

Ascertain of Rank for (NSE) Nifty 500 by Using Various Financial Filters with the Help of Python, Data from April 2021 to April 2022

Seema Sahai, Sharad Khattar and Namita Sahay

Department of Management
Amity International Business School
Amity University Uttar Pradesh
ssahai@amity.edu, seema_sahai@yahoo.com, skhattar@amity.edu, nsahay@amity.edu

Doredla Radha Krishna and Abhishek Kumar

Amity International Business School
Amity University Uttar Pradesh
doredlaradhakrishna@gmail.com, abhishekkkrsh39@gmail.com

Abstract

National Stock Exchange (NSE) is on the list of the TOP 10 stock exchange markets in the world. NSE builds trust among customers to invest in the companies which are listed with them. In India, NSE has a list of companies that are from distinct sectors like Financial, Health, Information Technology, Retail consumers, Real estate, Fast-moving consumer goods, etc. NSE has categorized the companies into Nifty 50, Nifty 100, Nifty 200, and Nifty 500. In this article, we developed a machine-learning algorithm for Nifty 500 companies that helps to create a better portfolio within the top 500 companies for the customers or clients by using various Python libraries and tools, packages which finally results to find out the rank. Lastly, this rank helps the user to figure out whether to invest in that particular company or not even though they were belongs to the top 500 in the nation.

Keywords

National Stock Exchange, Python, Nifty, India, Sectors, Rank

1. Introduction

National Stock Exchange of India Limited is located in Maharashtra, and it is the leading stock institute in India which is successfully lasting for decades with its performance, determination, and hard work to create better stocks for the users. In the year 2021, it was the world's largest derivatives exchange by the number of contracts traded based on the FIA statistics. In 1992 the NSE establishes as the first dematerialized electronic exchange in the nation. NSE is only the private body that provided the modern and fully automatic screen-based stock exchangesystem for the users to invest and perform other stock operations. So as surpassing the years people came across the stocks knowing from their friends, colleagues, relatives, or self-learning which resulted in to increase in the number of users who were investing in stocks.

These days numerous youthful age individuals were attempting to put resources into stocks as everybody's saying is to make some cash which assists them with using for different purposes in their everyday life or to meet their desires. So when we are putting away our well-deserved cash we expect consequently sum as benefits. To creates such benefits in stocks one should have to follow essential standards like breaking down, noticing, investigating, and too fundamental monetary proportions and their correlations. Many individuals lost their cash because of an absence of certainty because the amateurs don't have a clue about the essentials of how to contribute, how much need to contribute, how to choose a stock, or probably what are the contemplations to decide on a stock. People should be equipped for anticipating considering the ongoing information or investigation.

Financial Times appraises that as of April 2018, 6 crores (60million) retail investors had invested their savings in stocks in India, either through direct purchases of equities or through mutual funds.

To eradicate this problem, in this article we developed an algorithm that helps everyone to identify a good company to invest with based on the rank obtained by us using more than 100 financial filters like financial ratios, enterprise ratios, growth rate, etc with in the 500 as for this project our team chooses Nifty 500. Which is developed by using various

Python libraries and packages/tools. This document discusses in detail the methodology used for the algorithm and other concepts dealing with the obtained results.

2. Literature Review

Evaluating different contextual investigations and examining them from different foundations or elements like IPO (Initial Public Contributions), foreseeing the stock costs, understanding the financial backer's way of behaving with separate their order of the family they have a place with. Since numerous people in India who were putting resources into stocks are not rich and local area arranged so their way of behaving of examining the stocks and understanding them is unique and it varies from one companion to another. So essentially the comprehension of clients on how and what boundaries are they turning upward to for putting resources into a specific stock is genuinely significant when the analyst is concentrating on stocks. Testing the effectiveness of powerless or better execution of NSE (National Stock Trade) and BSE (Bombay Stock Trade) since people while effective money management ought to know about the essential Sensex and stages that help them not to lose cash. As NSE comprises Nifty 50, Nifty 100, Nifty 200, and Nifty 500 which states top organizations remembering all areas of India.

