

DATA MINING TECHNIQUES FOR IDENTIFYING THE CUSTOMER BEHAVIOUR OF INVESTMENT IN STOCK MARKET IN INDIA

DR. NAVEETA MEHTA*; MS. SHILPA DANG**

* Associate Prof., MMIT&BM, MCA Dept.,
M.M. University, Mullana
Distt. Ambala , State Haryana(132203),INDIA

** Lecturer, MMIT&BM, MCA Dept.,
M.M. University, Mullana
Distt. Ambala , State Haryana(132203),INDIA

ABSTRACT

The idea behind writing this research paper is to identify the customer behaviour of investment in Stock Market INDIA. Since in India, customer belonging to different demographic characteristics is investing their hard-earned money in stock market. Therefore, in the present research work the objective will be to form up the clusters of the customers as per their investment behaviours. To accomplish the above said objective Two-step clustering is best suitable.

KEYWORDS: Data Mining, Stock Market.

INTRODUCTION

Data Mining is the process of automated extraction of the patterns or trends representing meaningful knowledge from the ever increasing and unmanageable large data sets using its descriptive and predictive techniques like clustering, Statistics, nearest neighbourhoods, role of induction, decision tree and many more. Clustering is a useful technique for grouping data points such that points within a single group or cluster are similar, while points in different groups are dissimilar. In general, the greater similarity within a group and the greater differences between groups mean the better clustering results. Clustering can be used for data exploration and also to understand the structure of data. It is used to find similarities between observations and then group them. Clustering is a widely used technique in the area of marketing research especially market segmentation, market structure analysis and a study of customer behaviours. For example, in marketing customer segments (or clusters) are defined using demographic information. Different marketing strategies are then developed and applied to each customer segments for understanding market segments and buyer behaviours. The two major steps involved in cluster analysis are: 1) Selecting measures of similarities or dissimilarities, and 2) Selecting the procedures for cluster formation. To carry on the present work Two Step clustering method has been applied. Two Step cluster method is a scalable cluster analysis algorithm designed to handle very large data sets. It can handle both continuous and categorical variables

and attributes. It requires only one data pass. It has two steps 1) pre-cluster the cases (or records) into many small sub-clusters; 2) cluster the sub-clusters resulting from pre-cluster step into the desired number of clusters. It can also automatically select the number of clusters.

1. Two- Step Clustering

Auto-clustering: The "Auto-clustering statistics" as shown in table 1 examined the BIC and AIC change for all solutions. In auto-determining the optimal number of clusters, various criteria have been proposed such as the number of clusters with the smallest BIC (the Schwarz Bayesian Information Criterion), the number of clusters with the smallest AIC (Akaike Information Criterion), and others. By default the clustering algorithm uses a combination of BIC (though AIC may be selected by the researcher) and log-likelihood distance. For the better results the author pick a solution with a reasonably large "Ratio of BIC Changes" and a large "Ratio of Distance Measures." Simulation studies have shown this combined criterion works better than BIC or AIC alone It is also possible to enter the requirement in the algorithm; that how many clusters are wanted. This table is not provided when the researcher has asked for a fixed number of clusters as there is no "change."

Table 1: Auto-Clustering

| Number of Clusters | Schwarz's Bayesian Criterion (BIC) | BIC Change ^a | Ratio of BIC Changes ^b | Ratio of Distance Measures ^c |
|--------------------|------------------------------------|-------------------------|-----------------------------------|---|
| 1 | 11255.221 | | | |
| 2 | 10621.628 | -633.593 | 1.000 | 1.908 |
| 3 | 10457.637 | -163.990 | .259 | 1.458 |
| 4 | 10456.094 | -1.543 | .002 | 1.075 |
| 5 | 10479.225 | 23.131 | -.037 | 1.169 |
| 6 | 10550.109 | 70.884 | -.112 | 1.149 |
| 7 | 10657.686 | 107.576 | -.170 | 1.066 |
| 8 | 10780.463 | 122.777 | -.194 | 1.243 |
| 9 | 10948.227 | 167.764 | -.265 | 1.176 |
| 10 | 11143.735 | 195.508 | -.309 | 1.036 |
| 11 | 11344.656 | 200.921 | -.317 | 1.035 |
| 12 | 11550.671 | 206.015 | -.325 | 1.042 |

| | | | | |
|----|-----------|---------|-------|-------|
| 13 | 11762.606 | 211.935 | -.334 | 1.004 |
| 14 | 11975.035 | 212.429 | -.335 | 1.001 |
| 15 | 12187.548 | 212.514 | -.335 | 1.212 |

Further in table1 the superscript ‘a’ states that the changes are from the previous number of clusters in the table and superscript ‘b’ states that the ratios of changes are relative to the change for the two cluster solution. Similarly superscript ‘c’ states the ratios of distance measures are based on the current number of clusters against the previous number of clusters. Keeping the results on this table in view it may be noted that in all there are fifteen different clusters.

