

Survey on Investor's Perception in Mutual Fund Using Data Mining Techniques

Dhanalakshmi¹, Murugesan²

^{1,2}Data Analytics, Jayaraj Annapackiam College for Women, India

Abstract - A mutual fund is a kind of economic vehicle made up of a pool of money collected from many investors to capitalize in safeties such as stocks, bonds, money market instruments and other assets. Today there are many mutual fund schemes having investor's reviews and discussion about a good return rate. This paper provides the statistical data analysis through the investor's perception of mutual funds understanding about growth and the risk return of the investors. As the number of investor's high return index value, the number of reviews about the mutual fund scheme grows rapidly. This large amount of investor's return value has to be collected from the website, magazine and investors and it needs to be explored, analyzed, and prepared for select the better scheme on mutual fund. Hence, classification schemes provide the quick information about which funds are worth and show an application of clustering methods to the mutual funds historical data. This paper concentrations on the mutual fund data analytics and high return cost from mutual fund scheme.

Index Terms - Data mining, Classification algorithm, Clustering algorithm, Mutual fund.

I.INTRODUCTION

A mutual fund is a financial tool that allocates income gained by conveying the management of funds collected from individuals to a specific expert, according to their contribution amounts. Investment strategy and investment style are basis for classification. Considerable variation is observed in the definitions, return calculation methods and assumptions. Some earlier studies have proposed methods to cluster mutual funds based on the ancient performances and apply these methods to present mutual fund schemes.

In this investigation focus on investment similarity with return fund in various mutual fund schemes. The mutual fund categories are equity, debt, hybrid, and solution oriented mutual funds. This fund provides

very limited information to the investors' mainly periodic returns. Sources of data for the various websites, journals, and fact sheets of various mutual fund schemes published by them time to time. So, first classify the all-mutual fund schemes with return fund and clustering the return fund based on which scheme is most suitable for investor. The cluster analysis approach to classify mutual funds. The mutual funds are classified based on the asset classes they invest in the incentives fee, the risk liquidity of the investment strategy and the size. To increase the confidence of an investor, cluster algorithm can be implemented and more confidence will be extracted as the data analytics.

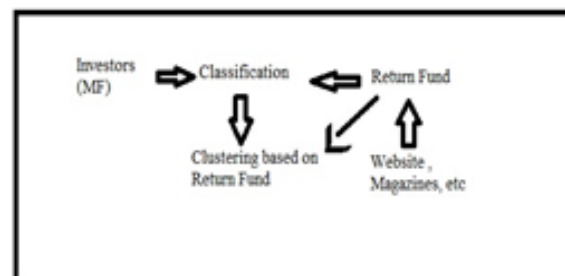


Fig.1 Analysis return fund using cluster and classification

A. Data Mining – Classification and Cluster
Data mining is compact on database in order to generate expected information. Data mining has including supervise algorithm and unsupervised algorithm. The classification algorithm which belongs to discrete and predefined set of classes from which further classified instances may be expected to extract. Classification models normally forecast definite class labels where model for prediction normally use functions that have continuous values. These functions are used for prediction based on some categories. Cluster technique used to place the data elements into their related groups. It is the process of partitioning the

data into the same class, the data in one class is more similar to each other than to those in other cluster.

B. Clustering Algorithm – K-Means

K-means algorithm falls into methods of unsupervised learning mechanism. This algorithm categorizes an available data unit over a specific quantity of groups. The procedure first starts with middle of the groups on random bases. Then the next step calculates the distance between the item and the middle point of all the groups. After that all points which belong to a given data unit are occupied and associated with the adjacent center and recalculation of the group center value is done.

The process continues with the intention of reducing an error with function like

$$J(v) = \sum_{i=1}^K \sum_{j=1}^{C_i} (\|x_{ij} - v_j\|)^2$$

Step 1 - Centers of the clusters $c_1 \dots c_k$ need to be selected randomly.

Step 2 - Distance from center of the cluster to every point of data is calculated with the help of distance matrix. Normally Euclidean distance is applied.

Step 3 - Data points which have the minimum distance from cluster center to data points are assigned.

Step 4 - The new center of cluster is again calculated.

Step 5 - Distance of every data point to newly calculated centers of cluster are again calculated.

Step 6- Stop when any data point is not assigned, else start again with step 3.

C. Neural Network

A neural network is a series of algorithms that endeavors to recognize underlying relationships in a set of data through a process that imitates the way the human brain operates. In this sense, neural networks refer to systems of neurons, either organic or artificial in nature. Neural networks can adapt to changing input; so the network generates the best possible result without needing to redesign the output criteria.

