Kleio Baxevani, Prem Chand, Desiderio Pilla
CISC 684
Project Proposal

Title: Predicting Stock Market using Machine Learning Techniques

The goal of this project is to predict, for a particular stock on a particular date, whether or not the price of the stock will increase or decrease one week from the date. This is a Boolean classifier.

Our data has 9 attributes: *Date, Symbol, Open, Volume, GISC Sector, Value, Price to Earnings Ratio, Debt to Equity Ratio,* and *Free Cash Flow Yield*. Every attribute is a real value (except GISC Sector is a class)

For every instance, we will have to define the target classification based on the price of the stock one week prior.

We can use SVM/Neural Networks/Logistic Regression to predict the classification based on the 9 attributes.

Datasets: https://www.kaggle.com/dgawlik/nyse/download

Files: Attributes

- Only the attributes that will be used are listed. The actual datasets contain more attributes.
- prices-split-adjusted.csv – date, symbol, open, close, volume
- fundamentals.csv - Ticker Symbol, Period Ending, Cash and Cash Equivalents , Common Stocks, Net Cash Flow-Operating, Total Assets, Total Liabilities, Earnings Per Share, Estimated Shares Outstanding
- securities.csv – Security, GICS Sector

The prices-split-adjusted.csv file has data for every stock with a distinct instance for each **day**.
The fundamentals.csv file has data for every stock with a distinct instance for each **year** (not starting on January 1st, but at some specified day every year).
The securities.csv file has data for every stock with a **single instance** for each stock.

The files would all be joined such that each instance contains the *daily* prices-split-adjusted values, the *most recent* fundamentals values, and the securities data. Joins will be on the date and symbol (for securities.csv, the "Security" label is the ticker symbol).

With the newly created dataset, the following attributes would be created for each instance:

- Value
- Price to Earnings Ratio
- Debt to Equity Ratio
- Free Cash Flow Yield

After these attributes are created, the only columns that will be saved are

- Date
- Symbol
- Open
- Volume
- GICS Sector
- Value
- Price to Earnings Ratio
- Debt to Equity Ratio
- Free Cash Flow Yield

Derived Attributes:

- $Value = \frac{close}{\frac{Total\ Assets - Total\ Liabilities}{Common\ Stocks}}$
  - The *value* of a stock is the ratio between a single share's actual price and its valued price. A value greater than 1 indicates a share is trading above its inherent value, while a value less than 1 indicates a share is trading below its inherent value.
  - $Value = \frac{Market\ Value}{Book\ Value}$
  - $Book\ Value = \frac{Shareholder\ Equity}{\#\ of\ common\ shares}$
- $Price\ to\ Earnings\ Ratio = \frac{close}{Earnings\ Per\ Share}$
  - The P/E shows whether a company's stock price is overvalued or undervalued compared to the earnings of a company. The value is almost always greater than 1, as earnings are expected to increase each year for a company. Higher P/E ratios indicate a stock is more expensive compared to other companies.
- $Debt\ to\ Equity\ Ratio = \frac{Total\ Liabilities}{Total\ Assets - Total\ Liabilities}$
  - This ratio puts a company's debt into perspective. The smaller the value, the larger financial leverage a company has.
- $Free\ Cash\ Flow\ Yield = \frac{Net\ cash\ flows(Operating) - Capital\ Expenditures}{Estimated\ Shares\ Outstanding * open + Total\ Liabilities - Cash\ and\ Cash\ Equivalents}$
  - This is the amount of cash that a company has available to them. Higher values are desired.
  - $Market\ Capitalization = Estimated\ Shares\ Outstanding * open$
  - $Enterprise\ Value = Market\ Capitalization + Total\ Liabilities - Cash\ and\ Cash\ Equivalents$
  - $Free\ Cash\ Flow = [Net\ cash\ flows(Operating)] - Capital\ Expenditures$

- $Free\ Cash\ Flow\ Yield = \frac{Free\ Cash\ Flow}{Enterprise\ Value}$

References

[1] P. D. Yoo, M. H. Kim, T. Jan, "Machine Learning Techniques and Use of Event Information for Stock Market Prediction: A Survey and Evaluation", *Proceedings of the International Conference on Computational Intelligence for Modelling, Control and Automation, and International Conference on Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC)*, 2005.

[2] J. Patel, S. Shah, P. Thakkar, K. Kotecha, "Predicting stock market index using fusion of machine learning techniques", *Journal of Expert Systems and Applications*, vol. 42, issue 4, pp. 2162-2172, 2015.

[3] C. F. Tsai, S. P. Wang, "Stock Price Forecasting by Hybrid Machine Learning Techniques", *Proceedings of the International Multi Conference of Engineers and Computer Scientists*, 2009.