

# Financial Analysis & Investment Decision-Making Process Using Machine Learning

**Eakam Sayal**

Modern School vasant vihar.  
Sayal.eakam@gmail.com

**Reetu Jain**

Mentor, On My Own Technology Pvt Ltd  
Reetu.jain@onmyowntechnology.com

**Abstract**—New trading tactics are booming in the algorithmic trading industry. Companies have invested billions in R&D and infrastructure so they can outperform the competition and dominate the market. However, it is widely known that the purchase-and-hold approach may beat many algorithmic methods, particularly over the long term. However, very few people have mastered the skill of identifying value in stocks. Investment choices may benefit from recent advancements in data science and machine learning. The quick enhancement of the ML algorithms has coincided with the expansion of micro-level data availability. In this paper the focus is on, Can a machine learning algorithm distinguish between stocks which perform well and the ones that don't? If so, might it utilise this knowledge to anticipate which stock would be a good investment and facilitate investors' decision-making processes. Additionally, the article focuses on doing an exploratory data analysis on a dataset that includes 200+ financial indicators that are typically included in the 10-K filings that each publicly traded firm produces each year for a variety of equities (on average, 4k stocks are listed in each dataset for the F.Y 2014-2018).

**Keywords**—*Investments, Investing, Machine Learning, Financial planning, stock Price, financial indicators.*

## I. INTRODUCTION

Accurate financial estimates and plans for wise and successful resource allocation are the main products of the finance department in modern firms. In situations where markets are volatile or move quickly, among other things, quick and precise forecasting and planning are crucial [1]. Excellent forecasting is among the characteristics that make an effective finance Department [2]. Thus, it should not be surprising that the majority of bigger organisations include FP&A (financial planning and analysis) teams within their finance divisions. Due to the growing accessibility of

big data and cutting-edge analytical methods, FP&A can now provide the organisation with more and better insights more quickly. a collection of methods known as machine learning that were created in computer science and statistics, seem incredibly well adapted to this situation. The findings are best suited for investing not trading as Buying or selling a financial item, such as stocks, bonds, or commodities, is referred to as trading. Executing a buy/sell decision is what it is. Trading is sometimes used interchangeably with speculating, which is short-term (or extremely short-term) dealing with the intention of making an immediate profit—in some cases without any specific long-term strategy or industry, nation, or asset class emphasis. The latter interpretation is what gives trading its mixed (if not outright negative) image. Investing is a longer-term commitment that adheres to a plan that was laid out at the beginning and approved by the investors.[3]. We have used a dataset that consists of 200+ financial indicators that help to identify those stocks that an investor should buy or not. This is how this paper is organized. Section 1 of this article reviews prior research and conclusions. In Section.2, we conduct EDA and quickly evaluate the datasets for FY 2014–2018. A concise, concentrated review of machine learning is given in Section 3. We specifically draw attention to the danger of conflating forecasting with planning. We discuss our findings and recommendations in Section 4.

## II. LITERATURE REVIEW

There are two reasons why there don't seem to be many articles on machine learning in Financial analysis and planning. First, time-series forecasting has received extensive coverage and study over a long period of time et al. [4]. Both from an academic and theoretical standpoint as well as from the perspective of practitioners, a wide range of tools for this goal have been created, including simple-to-use off-the-shelf software. et al.[5]. From a practical FP&A standpoint, these tools enable practitioners to reach conclusions that—on the whole—serve sufficiently to accomplish

the goal of establishing financial plans. They work in conjunction with the domain expertise of the specialists working in the FP&A function. Because of this, practitioners in particular can view machine learning as a "so-so" technology. et al.[6], not quite deserving of their (full) attention yet. As a result, the intrinsic motivation to explore new tools, such as machine learning, in financial planning and analysis is currently less strong than it is, for example, in stock market forecasting, where even a relatively little increase in predicting accuracy may have a significant economic impact. With the increased adoption of digitalization and the ensuing rise in data accessibility as outlined above, we think that this will alter. Given that the traditional procedure is very labor- and time-intensive, automated predictions powered by machine learning can also result in a significant decrease in expenses and enhanced flexibility in addition to enhancing the accuracy of financial projections. creating a long short-term memory (LSTM) classification model et al. [7] to predict the movement of market trends using adaptive stock technical indicators. For greater accuracy, technical indicators are used as model inputs. Prediction accuracy rises as the number of epochs rises, but after a certain number of epochs, over-fitting causes a decline in prediction accuracy. The LSTM model's optimality is then established. The outcomes showed that the best LSTM's prediction accuracy is higher than that of logistic regression models and support vector machines (SVM). By utilizing a variety of classification algorithms, et al. [8] looked at the value of time series descriptors obtained from technical analysis and news data to predict stock direction reversal for the upcoming few days. The effectiveness of several suggested categorization algorithms was examined. These included a collection of technical indicators, a set of news sentiment descriptors, and combinations of these. The findings showed that, on average, the tested combinations that combined historical news sentiment with stock price indicators performed the best. News stories are important while researching stock trend forecasts as a result.