D. Mathematical Formula

The Sharpe ratio and the Treynor ratio are two ratios used to measure the risk-adjusted rate of return. Both are named for their creators, Nobel Prize winner William Sharpe and American economist Jack Treynor, respectively. While they may help investors understand investments and risk, they offer different

approaches to evaluating investment performance. The Sharpe ratio helps investors understand an investment's return compared to its risk while the Treynor ratio explores the excess return generated for each unit of risk in a portfolio.

II.RELATED WORK

Takumasa et al.,[3] proposed two methods for classifying mutual funds based on the investment similarity and evaluate the proposed methods based on actual 551 Japanese mutual funds. In this research evaluate, clustering methods such as k-means and spectral clustering. It is confirmed that the k-means method can divide finely in the center part of the network whereas spectral clustering can divide finely in the outer part.

Robert J.Binanchi et al.,[4] introduced the global hedge fund industry uses a system of self-classification to define instrument styles. In this problem using the Gap statistic method of Tibshirani, walther and hastie (2001) discussed on only three styles in the global hedge fund for the period 1994 through 2001. In this paper three hedge fund styles can be defined as, quasilong equity, non-directional and global directional. This paper following some research questions. First what is the impact on style analysis when you include survivors and non-survivors? Second, how many hedge fund investment styles are there in long term and are there short dynamics? The conclusion that the Gap statistic provides from this study is that contradicts the academic literature and industry-based classification methodologies on hedge fund style analysis.

Nandita Das et al.,[5] explored on hedge fund performance may produce different results based on the chosen database and the results are difficult to compare as there are many different ways to classify and hedge fund. In this paper hedge funds in the ZCM/Hedge database are classified using cluster analysis. The attributes used for classification in this paper are those that influence the return characteristics of the hedge fund. These attributes will affect the hedge fund return, but the return will not affect the classification scheme.

Lisi Francesco et al.,[6] provide the show an application of clustering methods to the mutual funds historical data. Starting from the monthly time series of the Net Asset values of a specific style-based

category, namely the large blend US mutual funds. So, this work applies distance-based clustering methods twice on a set of return. In this research include database to apply two large clusters. The first one gathers maximum of the risk-adjusted return indexes at a very low level of distance. And second large cluster includes the remaining measures, but two subgroups are clearly visible, gathering the whole set of risk indexes and the measures of rewards.

Adem Kilicman et al.,[7] focuses on the equity mutual funds obtainable by three Malaysian banks, namely Public bank Berhad, CIMB, and Malayan Banking Berhad. This work find the optimal asset allocation in each cluster we develop a hybrid model of optimization and fuzzy based on return rates variance. This work done by maximizing the fuzzy return for a tolerable fuzzy risk and minimizing the fuzzy risk for a desirable fuzzy return separately at different confidence levels. The results of cluster analysis are observed the performance of funds in cluster2 and cluster3 dominates the other cluster.

Noney Lenin Kumar et al.,[8] presented the five years and performing the period of study (ie 2003-2007) are selected for the research. These research databases consist of 340 mutual funds belonging to 19 investment styles. In this work performance of selected funds is evaluated using rate of return of fund, standard deviation and coefficient of variation. This research executed the three investment styles funds have average returns greater than 50%, average returns are between 40% and 50% for equity index and equity MNC investment styles.

Santit Narabin et al.,[9] proposed an investor can estimate profit and loss rate of the mutual fund in his/her portfolio by using the net asset value change ratios (NAVCR) and value of each mutual fund used for clustering. These research data at different times from the set for the year 2010-2017 are used. The results of the experiment find that the diversified NAVCR portfolio management offers available approach for portfolio management. Empirical results show that they offer significantly better performance than the random portfolio management.

Wen-Tsao Pan et al.,[10] introduced a calculation method is applied in this study to predict the net value of domestic mutual funds. In that method including 17 open-end balanced stock funds data will be collected from domestic securities companies' websites. Then, the mutual fund net worth prediction model is built by

data mining methods including Back propagation, Neural Network and GABN, and the forecasting ability is compared with the traditional regression model. These research result had SBC Taiwan Safe and Rich and Sino Pac Balance is best performance in the six mutual fund net worth prediction models.

Rupel Nargunam et al.,[11] proposed the Gold exchange traded funds are aimed at tracking the price of physical gold in the financial market. To use data on the GETF of both banking and non-banking closing prices, from 22/03/2010 and daily return in the period from 22/3/2010 to 28/8/15 there used. In the test was performed for daily returns under the assumption that the observed number of runs approximately follows the normal distribution. These values display the standard normal variate the results of the runs test using daily returns.

Yushen Kong et al.,[12] presented the mutual fund performance evaluation using econometric models. In this paper six continuous years 2010-2011, 2012-2013, and 2014-2015 is take a mutual fund performance evaluation model utilizing the fast adaptive neural network classifier (FANNC) and compared resilient back propagation neural networks model. These results from ERBPNN indicating the significant difference by a factor of two or three. FANNC consume less than 1 second in terms of processing time, whereas requires at least a minimum of 15 seconds. RSME points of scatter diagram from FANNC are more focus and closer to 45-degree line comparatively by ERBPNN. This means that the FANNC approach is highly accurate within the Sharpe index classification than the ERBPNN approach.

