Automated stock recommendations using Financial Indicators and Machine Learning

Utkarsh Sharma
B. Tech CSE
ASET, Amity University, Noida
skhiearth@gmail.com

Simran Gogia
B. Tech CSE
ASET, Amity University, Noida
simrangogia20@gmail.com

Abstract

Stock market is suggested and regarded as one of the high-yielding long-term investments, yet a majority of people don't capitalize on the same. Dubious advice and attempts to 'beat the market' usually give rise to skepticism and distrust among first-time investors. This paper proposes a subjective, low-risk stock market advising platform that leverages Machine Learning clustering (K-Means) on basic Financial Indicators that are used to track the performance of stocks in the exchange to serve as an aid in investment decision, particularly for first-time investors. The results suggest that clustering-powered subjective recommendations can prove to be a low-risk advising tool.

Keywords: Stocks, Recommendations, Finance, Analysis, Machine Learning, Clustering Algorithms

Introduction

The fact that stock investments return well is well established. Stock investments are highly advisable (Blanchard et al. 1993) for all – due to low barrier of entry, trust and profitability. However, the lack of reliable information and advice regarding investments for beginners is a major hurdle in absolute adoption of stocks as a mainstream investment channel among those with limited capital.

Inspirations for this research include the financial problems concerning SMEs and individuals (Beauregard et al. 2020) generated due to the COVID-19 pandemic and review reports (Ramnath et al. 2008) (Morgan et al. 2003) addressing the forecasts made by financial analysts about stock recommendations. Provided the economic limitations of first-time investors, subjective stock recommendations pose themselves as a more viable option than risky advice – in that Machine Learning clustering aims to group stocks that have performed well within an exchange together.

Using key financial metrics, we can observe the trends and patterns in the financials of such companies, and with the help of Machine Learning – these metrics can be used for recommendations. Listed stocks from the most prominent exchanges from Asia-Pacific and Europe are used for this analysis. The following key financial indicators are considered while clustering the stocks in the different exchanges:

Gross Profit Ratio

Gross profit ratio is defined as the profitability ratio between gross profit and net sales.

$$Gross\ Profit\ Ratio = \frac{Gross\ profit}{Net\ sales}$$

The GP Ratio is a good indicator of the financial performance and effective resource allocation (Nariswari et al. 2020) of a company and helps in analyzing its profitability.

P/E Ratio

This indicator is included to determine the possible overvaluation or undervaluation of the ticker in question – reflecting the market's belief on the company's grossing capacity. A high P/E Ratio typically indicates over-valuation (Shen 2000).

$$P/E\ ratio = \frac{Market\ value\ per\ share}{Earnings\ per\ share}$$

Price cash flow ratio

Used for equating the company's market worth to the cash flow, PCF ratio provides a reliable sign of long-term returns.

$$PCF\ ratio = \frac{Share\ price}{Operational\ cash\ flow\ per\ share}$$

EPS diluted

It is an indication of the company's potential earnings per share if all eligible securities were converted.

$$EPS\ diluted = \frac{Net\ income - Preferred\ Stock\ Dividends}{Average\ Outstanding\ Shares + Dilutive\ Shares}$$

It is a constructive metric because it takes into account all convertible securities.

Price to book value ratio

Used for indication of a share's worth according to the book of accounts.

$$PBV\ ratio = \frac{Market\ price\ per\ share}{Book\ value\ per\ share}$$

Asset turnover ratio

A measurement of the ability of company management to effectively use the net assets available for sales revenue generation.

$$AT\ ratio = \frac{Net\ Sales}{Average\ Total\ Assets}$$

Dividend yield percentage

Indication of the expected return percentage in the means of dividends on the investments made at the prevailing market price.

Price earning to growth ratio

It draws the relationship between a stock's P/E ratio and projected earnings' growth rate over a period of time – giving an estimation about the actual value taking into account the predictable growth.

$$PEG\ ratio = \frac{P/E\ Ratio}{EPS\ Growth}$$

This paper is divided into the following sections: Related Work, Proposed Approach, Methodology, Results and Discussion, Conclusion and Future Scope, and References.

Related Work

In paper (Sanboon et al. 2019), the authors propose a model to capture long-term dependencies in stock price data using LSTM to develop a Deep Learning model for predicting buy and sell recommendations in the Thai Stock Exchange.

Using natural language processing techniques, Robert et al. (Schumaker et al. 2009) present a quantitative stock prediction system based on daily financial news. Decision support systems are proposed (Gottschlich et al. 2014) to help incorporate crowd recommendations into investment decision.

In paper (Barber et al. 2003), the author's primary concern is to assess and analyze the returns obtained on investments based on analysts' stock recommendations and (Bradshaw 2004) deals with addressing how analysts use their forecasts to generate such recommendations. (Nair et al. 2017) use Machine Learning based clustering for generating stock trading recommendations based on price time-series data.

Proposed Approach

In this paper, the K-Means clustering algorithm is used to group stocks from various exchanges around the globe into an appropriate number of disjoint clusters – and to extract information from such clusters. All exchanges are kept mutually exclusive to prevent cross-exchange clusters.

