Presentation Notes:

We implemented a Supervised Machine Learning Model to predict the Sector of a Stock based solely on its stock financial numbers:

In doing we applied the two fundamental classification algorithms: decision trees, and random forests and we also built a neuro feedback network.

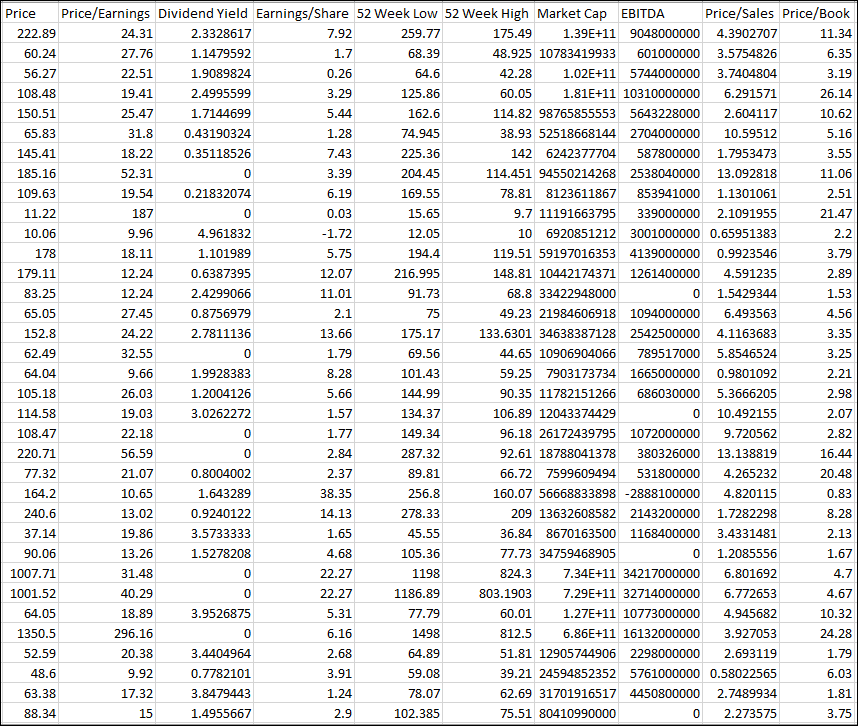
* Price
* Price/Earnings
* Dividend Yield
* Earnings/Share
* 52 Week Low
* 52 Week High
* Market Cap
* EBITDA
* Price/Sales
* Price/Book

The 11 different Sectors:

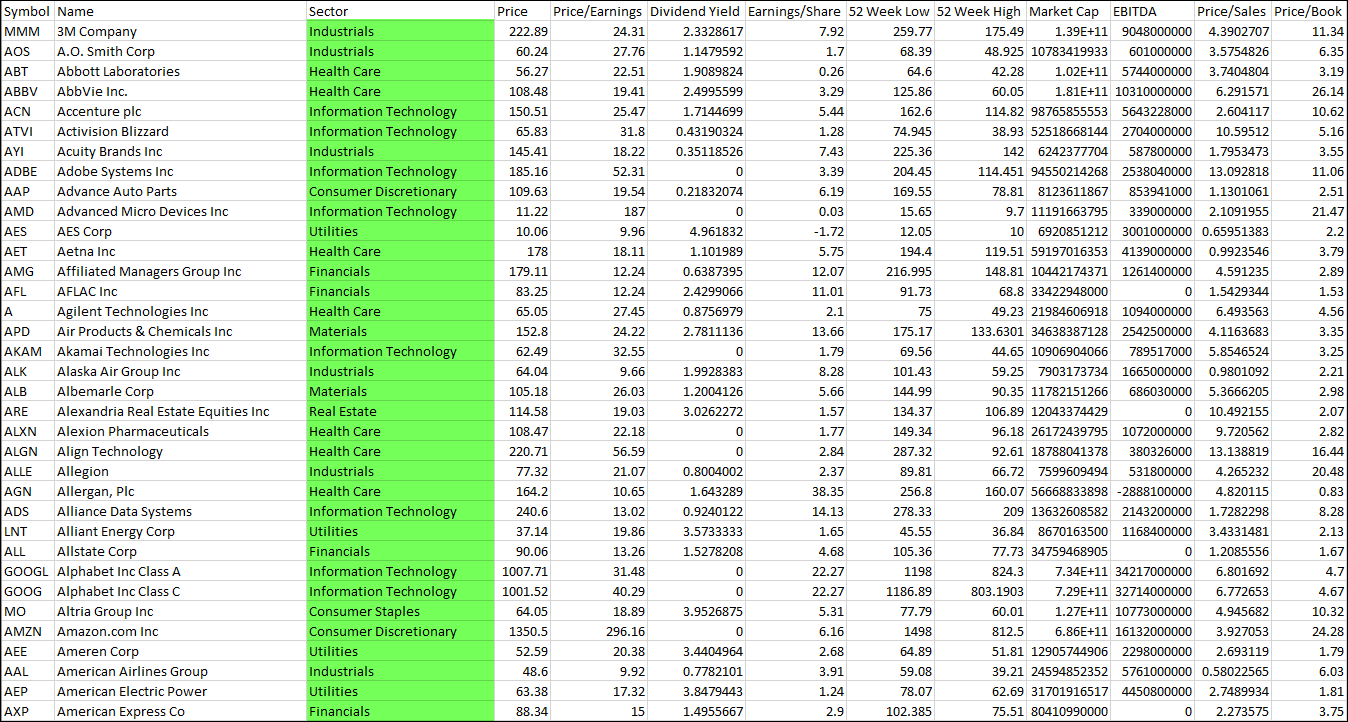
* Consumer Discretionary
* Consumer Staples
* Energy
* Financials
* Health Care
* Industrials
* Information Technology
* Materials