K.P.Sowmya et al.,[13] proposed the research provides the statistical data analysis for investor's growth and the risk return. In this work focused on the investor's point of view towards the mutual fund scheme which is predictable in the future rate of investment. The data analysis only Coimbatore city and sample size 508. The result was produced, chi-square analysis and Hypothesis.

Sakshi singh et al.,[14] explored, this paper in order to analyze the various instruments used by the people of different profiles and also applied fuzzy data mining technique to the demographic factors. This research mainly focused to apply data mining to find the investment patterns of a person based on their characteristic. The data was taken from the survey of consumer finance. The output of FP growth algorithm

specifies the investment patterns of the people who have same demographic factors as the user in 11 different financial instruments. All the percentage in the results were stated with financial assets of the user's input set as the base.

Vikas Choudhary et al.,[15] presented, this paper to analyze the performance of the growth-oriented equity diversified schemes on the basis of return and risk

evaluation. This research analysis was done by Average Return, Sharpe Ratio, Treynor Ratio, Standard Deviation Beta and Coefficient of Determination. The average returns 75% of the diversified fund scheme have higher return and sixty two percent have less risk return in the market. Seven out of eight funds have superior performance under the Sharpe Ratio as well as Treynor Ratio.

Author	Title of the paper/year	Issues Identified	Techniques/Tools
Takumasa et al[3]	Clustering mutual funds based on investment similarity (2015)	Performed analysis only on 551 Japanese mutual funds	K-means and Spectral clustering
Robert J.Binanchi et al[4]	An analysis of Hedge Fund Styles using Gap statistic(2005)	Finding the performance of three styles in the global hedge fund for the period 1994 through 2001.	Classification (Likelihood ratio, Gapstatistic)
Nandita das et al[5]	Hedge fund classification using K-means clustering method(2003)	Hedge funds in the ZCM/Hedge database are classified using cluster analysis.	Classification and Cluster analysis.
Lisi Francesco et al.,[6]	Double clustering for rating mutual funds(2015)	The Analysis of monthly time series of the net asset values of large blend US mutual funds.	Distance based Cluster Analysis.
Adem kilicman et al.,[7]	Portfolio Optimization of Equity Mutual Funds – Malaysian Case Study(2010)	The data base analysis of Malaysian bank, public bank berhad, CIMB and Malayan banking berhad bank and find the return rate variance.	Hybrid model of optimization and fuzzy model.
Noney lenin kumar et al.,[8]	Cluster Analysis of Mutual Funds(2011)	The database analysis of 340 mutual funds belonging to 19 investment styles.	Standard deviation and coefficient of variation.
Santit Narabin et al.,[9]	A Cluster Analysis of Mutual Funds Data(2018)	An investor can estimate profit and loss rate of the mutual fund database(2010-2017) by using NAVCR(Net Asset value change ratios)	Cluster Analysis
Wen-Tsao Pan et al.,[10]	Prediction of Mutual fund net value based on data mining model (2014)	The analysis of data, which scheme is best performance within 17 open-end balanced stock fund data from companies' website.	Back Propagation, Neural Network, GABN.
Rupel Nargunam et al.,[11]	Market efficiency of Gold Exchange-Traded Funds in India (2017)	The gold exchange traded funds are aimed at tracking the price of physical gold in the financial market. It's analysis of data (daily return) GETF of banking and non-banking from 22.03.2010 to 28.08.2015	Cluster with Normal distribution.
Yushen kong et al.,[12]	Evaluation of the robusticity of mutual fund performance in Ghana using Enhanced Resilient Backpropagation Neural Network(ERBPNN) and Fast Adaptive Neural Network Classifier (FANNC)	Performed analysis using econometric models and using database of six years (2010-2015)	FANNC & Back Propagation Neural Network Model.
K.P. Sowmya et al.,[13]	Mutual Funds – A Pragmatic Research on Investors Perception towards Risk-Return Pattern with special reference to Coimbatore city	Statistical data analysis for investor's growth and risk return in Coimbatore.	Chi square analysis and Hypothesis.
Sakshi singh et al.,[14]	Prediction of Investment Patterns Using Data mining Technique	Focused to apply data mining to find the investment pattern of a person based on their characteristic. Data base of 11 different mutual fund scheme.	FP Growth Algorithm

Vikas Choudhary et al.,[15]	Performance Evaluation of Mutual funds: A Study of selected Diversified Equity Mutual Funds in india.	The performed of the growth oriented equity diversified schemes	Average return, Sharpe Ratio, Treynor Ratio, Standard Deviation, Beta and Coefficient of Determination.
-----------------------------	---	---	---

