UtilitasMathematica

ISSN 0315-3681 Volume 122, 2025

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

1.S.DHANALAKSHMI, 2. R.MURUGESAN

1.*Department of Computer Science, Madurai Kamraj University, Madurai, India*.
dhanamdhanah99@gmail.com
2.*Department of Computer Science, C.P.A College, Bodi, India.*
rmncpa90@gmail.com

ABSTRACT

A mutual fund is a professionally run investment program that invests in stocks, bonds, and other assets by pooling the money of several individuals. Asset management companies usually handle these funds. 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

When a collection of depositors pond their money with a specific asset neutral, a mutual fund acts as a monetarist bridge that allows the fund management to invest the combined capital in certain assets. Mutual funds are single of the maximum popular investment vehicles among investors due to their many advantages, which include professional investment management, high capital flow liquidity, risk spreading, legal tax savings, and a range of investment objectives. This article uses a clustering algorithm, a novel evolutionary intention technique, to forecast the net asset value of Indian mutual funds and evaluate rate of return. In this work, we initial gather data from Indian open-end balanced funds in order to do Principal Component Analysis and Hierarchical Clustering. Our investment objectives are funds with a technical performance value of 1, and between May 2020 and June 2022, we gather the actual fund's net market value. PCA and hierarchical clustering are then used to generate the mutual fund net worth clustering model.

LITERATURE REVIEW

Fund classification serves as the basis for fund evaluation. Numerous fund kinds necessitate dissimilar analysis approaches and estimation standards because of their separate features, such as risk and return. As a result, fund classification guarantees effective and comparable fund evaluation. Fund classification can be classified into two categories: ex ante and ex post. The ex-ante classification procedure creates the fund category based on the investing goals and strategies of the fund, which are described in the fund issuance announcement. However, the defined information typically differs from the process's original agreements. More than 40% of stock funds are misclassified, according to DiBartolomeo et al. [6], who used William Sharp's attribution methodology to regress the fund's net worth. They argue that the ex-ante categorization method's imprecision and fund managers' ex post deviation manipulation due to peer pressure are the primary causes of the misclassification. Luo et al. [7] categorize 54 Chinese funds using feature and cluster analysis and find that almost 40% of the assets do not adhere to the investing strategy specified in their prospectus. On the other hand, ex post classification groups various fund kinds conferring to their routine after fund operations and the characteristics mentioned in the issuance announcement. This categorization technique is improved by Brown et al.'s [8] use of a factor model to translate the nonlinear features of fund results into the ordinary of investment managers' chic. Kim et al. [9] employ principal component analysis (PCA) to choose other market characteristics and classify funds based on these recently identified parameters. The ex post classification method has several drawbacks, one of which is the collinearity of components in numerous regressions.

The drawbacks of traditional fund classification techniques are thankfully lessened by machine learning, which can capture nonlinear characteristics and is independent of data factors like illustration size below unsupervised learning. Marathon et al. [10] classified funds using K-means clustering and found that 43% of the fund trials did not match the investment categories for which the funds were first defined. Furthermore, they novelty that a large number of fund categories that are categorized using traditional methods have strikingly parallel risk and return appearances, suggesting that categorization analysis will make fund management easier. According to Lajbcygier et al. [11], the restrictions of funds of different sorts should be continuous rather than strictly divided.

They therefore use a elastic clustering way based on fuzzy C-means and discovery that it can yield improved classification results. In Menardi et al.'s [12] two-step clustering technique, 1436 public funds are divided into 24 types of features by means of hierarchical clustering after PCA is used to reduce the dimensionality of 24 fund features. Moreno et al. [13] use a self-organizing planning neural network (SOM) to classify 1,592 funds from the Spanish marketplace in order to extract nonlinear features. They find that, in comparison to K-means clustering, SOM may dramatically decrease misclassification.

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 is 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.

Morningstar rates mutual funds according to their fund manager on a scale of one to five stars. After accounting for risks and expenses, these rankings are based on the fund's performance relative to other funds in the same category.

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: An actively managed portfolio should, in the investor's opinion, have an expense ratio of 0.5% to 0.75%. An expenditure ratio greater than 1.5 percent is considered high. In general, mutual funds have a higher expense ratio than exchange-traded funds (ETFs).

