
CAP 6610 Project Proposal: Stock Growth Trend Classification with Machine Learning

Jonathan Wozny

Department of Computer and Information Science
University of Florida
Gainesville, FL 32611
jwozny@ufl.edu

1 Introduction

Finding acceptable and trustworthy methods for investing in stocks is a subject of great interest. There are many different strategies, algorithms, and investing approaches to get acceptable returns. Stock prices can be volatile with many unpredictable fluctuations, and making any consistent correct predictions in the short-term (i.e. daily and weekly) is not likely. Rather than trying to predict short-term price fluctuations, I will attempt to classify the long-term price growth trends of stocks using key financial metrics that are calculated from company financial statements. Key financial indicators, such as a companies earnings per share, price-to-earnings ratio, quick ratio (assets divided by liabilities), return on equity, can be used to analyze or get an idea of the intrinsic value and how risky or safe an investment might be in a company.

I find this problem area interesting as it is a challenging and seems to be a quite open-ended machine learning problem with many variables. Not only will it give me good practice in many areas of machine learning (gathering data, creating a model, feature selection), but it is a popular problem that is applicable to the finance industry. I have decided to analyze the long-term growth rather than looking at the short-term because I do not believe, given the time constraints, that accurate short-term price predictions are feasible, or they would require an extremely complex model that would take an extensive amount of time and research to produce anything of value.

2 Resources Summary

The article I plan to study [1] is entitled *Equity forecast: Predicting long term stock price movement using machine learning* and uses a variety of machine learning models (and compares their results) to predict whether the price of a given stock will grow by at least 10% in the next year. The author uses many features (and later pares down the number of features used) such as earnings per share, price-to-earnings ratio, current ratio, etc., for about 1800 stocks. Each row of the dataset contains these ratios for the final quarter of year x with the label being a 1 if the price grew by 10% from year x to year $x+1$ or -1 if it did not achieve this growth. Each stock had multiple years of data, so this resulted in about 4500 data points (3-4 years for each). The author then performs manual feature selection in order to simplify the model and use only the most important features. The author uses precision, recall, and F-score as his metrics. Overall, this article provides insight into how to organize and clean the dataset, what features and models are promising, and how to approach the problem.

30 **3 Approach**

31 I will take on an approach similar to that done in the article [1]. As evidenced by much of my research,
32 I will use SVM for this classification task. In many of the articles I read, it tends to outperform many
33 models in similar endeavors, and as the market can be volatile, it handles outliers well.

34 The dataset I will be using will be scraped from macrotrends.net, which contains many key financial
35 metrics in tabulated form for each quarter starting from 2009, if applicable. I will select a variety
36 of good and bad stocks (all from the same industry) determined by the price growth or decline, and
37 each row will correspond to the mentioned features for a given quarter or year for each stock (there
38 will be multiple rows for each stock). I am not sure whether it will be more fruitful to predict the
39 next quarter's price trend based on the previous quarter's data, or whether to predict the next year's
40 price trend based on the previous year's data, but I will make a choice through experimentation. As
41 above, the label will be whether a stock achieved a certain amount of growth (+1) or not (-1) in the
42 next quarter or year. I would like to note that I have already written code to scrape the website and
43 retrieve this data.

44 **References**

45 [1] Milosevic, Nikola. (2016). *Equity forecast: Predicting long term stock price movement using machine*
46 *learning*. CoRR abs/1603.00751. (2016).