2. Cluster Distribution Table:

This algorithm simply shows how many cases are there in each cluster, enabling the researcher to judge if there are skewed splits among the clusters. Here table 2 examined three such clusters in addition to one outlier. Cluster 2 has 51.1% intra class similarities which are maximum and cluster 3 has 14.4% intra class similarities which are minimum. Out of the response of 229 respondents; it is also examined that 2.6% responses does not belong to any cluster, hence forms to outliers.

Table 2: Cluster Distribution

| | N | % of Combined | % of Total |
|--------------|-----|---------------|------------|
| Cluster 1 | 73 | 31.9% | 31.9% |
| 2 | 117 | 51.1% | 51.1% |
| 3 | 33 | 14.4% | 14.4% |
| Outlier (-1) | 6 | 2.6% | 2.6% |
| Combined | 229 | 100.0% | 100.0% |
| Total | 229 | | 100.0% |

After executing the algorithm; it is evaluated that how much clusters are finally framed and how one cluster is differ from the other. Table 2 offers numerous displays the composition of the clusters and the importance of each variable in determining the cluster. By applying two step clustering algorithm, it was found that increasing the number of clusters by one in all cases produced essentially a splitting of one of the clusters, with almost no change to the other clusters. This has been depicted in Table 2 which construes that cluster 1, 2 and 3 occupies 31.9%, 51.1% and 14.4% spaces respectively. This shows that 117 respondents has the common views and forms the cluster 2 which is bigger in all and 33 respondents from the population have the common ideas which forms the cluster 3 which is smaller in all. The percentage of other cluster lies in between the said two cluster. In addition to this it is also observed that there are 6 respondents whose perceptions do not match with the other population under study. Therefore these outliers are formed as outlier cluster.

3. Frequencies Table:

This is the descriptive output for the categorical variables as shown below in Table 3. There is one such table for each categorical variable. The table shown below depicts the contribution of each and every variable for framing up the clusters. For example if we look on the respondent age then we can easily conclude that respondent belonging to 18-27 years of age frame cluster 2 (as the frequency and %age is maximum i.e. 18 & 44.4% resp.) In the same way intra-class similarities of the clusters can be easily examined by looking up the various scales of the variable under study.

Table 3: Frequencies (Categorical Variables)

| Respondent Age | | | | | | | | | | |
|----------------|-----------------------|-------|-----------------------|-------|-----------------------|-------|----------|-------|----------|-------|
| | 18 years- 27 years | | 28 years- 37 years | | 38 years- 47 years | | 57 years | | 67 years | |
| Cluster | Freq | %age | Freq | %age | Freq | %age | Freq | %age | Freq | %age |
| 1 | 10 | 34.5% | 28 | 42.4% | 13 | 22.0% | 18 | 38.3% | 4 | 14.3% |
| 2 | 13 | 44.8% | 29 | 43.9% | 32 | 54.2% | 25 | 53.2% | 18 | 64.3% |
| 3 | 36 | 70.7% | 7 | 10.6% | 11 | 18.6% | 3 | 6.4% | 6 | 21.4% |
| Outlier | 0 | .0% | 2 | 3.0% | 3 | 5.1% | 1 | 2.1% | 0 | .0% |

| | | | | | | | | | | |
|------------------------------|------------------------|--------|--|--------|--|--------|--|--------|---|--------|
| Combined | 29 | 100.0% | 66 | 100.0% | 59 | 100.0% | 47 | 100.0% | 28 | 100.0% |
| Respondent Education | | | | | | | | | | |
| | Matric | | 10+2 | | UG | | PG | | Other | |
| Cluster | Freq | %age | Freq | %age | Freq | %age | Freq | %age | Freq | %age |
| 1 | 0 | .0% | 8 | 28.6% | 44 | 37.6% | 18 | 31.0% | 3 | 23.1% |
| 2 | 8 | 61.5% | 15 | 53.6% | 54 | 46.2% | 34 | 68.6% | 6 | 46.2% |
| 3 | 2 | 15.4% | 4 | 14.3% | 18 | 15.4% | 6 | 10.3% | 3 | 23.1% |
| Outlier | 3 | 23.1% | 1 | 3.6% | 1 | .9% | 0 | .0% | 1 | 7.7% |
| Combined | 13 | 100.0% | 28 | 100.0% | 117 | 100.0% | 58 | 100.0% | 13 | 100.0% |
| Respondent Income | | | | | | | | | | |
| | Below Rs 75000 p.a. | | Rs. 75000 p.a. – Rs. 150000 p.a. | | Rs. 75000 p.a. – Rs. 150000 p.a. | | Rs. 150000 p.a.– Rs. 225000 p.a. | | Rs. 225000 p.a. – Rs. 300000 p.a. | |
| Cluster | Freq | %age | Freq | %age | Freq | %age | Freq | %age | Freq | %age |
| 1 | 22 | 37.3% | 10 | 19.2% | 23 | 44.2% | 13 | 26.0% | 5 | 31.3% |
| 2 | 27 | 45.8% | 31 | 59.6% | 22 | 42.3% | 28 | 56.0% | 9 | 56.3% |
| 3 | 8 | 13.6% | 8 | 15.4% | 6 | 11.5% | 9 | 18.0% | 2 | 12.5% |
| Outlier | 2 | 3.4% | 3 | 5.8% | 1 | 1.9% | 0 | .0% | 0 | .0% |
| Combined | 59 | 100.0% | 52 | 100.0% | 52 | 100.0% | 50 | 100.0% | 16 | 100.0% |
| Respondent Occupation | | | | | | | | | | |
| | student/ idle/ | | farmer | | professional | | service | | business man | |