NAV: NAV: The NAV per unit indicates the success of a mutual fund strategy. The marketplace worth of the securities in a scheme is separated by the total number of parts in the scheme as of a explicit date to regulate the NAV each unit.

Standard deviation: An indicator of how much a mutual fund scheme's returns are expected to diverge from its average yearly returns is the standard deviation, which is expressed as a percentage.

Jenson's measurement, sometimes referred to as Jenson's alpha, is a risk-adjusted performance metric that assesses whether, given the beta of the portfolio and the overall market return, the average return of an investment is higher or lower than that anticipated by the capital asset pricing model (CAPM).

Beta is calculated by dividing the portfolio's returns by the risk-free rate of return. The investor earns an extra return above and above the earnings obtained without taking any risks. The risk-free investment return is thought to be a Treasury bill or other government security.

Research Methods

As part of the model framework for the investigation, we create the prediction model and the clustering model.

Principal component analysis and hierarchical clustering techniques are the data clustering models used in this work. In this study, PCA was used to dimension-minimize systematic produce data from funds with varying maturities that were utilized in the first classification stage.

PCA

The unique feature space is transformed into a whole new feature space using the dimension reduction technique known as principal components analysis (PCA). It functions as a data preprocessor most of the time. There are n fund samples in the scenario, and each fund sample has 2100 attributes. These include the reduction in computational resources and the massive amount of noise. Table 4 lists the number of eigenvalues employed and the reducibility of the PCA. The PCA approach is quite reducible, and its reducibility increases with the number of rolling days used to maintain a constant number of eigenvalues.

Hierarchical Clustering

Without knowing how many clusters there will be beforehand, it is used to group data into clusters according to the separation between data points. Several approaches are used in the data analysis and clustering process to help find patterns and structures. Scaling the data is one technique to ensure that different features or variables are comparable and understandable. This procedure is used to standardize and normalize the data.

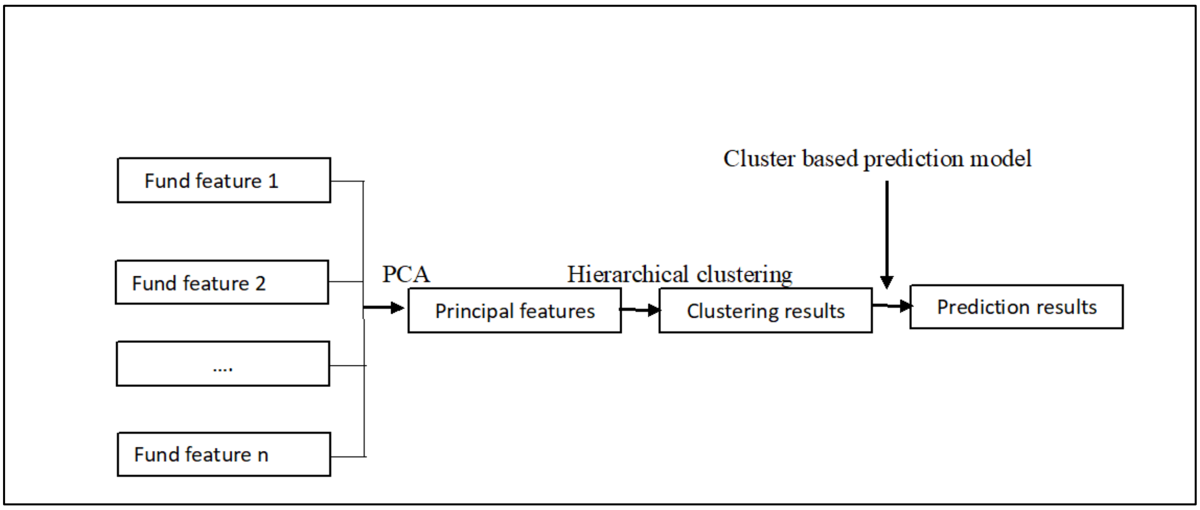

Figure 1: Model Framework

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

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

The optimum number of separations in hierarchical clustering is four. Figure 2 shows a coloured line representing respectively cluster to which the fund goes, and Table 2 shows the mean and variation of return for each fund cluster. A cluster is shown in Figure 2 that comprises The color-lines for the small, mid, and large caps are blue, grey, orange, and green, respectively. Each fund group has been separated into a subgroup based on hierarchical clustering, which uses the k-means algorithm to merge them into a single group. Furthermore, note that there is a significant variance in returns even though the means of returns for large cap funds vary.

