Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 11, No. 2, November : 2023

CLUSTER BASED GROWTH-ORIENTED EQUITY DIVERSIFIED SCHEMES OF MUTUAL FUND ON THE BASIS OF RETURN AND RISK

Ms. S. Dhanalakshmi , Research Scholar, Department of Science, Madurai Kamaraj University Madurai – 625 021. Tamilnadu. Email ID: dhanamdhanah99@gmail.com
Dr. R. Murugesan , Associate Professor, Department of Science, CPA College, Bodinayakanur - 625 513. Tamilnadu. Email ID: rmncpa90@gmail.com

ABSTRACT

A mutual fund is a professionally managed investment program that pools a group of investors' funds and invests them in stocks, bonds, and other securities. These funds are typically handled by asset management firms. Clustering techniques are now frequently employed in the banking sector to address financial issues. This study combines the clustering techniques of hierarchical clustering and principal component analysis to examine the performance of a growth-oriented mutual fund scheme under Net Asset Value. On the basis of the clustering results, a clustering-based model is built at the same time to forecast the NAV points of an Indian mutual fund scheme. This paper's main contribution is to present the top mutual fund investment plans based on NAV and mutual fund type (Large Cap, Mid Cap, Small Cap, Multi Cap).

INTRODUCTION

A mutual fund functions as a financial bridge that enables a group of investors to pool their funds with a defined investment goal, and the fund management then invests this aggregated capital in particular assets. The benefits of mutual funds include expert investment management, high liquidity of capital flow, spreading investment risks, legal tax savings, and a variety of investment aims, among others, making them one of investors' preferred investment vehicles. This article predicts the Net Asset Value of Indian mutual funds and evaluates rate of return using a clustering algorithm, a new evolutionary calculating method. In order to do Principal Component Analysis and Hierarchical Clustering in this study, we first gather data from Indian open-end balanced funds. As investment objectives, we choose funds with a technical efficiency value of 1, and from May 2020 to June 2022, we gather the underlying fund's net market value. Then, using hierarchical clustering and PCA, the mutual fund net worth clustering model is generated.

LITERATURE REVIEW

The foundation for fund evaluation is fund classification. Due to their unique characteristics like risk and return, many types of funds require various analysis techniques and evaluation criteria. So, fund classification assures that fund evaluation is efficient and comparable. Ex ante classification and ex post classification are the two types of fund classification. According to the fund's investment objectives and strategies, which are detailed in the fund issuance announcement, the ex-ante classification process establishes the fund category. The defined information, however, usually deviates from the initial agreements in the process. After regressing the fund's net value using William Sharp's attribution methodology, DiBartolomeo et al. [6] find that more than 40% of the stock funds are misclassified. They contend that the main causes of the misclassification are the imprecision of the ex-ante classification method and the ex post deviating manipulation by fund managers as a result of peer pressure. Through factor and cluster analyses, Luo et al. [7] categories 54 funds listed in China, and they discover that roughly 40% of the funds do not follow the investing strategy outlined in their prospectus. Ex post classification, on the other hand, categorizes different types of funds based on their performance following fund operation and the features listed in the issuance announcement. Brown et al.'s [8] use of a factor model to map the nonlinear aspects of fund returns into the mainline of investment managers' style is an enhancement to this categorization method Principal component analysis (PCA) is used by Kim et al. [9] to choose additional market features and categorize funds

using these newly discovered factors. The collinearity of components in multiple regressions is one shortcoming of the ex post classification method, among others.
Since machine learning can capture nonlinear variables and is independent of data factors like sample size under unsupervised learning, it fortunately mitigates the shortcomings of the classic fund classification methods. By using K-means clustering to categorize funds, Marathon et al. [10] discovered that 43% of the fund samples were at odds with the investment kinds for which the funds were first characterized. Additionally, they discover that many fund categories classified using conventional approaches have remarkably comparable risk and return characteristics, which indicates the use of categorization analysis will simplify fund management. According to Lajbcygier et al. [11], rather than being rigorously split, the boundaries of funds with various types should be continuous. As a result, they employ a flexible clustering technique based on fuzzy C-means and discover that this technique can produce improved classification outcomes. Menardi et al.'s [12] two-step clustering method involves first utilizing PCA to lower the dimensionality of 24 fund characteristics, and then using hierarchical clustering to divide 1436 public funds into those 24 categories of characteristics. Moreno et al. [13] categorize 1,592 funds from the Spanish market using self-organizing mapping neural network (SOM) for the extraction of nonlinear properties and discover that SOM may significantly minimize misclassification when compared to K-means clustering.