An automated platform is set-up to aid on-the-flow cluster generation as the stocks move over time, accounting for new changes in the financial indicators of the companies.

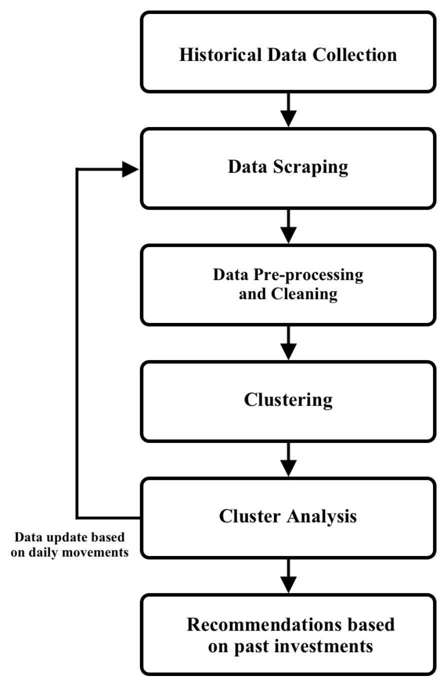

Figure 1. Generalized Framework

Methodology, Results and Discussions

Tools used

For implementation of the clustering algorithm and cluster analysis, Python (Sklearn) and R (dplyr, caTools, e1071, factoextra and ggplot2) are used, whereas the Pandas, requests and NumPy libraries of the Python language are used for scraping and data pre-processing. Python and R were selected due to the availability of well-kept popular libraries.,

Dataset, Cleaning and Pre-processing

Due to unavailability of an exhaustive dataset that comprised of all the financial indicators, and crucially, targeted the time-frame concerning the pandemic, Web Scraping was used to collect the data.

The scrapped dataset comprised of more than 20,000 stock tickers from over ten stock exchanges from Asia Pacific and Europe – covering the different continents and eliminating location bias (Pirinsky et al. 2006). All stock data, across exchanges, was scaled (Standard Scaling) to have a mean value of zero and a standard distribution of unity. The standard score of a sample stock ticker, x is computed as:

$$z = \frac{(x-u)}{s}$$

where, u is the mean of the samples, and s is the standard deviation of the samples.

Clustering

We use K-Means Clustering to assemble similar tickers into a pre-defined set of clusters – to partition stocks based on financial patterns. All these clusters are disjoint and limited to the scope of one stock exchange. K-Means is the widely used iterative clustering algorithm. The algorithm requires the number of clusters K to be provided as input (Lloyd. 1982). Financial indicators are used as independent variables.

To initialize the number of clusters, k, two techniques were used: Elbow Method and Silhouette Method. In the former, the number of clusters corresponding to a severe cliff in the sum of squares of the distances of each data point in all clusters to their respective centroids are considered as the optimal number of groupings. The silhouette method computes silhouette coefficients - a degree of how comparable a stock is to its specific group (cohesion) compared to other clusters - of each stock that measure how much a stock is similar to its own cluster paralleled to other groups.

For several exchanges, to boost the number of clusters, Principal Component Analysis was employed. It is vital to ensure a health clustering of data as it is known that the statistical power of results increase with an increase in number of groups – to better analyze the inter-cluster disparity. The ideal number of clusters varied across exchanges. The following table illustrates the number of clusters and the corresponding silhouette scores for the optimal number.

Stock Exchange	Number of Clusters	Silhouette Score	PCA Components
NSE	3	0.66	1
ASX	2	0.68	1
MCX	4	0.68	1
HKSE	2	0.70	0
Osaka	2	0.74	0
Lisbon	6	0.72	1
Brussels	7	0.74	0
Amsterdam	7	0.73	0
XETRA	2	0.68	1
LSE	2	0.76	0

Table 1. Ideal number of clusters across exchanges

A silhouette score of positive unity indicates perfect distribution of data points into disjoint clusters. For most of the exchanges, the silhouette score can be termed as satisfactory provided the high level of variations among stocks of different industries and maturity.

Cluster Analysis

Financial Analysts can utilize patterns in the clusters formed by K-Means clustering to better study the co-relation of the financial indicators in questions across different stocks.

Financial Indicator	Cluster #1	Cluster #2	Cluster #3
Gross Profit ratio	0.277	0.753	0.538
P/E ratio	33.242	72.948	30.499
Price Cash Flow ratio	68.632	98.680	9714.396
EPS diluted	5.049	19.547	24.351
Price to book-value ratio	7.416	7.873	15.340
Asset turnover ratio	1.080	0.717	1.010
Dividend yield percentage	1.299	0.815	1.239
Price earnings to growth ratio	1.813	8.866	-10.847

Table 2. Mean values of financial indicators in NSE clusters

Financial Indicator	Cluster #1	Cluster #2
Gross Profit ratio	0.208	0.793
P/E ratio	11.736	53.299
Price Cash Flow ratio	50.722	75.744
EPS diluted	0.210	0.111
Price to book-value ratio	5.165	9.381
Asset turnover ratio	1.832	0.875
Dividend yield percentage	2.738	2.639
Price earnings to growth ratio	209.035	340.632

Table 3. Mean values of financial indicators in Australian Securities Exchange clusters

Similar trends and patterns can be found in all exchanges across the three continents – providing a useful tool for financial analysts and individuals alike to evaluate stock recommendations.

