Senior Project – [CS/ECO ID] – [2015]

Predicting Stock Price Trends Using Machine Learning Techniques

Conor Carey

Advisor – Prof. Cass, Prof. Yaisawarng

Abstract

The stock market attracts risk takers who are willing to bet their money on whether a stock price will increase or decrease. What if you had an edge in this gambling mecca? Machine learning techniques are among the newest methods of stock market price trend prediction. Using a support vector machine (SVM), as implemented by WEKA, and eleven attributes in each data point, the data was classified as either BUY or DNB (Do Not Buy), depending on whether the price would increase or decrease. Many previous experiments' goals were to find attributes that correlated with future stock price trends. These attributes were used to create the dataset used in this study, along with a nominal sector attribute, to determine if the addition of this variable improves the accuracy of the stock price trend prediction model.

Attributes

Independent Variables	Effective Tax Rate, Inventory, Gross Profit Margin, Dividend Yield, Price, US Industrial Production, US Inflation Rate, US Long-Term Interest Rates, Labor Force and Sector
Dependent Variables (Class Variable)	Recommendation (BUY/DNB)

Methods

The data used in this experiment was collected from Ycharts.com. The data consisted of quarterly reports from each of the 50 companies. The quarterly price was then run through a Python script to give the classification variable of Buy or DNB, which was determined by whether the price increased or decreased from the current price to the price in the next quarter. After aggregating all 50 companies into one dataset with the correct classifications, a SVM, as implemented by WEKA, was then used to test whether the particular sector variable improved the accuracy of the model. The dataset was partitioned in many different ways to attempt to find a relationship between the sector variable and the prediction accuracy.

Results

Table 1. SVM output on data without sector attribute
59.21% Accuracy

Without Sector Attribute	Buy	DNB
Correctly Classified	2423	64
Incorrectly Classified	1700	13

Table 2. SVM output on data with sector attribute
59.80% Accuracy

With Sector Attribute	Buy	DNB
Correctly Classified	2417	95
Incorrectly Classified	1669	19

Table 3. Basic Materials sector evaluated alone vs. within full dataset
58.50% vs. 58.16% Accuracy

	Buy (Alone)	Buy (Full)	DNB (Alone)	DNB (Full)
Correctly Classified	688	676	0	8
Incorrectly Classified	488	480	0	12

Conclusions

After running several experiments, no correlation between the nominal sector variable and increased accuracy of the prediction model could be determined.

Future Work

This experiment involved 4200 data points collected from 50 different companies, each with 11 attributes. Future work might include more data points, more ways to partition the data to further test how sectors might effect the prediction outcome, and different company stocks. The companies used for this thesis were collected by highest trade volume, which could have an adverse affect on the prediction outcome. Another subject for future work would be to isolate the reason why the SVM heavily favored the BUY recommendation. When sectors were tested individually, the SVM selected BUY every time. When tested with the entire dataset, the SVM would only select DNB fewer than 2% of the time.