

## A Study on Mining Financial Time Series Databases in Mutual Equity Funds Using Computing Techniques

B.Sharmila<sup>1</sup>, Dr.R.Khanchana<sup>2</sup>

Research Scholar, Department of Computer Science, Sri Ramakrishna Arts and Science College for Women,  
Coimbatore, India<sup>1</sup>

ksharmilab@gmail.com

Research Supervisor, Department of Computer Science, Sri Ramakrishna Arts and Science College for Women,  
Coimbatore, India<sup>2</sup>

khanchanacs@srcw.ac.in

---

**Abstract:** With the plethora of schemes available in the Indian markets, an investors needs to evaluate and consider various factors before making an investment decision. The present investigation is aimed to examine the performance of safest investment instrument in the security market in the eyes of investors i.e., mutual funds by specially focusing on systematic investment plan in mutual fund schemes. The study of research helps the people and financial analysts to analyze various securities or funds while selecting the best investment alternative out of the galaxy of investment alternatives. Time series analysis in data mining helps to find the market movement in order to find out the comparative analysis report of various funds.

**Keywords:** Data mining, Forecasting, Portfolio, Time series, Artificial Neural Network, Feature selection.

---

### I. Introduction

Apart from the many advantages that investing in mutual funds provides good returns over a period of time and the helps as the best investment option. A Mutual Fund is a pure intermediary that performs a basic function of buying and selling securities on behalf of its unit holders. Mutual Fund is a body corporate which pools up the money from different types of investors and invests those funds on behalf of the investors in diversified securities. In other words, a mutual fund allows an investor to take a position indirectly in a basket of assets. A majority of investors are quite content in simply analyzing the appreciation in the net asset value (NAV) of their investment. They are not much more concerned about the risk associated with the investment alternative. Risk measure mostly deal with the character of a fund's returns and the manner in which these returns have been achieved. The present research will explore the measures of risk and return for the selected mutual fund schemes.

### II. Objectives Of The Study

The present study aims to achieve the following objectives by considering the main objective as to select the best Equity-Diversified Mutual fund among selected eight funds during the period of study[1]:

- To compare and analyze the Equity-Diversified Mutual Fund schemes of select mutual fund players.

To compare the growth in Equity-Diversified Mutual Fund schemes with Industry average. To find out the best Equity-Diversified Mutual Fund scheme in terms of return over the selected period of study. To suggest the public to improve their return by investment in mutual funds.

The main objective of investment is to get return from investment from the mutual funds. An investor should take following points into consideration to earn good returns. An investor should assess the portfolio risk profile before investing in any fund. To select a fund with good past records of returns which is less volatile. An investor should select a portfolio of three to five funds which are less volatile in nature, and a good track record of consistent returns. For selection of a good fund, investor can compare the return of the fund with the industry average and benchmark indices. The fund which outperforms the both can be selected for investment. Investors should review their portfolio of mutual funds from time to time. Investors should try to keep their investment for a longer period of time so as to ensure that they can beat market volatility. Last but not the least; investor can withdraw funds according to his needs and purpose. The fund house should be professional, with efficient management and administration.

To carry out the research following methodology is adopted:-

Data Collection in the present research is a study of examining and analyzing selected mutual fund schemes by using different financial and statistical tools. The schemes taken for this purpose are Equity-Diversified Mutual Fund Schemes (Systematic Investment Plan). This study compares diversified funds launched by public sector, private sector, in India. The schemes have been selected using deliberate sampling method subject to the criteria mentioned as under:

- a) Corpus size > 500 crores.
- b) Returns of 5 years.
- c) Top 8 schemes ranked on the basis of 5 years

Closing Net Asset Values of the selected schemes are taken on Monthly basis for calculating the desirable results [1]. The study of research is exclusively based on secondary data, which has been collected from various websites, journals and fact sheets of various mutual fund schemes published by them time to time.

### III. Specifics Of Data Mining In Finance

**TOOLS AND TECHNIQUES:** The collected data will be analyzed on basis of returns of last one month, six months, one year, three years, and five years. Various statistical and financial techniques namely, Standard Deviation and Sharpe ratio have been used to measure volatility of returns, and returns per unit of risk respectively[2].