### III. DATASET

The Machine Learning Model Training and data analysis on a few indicators is done on datasets for FY 2014–2018 Each dataset includes more than 200 financial indicators for a wide range of companies that are typically included in the 10-K filings that each publicly traded firm submits each year (on average, 4k stocks are listed in each dataset). Using the pandas DataReader and Financial Modeling Prep API, this

dataset was created The last column, class, lists a binary classification for each stock, where for each stock, if the PRICE VAR [%] value is positive, class = 1. From a trading standpoint, the 1 shows which stocks a fictitious trader should BUY at the beginning of the year and SELL for a profit at the conclusion of the year. For each stock, if the PRICE VAR [%] value is negative, class = 0. A hypothetical trader should NOT BUY the stocks marked with a zero from a trading standpoint since they will lose money if their value drops.

## IV. EXPLORATORY DATA ANALYSIS

### A. Ratio Analysis

All analyses were conducted on a single organization (Procter & Gamble). We conducted each analysis in a standardized way. A company's ticker symbol, which is listed in the column, can be used to identify it. A financial ratio is a simple approach to assessing the status of any business. There are around 59 different financial ratios. Rapid Ratio: Inventory current assets and current liabilities (This is the easiest method to comprehend an organization's liquidity, or how soon it can settle all of its debts.)

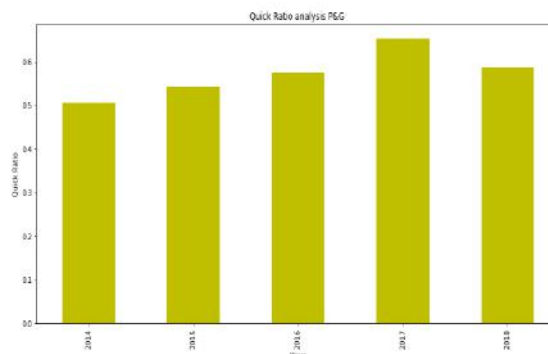


Figure 1: Ratio analysis of P&G

Before we can draw any conclusions regarding short-term commitments, we must first assess the current ratio, which P&G has at 0.57. This indicates that the business is having liabilities that are more than twice as great as its current assets. also Current Assets / Current Liabilities: Current Assets plus Current Inventory A good current ratio is between 1 and 2, and if it exceeds 2, it indicates that the firm cannot utilize its inventory.

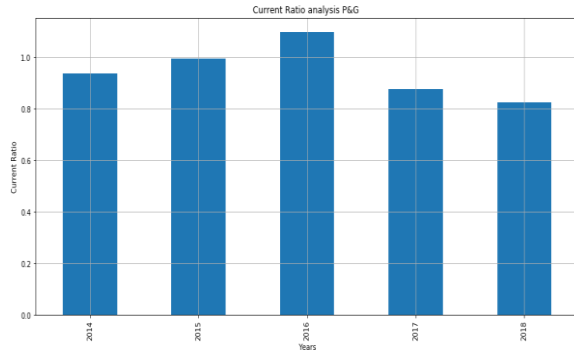


Figure 2: Current ratio analysis of P&G

It is a positive indicator for investors: Most investors like businesses with debt ratios between 0.3 and 0.6. Let's examine The turnover of inventories The number of times inventory is sold or utilized during a given time frame, such as a year, is known as inventory turnover. It is computed to determine whether a company has an excessive amount of inventory in relation to its level of sales.  $\text{Sales} / \text{Average Inventory}$  equals inventory turnover.  $\text{Average Inventory}$  is equal to  $(\text{Starting Inventory} - \text{Finishing Inventory})/2$ .

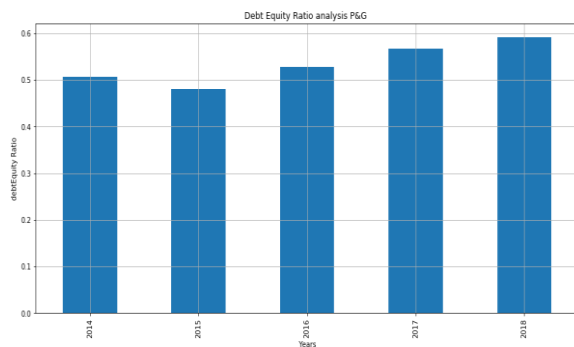


Figure 3: Debt Equity ratio analysis of P&G

Let's examine each one Short-Term Asset of P&G: considering the following Parameters Cash and Cash Equivalent Cash and Short-Term Investment Inventories Average Receivables Investment

