So Just how value is value? (and is value the value you value…and by value I mean value not valuation value)

At its core, the asset manager’s job is fairly straightforward. We buy low…and Sell high. “So easy a caveman can do it,” to quote the famous insurance commercial. Afterall, the odds are in our favor. The first passive investment instrument hit the market in 1976. Since then, the Vanguard 500 index fund averages an 11.09% annual return. In fact, if Mom Dad would have put $1,000 in that fund for me on the day I was born, I would be about $90,713.66 wealthier…thanks Mom and Dad. Obviously, things are never that simple. Let’s say Mom and Dad did put $1000 in that fund for me on my birthday. Who is to say my parents would have been able to withstand the ups and downs that come with stock market investing? Who is to say I wouldn’t have been content with taking my $14,639 (Pre tax end of day value) on my 18th birthday and buying a really nice stereo, or a new Ford Escort? Or what if Mom and Dad gave it to one of their many “Broker” friends instead (I went to a private Jewish Elementary school in Los Angeles, we had Lawyers, Doctors, CPA’s, Movie producers, and Brokers…good luck finding a mechanic).

To make life easier for Mom and Dad, in 1992 Morningstar came out with a new tool, the style box. The style box made this simple task of investing even simpler. As it turns out, stocks, much like everything else in life fit nicely and neatly into a set of predefined categories that have very clear and delineated features…how could one go wrong now? After all, we know that we have to buy low, and sell high, and we know that to gage value we just need to look at multiples, and now there is this extremely easy to use tool that does it all for us. But wait, there is more. The saintly ladies and gentlemen on Wall Street took it upon themselves to make life that much easier for Mom and Dad. They created ETF’s based on that really great and always accurate style box tool! Now if Mom and Dad just want to buy S&P 500 Value Stocks they can go ahead and buy SPYV and POW!!! Mom and Dad have a portfolio of 386 value stocks, or they could buy SPYG and BOOM!!! Mom and Dad now have a portfolio of 282 Growth stocks. So now Mom and Dad can even buy both SPYG and SPYV and own all 668 stock in the S&P 500…Hmmmm, 686 stocks, how does that work? How can we divide the S&P 500 into two distinct segments and end up with about 32% more stocks? Clearly we cannot, there is obviously a lot of overlap, so what exactly are we buying here? I am going to use some of my new toys to investigate.

Machine learning, a sub-section of Artificial Intelligence (really just a cool way of saying statistics)

Broadly speaking, the Machine learning world is divided into three sections, supervised learning, unsupervised learning, and DEEP learning.

Supervised learning models learn to predict target values from labelled data. In other words, we take a bunch of information we know, and line it up with known outcomes based on the same info. Once that’s done, we tell the computer to “learn” the relationships that exist between the data and the known outcome. The relationship can then be applied to out of sample data.

Unsupervised learning models work a little differently. With unsupervised models we don’t have any labels so there is nothing to predict. So unlike the supervised models where I give my computer the problems and the answer to study from. Here I give my computer data and ask it to find relationship in that data. Unsupervised models are looking to do one of two things. Identify structure/clusters within data, or transform/ reduce data into simpler terms.

Deep Learning is the most advanced and complex form of machine learning. The primary set of models used for deep learning revolve around Artificial Neural Networks. Much like they sound, Neural networks are designed to work a lot like a human brain works. When a model receives information, it is fed to a series of “neurons”, these neurons apply some sort of transformation to the information and send the transformed information to the next set of neurons and so on until we reach an output stage. While Neural networks are very common tools for things like facial recognition, fraud detection, visual art processing and others, applying them to some value ETFs is probably overkill. So we will save this toy for another time.

Back to Our Questions…(How value is value? and is value the value you value?)

The first model I have selected is called a Gradient Boosted Decision Tree. Gradient boosted decision trees are supervised machine learning algorithm. Decision trees work a lot like call centers do. We start with a question called a root node, and based on the answer we branch out to one of a number of next questions, and so on and so forth until we reach what’s called the leaf node - normal people call this “the answer,” data scientists call it the classification. Once the first “tree” is built the algorithm takes a look at how many answers the tree got wrong and starts building another one to improve on the first one, and so on until the model cannot improve on itself.

The second model I have selected is called the KMeans clustering algorithm. This unsupervised model aims to split unlabeled data into a prespecified number of groups. It does this constantly adjusting the area encompassing a cluster as each new item is added to the calculation. The model adjusts the area by continuously calculating points called centroids for each cluster.

What goes into the model?