Table 1 Summary of Chart

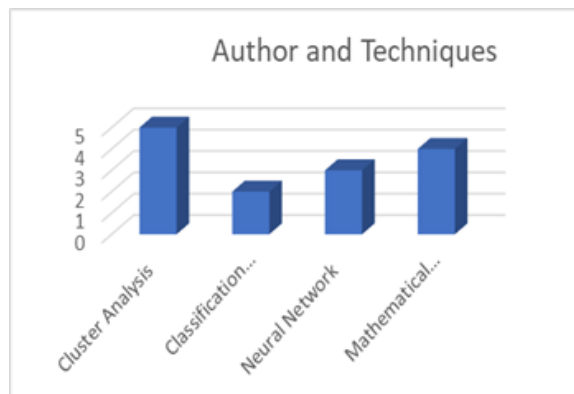


Fig. 2 Chart of Author and Techniques

III.CONCLUSION

This study offers a survey of approaches for extracting and analyzing mutual fund data from websites and magazines. This strategy is beneficial in determining the areas and methods to be investigated. This will be extremely beneficial to new mutual fund researchers. Investors can verify the good rating of mutual fund schemes in which one is available while investing in mutual funds. The level of investor confidence can be significantly improved as a result of the scheme's positive return. Cluster analysis was commonly utilized in mutual fund data studies.

REFERENCES

[1] Guidebook on Mutual funds. www.elearnmarkets.com

[2] Jiawei Han, Data Mining Concepts and Techniques B

[3] Takumasa Sakakibara, Tohgoroh Mtsui, Atsuko Mutoh, Nobuhiro Inuzuka, "Clustering Mutual Fund Based on Investment Similarity", 19th International Conference on Knowledge Based and Intelligent Information and Engineering Systems. Procedia computer science 60(2015)881-890.

[4] Robert J.Binanchi, Michael E.Drew, Madhu Veeraraghavan and Peter Whelan, "An Analysis

of Hedge Fund Styles using the Gap Statistic" www.hennesseegroup.com.

[5] Nandita Das, "Hedge Fund Classification using K-means Clustering Method", 9th International Conference on Computing in Economics and Finance, July 11-13,2003, U of Washington, Scattle.

[6] Lisi Francesco, Menardi Giovanna, "Double Clustering for rating mutual funds", Electronic Journal of Applied Statistical Analysis. Vol.08, Issue 01, April 2015, 44-56.

[7] Adem Kiliman and Jaisree Sivalingam, "Portfolio optimization of Equity Mutual Funds-Malaysian Case Study", Hindwai Publishing corporation Advance in Fuzzy Systems, volume 2010, Article ID 879453,7pages.

[8] Mr.Nooney Lenin Kumar, Dr.V. Rama Devi,"Cluster Analysis of Mutual Funds", International Journal of Multidisciplinary Research, Vol.1 Issue 5, September 2011,ISSN 2231 5780.

[9] Santit Narabin, Laor Boongasame, "A Cluster Analysis of Mutual Funds Data", 2018 International Conference on Big Data Artificial Intelligence.

[10] Wen-Tsao Pan, Shi-Zhuan Han, Hui-Ling Yang, Xue-Ying Chen, "Prediction of Mutual fund net value based on data mining model", cluster computing, <https://doi.org/10.1007/s10586-018-2272-2>.

[11]Rupel Nargunam, N.Anuradha, "Market efficiency of gold exchange-traded funds in india", Nargunam and Anuradha Financial Innovation(2017) 3:14 DOI 10.1186/s40854-017-0064-y

[12] Yushen Kong, Micheal Owusu-Akomeah, Henry Asante Antwi, Xuhua Hu and Patrick Acheampong, "Evaluation of the robusticity of mutual fund performance in Ghana using Enhanced Resilient Backpropagation Neural Network (ERBPNN) and Fast Adaptive Neural Network Classifier (FANNC).

- [13] K.P.Sowmya, Sadiya,C.Shanthini, “Mutual Funds- A Pragmatic Research on Investors Perception towards risk-Return pattern with special reference to coimbatore city”. International journal of innovative technology and exploring engineering (IJITEE), ISSN:2278-3075, volume-8 Issue-12S.
- [14] Sakshi Singh, Harsh Mittal and Archana Purwar, “Prediction of Investment patterns Using data mining Technique”, International Journal of Computer and Communication Engineering, Vol.3, No 2.
- [15] Vikas choudhary and Preeti Sehgal Chawla,”Performane Evaluation of Mutual funds: A study of selected diversified equity mutual funds in india, International conference on business, law and corporate social responsibility(ICBLCSR’14).
- [16] Reference<https://www.investopedia.com/terms/n/neuralnetwork>