Understanding the various plots connected with stocks assists everybody with picturing the information and assists with distinguishing the investigation of forecasts of the stock costs. Unfamiliar monetary organizations assume a significant part in changing the Indian NSE stock cost. Without a doubt, it is a recipient for us to realize what elements show effect and how the outcomes lead to be a representation. (Goudarzi et al., 2011) Noticing the market represents many retail financial backers. For example, assuming the financial backers begin feeling apprehensive or dread putting resources into a specific area and that stock then the theory begins in a pessimistic manner that prompts lessening of the organization's stock cost. In different cases organizations begin neglecting to stay true to their commitments or lost their future tasks, any catastrophic event occurs, flighty circumstances happen like Coronavirus, Russia - Ukraine war, and so forth likewise prompts an uncommon fall in stock costs.

Finally, by this literature review, we could be able to understand the predictions and analyze the stocks and technically how the actual algorithm helps the users to invest (Hiransha et al 2011, Pandey 2013).

3. Methods

To get a more solid presentation of the extremely settled acquired position of the organizations inside the Clever 500 from the informational index that is gathered from the ticker tape. A few methods and calculations are utilized. Information can be better examined by utilizing connection plots. Before applying procedures and calculations should check whether the information is in fixed mode or nonstationary model.

3.1 Correlation Plot

Correlation is a statistical procedure accustomed to assessing a potential linear association between 2 continuous variables. it's easy each to calculate and interpret. (Ezekiel, 1930) Helps the data to spot and analyze however one variable relies on another variable (Figure 1).

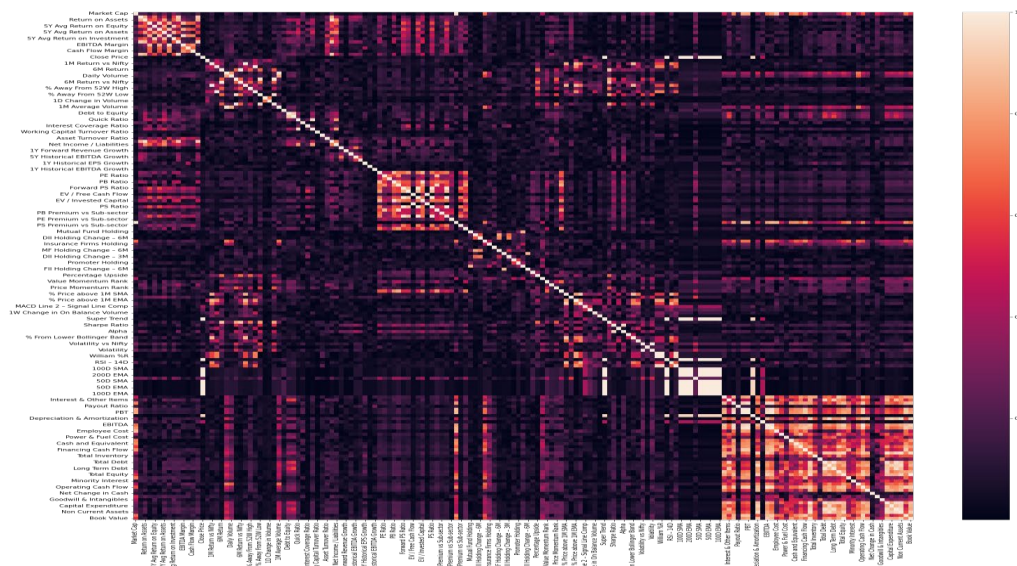


Figure 1. Heatmap of correlation

From the above figure 1, one can observe the diagonal with dots which shows it correctly mapped all the values Correlation heatmap of overall dataset after filling and removing missing or null value. Here, blocks shown in white color denotes value 1 i.e., those two corresponding variables have correlation coefficient of 1, similarly black color block denotes correlation coefficient equal to 0 means those two features are totally unrelated. While, other colors in between white and black denotes intermediate correlation coefficient between 0 and 1.

3.2 Heatmap of Correlation

Graphical representation of data where values are depicted by color. It is obvious to envision complex data also. It reveals how each variable is correlated to every individual variable. If the correlation value is high betwixt two variables, then it says that two variables are strongly correlated with each other.

The first technique in this methodology is correlation which helps to reduce the filters(columns) to upgrade the model for better output.

Pearson Correlation Method.

