

# MINING THE INVESTOR'S PERCEPTION ABOUT DIFFERENT INVESTMENT OPTIONS USING CLUSTERING ANALYSIS

Gunjan Batra

Bharati Vidyapeeth College of Engineering and technology  
New Delhi, India

[gunjan\\_batra27@yahoo.co.in](mailto:gunjan_batra27@yahoo.co.in)

Vijaylaxmi

Bharati Vidyapeeth College of Engineering and technology  
New Delhi, India

[tyagi.vijaylaxmi@gmail.com](mailto:tyagi.vijaylaxmi@gmail.com)

Anisha Gupta

Bharati Vidyapeeth College of Engineering and technology  
New Delhi, India

[anishagupta2@gmail.com](mailto:anishagupta2@gmail.com)

## Abstract

Investors' expectation is a very significant factor that needs to be evaluating by all investment alternatives. The achievement of any investment policy depends on how successfully it has been able to convene the investors' expectation. But, the organizations are facing the difficulty of variation and muddled behavior of customers, the lack of adequate information. Human analyst's deficient a perceptive of the hidden patterns in business data, thus, can miss corporate business opportunities. In order to embrace all business opportunities, develop the competitiveness, finding of hidden knowledge and unpredicted patterns from large databases have provided a feasible solution for several decades. To overcome the organization current concern, the new variety of method is requisite that has intelligence and potential to solve the knowledge insufficiency and the method is called Data mining. The objective of this paper is to identify the investors' perception about different investment options by one of the data mining technique – customer clustering. The study focuses on quantifying the investors' expectation and their predilection. It also attempts to estimate the factors that they take into consideration before making any investment in mutual fund as well as the consciousness level among individual investors regarding mutual fund investment. The sample survey has been conducted in Delhi.

**Keywords:** Data Mining; Cluster Analysis; SPSS tool; Dendogram.

## 1. Introduction

Recent developments in information technologies have resulted in provision and accessibility of rich information about customers in different fields. This is a key factor to approach real market segmentation. Market segmentation as a fundamental concept in marketing was first presented by Wendel Smith in 1956 [1]. Market segmentation strategy plays a significant role in creating competitive advantage [2]. Perception of differences between customers according to their needs and their responses to marketing mix plays an important role in customer relationship management (CRM) [3]. Market segmentation is an appropriate method for realizing needs of different groups and providing people a chance to be observed and to realize their needs. This method facilitates one-to-one marketing. Nowadays various sciences and techniques have made considerable attention to individual characteristics of people in their field of activities. From a methodological point of view, a clustering procedure is often employed to form segments of consumers with similar preferences. Classic clustering methods has some disadvantages such as changing initial seeds. Therefore, an analytical method is needed to segment market without any biased information which is brought by these changeable initial seeds. There is also a need to have competency to build high complex models for systems and teach them. These clustering algorithms are in general heuristics in that they assign objects in to clusters based on some distance measures between an object and the centroid of the cluster.

## 2. Objective of the Study

### 2.1. Primary Objective

To Study the various investment avenues and the investors risk preference towards it.

### 2.2. Secondary Objectives

- To find out the general demographic factors of the investors dealing in capital market.
- To find out the preference level of investors on various Capital Market instruments.
- To find out the type of risk which are considered by the investors
- To find out the ways through which the investors minimize their risk
- To find out the preferences of Investors in derivatives market
- To divide the data into different clusters based on demographic similarities and then to analyze which investment option is preferred by different clusters.

**2.3. Data Collection:** Primary data is collected through structured questionnaire. The questionnaire is administered in Delhi, India.

**2.4. Research Methodology:** Cluster analysis through SPSS is used for an objective to form clusters of different choices of investors based on their age, profession and income level and one way Anova to check the significance of clusters formed.

## 3. Cluster analysis

It is a class of techniques used to classify cases into groups that are relatively homogeneous within themselves and heterogeneous between each other, on the basis of a defined set of variables. These groups are called clusters.

Start out with a number of cases and want to subdivide them into homogeneous groups. First, we have to choose the variables on which we want the groups to be similar. Next, we must decide whether to standardize the variables in some way so that they all contribute equally to the distance or similarity between cases. Finally, we have to decide which clustering procedure to use, based on the number of cases and types of variables that we want to use for forming clusters.

Steps to conduct a cluster analysis:

1. Select a distance measure
2. Select a clustering algorithm
3. Determine the number of clusters
4. Validate the analysis.

### 3.1. Application of cluster analysis

The type of clustering used in this project is hierarchical clustering. The data is to be clustered on the basis of similarity in age, occupation and income level of respondents.

Use of SPSS tool for cluster analysis:

The variables taken are:

1. Age
2. Occupation
3. Income level

Statistics used are Agglomeration Schedule, Plot is Dendrogram, Cluster method is Ward's Method and the Interval Measure taken is Squared Euclidean Distance.

Table 1: Agglomeration Schedule

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	185	187	.000	0	0	3
2	124	186	.000	0	0	57
.	.	.	.	.	.	.
183	13	72	226.791	179	172	185
184	10	17	281.900	180	181	186
185	1	13	389.353	182	183	186

Looking at the agglomeration schedule, shows that the last 3 coefficients have major difference between them so it represents that appropriate no. of clusters will be 3, taking more clusters will make no difference among the clusters. These 3 clusters have homogeneous data within each cluster but they are heterogeneous among themselves.

After this, the Dendrogram shows which all respondents are in the similar cluster.

To check the validity and reliability of these clusters one way ANOVA is used. The SPSS output for one way ANOVA is:

Hypothesis undertaken for the ANOVA was:

Null hypothesis ( $H_0$ ): There is no significant difference between means of 3 clusters.