To do this we pulled these specific performance measures for each of the stocks under the S&P 500.

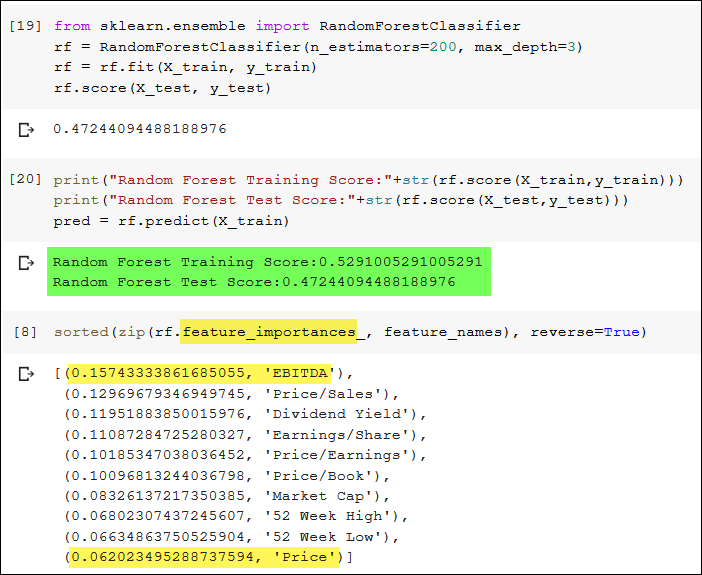
Partial Screenshot of dataset of what the machine model will look at to make predictions (imagine a human trying to predict the sector based on these numbers alone):



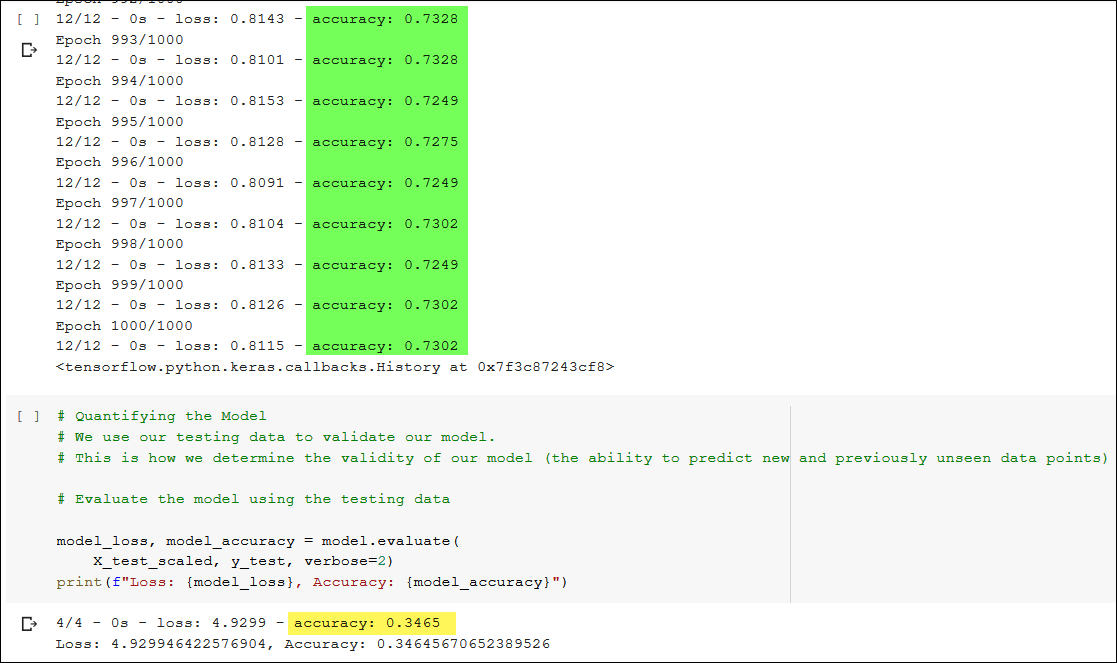
Partial screenshot of dataset with the Sector information that will be used to train the model:



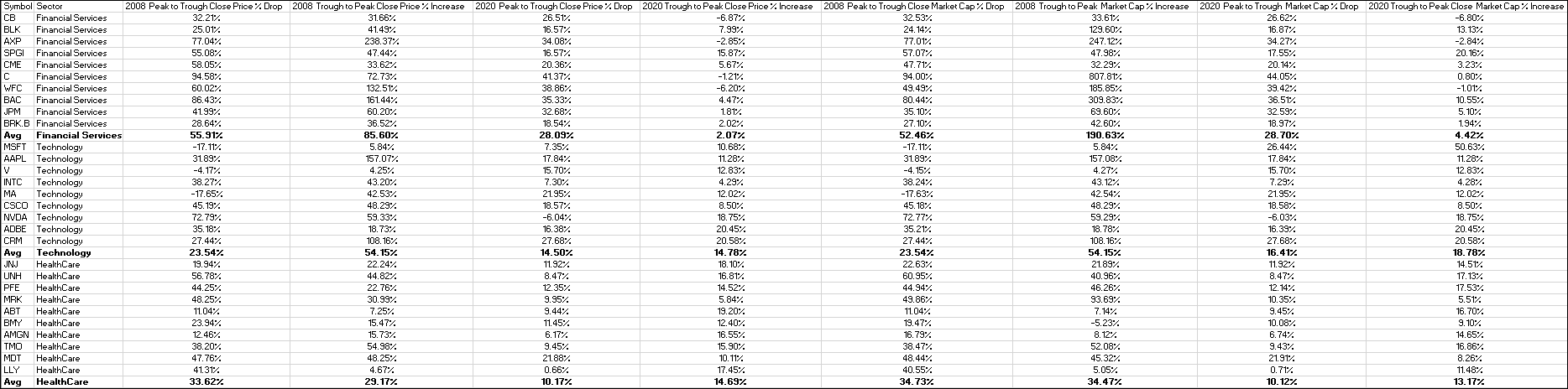
We were able to implement a Random Forest Classification Model with an accuracy of about 50%. EBITDA has the highest feature importance whereas Price had the least importance.



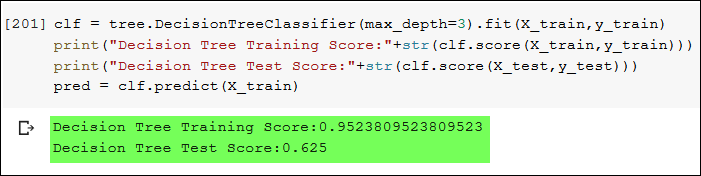
We also built a Neural Network Model that went through 1,000 iterations that got up to 73% accurate but in testing for making predictions through this model the score was only around 35% so the Random Forest Model was actually better for making predictions.



Next we pulled Peak to Trough and Trough to Peak performance measures on select stocks from three different sectors to see if to see how well a machine model can make predictions of the sector based on the performance changes. We started out by only using Price and Market Cap as our feature measures.



We were able to build, train and test a Decision Tree model with a test score of about 63%



**We liked the decision tree model for this classification as it’s easy to interpret and doesn’t require feature scaling.** **But we do understand a flaw can be that its prone to overfitting so we set a maximum depth of 3. Tree depth is a measure of how many splits a tree can make before coming to a prediction. Without putting a limit on the splits can result in the model having a very deep classification tree with many nodes which often leads to overfitting on the training dataset.**

**Next we added some additional feature measures to the model including Earnings per Share and EBIDTA.**

**It turned out that by adding the additional features that we were able to increase the accuracy to 75% and by using a Random Forest model instead of a Decision Tree Model. We found the Random Forest Model was better when using more features.**