DATA

The majority of the data for this article came from an Indian mutual fund database. The information gathered includes complete samples of the Indian mutual fund classification's stock funds, hybrid funds, and bond funds. From May 2020 to June 2022, a total of 3535 funds are used. The information was also gathered from numerous AMC, AMFI, money control, etc. websites. Over a two-year period, the NAVs of the sample mutual fund schemes were gathered on a monthly basis. The attributes include scheme name, plan category, crisil rank, starting year, shareholder level, fund manager rank, expense ratio, NAV, standard deviation, beta, sharpe ratio, jenson's alpha, and treynor's ratio as input values.

ANALYSIS OF DATA

Using categories like Large Cap, Small Cap, Mid Cap, and Multi Cap, one can categorize mutual funds based on their key investment features and investment objectives. Shareholders are those who own at least one share of a company's stock or investment in a mutual fund.
Mutual funds are rated by Morningstar on a scale of one to five stars based on their fund manager. These rankings are based on how the fund has done in comparison to other funds in the same category after taking risks and expenses into account.
Ratio of institutional investment - This ratio is important because it can demonstrate investors the consistency of returns over a given time period and help them determine the skill level of an asset or investment manager.
Expense Ratio - From the perspective of the investor, an actively managed portfolio should have an expense ratio of between 0.5% and 0.75%. A high expense ratio is one that is larger than 1.5 percent. Generally speaking, the expense ratio for mutual funds is higher than the expense ratio for ETFs.
NAV - A mutual fund scheme's success is shown by its NAV per unit. The NAV per unit is calculated by dividing the market value of the securities in a scheme by the total number of units in the scheme as of a particular date.
Standard deviation: A measure of how much a mutual fund scheme's returns are likely to vary from its average yearly returns is a number (given as a percentage) called the standard deviation.
Jenson's measurement, also known as Jenson's alpha, is a risk-adjusted performance indicator that measures whether an investment's average return is higher or lower than that predicted by the capital asset pricing model (CAPM), given the portfolio's beta and the overall market return.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 11, No. 2, November : 2023

Beta : The portfolio returns minus the risk-free rate of return divided by the portfolio beta value. Over and beyond the earnings earned risk-free, the investor receives an excess return. A Treasury bill or government security is regarded as the risk-free investment return.

Research Methods

We develop the clustering model and the prediction model as part of the model framework for the inquiry.

Models for data clustering they models employed in this paper are principal component analysis and hierarchical clustering approaches. In this work, rolling yield data from funds with different maturities that were used in the initial classification step were dimension-minimized using PCA.

PCA

Principal components analysis (PCA), a dimension reduction approach, converts the original feature space into a whole new feature space. Usually, it works as a data preprocessor. A scenario with n fund samples and 2100 qualities per fund sample is presented. Among these are the tremendous amount of noise and the decrease in computational resources. The number of eigenvalues used and the PCA's reducibility are listed in Table 4. The PCA method is highly reducible, and the more rolling days are selected to keep the number of eigenvalues constant, the higher the reducibility.

Hierarchical Clustering

It is used to classify data into clusters based on the distance between data points without having to know how many clusters there will be in advance. The data analysis and clustering process employs a variety of methodologies to aid in the discovery of patterns and structures. One way to make sure that various features or variables are comparable and easy to understand is to scale the data. The data are normalized and standardized using this method.