Alternate Hypothesis ( $H_1$ ): There is significant difference between means of 3 clusters.

Table 2: ANOVA table

		Sum of Squares	df	Mean Square	F	Sig.
Occupation	Between Groups	474.787	2	237.393	495.921	.000
	Within Groups	88.079	184	.479		
	Total	562.866	186			
Income Level - Family income (in case you are a Student)	Between Groups	19.978	2	9.989	12.243	.000
	Within Groups	150.119	184	.816		
	Total	170.096	186			
Age	Between Groups	106.940	2	53.470	225.126	.000
	Within Groups	43.702	184	.238		
	Total	150.642	186			

The significance level for all the 3 factors is less than  $.05(\alpha)$ , so it rejects the null hypothesis or say accept the alternate hypothesis which states that there is there is significant level of difference between the clusters. So the clusters formed are reliable and analysis could be found on its basis.

### 3.2. Cluster Analysis Results

From the Dendrogram following 3 clusters are formed:

**3.2.1. Cluster 1** is the largest cluster amongst all the clusters. It has approximately 44% respondents (investors) including Students in the age group '20-30 years' with annual family income level less than 5 lakh Rupees.

According to the responses collected, students mostly prefer to invest in mutual funds, public provident fund and Fixed Deposit. Their purpose of investment is mainly growth of money and future education. They are aware about all the investment options available to them. Before investing they prefer to take into consideration both, the risk as well as the returns from an investment option.

Ranking of the different options on a scale of 1 to 5 (1 being the most preferred) according to this cluster is as follows:

1. Fixed deposits
2. Mutual funds
3. Public Provident fund
4. Recurring Deposit
5. Post office deposit

**3.2.2. Cluster 2** has 35 respondents (investors) including majorly the Entrepreneurs in the age group '20-30 years' with annual family income level less than 5 to 15 lakh Rupees.

According to the responses collected, entrepreneurs mostly prefer to invest in Mutual Funds, Fixed Deposits, Public Provident Fund and Equity Market. Their purpose of investment is mainly Growth of Money, Tax Saving and also for buying a house. They are aware about all the investment options available to them. Before investing they prefer to take into consideration both, the risk as well as the returns from an investment option.

They prefer to invest either quarterly or semi annually.

Ranking of the different options on a scale of 1 to 5 (1 being the most preferred) according to this cluster is as follows:

1. Fixed deposits
2. Public Provident fund
3. Mutual funds
4. Recurring Deposit
5. Post office deposit

**3.2.3. Cluster 3** has 39 respondents (investors) including **Employed People** in the age group '20-30 years' with annual income level less than 5 lakh Rupees and almost 40% are of income level between 5 to 15 lakh.

According to the responses collected, students mostly prefer to invest in, Public Provident Fund, Fixed Deposit and Mutual Funds. Their purpose of investment is mainly Growth of Money, Tax Saving and Retirement Purposes. They are aware about all the investment options available to them. Before investing they prefer to take into consideration both, the risk as well as the returns from an investment option.

They prefer to invest either yearly.

Ranking of the different options on a scale of 1 to 5 (1 being the most preferred) according to this cluster is as follows:

1. Fixed deposits
2. Mutual funds
3. Post office deposit
4. Public Provident fund
5. Recurring Deposit.

#### 4. Conclusion and Future Work

The research included Cluster analysis (using SPSS) which divided the whole sample of 188 respondents into 3 Clusters based majorly on the occupation level and income level.

From the analysis done on 3 clusters, it is clear that fixed deposit is common among all types of investors and even preferred more as against all other options available because of its features of being risk free and fixed rate of interest. Even investors preferred to go for mutual funds but majorly those investors include students and employed people but on the other hand entrepreneurs prefer to go for Public provident fund as their main purpose of investment is tax saving but they aren't that much aware about the tax saving through ELSS mutual funds which serve the same purpose of tax saving as well as give better returns as compared to PPF. Growth of money is the primary purpose of each investor; employed people prefer to save for retirement purposes as well as for buying a house. The data is of age group under 30 because this is the segment which the company will be focusing in future.

We observed that many of the respondents have fear of Mutual Fund. They think their money will not be secure in Mutual Fund. They need the knowledge of Mutual Fund and its related terms. Many of people do not have invested in mutual fund due to lack of awareness although they have money to invest. As the awareness and income is growing the number of mutual fund investors are also growing. Similarly the main purpose of investment for most of the respondents is tax saving and they are preferring PPF as their investment option even though mutual funds offer ELSS option which is not so popular, so appropriate measures should be initiated to increase the awareness about the mutual funds to the investors and the firm should offer the mutual fund according to the need or purpose of the investor.

The investors prefer to go for Fixed deposits in bank because of their less risky nature, some of the investors treat mutual funds risky but which is not supported by our calculations as shown above mutual funds always give good returns if investment is for long time period .

The firms should target for more and more young investors as well as for persons at height of their career. They should try to highlight the benefits of new investment policies such as rupee cost averaging, power of compounding, etc.

#### References

- [1] Kuo, R.J.; Wang, H.S; Hu, Tung-L. & Chou, S.H.(2005).Application of Ant K-means on clustering analysis .Computer & mathematics with applications, 50; 1709-1724.
- [2] Danneels, E. (1996).Market Segmentation: Normative models versus Business reality. European Journal of Marketing.30, 6:36-51.
- [3] Kim, K-j & Ahn, H. (2008).A recommender system using GA K-means clustering in an online shopping market. Expert Systems with Applications, 34: 1200-1209.
- [4] Alexander Kraskov, et al (2003) Hierarchical Clustering Based on Mutual Information, Journal of Educational Psychology
- [5] Johnson, Stephen C (1967) Hierarchical Clustering Schemes, Psychometrika, the Canadian journal of Statistics.