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}$$

The Pearson's correlation method is used when you have two quantitative variables, and you wish to see whether there is a linear relationship between those two or not. It also measures the strength of the linear relationship between two variables. It has values betwixt -1,0, +1. -1 means total negative correlation, 0 means no correlation, and +1 means total positive correlation.

After doing correlation, based the person values which have more than 0.95 value are eliminated from the data set.

Result: - the total number of (Filters)columns now is 128, and the total number of rows is 204.

The Second Technique in this methodology is "Normalizing" the updated data set which consists of 204 rows and 128 columns. By using a statistical library named SciPy which is prominently used to import statistics data. Finally helps to normalize the data. (Caloia et al., 2019)

Normalization:- It is the process of organizing the data in the given data set. The two main objectives of the normalizing process are to eliminate the redundant data and make sure that data dependency makes some sense.

Here, normalization helps the model to analyze the data on a unique scale. As the distinct filters consist of various measurements. So by using this one can assess the information ease with the help of this scale.

$$Z = \frac{x - \mu}{\sigma}$$

For example, the following image shows detailed information about the correlation plots and normalized values. Likewise, profiling consists of data related to all filters. This helps the user to see which company performs well or poorly based on the required filter. Since retail investors can also be categorically divided into two types, they are part-time investors and full-time investors. So when it comes to day trading, full-time investors will consider stocks based on many parameters, but these plotted charts will also help beginners understand them and analyze them thoroughly to decide whether to invest or not (Figure 2).

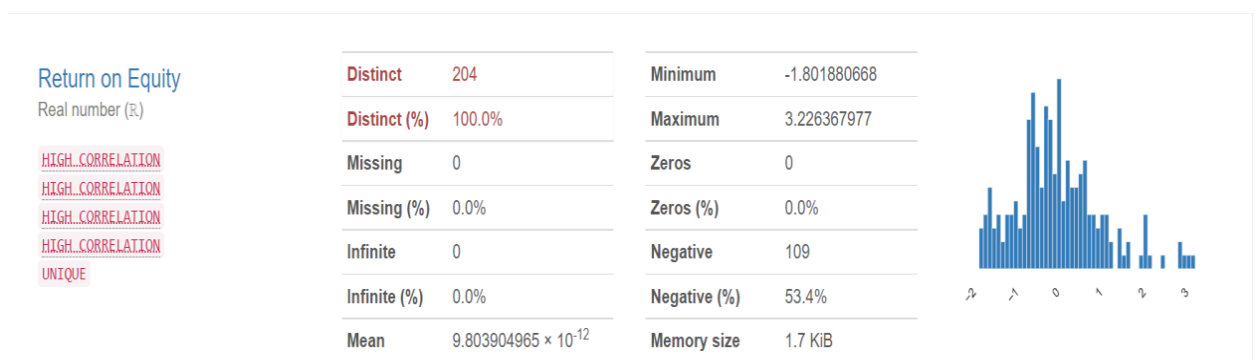


Figure 2. Return on Equity

Return on Assets

Real number (R)

HIGH CORRELATION
 HIGH CORRELATION
 HIGH CORRELATION
 HIGH CORRELATION
 UNIQUE

Distinct	204	Minimum	-1.612844265
Distinct (%)	100.0%	Maximum	2.918967195
Missing	0	Zeros	0
Missing (%)	0.0%	Zeros (%)	0.0%
Infinite	0	Negative	111
Infinite (%)	0.0%	Negative (%)	54.4%
Mean	$-4.353815783 \times 10^{-18}$	Memory size	1.7 KiB

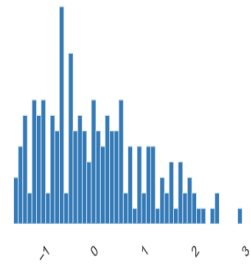


Figure 3. Return on Assets

The above diagrams (Figure 3) are representative of summary function of a particular feature as for e.g. we have taken return on assets and return on equity features. Summary function gives information regarding mean, maximum, minimum values of particular column.

The third technique in this methodology is "data criteria" which shows us what the individual needs to maximize or minimize to analyze the company based on the filters. Let's consider the return on investment (ROI) as a variable (filter) to analyze companies. ROI should be maximized whenever a company is to be presented as profitable or doing well in comparison to others, forcing investors to make a decision. So now the ROI is rated MAX cedar. Now let's look at the liabilities given below. If a company has more debt, investors won't show much interest in investing. The responsibility lies under the MIN cedar. It divides all filters into MAX and MIN based on their basic definition and meaning.

