IOSR Journal of Engineering (IOSRJEN) **www.iosrjen.org**
ISSN (e): 2250-3021, ISSN (p): 2278-8719
PP 37-39

Prediction of Investment Options using Optimal Decision Tree Algorithm

B.Sharmila[1], Dr.R.Khanchana[2]
Research Scholar, Department of Computer Science, Sri Ramakrishna Arts and Science College for Women, Coimbatore, India[1]
Assistant Professor, Department of Computer Science, Sri Ramakrishna Arts and Science College for Women, Coimbatore, India2

Abstract: *Investment decision is a major issue for every individual. The spectrum of investment is extremely wide. Many investment options are available for the investor. People are not aware of best saving scheme for their investment. This paper deals with information from various domain to suggest the investor best investment option for his investment. The decision is based upon various parameters of the investor. A refined algorithm helps the investor to make an effective decision for his investment which suits their requirement. Expert's rules and feature reduction technique have been applied to this data set to convert it into an optimal dataset. Decision tree technique is applied to make investment decisions.*

Keywords: *Data mining, Decision tree, Expert rules, gain ratio, feature reduction.*

I. Introduction

Investment refers to the commitment of funds at present, in anticipation of some positive rate of return in future. Investment is very important because it helps to grow our money. But, the major in India is lack of awareness. So many investment options like bank deposits, mutual funds, equity, shares, gold, post office investments etc. are available in market but people are not aware of best scheme for their investment. In this study, investment options considered are short term mutual funds (MFS), long term mutual funds (MFL), short term fixed deposits (FDS), long term fixed deposits (FDL), share market (SM), and public provident fund (PPF).

Mutual funds are those funds that are collected from many investors to invest in securities such as stocks, bonds and money markets [4]. They are operated by money managers. One major benefit of mutual funds is that it is suitable for those investors that do not have awareness of market trends but have risk tolerance power. Mutual fund schemes can be classified as short term mutual funds termed as MFS and long term mutual funds termed as MFL. Selection between MFS and MFL depends upon various attributes such as Investment purpose like (income, growth, tax saving) and loan facility requirement because loan facility is generally not available in many MFS schemes.

A fixed deposit is a financial instrument provided mainly by banks [3]. This is beneficial for savings because it gives saving and investment option for short and long period. Fixed deposits can be short term fixed deposits termed in this study as FDS and long term fixed deposits termed as FDL. Tenure of FDS varies from 7, 15 or 45 days to 1.5 year. FDL have tenure more than 1.5 year to 10 years. Some FDL's having tenure of 5 years and above also gives tax rebates to investors.

Public Provident Fund termed as PPF is another ideal investment scheme. PPF account is opened for the tenure of 15 years. It gives tax saving benefits to the investors. Another benefit of PPF is safe deposits and loan facility available against your deposits.

II. Data Mining Techniques

Data Mining H. Trevor, T. Robert, and F. Jerome[1] Data mining is useful to discover new patterns from the present data by using different algorithms that have been already purposed. Comparison of different tools of data mining is done by Nguyen Thai Nghe, P. Janecek, and P. Haddawy,[1]. They have suggested WEKA the best tool for data mining. Because of its computational speed and support to large datasets[2]. Data mining is helpful to analyze the representation of data which is used to discover new and better findings. A large dataset is analyzed using various data mining tools such as text mining tools, traditional data mining, and dashboards to find the relationship among data. The job of data mining is not just to collect and manage the records; it focuses mainly on new discoveries. Data mining uses many arguments to make evaluations. Some mostly used are:

.

Sequence: Second argument used in data mining is Sequence it refers to the order of events in which they occur.

***Classification**:* In this technique data is divided into different groups based upon predefined classes. For instance a vehicle producing company classifies their product into classes like high demand, mild demand, and low demand. A model is derived based upon some features like price, mileage, gender of customer, brand etc... Classification further uses some technique like neural network and k-nearest neighbor classifier. Neural network is based upon the technique of back-propagation. In back-propagation, the datasets are processed, the results are evaluated with the predicted outcome and the error rate is returned back. The main job of back-propagation is to minimized mean squared error. K-nearest neighbor is based upon the concept of learning by analogy. In this technique the data is compared to the training data. Classification technique is used in this study to automatically evaluate best saving schemes for the investors.

Clustering: The process of dividing data into different groups called clusters is called clustering. In clustering the classes are not predefined. The objects of one cluster are same to other objects in that cluster but different from objects of another cluster. The benefit of clustering is that they are easy adaptable to changes.

III. Decision Tree Classification

Han Kamber[1]Decision tree is a flow chart like tree structure, where each node denotes a test on an attribute value, each branch represents an outcome of the test, and tree leaves represent classes. Decision trees are very useful in classification. Whenever class of any tuple is not known, it is inserted at the root, by tracing path from root to leaf; class of that tuple is identified. ID3, CHAID and C4.5 are some decision tree algorithms that are widely used. In this study, C4.5 is applied to the dataset after making the dataset an optimal dataset.In decision tree approach, tuples of training data are split by selecting an attribute that best discriminates these tuples according to class. Selection of this attribute is done by Information Gain Ratio or Gini Index method.