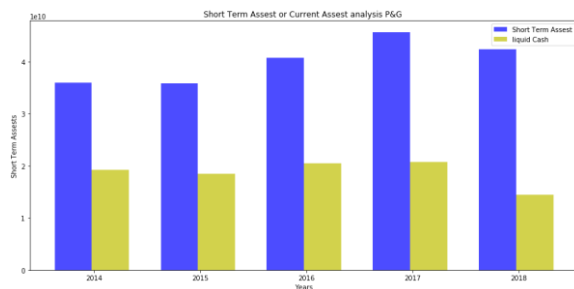


Figure 4: Short-term asset analysis of P&G

Long Term Asset: considering following parameters Property, Plant and Equipment Goodwill and Intangible Assets Long-Term Investment.

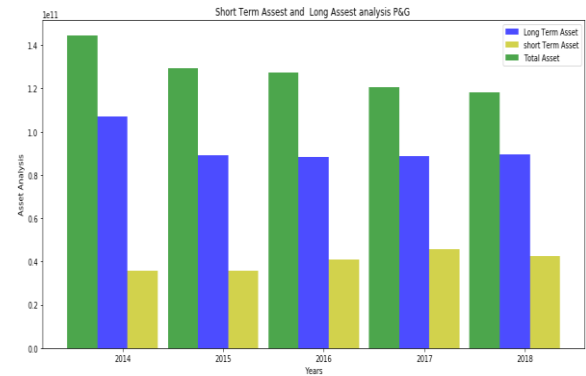


Figure 5: Long-term asset analysis of P&G

### B. Balance Sheet Analysis

Two sections make up the balance sheet. Assets and obligations There are two choices under Asset Section: short-term assets and long-term assets. There are also two choices in the liabilities section: Long-term obligations against short-term obligations Cash and cash reserves, cash equivalents, inventories, accounts receivable, securities, etc. are examples of short-term assets. Property, plant, and equipment are all examples of long-term investments that are intangible assets. Short-term obligations: Customer deposits, dividend and trade account payables, short-term obligations, and the present level of long-term debt Long-term liabilities include loans, delayed income, deferred wages, and other obligations. Let's examine each one Short-Term Asset of P&G: considering the following Parameters Cash and Cash Equivalent Cash and Short-Term Investment Inventories Average Receivables Investment

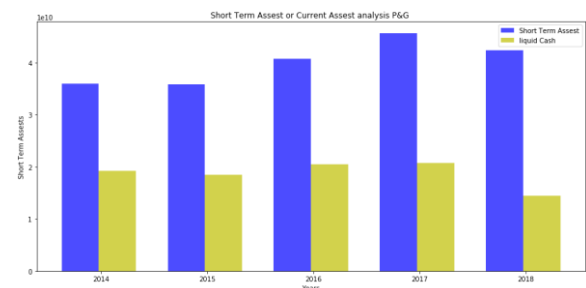


Figure 4: Short-term asset analysis of P&G

Long Term Asset: considering following parameters Property, Plant and Equipment Goodwill and Intangible Assets Long-Term Investment.



Figure 5: Long-term asset analysis of P&G

### V. MACHINE LEARNING

Machine learning may be characterized as a group of techniques that automatically create predictions from complicated data, while there is no one definition for it [9]. Machine learning essentially uses a function-fitting technique to try to find a suitable approximation of the function underlying the predicted link between input and output data [11]. Machine learning uses statistical methods and algorithmic approaches from computer science in this search for patterns in data [10], which is largely carried out automatically. Machine learning specifically seeks to handle the issue of data with a high dimension. When there are more input factors (independent variables, features) than observations available to predict the output variable (dependent variable), this is known as high dimensionality. standard statistical In this situation, strategies are ineffective [9]. Supervised learning, unsupervised learning, and reinforcement learning are the three main types of machine learning. The goals of supervised learning include estimating an output variable's value from the values of a group of input variables. To do this for each data point, supervised learning uses a collection of input and output variables that are collectively viewed. [11]. For building our model we have used the Support vector Machine algorithm which is a supervised machine learning algorithm, the reason we used SVM as it is preferred for dense data( where the Number of columns is more than the number of rows ).