RULE: Tool/Algorithm/Function/Method = "WEIGHTED SUM"

Let A be the attribute that we want to maximize

$[a_1, a_2, a_3 \dots a_n]; 1 \leq i \leq n$ Maximize(a_i)= maximizeFunc(a_i, A)

Maximize(a_i)= maximizeFunc(a_i, A)

$$\text{Where maximize fun}(a_i, A) = \begin{cases} \frac{a_i}{\text{sum}(A)}, & \text{sum normalization} \\ \frac{a_i}{\text{max}(A)}, & \text{max normalization} \\ \frac{a_i - \text{min}(A)}{\text{max}(A) - \text{min}(A)}, & \text{max - min scaling} \\ \dots & \dots \end{cases}$$

Lets A be an attribute that we want to minimize

$[a_1, a_2, a_3 \dots a_n]; 1 \leq i \leq n$

minimize (a_i)= minimizeFunc(a_i, A)

$$\text{minimizeFunc}(a_i, A) = \begin{cases} \frac{1}{\text{maximizeFunc}(a_i, A)}, & \text{inverse} \\ 1 - \text{maximizeFunc}(a_i, A), & \text{subtract} \\ \dots & \dots \end{cases}$$

Apply the weights.

Weight Modification: Let W be the vector that denotes the vector weighing each attribute

Weight vector = $[\omega_1, \omega_2, \omega_3 \dots \omega_m]; 1 \leq i \leq m$

$$\omega_i = \begin{cases} \omega_i, & \text{as it is} \\ \frac{\omega_i}{\text{sum}(W)}, & \text{sum normalization} \\ \frac{\omega_i}{\text{max}(W)}, & \text{max normalization} \\ \dots & \dots \end{cases}$$

4. Data

Initially, the data was collected from the TickerTape platform for Nifty 500 which comprises 1 year of data from April 2021 to April 2022. The data set has entirely 501 rows and 218 columns. Columns are namely filters: Profitability filters, Financial Ratios, price and Volume, Growth, Valuation, Analyst Ratings, Future and Options, Ownership, Technical Indicators filters, Income statements, Balance sheets, Cash Flows, etc. (All together by summing each internal filters the raw data set consists of more than 200 filters).

For instance, let's consider Profitability as it is one of the filters. The internal filters which are under the profitability are namely: Return on Equity, Return on Assets, Net Profit margin, Return on Investment, EBITDA, Cash flow margin, Costs of goods sold, ROCE, etc. Similarly, all the above filters consist of numerous internal filters which all together sum up to more than 200 resulting in initially 218 columns.

Rows: consists of 500 companies.

5. Results and Discussion

This is unsupervised data, the data set exists with a greater number of empty cells which means missing data is high for some companies (Rows). To analyze the large data by considering many variables then it would be difficult to perform the analysis and develop the algorithm also.