Automated Recommendations

As time passes, the financial indicators of the companies in question are bound to change – affecting the relationship between stocks. To incorporate for such market movements, a platform should be set up that updates the considered data for every quarter (Kiger 1972) or daily movements – thereby updating the clusters of tickers and providing the most accurate disjoint groups of related stocks.

As part of a POC, a Python-based script was used for this continuous updating process and deployment of clusters.

Conclusion and Future Scope

Recommendations across a wide range of exchanges and the stock token therein are ought to be beneficial to first-time investors and those with superficial or no knowledge about the market. The simple input sets used in the model improve transparency and the subjective nature of the proposed approach helps reduce risk. The combined automated employed can be deployed in the form of a consumable REST API, leveraged by end-user application – a mobile app was developed as part of our POC. The source code for the clustering algorithms, the API, the end-user application and the original and clustered datasets are open-sourced and forthcoming researches are welcome to experiment on the same (Simran et al. 2020).

A similar recommendation and clustering engine can be extended beyond listed stocks to Mutual Funds, Bonds and Real estate as well – offering more diverse investment options. The lack of reliable data in these sectors and the lack of domain knowledge are the prime restraining factors for the authors as of writing.

Deeper domain knowledge can be used for refining the inputs to the Machine Learning models. We propose feature manipulation and selection of a different array of indicators. Although advances in objective stock recommendations have been largely unsatisfactory so far, advances in Neural Networks can be leveraged to build transparent recommendation engines based on similar input tuples.

References

Barber, Brad, Reuven Lehavy, Maureen McNichols, and Brett Trueman. "Reassessing the returns to analysts' stock recommendations." Financial Analysts Journal 59, no. 2 (2003): 88-96.

Beauregard, Remy, Jose A. Lopez, and Mark M. Spiegel. "Small Business Lending during COVID-19." Small 2020 (2020): 35.

Blanchard, Olivier, Changyong Rhee, and Lawrence Summers. "The stock market, profit, and investment." The Quarterly Journal of Economics 108, no. 1 (1993): 115-136.

Bradshaw, Mark T. "How do analysts use their earnings forecasts in generating stock recommendations." The Accounting Review 79, no. 1 (2004): 25-50.

Jörg Gottschlich, Oliver Hinz, "A decision support system for stock investment recommendations using collective wisdom," Decision Support Systems, Volume 59, 2014, Pages 52-62, ISSN 0167-9236.

Kiger, Jack E. "An empirical investigation of NYSE volume and price reactions to the announcement of quarterly earnings." Journal of Accounting Research (1972): 113-128.

Lloyd, Stuart. "Least squares quantization in PCM." IEEE transactions on information theory 28, no. 2 (1982): 129-137.

Morgan, John, and Phillip C. Stocken. "An analysis of stock recommendations." RAND Journal of economics (2003): 183-203.

Nair, Binoy B., PK Saravana Kumar, N. R. Sakthivel, and U. Vipin. "Clustering stock price time series data to generate stock trading recommendations: An empirical study." Expert Systems with Applications 70 (2017): 20-36.

Nariswari, Talitha Nathaniela, and Nugi Mohammad Nugraha. "Profit Growth: Impact of Net Profit Margin, Gross Profit Margin and Total Assests Turnover." International Journal of Finance & Banking Studies (2147-4486) 9, no. 4 (2020): 87-96.

Pirinsky, Christo, and Qinghai Wang. "Does corporate headquarters location matter for stock returns?." The Journal of Finance 61, no. 4 (2006): 1991-2015.

Ramnath, Sundaresh, Steve Rock, and Philip B. Shane. "A review of research related to financial analysts' forecasts and stock recommendations." Available at SSRN 848248 (2008).

Robert P. Schumaker, Hsinchun Chen, “A quantitative stock prediction system based on financial news,” Information Processing & Management, Volume 45, Issue 5, 2009, Pages 571-583, ISSN 0306-4573.

Shen, Peter. "The P/E ratio and stock market performance." Economic review-Federal reserve bank of Kansas city 85, no. 4 (2000): 23-36.

T. Sanboon, K. Keatruangkamala and S. Jaiyen, "A Deep Learning Model for Predicting Buy and Sell Recommendations in Stock Exchange of Thailand using Long Short-Term Memory," 2019 IEEE 4th International Conference on Computer and Communication Systems (ICCCS), Singapore, 2019, pp. 757-760, doi: 10.1109/CCOMS.2019.8821776.

Simran Gogia, and Utkarsh Sharma. (2020) Bullz and Bearz. GitHub. https://github.com/simmsss/Bullz-and-Bearz.