756



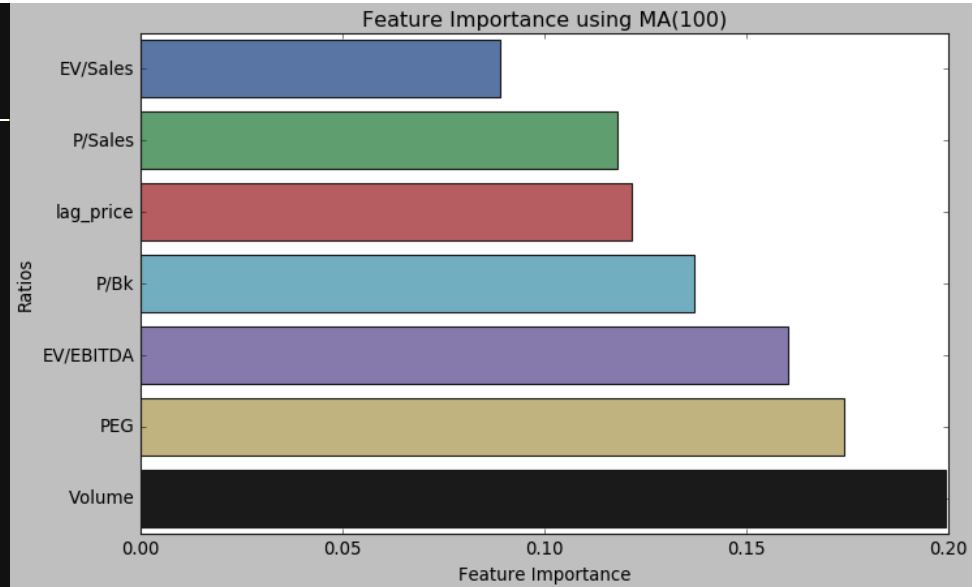
Accuracy = 0.92
Precision = 0.93
Recall = 0.96

The mean accuracy is: 0.913
The std. dev. of accuracy is: 0.016

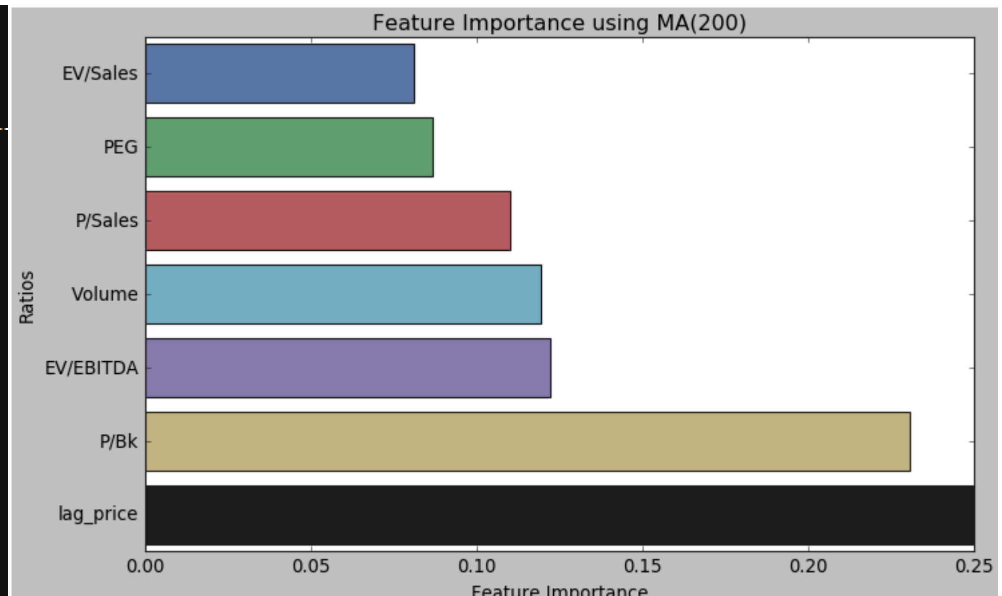
Features	Feature_importance
EV/Sales	0.1075
P/Bk	0.1269
P/Sales	0.13119
lag_price	0.14228
Volume	0.14454
EV/EBITDA	0.15325
PEG	0.19433



```
Accuracy = 0.96
Precision = 0.97
Recall = 0.98
----------------------------------------
The mean accuracy is:  0.94
The std. dev. of accuracy is:  0.022


+------------+----------------------+
| Features   |   Feature_importance |
|------------+----------------------|
| EV/Sales   |              0.08915 |
| P/Sales    |              0.11822 |
| lag_price  |              0.12157 |
| P/Bk       |              0.1371  |
| EV/EBITDA  |              0.1604  |
| PEG        |              0.17411 |
| Volume     |              0.19945 |
+------------+----------------------+
```




```
Accuracy = 0.97
Precision = 0.97
Recall = 0.98
----------------------------------------
The mean accuracy is:  0.959
The std. dev. of accuracy is:  0.016


+------------+----------------------+
| Features   |   Feature_importance |
|------------+----------------------|
| EV/Sales   |              0.08108 |
| PEG        |              0.0867  |
| P/Sales    |              0.11019 |
| Volume     |              0.11946 |
| EV/EBITDA  |              0.12241 |
| P/Bk       |              0.23053 |
| lag_price  |              0.24963 |
+------------+----------------------+
```



REFERENCES

Beaver, W. H. (1966). Financial Ratios as Predictors of Failure. *Journal of Accounting Research*, 77-111.

Research, M. E. (2020). *Energy Select Sector SPDR Fun*. Retrieved from ETFdb.com: https://etfdb.com/etf/XLE/#etf-ticker-profile

TILGNER, M. (2019, September 25). *Time series forecasting with random forest*. Retrieved from STATWORX Blog: https://www.statworx.com/at/blog/time-series-forecasting-with-random-forest/