So to set up our supervised machine learning Gradient decent tree model we need a few ingredients:

1. An algorithm
   1. An algorithm is nothing more than a series of steps taken to solve a problem. We chose our algorithm based on the problem we are trying to solve, and the information we have.
2. Training data
   1. Think of this as a question bank for an exam. In any question bank there should be enough questions to ensure we understand the concepts, and all the answer must be available so we can check our work
3. Testing data
   1. Think of this as a practice exam for the same exam. So after studying for the exam, in order to make sure we are prepared, we will use a practice test. The practice test should test the same exact concepts and if we’ve done a good job studying we should do well
4. Evaluation data
   1. This is the big test. If we have a good model, we should do well.
5. A way to evaluate our results
   1. This could be a metric like accuracy, precision, recall, and others.
   2. Remember we are not always interested in the most accurate model. Sometimes other factors are more important. For example, If evaluating the results from a COVID – 19 test, I would rather have a model that captures 99% of True Positives at the expense of more false positives with an accuracy score of 85% vs a model that captures 90% of True positives and 90% of true negatives with an accuracy score of 90%. Why? Because the goal is to identify as many sick people as possible, and if someone has a false positive that should resolve itself in subsequent re-tests.

And now our models

Gradient Boosted Decision Tree

Our features:

Keeping in line with the industry norms, we have selected the following features to classify stocks.

* Price/Book - The price-to-book, or P/B ratio, is calculated by dividing a company's stock price by its book value per share.
* Forward price-to-earnings (forward P/E) - is a version of the ratio of [price-to-earnings](https://www.investopedia.com/terms/p/price-earningsratio.asp) (P/E) that uses forecasted earnings for the P/E calculation.
* EV/Cash Flow from Operations - Enterprise Value to Operating Cash Flow compares the total value of the company with its ability to generate cashflow from its business operations.
* GICS Sector Codes – This is a numerical representation of a stock’s Sector classification.

Our Training/Test Data:

Our training/test data consists of all the stocks in the S&P 500 Pure Value, and Pure Growth indices. We chose to use these stocks as they are the best representation of what a value/growth stock look like fundamentally. It’s important to note that we removed all real estate and financial stocks form the data set because the most of the valuation metrics used in the model are not usually applicable to these sectors. To generate our training/testing data properly, we randomly assigned all the stocks in the data set to be part of the training subset (75%) or part of the test subset (25%).

Our evaluation data:

Our evaluation data consists of all the stocks in the S&P 500. Here too we removed financials and real estate sectors to ensure that our model can evaluate stocks properly based on the metrics we chose.

Transformations applied to the data:

* Missing data was removed from the set – Not always the best solution but based on the variance of metrics was too high at all levels to consider alternative methods of handling missing data.
* Winsorization – Winsorization helps us deal with extreme observations (outliers). Outliers can skew results unnecessarily and result in poor classifications.

Results

* Price to Book and Forward P/E were of about equal importance in the classification process, GICS sector and EV/OCF were less important.
* Our Model’s accuracy score on the training set was a 93.1%. This means, that during training the model identified about 93% of value stocks accurately based on our input features.
* Our Model’s accuracy score on the test set was 91.6%. This means that the model successfully identified value stocks 91.6% of the time based on input features. This similarity in accuracy scores is also indicative of a properly fitted model.
* Our Model’s accuracy score on the evaluation data was 52.03%. This is a significant drop in model accuracy from the training and testing scores.

Interpretation of results

* Our model is great for identifying stocks that are CLEARLY value or growth stocks as the case in the pure value and growth portfolios. Our model failed to perform where the delineation between value and growth was less obvious.
* Our model had a particularly difficult time with growth classifications. When classified a stock as growth, it was wrong 46% of the time, whereas when a model predicted a stock was value, it was only wrong 2.1% of the time. This could indicate significant disparities between features within the value stocks as they are classified by the industry.

Given the results of the model above, it is clear that other factors not captured by valuation have gone into the construction of value and growth indices. Furthermore, the model makes clear that value vs growth is not always as cut and dry as it feels, there is plenty of grey. Some potential improvements to our supervised model include.

* Reverse the process – instead of using valuation factors to classify stocks into value and growth (more accurately “not value”), Use growth metrics for the classification process.
* Identify different features – The list of multiples used to build the model is far from exhaustive. Including other features may result in a stronger model.
* Time – cross sectional data (lots of data but all from one point in time) can be funny sometimes particularly in the financial markets. We may have better success with average/median multiples over time for each company.