- Data mining in financial sector forecast multidimensional time series with high level of noisy data.
- accommodate specific efficiency criteria (e.g., the maximum profit ) in addition to prediction accuracy;
- Make coordinated multi resolution forecast (weeks, months, and years).
- incorporate a stream of text signals as input data for forecasting models are be able to explain the forecast and the forecasting model (“black box” models have limited interest and future for significant investment decisions)
- Incorporate the impact of market players on market regularities.

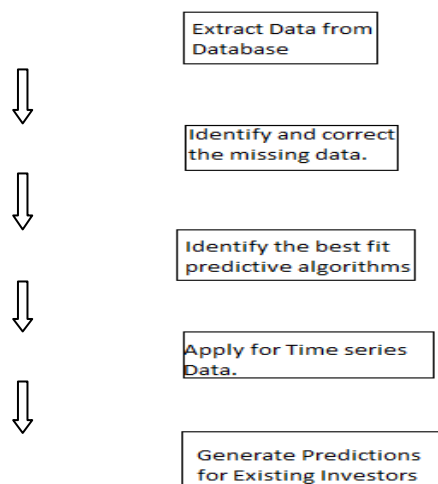
#### Time series analysis

A temporal dataset  $T_d$  called a time series is modeled in attempt to discover its main components such as Long term trend,  $L(T_d)$ , Cyclic variation,  $C(T_d)$ , Seasonal variation,  $S(T_d)$  and Irregular movements,  $I(T_d)$ . Assume that  $T_d$  is a time series index from moment 0 to current moment  $k$ , then the next value of the time series  $T_d(k+n)$  is modeled by formula[2] :

$$T_d(k+n) = L(T_d) + C(T_d) + S(T_d) + I(T_d) \quad (1.1)$$

Traditionally classical ARIMA models occupy this area for finding parameters of functions used in formula 1.1 Potentially data mining methods can be used to build such models to overcome ARIMA limitations. Decision tree methods are very popular in DataMining applications in general and in finance specifically. They provide a set of human readable, consistent rules, but discovering small trees for complex problems which is very challenge in finance (Kovalerchuk and Vityaev, 2000).

#### Methodology Frame Work For Prediction



#### **IV. Portfolio Management And Neural Networks**

The neural network most commonly used by financial institutions is a Multilayer Perceptron (MLP) with a single hidden layer of nodes for time series prediction. The peak of research activities in finance based on neural networks by Trippi and Turban, 1996 explains the market issues in the financial sectors. Other neural networks used in prediction are time delay networks, Elman networks, Jordan networks, GMDH, multi-recurrent networks (Giles et. al., 1997). The below steps of research methods explain the method of finding the portfolio management using the neural network that helps to forecast the return values [2].

- Collect data of Top 8 schemes ranked on the basis of 5 years.
- Build a neural network NN1 for predicting the return values.
- Repeat steps 1 and 2 for every fund  $F_i$ , that is monitored by the investor.
- Forecast return  $F_i(X + Y)$  where  $X$  = Index return;  $Y$  = fund return
- Select  $n$  highest  $F_i(X+Y)$  values of predicted fund return.
- Recomputed NN1 model with new arrived data to the training set. Repeat all steps for the next portfolio adjustment.

The heterogeneous investor data is collected from the database and was pre-processed to remove irregularities. The raw data that is generated in the databases is time-series in cross-sectional format, which is not very valuable as these data points cannot be applied directly for developing predictive models. These raw measurements are not useable for data modeling, and the issues with the raw data are as follows:

1. Disturbances due to noises
2. Missing observations in the data points
3. Different scales of the data
4. Imbalanced distribution of data
5. Sparsity in the data

These irregularities in the data are removed through preprocessing.

Financial industries comprise on huge amount of data. It is tedious to handle them and in order to reduce the difficulty the following Feature selection methods are examined by various researchers such as

- Information gain using Ranker search
- Correlation based feature selection using best first search.
- Relief Feature selection using Ranker search

##### **A. Information gain using Ranker search:**

Information gain is one of the attribute selection measure. This method is used on ranker search evaluation function.

##### **B. Correlation based feature selection using best first search:**

This method can be used to detect redundancies. For nominal data can only use Chi-square test and numeric attributes can be used for correlation coefficient and covariance.

##### **C. Relief Feature selection using Ranker search**

The method is based on distance measure. Distance function is a function that defines a distance between each pair of elements of a set. A most familiar distance measurement is Euclidean Distance method is used in data mining.

