Key Financial Indicators: Unlocking Stock Performance Insights

Radjamin Hukom

rhukom@bellarmine.edu

1/16/2025

Executive Summary

This project seeks to identify and rank the financial ratios most significant in predicting the performance of Fortune 500 stocks, measured by the percentage change in stock prices from December 31, 2019, to December 31, 2024. Financial ratios such as P/E, P/B, EPS, ROE, and debt-to-equity, among others, are analyzed to assess their influence on stock performance. By narrowing the focus to the most impactful ratios, this project aims to simplify decision-making for long-term investors without requiring exhaustive analysis of all financial data.

To achieve this, advanced machine learning models like Random Forest, Gradient Boosting Machines, and Support Vector Regression are employed alongside Multiple Linear Regression as a baseline for comparison. These methods allow for handling non-linear relationships, ranking feature importance, and ensuring predictive accuracy. Data collection is automated using Python libraries (yfinance, Quandl) and Google Sheets, while Google Colab is used for programming, visualization, and analysis. The findings will provide a ranked list of key financial ratios, offering investors actionable insights for prioritizing their research. This approach seeks to empower investors by streamlining the evaluation process, focusing on metrics that matter most for stock performance, and leveraging automated tools and advanced models for robust analysis.

Project Idea

This project aims to rank the significance of the different indicators of the financial health of the stocks in terms of the stock's performance. The various indicators specified would be financial ratios such as P/E ratio (price-to-equity), P/CF ratio (price-to-cash flow), P/S ratio (price-to-sales), P/B ratio (price-to-book), Beta (a metric of a stock's volatility compared to the overall market with 1 being the market), market cap (represents its market size), EPS (Earnings Per Share), dividend yield (dividends compared to stock price), ROE (return on equity), ROA (return on assets), gross margin (shows production efficiency), operating margin (shows operational efficiency), net profit margin (shows overall profitability), debt-to-equity, cash ratio (shows its ability to cover short-term liabilities with cash), current ratio (shows short-term financial health), quick ratio (shows its ability to cover short-term obligations with current assets), debt/EBITDA (indicates debt burden), Asset Turnover (measuring efficiency of assets), inventory turnover (measuring inventory efficiency), receivables turnover (efficiency of receivables used). The stock's performance would be signified by the percent change in the stock's price from December 31, 2019, to December 31, 2024, which is a period of 5 years. The stocks used in this data would be the ones from Fortune 500 companies due to the relatively high availability of data. In the case of insufficiency of data, columns would be

dropped, for example, dividend yield can be hard to get due to fewer companies in the Fortune 500 still paying dividends compared to before.

Background

There are tons of stocks in the market. Looking at the Fortune 500 alone, which is already the cream of the crop most of the time, there are about 500 stocks that a would-be investor would need to comb over. Making an educated decision is nigh impossible without reviewing each company's stock prospectus and 10-K, which contains the company's financial sheets. Now, it is hard to make sense of the financial sheets without using financial ratios which summarize the financial sheets and present them in numbers that compare the different values contained within the financial sheets to show the company's financial health. There is other important information contained within the 10-K form, such as liabilities, vision, challenges to growth, and plans for the future among other things, however due to the limited amount of time provided and the lack of numerical metrics to be measured, only financial ratios would be measured in the scope of this project. Even only considering financial ratios, looking at the Fortune 500 companies, a would-be investor would need a lot of time which he or she might not have if they are only looking to put their money to grow in the long-term, not necessarily able to trade hourly or daily. Hence, this project seeks to further cut down on the ratios and see which ratios are the most significant in determining the stock's growth.

Modeling

Random Forest can be useful in modeling this data as it is a robust ensemble learning method, handling complex, non-linear relationships and ranks feature importance, which is ideal for dealing with potentially correlated financial indicators. However, it is less interpretable compared to linear regression sometimes. Gradient Boosting Machines would also be used as they are powerful predictors of complex relationships in structured data, excellent for ranking the importance of financial indicators, and can effectively handle missing values. They give highly accurate predictions and also reduce overfitting, however, they are computationally more intensive compared to other simpler models. Support Vector Regression can also be used as they are effective in predicting continuous outcomes, especially when the predictors and the target variable's relationship might not be linear. It also works well with high-dimensional data and in capturing complex patterns. It works well with both small and large datasets. However, it is computationally expensive for large datasets. As a baseline, I might use Multiple Linear Regression (MLR) to quantify the relationship between the dependent variable (stock price change) and the multiple independent variables (financial ratios). It provides straightforward interpretability and makes it easy to identify which variables have

significant effects. However, it assumes linear relationships between the variables, which might not necessarily be the case here.

Tools

To make it easy on myself, Google Sheets would be used over Microsoft Excel due to the availability of the function googlefinance() in Google Sheets which directly pulls metrics from Google Finance. However, that alone would not be sufficient to pull historical data of the financial ratios of the various stocks, which would require yfinance library, pandas_datareader, or Quandl which will all be accessed through Python. They would be pulling financial datasets from Yahoo Finance, Alpha Vantage, or Quandl. Google Collab would also be used for faster programming, automating, data collecting, data preprocessing, data processing, and graphic visualization for presentation. Lastly, either Canva or Microsoft PowerPoint would be used to present the project's findings.

Conclusion

This project demonstrates a data-driven approach to identifying the most impactful financial ratios influencing stock performance over a five-year period. By leveraging advanced machine learning models and automating data collection and analysis, it simplifies the process of evaluating Fortune 500 stocks. The ranked list of financial ratios derived from the analysis will enable investors to focus on the metrics that matter most, reducing the time and effort needed for decision-making.

Ultimately, the project bridges the gap between financial complexity and actionable insights, providing long-term investors with a practical framework to assess stock health and growth potential efficiently. The findings not only support more informed investment decisions but also highlight the value of integrating financial analytics with modern technology.

References

Alpha Vantage Inc. (n.d.). *Alpha Vantage API Documentation*. Retrieved from https://www.alphavantage.co

Google. (n.d.). *Google Finance*. Retrieved from https://www.google.com/finance

Nasdaq. (n.d.). *Quandl: Financial and Economic Data*. Retrieved from https://www.quandl.com

Yahoo. (n.d.). *Yahoo Finance*. Retrieved from https://finance.yahoo.com