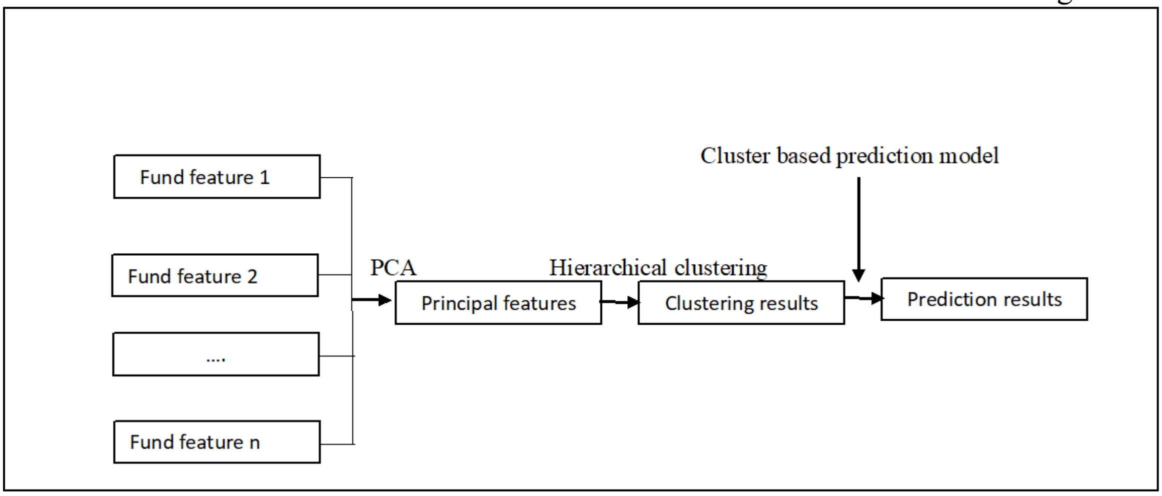

Figure 1: Model Framework

Results

Clustering has been done, and the outcomes are documented, for each model and set of hyperparameters. On the basis of the NAV points, we contrast these models.

Item Reliability Statistics

				If item dropped	
	Mean	**SD**	**Item-rest correlation**	**Cronbach's α**	**McDonald's ω**
NAV	28.47	5.10	-0.220	-0.401	0.00
Fund_Manager_Rank	4.72	2.79	-0.220	-0.120	0.00

Table 1. Reliability Statistics **PCA**

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 11, No. 2, November : 2023

The categorization outcomes of the initial PCA from several dimensions are displayed in Table 3. Starting with each category, we look at the average ratio between return value and risk measurement of fund value.

Hierarchical Clustering

In hierarchical clustering, four divisions are the ideal number. Each cluster to which the fund belongs is represented by a colored line in Figure 2, and Table 2 displays the mean and variance of return for each cluster of funds. Figure 2 depicts a cluster that includes The color-lines for the Small cap are blue, the Mid cap is grey, the Multi cap is orange, and the Large cap is green. According to hierarchical clustering, each fund group has been divided into its own subgroup after being integrated into one using the k-means algorithm. Observe moreover that although the means of returns differ for large cap funds, the variation of returns is substantial.

Descriptives								
	Investment_Style	**NAV**	**Standard_Deviation**	**Beta**	**Sharpe_Ratio**	**Jension's_Alpha**	**Treynor's_Ratio**	**Crisil_Rank**
N	3035	3035	3035	3035	3035	3035	3035	3035
Missing	11	11	11	11	11	11	11	11
Mean	1.50	198	21.0	0.901	0.767	4.40	0.180	4.33
Median	2	75.1	20.8	0.920	0.700	2.92	0.160	4
Standard deviation	0.502	346	1.28	0.0639	0.267	4.68	0.0689	0.473
Minimum	1	17.0	16.7	0.720	0.160	-7.16	0.0400	4
Maximum	2	2315	25.3	1.04	1.69	22.6	0.440	5