#### **V. Conclusion**

Mutual fund investment helps the investors who invest the savings in the form of share allocation in common financial goal, like the capital appreciation and dividend earning. The money thus collect is then invested in capital market instruments such as shares, debenture, and foreign market. Investors invest money and get the units as per the unit value which we called as NAV (net assets value). Mutual fund is the most suitable investment for the common man as it offers an opportunity to invest in diversified portfolio management. The data mining learning methods such as neural networks, the nearest neighbor methods, and decision trees dominate in financial applications for analyzing the report that helps the investor to select the best portfolio in the market. These methods are relatively simple, efficient, and can handle noisy data.

#### **Acknowledgment**

I would like to express my deep gratitude to Dr. R. Khanchana for the encouragement, comments and help. She helped me to write the research paper and given me the necessary requirements to build my own identity in the research area.

### References

- [1]. Deepika Sharma, Poonam Loothra, Ashish Sharma “Comparitive study of selected equity diversified mutual fund schemes” on International Journal of Computer Science & Management Studies, Vol. 11, Issue 01, May 2011.
- [2]. Boris Kovalerchuk and Evgenii Vityaev “Data Mining For Financial Applications” pp1-22 Agrawal V. P. Financial Markets Operations; Sahitya Bhawan Publications; Ed. 2005, p. 239-240.
- [3]. Assel K.R.Icfai Reader July 2006- Information about the volatility trend in stock market.
- [4]. Banz,R.W; Relationship between Return and Market Value of Common Stocks, Journal of Finance, Vol.9 pp3-18.
- [5]. Gogal K.R.(2005), Journal of finance, Feb.2006 Various methods to analyse the Fund Performance pp23-25.
- [6]. Jagric, T., Podobnik, B., and Kolanovic M., 2004, “Testing Efficiency Market Hypothesis for Six Transition Economies”, Bilten EDP, 3-4, 60-80.
- [7]. Jensen, M., 1968, “The Performance of Mutual Fund in the period 1945-1964”, Journal of Finance, 23, 389-416.
- [8]. Joshy Andrews, the management accountant, February 2007, pp 152.
- [9]. Henning, C.N. and others, Financial Markets and the Economy, Prentice Hall, 1978, p.p. 145-149.
- [10]. Mohantu P(jan.2007) Vikalpa-stock returns in India pp25-28
- [11]. Sanjay Kant Khare, Southern Economist, January 15, 2007. pp21.
- [12]. Sarma V.V.S.; Management Accountant; April, 2005, Vol. 40; No. 4; pp 274-275.
- [13]. Azoff, E., Neural networks time series forecasting of financial markets, Wiley, 1994  
Back, A., Weigend, A., A first application of independent component analysis to extracting structure from stock returns. Int. J. on Neural Systems, 8(4):473– 484, 1998.
- [14]. Becerra-Fernandez, I., Zanakis, S., Walczak, S., Knowledge discovery techniques for predicting country investment risk, Computers and Industrial Engineering Vol. 43 , Issue 4:787 – 800, 2002.
- [15]. Berka, P. PKDD Discovery Challenge on Financial Data, In: Proceedings of the First International Workshop on Data Mining Lessons Learned, (DMLL2002), 8-12 July 2002, Sydney, Australia
- [16]. Bouchaud, J., Potters, M., Theory of Financial Risks: From Statistical Physics to Risk Management, 2000, Cambridge Univ. Press, Cambridge, UK.
- [17]. Bratko, I., Muggleton, S., Applications of Inductive Logic Programming. Communications of ACM, 38(11): 65-70, 1995.
- [18]. Kovalerchuk, B., Vityaev E., Ruiz J.F., Consistent and Complete Data and "Expert Mining" in Medicine, In: Medical Data Mining and Knowledge Discovery, Springer, 2001, 238-280.
- [19]. Giles, G., Lawrence S., Tshoi, A. Rule inference for financial prediction using recurrent neural networks, In: Proc. Of IEEE/IAAFC Conference on Computational Intelligence for financial Engineering, IEEE, NJ, 1997, 253-259.
- [20]. Cheng ,Li,cho and chen Hong, “A Hybrid Feature selection Method for Microarray classification”, International Journal Of Computer Science, 2008.
- [21]. Deepika Sharma, Poonam Loothra, Ashish Sharma “Comparitive study of selected equity diversified mutual fund schemes” on International Journal of Computer Science & Management Studies, Vol. 11, Issue 01, May 2011.