5.1 Graphical Results

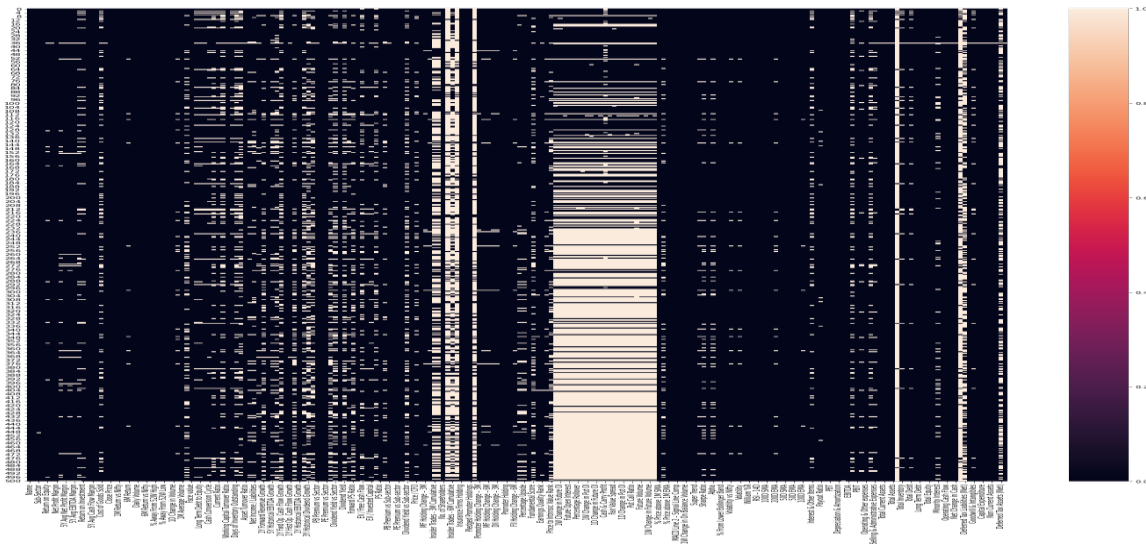


Figure 4. Missing values heat map

Overall representation of missing or void cells within dataset. Here, white color block represents empty cells i.e. missing value while, black color blocks represents cells having any values filled within it (Figure 4).

So to defeat this the information pre-handling outfits a more reliable portrayal of the achievement of the model. So the information cells which are vacant of more than 20% we eliminated to make a superior model. (Nayak 2016). Then, at that point, the number of segments coming about is equivalent to 169 which comprises 166 channels. After eliminating the invalid upsides of the information then the out number of organizations is 204.

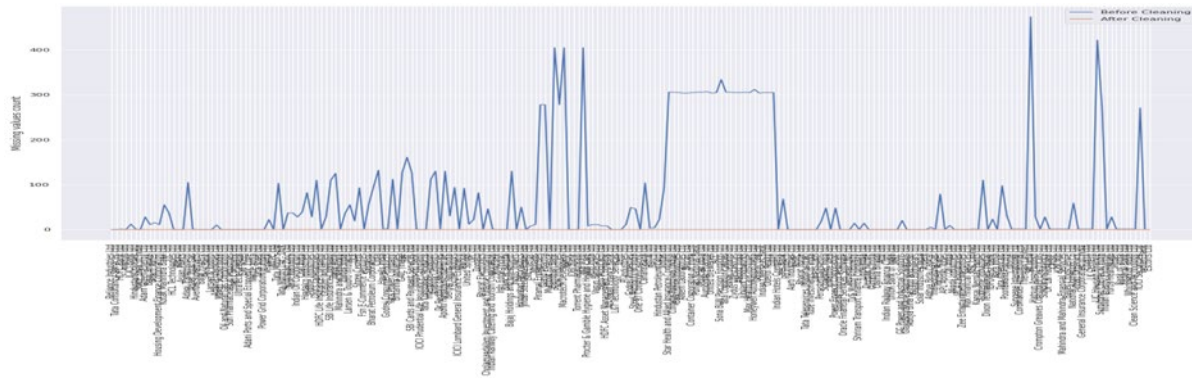


Figure 5. Missing Values Count

The above Figure 5, line diagram represents the count of missing i.e empty cells in overall dataset representing names of features or columns on x-axis and empty cells frequency of that column on y-axis

If we observe the above figure there is missing data in many columns which is the reason for not accepting the flow and it does not lead to obtaining an efficient model. To eradicate this issue, data can be filled by interpolation methods or imputing data by using mean, median, and mode.

Here the Mean is used to impute NAN values in the current data set (Figure 6).