Decision tree algorithm only helps to classify data. According to existing decision tree algorithms, if a person X having some attributes invest in plan P1 then another person Y having the same attributes must invest in plan P1. But, in real sense, it might be possible that X's decision was wrong. X invested in wrong plan. In our research, the data is firstly converted to an optimal dataset by applying Expert rules and then feature reduction is performed on this dataset. Finally decision tree has been constructed from this optimal dataset. The steps followed in this research are as under:

The steps followed in this approach are as under:

- Primary data can be from people through questionnaires and interviews.
- Training dataset has been created from this data by randomly.
- Knowledge eliction from domain experts.
- From this knowledge, rules have been created in PL/SQL code.
- These rules have been applied to the data set.
- Feature reduction is performed to this dataset by calculating gain ratio of every attribute and three attributes having minimum gain ratio have been deducted which results in an optimal dataset.
- The optimal decision tree has been constructed.
- Finally, comparison has been made between existing approach and our approach.

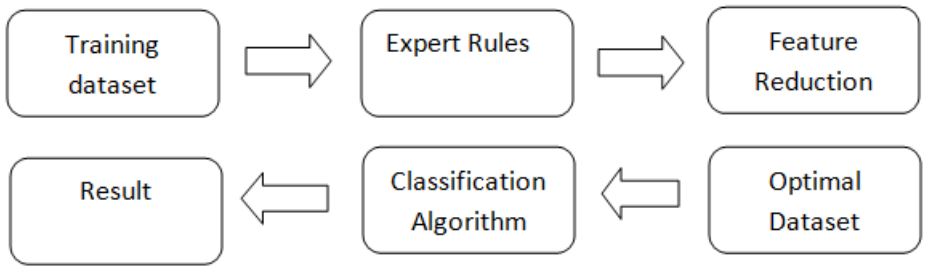

Fig 1: Representation of approach

Training Dataset: training dataset are randomly selected from the dataset which has been collected by questionnaires and interviews from the present investors. The attributes considered for the analysis of investment decision are: Age of customer, Income of investor, Risk tolerance, Loan facility required, Gender, Awareness of investment options, Purpose of investment and Time horizon. Attribute 'Age' has domain values {young, middle, old}. Domain value 'young' has range 20-35 yrs and the range of domain values 'middle' and 'old' lies between 36-50 and 51 onwards respectively.Another attribute 'Risk_tolerance' describes the risk taking ability of the investor. According to experts, people with high risk tolerance ability prefer to invest their money in stock market or mutual funds. Domain values of this attribute are {yes, no}.'Loan_facility' is another

attribute having domain values {yes, no}. It specifies whether investor wants to take loan against his investment or not. Another attribute 'Gender' has domain values {M, F}. This attribute specifies the sex of an investor. 'Awareness' is another attribute which specifies awareness of the customer to the present investment options. {yes, no} are the domain values of this attribute.

IV. Conclusion And Scope For Future

This optimal algorithm is improved new features such as implementation of Expert rules and feature reduction technique on large dataset. Several attempts have been made to design and develop the generic data mining system but no system found completely generic. Thus, for every domain the domain expert's assistance is mandatory. The domain experts shall be guided by the system to effectively apply their knowledge for the use of data mining systems to generate required knowledge. By applying this optimal approach on Investor's data, we can find out the important features that will influence investment decisions. Moreover, supervised learning has been implemented. A user interface has been provided to the user to input various parameters. Based upon these parameters an output showing best investment scheme for the investor is displayed on the screen. This approach is better to an existing approach. The accuracy of decision has been improved by 9.6%. Computation time and memory requirement has also reduced. This approach is capable to handle multivariate data which makes it suitable for many other applications. This research is not bounded to particular area. In this work, this optimal technique is applied to improve accuracy of investment decisions. But, other applications like Analysis of education patterns, Human talent management, risk evaluation etc can be implemented by this approach also.

References

[1]. Weka, University of Waikato, New Zealand, http://www.cs.waikato.ac.nz/ml/weka.
[2]. Income Tax and Investment Journal – (AY-2008-09)–by A.N. Agarwal (Income tax expert), Rajesh Agrwal(CA), Sanjay Kulkari (CA), and Dr. Gajanan Patil.- ABC Publication- Nagpur.
[3]. Dr.Binod Kumar Singh" A study on investors' attitude towards mutual funds as an investment option", International Journal of Research in Management ISSN 2249-5908 Issue2, Vol. 2 (March-2012).
[4]. S. Saravana Kumar in his article "An Analysis of Investor Preference Towards Equity and Derivatives" published in The Indian journal of commerce, July-September 2010
[5]. Mohammed M Mazid, A B M Shawkat Ali, Kevin S Tickle, 'Improved C4.5 Algorithm for Rule Based Classification', Recent Advances in Artificial Intelligence, Knowledge Engineering and Data Bases.
[6]. Chotirat "Ann" Ratanamahatana and Dimitrious Gunopulos, 'Scaling up the Naïve Bayesian Classifier: Using Decision Tree for Feature Selection'