One notable result is that the SVM model is only tenths of a percentage point better than simple random guessing when it comes to predicting price direction one day ahead ($m=1$). This has several important implications. First, it strongly reinforces the Efficient Markets Hypothesis. If a model that incorporates several types of historical data, including some features like momentum that economists have demonstrated are present in stock price data, is not able to do better than a coin flip when it comes to predicting the next day's price direction, then this is very strong evidence that prices follow a random walk. Prices that already reflect available information will change only based on new information, so tomorrow's price direction will be dependent only on new information that arrives tomorrow. A model like ours which analyzes only historical data should not be able to predict the price direction, as all this historical data should already be incorporated into the price. Therefore, our model's difficulty in predicting the next day's stock price supports EMH.

Several trends emerge as we increase the parameter $m$. The most striking is that the mean and median returns increase when $m=5,10,20$ but then decrease slightly when $m=90$ and $m=270$. Figure \ref{fig: meanvsm} shows the mean prediction accuracy against the parameter $m$, where mean prediction accuracy here is the mean of the mean accuracies for all 25 combinations of $n\_1, n\_2$ with a fixed $m$, where each combination itself reports the mean accuracy across the 34 stocks.

This number does not tell the whole story, however. Looking at the full results in \ref{sec: appendix}, we see that when $m$ is small, varying $n\_1$ and $n\_2$ has little effect. For example, when $m=1$ mean accuracy is within 49.5\% and 50.53\% for every combination of $n\_1, n\_2$. It is interesting, however, to note that for prediction when $m=1$ very small values or very large values of $n\_1, n\_2$ work best. Mean and median accuracy are highest when at least one of the two is 5, or when both are 90 or 270. This implies that very short-term trends, or long-term trends, are best for predicting the next day's price direction. Trends across two weeks or a month, however, perform worse than simple random guessing.

The parameters $n\_1$ and $n\_2$ start to become more important as we increase $m$. When $m=10$ the mean prediction accuracy varies between 53.3\% and 56.8\%, a much larger range than at $m=1$. This disparity is exaggerated when $m=90$, that is, when we try to predict the price direction across the next quarter. Some combinations of $n\_1, n\_2$, such as $n\_1=10, n\_2=90$ actually result in less than 50\% accuracy, which means one would be better off flipping a coin. However, other combinations have very high accuracies. For example, $n\_1 = 270, n\_2 = 5$ results in a 61.5\% prediction accuracy.

Such discrepancies indicate that as the model tries to forecast farther into the future, the input parameters and historical data have drastically more influence. This again reinforces the EMH and the idea of stocks as a random walk. Regardless of the historical data used, short-term changes are difficult to predict, indicating that trends do not matter much in the short run, but may still offer some predictive ability. Long-term changes, however, are subject to seasonality and the trends discussed in \ref{subsec: stock}. With the right input data, the model can take advantage of these trends. If the training data reflects similar conditions to those in the test dataset, the model will have high predictive power. However, the wrong training dataset can skew the model and even lead to worse than 50\% prediction accuracy. Figure \ref{fig: meansforquarter} shows the prediction accuracies for each combination of $n\_1, n\_2$ when $m=90$.

Another noticeable trend is that in general, $n\_1 = 270$ results in the highest prediction accuracy for all $m$. That is, the overall sector’s historical data over the past year tends to be helpful in predicting the price direction at any point in the future. It is more helpful as we try to predict further into the future, but even in the short-term, it appears that the general sector trend is a better predictor than the short-term sector trend.

Almost the opposite is true for the parameter $n\_2$. Mean prediction accuracy tends to be highest when $n\_2 = 5$, the smallest value, but this is affected by parameter $m$. When $m$ is small too, choosing $n\_2=5$ has most effect; when $m$ is small it still performs better than other values of $n\_2$, but not as much. This makes intuitive sense; a stock’s recent momentum is better than other indicators since it reflects most recent information. However, recent momentum is not as helpful when trying to predict farther into the future.

When $m$ is small, the choice of parameter $n\_2$ has more effect than choice of parameter $n\_1$, and here performance is best when $n\_2 = 5$. However, as $m$ increases, the choice of parameter $n\_1$ becomes more important. Again, this makes intuitive sense. When trying to predict price direction in the short term, the particular stock’s momentum can provide some indication (but not much). However for predicting long-term price direction, the long-term sector trends are most helpful. Long-term sector trends reflect the market’s overall macro-conditions, and are not swayed by small fluctuations. The particular stock’s data, however, reflects only idiosyncratic conditions. This indicates that predicting long-term price direction for a particular stock might depend more on the overall market’s direction, and larger macro-conditions, than on the stock’s own trends. Once again, this supports EMH. If long-term stock direction were predictable based solely on that stock itself, this would mean prices react not only to new information but also to existing information, violating EMH.

It is also important to note that for short-term periods, simply being able to somewhat accurately predict tomorrow's price direction does not necessarily translate into trading profits. Treating the problem as binary classification allows us to more easily construct models, but it also means that we cannot predict the magnitude of the price change. Stock transactions have transaction costs, and the profit gained from small trades might not outweigh the costs. Even though we did not find predictive ability for the next day, we were able to find some predictive ability within the next week or month. Nevertheless, translating this small predictive advantage into trading profits is entirely another issue.

Finally, we note that the long-term predictive ability also may not directly translate to long-term profits, although not due to trading costs. When $m=90, 270$ the overall mean predictive accuracy is less than in the short term, but with the right choice of parameters $n\_1, n\_2$ the accuracy can be higher than 60\%. This number, however, is itself an average of the prediction accuracies for each of the 34. As discussed above and seen in \ref{sec: appendix}, the range of prediction for stocks increases as we increase $m$, and becomes very large when $m=90, 270$. We see that the model is able to predict price direction for some stocks with greater than 80\% accuracy, but for others cannot predict with more than 30\% accuracy. The problem is that we do not yet know ahead of time which stocks the model will be able to predict accurately and which it will not, so profiting off the model is still difficult without more experimentation.

**FUTURE WORK**

\subsec{Adding Granularity}

One limitation of this study is that we only looked at daily stock price data. As a result, our momentum and volatility parameters were calculated over weeks. This is different from actual hedge funds and quantitative trading institutions, which study price data at a far more granular level, on the order of minutes or even seconds. We can further this study by looking at intra-day trading data in addition to closing price data. By observing intra-day trends, we can create more robust models that capitalize on sudden changes in momentum during intra-day trading. Individual stocks often have price swings that last only for a couple minutes, which an intra-day model can analyze, and then use to predict price direction in the next few seconds or minutes. This is how such financial institutions make trading profits.

\subsec{Feature Selection}

Feature selection often has the biggest impact on a machine learning model’s accuracy. Another area for future work is to add to our feature list. Here we have looked at price volatility and momentum for the particular stock and for the technology sector. Future work would involve adding features related to the specific company and related to broader macroeconomic factors. Features related to the specific company include its Price/Earnings ratio, its market cap, working capital ratio, cost-of-capital, etc. Features related to broader factors include interest rate, inflation, GDP growth rate, etc.

\subsec{Analyzing Other Sectors and Companies}

Another area for future growth is to apply our model to stocks in other sectors. In this study we have only focused on the technology sector, specifically 34 mature technology companies. We can apply our methods to other sectors such as healthcare, retail, service, etc. We can also apply our methods to mid-size and startup companies, to see how the size of a company affects our model’s predictive ability.