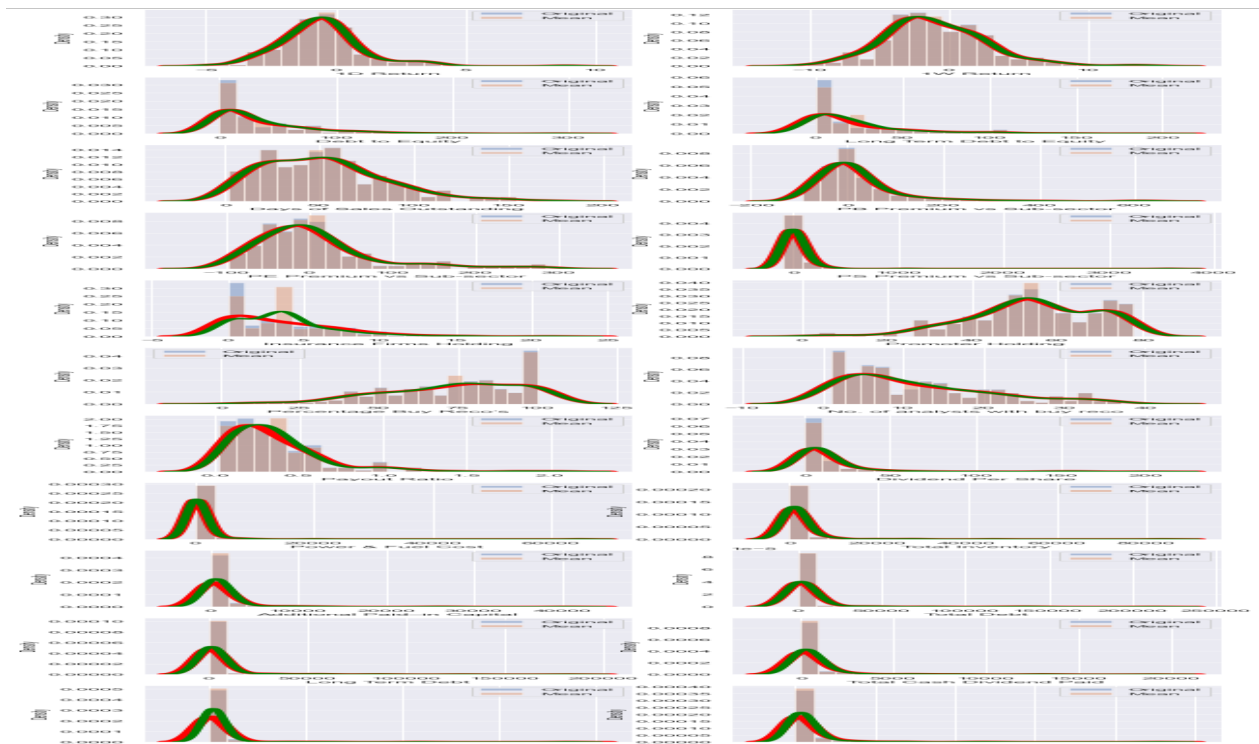


Figure 6. Pattern of data after 0 value imputation with mean

Overall distribution of values in each column after filling of cells having values equal to 0 with means of respective column or various filters

As shown in the above figure 3 there are no missing values in the data, the plot helps us to analyze the respective filter following whether any null values exist or not (Figure 7).