Table 2. Mean and variation of fund returns

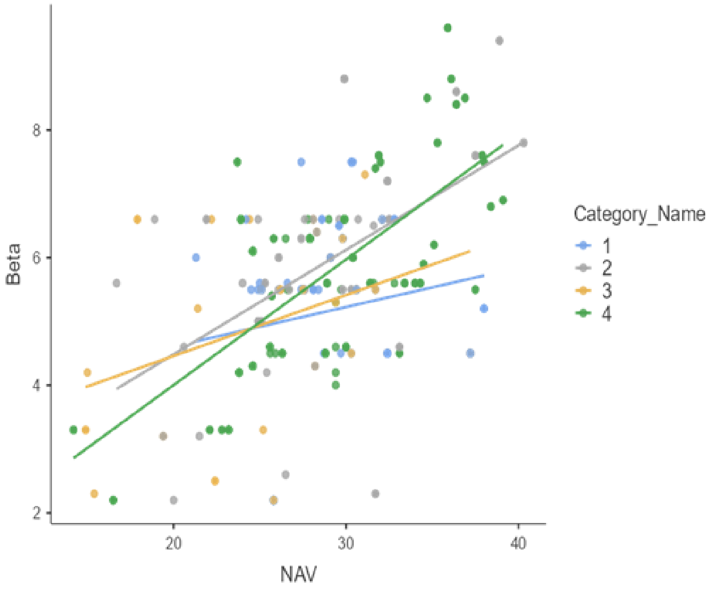

Figure 2. Hierarchical Clustering Model

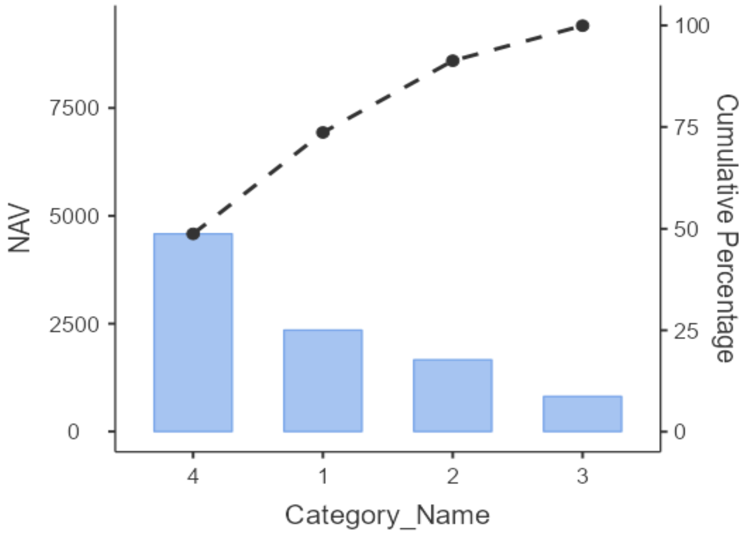

Figure 3. Hierarchical Model for comparing NAV and Category Type.

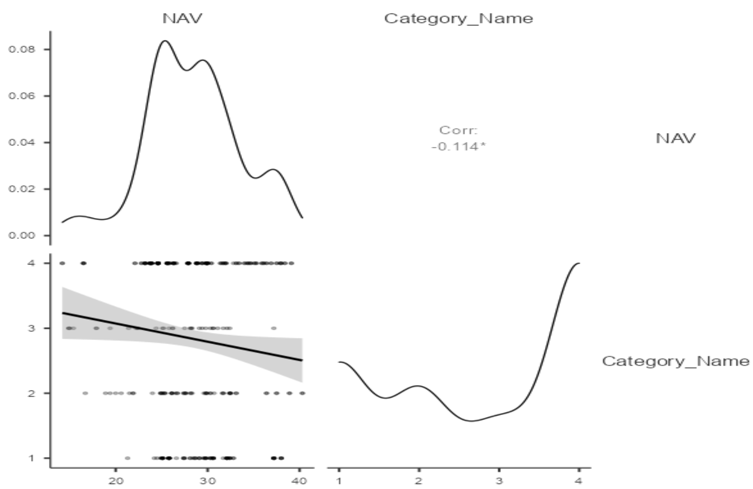

Figure 4. Correlation for category wise NAV points

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 11, No. 2, November : 2023

Model Coefficients

		95% Confidence Interval					
Predictor	Estimate	Lower	Upper	SE	Z	P	Rate ratio
Intercept	3.3929	3.353	3.4334	0.0206	164.50	< .001	29.753
Category_Name:							
2 – 1	-0.0375	-0.100	0.0253	0.0320	-1.17	0.242	0.963
3 – 1	-0.1578	-0.238	-0.0781	0.0407	-3.88	< .001	0.854
4 – 1	-0.0505	-0.100	-8.07e–4	0.0254	-1.99	0.046	0.951

Table 3. Prediction result for categories

Eigenvalues

Component	Eigenvalue	% of Variance	Cumulative %
1	2.15585	53.8962	53.9
2	1.44720	36.1799	90.1
3	0.39472	9.8681	99.9
4	0.00224	0.0559	100.0

Table 4. Number of eigenvalues

Conclusion and future work

We focused on the mutual fund returns in this paper and present two clustering methods based on investment similarity. Mutual funds that employ a different operation from the categorized technique can be found using the suggested procedures. One benefit of the suggested method is that new mutual funds can be classified because historical performance data is not required. In this work, we effectively clustered mutual funds using a variety of clustering algorithms, including Hierarchical clustering and PCA. The key findings of this study are outlined in comparison to how the PCA method and hierarchical clustering method produce probabilities that the funds fall into particular categories. With this strategy, funds are categorized according to their risk and return, and the result is a large capital fund with the highest return value. The following stage will be to create an effective clustering model to analyze and group users into distinct groups based on the similarities observed in their data points. To establish which of the offered approaches performs best on the provided dataset, further analysis of the suggested methods is required. Based on shared features or characteristics across users within a particular feature space, the best clustering algorithm can successfully group people together.