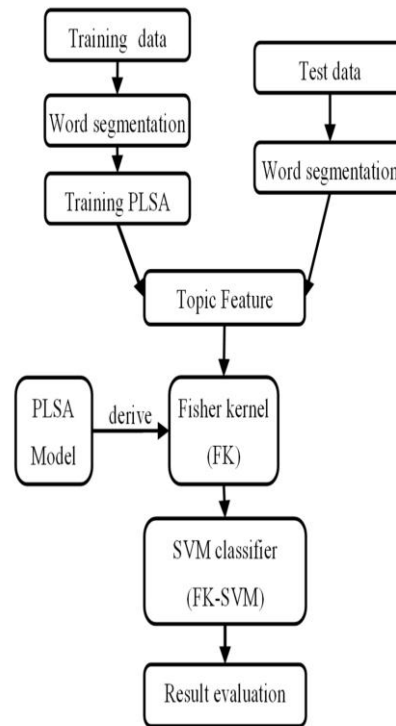


Figure 7: SVM FlowChart

One of the most popular supervised learning approaches, Support Vector Machine, or SVM is used to handle classification and regression issues. Nevertheless, Machine Learning Classification issues are where it is most frequently employed. To swiftly classify new data points in the future, the SVM approach aims to build the best line or decision boundary that can split n-dimensional space into classes. This best option boundary is known as a hyperplane. SVM is used to choose the extreme vectors and points that contribute to the hyperplane. The method is known as a support vector machine because these uncommon circumstances are known as support vectors. Look at the illustration below, where two different groups are divided by a decision boundary or hyperplane.

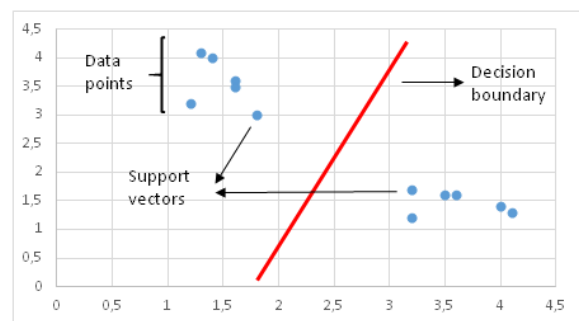


Figure 7: Working of SVM

Hyperparameters are some parameters that cannot be discovered by direct learning. Prior to starting training, individuals frequently choose them based on some intuition or trial and error. By enhancing the model's functionality, such as increasing the model's complexity or learning rate, these parameters demonstrate their significance. Models may have a large number of hyper-parameters, and determining the ideal set of parameters may be approached as a search issue. Finding the best hyperparameter for an SVM model is a highly challenging issue. These hyper-parameters include things like what C or gamma values to utilize. But it may be discovered by simply attempting all possible combinations and determining which inputs are most effective. The model had a 92% accuracy rate.

### CONCLUSION

Profit and liquidity management are crucial while making financial decisions. Businesses that successfully balance profitability and liquidity indicators achieve their most important financial results. Big data offers the financial planning and analysis function enormous possibilities, particularly when it combines vast volumes of data with the right instruments for analysis. This article provides an introduction to machine learning in this context. We have shown how machine learning may be helpful for Financial analytical & Planning practitioners by addressing numerous pertinent theoretical parts of it and talking about the findings of simulation research.

### REFERENCES

- [1] Mahlendorf, M. D., Schäfer, U., and Thaten, M. Becker, S. D. (2016). creating a budget during a financial crisis. 33, 1489–1517, Contemporary Accounting Research.
- [2] Roos, A., Tucker, J., Rodt, M., Stange, S., Ego, P., Boudadi, A., & Sheth, H. (2020). Lessons from best-in-class CFOs. Boston Consulting Group.
- [3] Review of machine learning experiments for equities investing decision-making: Why the majority of published research results fall short of expectations in practise Wojtek Buczynski, Fabio Cuzzolin Barbara Sahakian 3 June 2020
- [4] De Gooijer, J. G., & Hyndman, R. J. (2006). Time series forecasting for 25 years. *International Forecasting Journal*, 22, 443–473.
- [5] Kesters, U., McCullough, B. D., & Bell, M. (2006). Forecasting software: past, present and future. *International Journal of Forecasting*, 22, 599–615.
- [6] Acemoglu, D., & Restrepo, P. (2018). Artificial intelligence, automation, and work. *The economics of artificial intelligence: an agenda* (pp. 197–236).
- [7] Agrawal, M., Khan, A. U., & Shukla, P. K. (2019). Stock price prediction using technical indicators: A predictive model using optimal deep learning.
- [8] Attanasio, G., Cagliero, L., Garza, P., & Baralis, E. (2019, November). Combining news sentiment and technical analysis to predict a stock trend reversal. In *2019 International Conference on Data Mining Workshops (ICDMW)* (pp. 514–521). IEEE.
- [9] Taddy, M. (2019). *Business data science: combining machine learning and economics to optimize, automate, and accelerate business decisions*. McGraw-Hill Education.
- [10] Bishop, C. M. (2006). *Machine learning and pattern recognition. both statistics and information science*. Springer (Softcover published in 2016).
- [11] Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: data mining, inference, and prediction* (2nd ed.). Springer-Verlag.