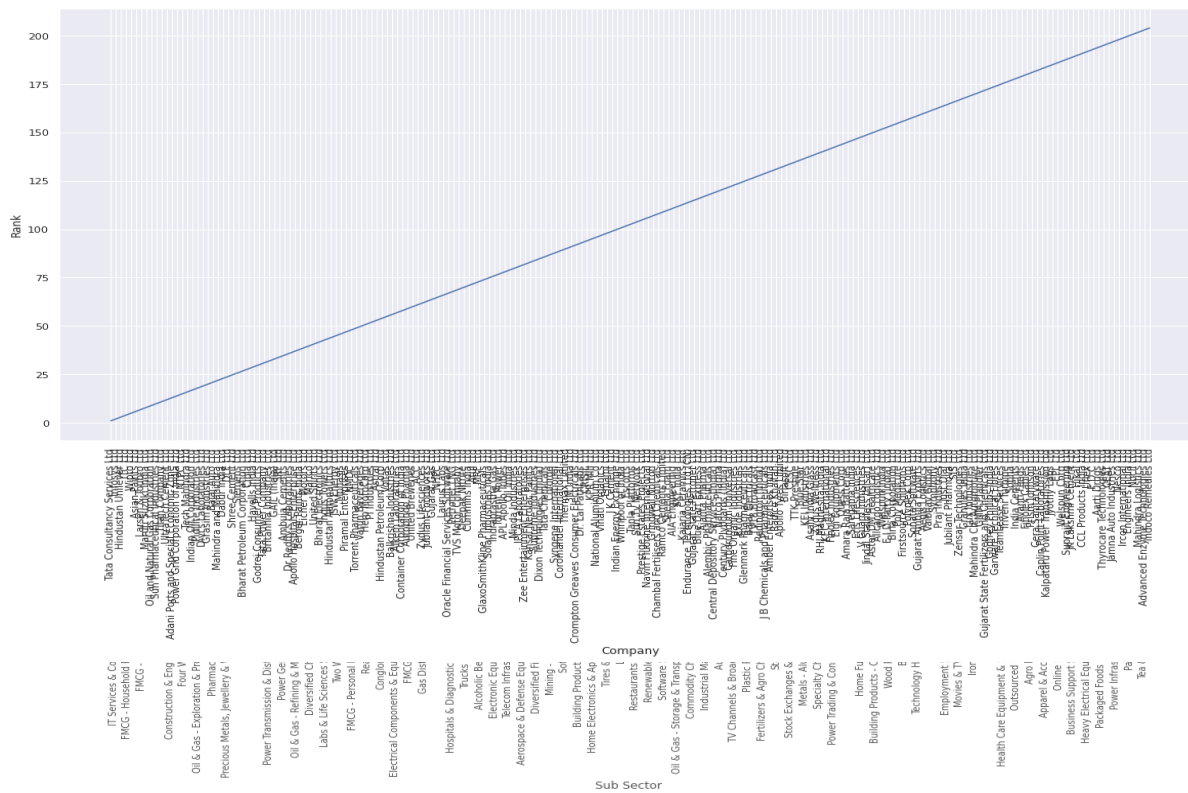


Figure 7. Sector-wise rank

The above diagram (Figure 7) displays ranks of various sectors present within the dataset. Each sector consists summarized/average rank of all companies belonging to particular sector. After done with the weighted sum as stated in the methodology then finally rank is obtained with help of all the above techniques and step by step analyses. The chart above depicts the rankings of several sectors within the Nifty 500. Because this graph is simple enough for a layperson to grasp, a retail investor may simply assess the data. The present statistic depicts which sectors an investor may truly invest their savings or hard-earned money in order to grow their wealth. The image above depicts the firm rank derived after employing several methodologies. This business ranking assists investors, particularly those who are new to investing and want to invest in stocks with a large quantity of money. As one can simply grasp the line graph that shows-axis-rank, x-axis-company name, one can quickly decide in which firm they want to invest based on their financial history.

5.2 Future Scope

As our mission and aim are to encourage more young people to participate in stocks, we are always striving to improve the data display and existing algorithm. As our future scope is to forecast organizations utilizing different financial-related parameters (Jiao et al., 2017) As a result, it aids consumers in planning their portfolios (Satyaprasad 2018) ahead of time.

6. Conclusion

Individual or retail investors, because many investors nowadays are young. If someone wants to invest in stocks, they need to understand the fundamentals of stock analysis. (Padhan,2007). However, being a youth conveys that the population is just focused on achieving their want, which is to get money, but they will not focus on the research portion and making themselves by educating or learning the information. So, based on the RANK we collected, our results assist them in finalizing the sector and specific they are interested in. The

goal is to encourage young people to invest in equities and not to let them go broke. In terms of our goal, we will always strive to improve the approaches that make the user interface more user pleasant and continue to work on data visualization to make it easier in response to feedback in the upcoming phase of the execution.

References

- Aparna Nayak, M. M. Manohara Pai, Radhika M. Pai, Prediction Models for Indian Stock Market, *Procedia Computer Science*, Volume 89, Pages 441-449, ISSN 1877-0509, 2016.
- Caloia, F. G., Cipollini, A., and Muzzioli, S., How do normalization schemes affect net spillovers? A replication of the Diebold and Yilmaz study. *Energy Economics*, 84, 104536, 2019.
- Ezekiel, M. (Methods of correlation analysis. Wiley. 1930.
- Goudarzi, H. and Ramanarayanan, C. S., Empirical Analysis of the Impact of Foreign Institutional Investment on the Indian Stock Market Volatility during World Financial Crisis 2008-09, 3(3),214–226, 2011.
- Hiransha M, Gopalakrishnan E.A., Vijay Krishna Menon, Soman K.P., NSE Stock Market Prediction Using Deep-Learning Models, *Procedia Computer Science*, Vol.132, pp. 1351-1362, 2018.
- Padhan, P. C., The nexus between the stock market and economic activity: an empirical analysis for India. *International Journal of Social Economics* , 34(10),741-753, 2017.
- Pandey, Anand, Efficiency of Indian Stock Market, 2003.
- Satyaprasad, B. G., and H, A. P., A Study On Optimal Portfolio Construction Of Fmcg And Pharmaceutical Sector Stock With Reference To BSE. *International Journal of Social Sciences and Economic Environment*, 3(1), 16–23, 2018.
- YJiao and J. Jakubowicz, "Predicting stock movement direction with machine learning: An extensive study on SandP 500 stocks," 2017 IEEE International Conference on Big Data (Big Data), 2017, pp. 4705-4713, 2017.