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

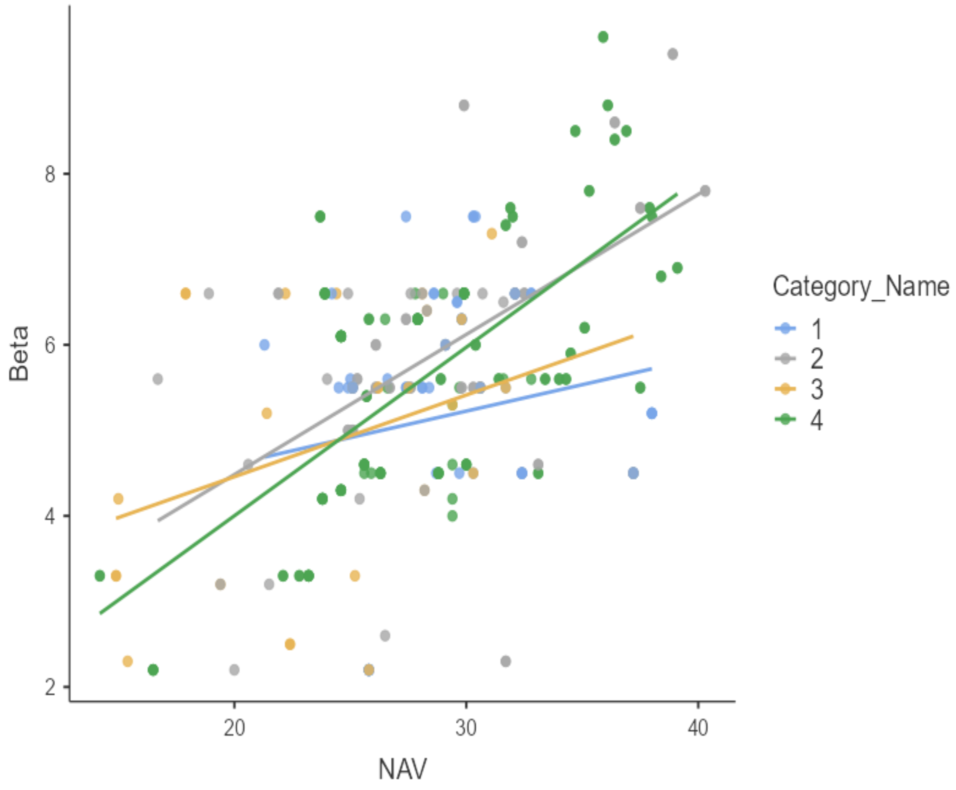

Figure 2. Hierarchical Clustering Model

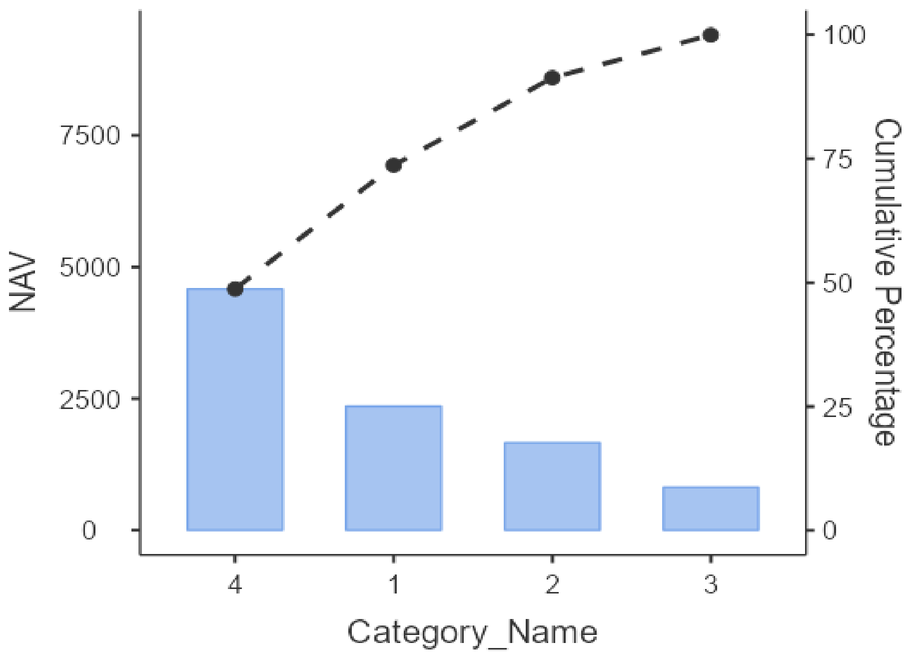

Figure 3. Hierarchical Model for comparing NAV and Category Type.

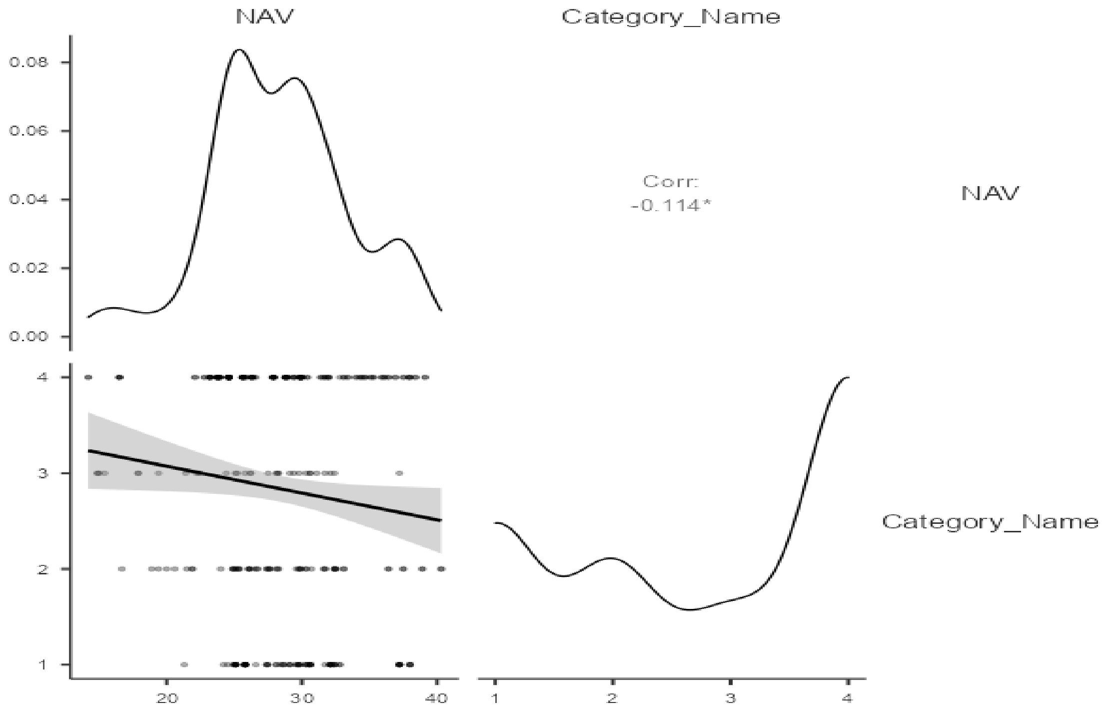

Figure 4. Correlation for category wise NAV points

Model Coefficients

Predictor	Estimate	95% Confidence Interval Lower	95% Confidence Interval 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

Eigenvalues

Component	Eigenvalue	% of Variance	Cumulative %
4	0.00224	0.0559	100.0

Table 4. Number of eigenvalues

Conclusion and future work

In this study, we describe two clustering algorithms based on investment similarities, with a focus on mutual fund returns. The recommended approaches can be used to identify mutual funds that use an operation different from the described technique. 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]