Reference

1. Strader T. J., Rozycki J. J., Root T. H., Huang Y. H. J. Machine learning stock market prediction studies: review and research directions. *Journal of International Technology and Information Management* . 2020;28(4):63–83. [Google Scholar]

2. Yun K. K., Yoon S. W., Won D. Prediction of stock price direction using a hybrid GA-XGBoost algorithm with a three-stage feature engineering process. *Expert Systems with Applications* . 2021;186 doi: 10.1016/j.eswa.2021.115716.115716 [CrossRef] [Google Scholar]
3. Zhong X., Enke D. Predicting the daily return direction of the stock market using hybrid machine learning algorithms. *Financial Innovation* . 2019;5(1):p. 4. doi: 10.1186/s40854-019-0138-0. [CrossRef] [Google Scholar]
4. Pierdzioch C., Risse M. A machine-learning analysis of the rationality of aggregate stock market forecasts. *International Journal of Finance & Economics* . 2018;23(4):642–654. doi: 10.1002/ijfe.1641. [CrossRef] [Google Scholar]
5. Jiang W. Applications of deep learning in stock market prediction: recent progress. *Expert Systems with Applications* . 2021;184 doi: 10.1016/j.eswa.2021.115537.115537 [CrossRef] [Google Scholar]
6. Dibartolomeo D., Witkowski E. Mutual fund misclassification: evidence based on style analysis. *Financial Analysts Journal* . 1997;53(5):32–43. doi: 10.2469/faj.v53.n5.2115. [CrossRef] [Google Scholar]
7. Luo Z., Zhang Z. How to classify funds: a new research perspective. *Financial Science* . 2004;(03):18–20. [Google Scholar]
8. Stephen J., Brown Mutual fund styles. *Journal of Financial Economics* . 1997 [Google Scholar]
9. Kim M., Shukla R., Tomas M. Mutual fund objective misclassification. *Journal of Economics and Business* . 2004;52(4):309–323. [Google Scholar]
10. Marathe A., Shawky H. A. Categorizing mutual funds using clusters. *Advances in Quantitative Analysis of Finance and Accounting* . 1999;7(1):199–204. [Google Scholar]
11. Lajbcygier P., Yahya A. *Soft Clustering for Funds Management Style Analysis: Out-of-Sample Predictability* . New York, NY, USA: Social Science Electronic Publishing; 2008. [Google Scholar]
12. Menardi G., Lisi F. *Double Clustering for Rating Mutual Funds* . 2015. [Google Scholar]
13. Moreno D., Marco P., Olmeda I. Self-organizing maps could improve the classification of Spanish mutual funds. *European Journal of Operational Research* . 2006;174(2):1039–1054. doi: 10.1016/j.ejor.2004.12.018. [CrossRef] [Google Scholar]
14. Schneburg E. Stock price prediction using neural networks: a project report[J] *Neurocomputing* . 1990;2(1):17–27. [Google Scholar]
15. Kimoto T., Asakawa K., Yoda M. Stock market prediction system with modular neural network. Proceedings of the 1990 IJCNN International Joint Conference onNeural Networks; June 1990; California, CA, USA. IEEE; [Google Scholar]
16. Khashei M., Hejazi S. R., Bijari M. A new hybrid artificial neural networks and fuzzy regression model for time series forecasting. *Fuzzy Sets And Systems* . 2008;159:769–786. [Google Scholar]
17. Liu S., Chao Z., Ma J. CNN-LSTM neural network model for quantitative strategy analysis in stock markets. Proceedings of the International Conference on Neural Information Processing; November 2017; Guangzhou, China. Cham: Springer; [Google Scholar]
18. Li C., Zhang X., Qaosar M. Multi-factor based stock price prediction using hybrid neural networks with attention mechanism. Proceedings of the International Conference on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech; August 2019; Fukuoka, Japan. [Google Scholar]