Biographies

Dr Seema Sahai is a Professor of Management and a distinguished academician, researcher, trainer and management consultant with a brilliant track record throughout her education and career. She has an experience of 26 years covering areas of strategic management, international business, project execution, project management, behavioral sciences, ICT, consultancy, research, teaching and training. She has been teaching papers in the field of Information systems and Decision Sciences and General Management. She is also HOD IT and Operations. Her experience varies from Corporate to Academics. She has worked on various prestigious projects like Computerization of the International Banking System of Oriental Bank of Commerce and State Bank of Patiala. She has presented many papers in International and National Conferences, including one at Beijing University. She has done her Phd in e-Governance using IT tools. She also has many Scopus and UGC publications to her credit.

Col (Dr) Sharad Khattar is currently serving as a Professor in Amity University, Noida since Dec 2011. His fields of interest are Operations Management , Decision Sciences, Supply Chain Management and Service quality Management. He has number of research papers, copy rights and patents against his name. His educational qualifications includes PhD in Management (Amity University), M.Tech (Industrial Engineering and Management) from IIT, Kharagpur, MBA (Finance) , B.Tech (Electronics) from Military College of Electronics and Mechanical Engineering, B.Sc from JNU, Delhi and PG Diploma in Management from IMT, Ghaziabad. Before taking up his present appointment he has served in Armed Forces as an equipment advisor in engineering corps, scientist in Defense and Research Development Organization (DRDO) and also had a stint in the Corporate.

Prof.(Dr.) Namita Sahay is a Professor and Head of Department of Finance at Amity International Business School, Amity University, Noida and has over Twenty Four(24) years of teaching experience to her credit. She is an Editorial Board Member of Scopus and ABDC Indexed Journal-Finance India and India International Journal of Advanced Multidisciplinary and Scientific Research (IJAMSR). She has sought her education from some of the premier institutions in India – University of Pondicherry; University of Allahabad; St. Mary’s Convent Inter College (Allahabad). She has over sixteen (16) courses like Financial Management; Mergers and Acquisitions; Management of Financial Services; International Finance and Forex Management; Security Analysis and Portfolio Management; Behavioral Finance; Financial Engineering; Banking Environment; Banking and Financial Institutions; Indian Financial System; Financial Risk Management and Derivatives; Insurance and Risk Management using a very unique teaching pedagogy of lectures, case studies and videos. She has also taken online lectures for Pan African E-network Project, of Amity University, Noida. Student feedback for her lectures has been extremely positive.

She was also a visiting faculty as a part of 3C Programme with California State University, San Bernardino in June 2016.

She has authored and co-authored around thirteen (13) books in Finance, published around twenty (25) research articles in a number of national and indexed/peer reviewed international Journals and has seven (7) book reviews to her credit.

She has conducted three (3) Management Development Programmes, five (5) Workshops in Finance for Wharton Business School Investment and one (1) value added course (VAC) of 30 hrs. She has also organized three (3) Finance Summits with National and International speakers and many Industry Interaction Sessions with senior official from RBI, Television analyst and Vice President (Finance).

Doredla Radhakrishna, A philonoist and passionate graduate student to learn something new and to adopt it. Focused on implementing innovative ideas and working on real-time projects which can contribute to society.

Abhishek Kumar, an MBA student with strong analytical capabilities along with management skills specialized in business intelligence and Data analytics. Aiming to work for Data analytics field in healthcare sector to get practical hands-on real life projects, improve my skills and contribute to society.