| | at home | | | | | | | | | |
|-----------------|---------|--------|------|--------|------|--------|------|--------|------|--------|
| Cluster | Freq | %age | Freq | %age | Freq | %age | Freq | %age | Freq | %age |
| 1 | 25 | 38.5% | 3 | 11.1% | 6 | 21.4% | 25 | 35.7% | 14 | 35.9% |
| 2 | 28 | 43.1% | 19 | 60.4% | 18 | 64.3% | 31 | 44.3% | 21 | 53.8% |
| 3 | 10 | 15.4% | 2 | 7.4% | 4 | 14.3% | 13 | 18.6% | 4 | 10.3% |
| Outlier | 2 | 3.1% | 3 | 11.1% | 0 | .0% | 1 | 1.4% | 0 | .0% |
| Combined | 65 | 100.0% | 27 | 100.0% | 28 | 100.0% | 70 | 100.0% | 39 | 100.0% |

Respondent Sex

| | Male | | Female | |
|-----------------|------|--------|--------|--------|
| Cluster | Freq | %age | Freq | %age |
| 1 | 26 | 36.1% | 47 | 29.9% |
| 2 | 32 | 44.4% | 85 | 54.1% |
| 3 | 11 | 15.3% | 22 | 14.0% |
| Outlier | 3 | 4.2% | 3 | 1.9% |
| Combined | 72 | 100.0% | 157 | 100.0% |

Respondent Marital Status

| | Single | | married | |
|----------------|--------|-------|---------|-------|
| Cluster | Freq | %age | Freq | %age |
| 1 | 16 | 34.8% | 57 | 31.1% |
| 2 | 22 | 47.8% | 95 | 51.9% |
| 3 | 8 | 17.4% | 25 | 13.7% |
| Outlier | 0 | .0% | 6 | 3.3% |

| | | | | |
|-------------------------|-------|--------|-------|--------|
| Combined | 46 | 100.0% | 183 | 100.0% |
| Respondent Place | | | | |
| | Rural | | Urban | |
| Cluster | Freq | %age | Freq | %age |
| 1 | 10 | 21.3% | 63 | 34.6% |
| 2 | 28 | 59.6% | 89 | 48.9% |
| 3 | 5 | 10.6% | 28 | 15.4% |
| Outlier | 4 | 8.5% | 2 | 1.1% |
| Combined | 47 | 100.0% | 182 | 100.0% |

Table 4: Customers invest in stock market to gain long term financial gains.

| | Can't Say | | Strongly Disagree | | Disagree | |
|--------------|-----------|---------|-------------------|---------|-----------|---------|
| | Frequency | Percent | Frequency | Percent | Frequency | Percent |
| Cluster 1 | 12 | 24.0% | 21 | 28.0% | 31 | 36.9% |
| 2 | 31 | 62.0% | 39 | 52.0% | 41 | 48.8% |
| 3 | 6 | 12.0% | 13 | 17.3% | 11 | 13.1% |
| Outlier (-1) | 1 | 2.0% | 2 | 2.7% | 1 | 1.2% |
| Combined | 50 | 100.0% | 75 | 100.0% | 84 | 100.0% |

Table 4: Customers invest in stock market to gain long term financial gains.

| | | Agree | | Strongly Agree | |
|---------|--------------|-----------|---------|----------------|---------|
| | | Frequency | Percent | Frequency | Percent |
| Cluster | 1 | 9 | 47.4% | 0 | .0% |
| | 2 | 6 | 31.6% | 35 | 65.0% |
| | 3 | 3 | 15.8% | 0 | .0% |
| | Outlier (-1) | 1 | 5.3% | 1 | 100.0% |
| | Combined | 19 | 100.0% | 1 | 100.0% |

Table 5: Investors buy the share polices for self purposes.

| | | Can't Say | | Strongly Disagree | | Disagree | |
|---------|--------------|-----------|---------|-------------------|---------|-----------|---------|
| | | Frequency | Percent | Frequency | Percent | Frequency | Percent |
| Cluster | 1 | 21 | 37.5% | 21 | 23.3% | 25 | 37.9% |
| | 2 | 28 | 50.0% | 51 | 56.7% | 32 | 48.5% |
| | 3 | 4 | 7.1% | 17 | 18.9% | 8 | 12.1% |
| | Outlier (-1) | 3 | 5.4% | 1 | 1.1% | 1 | 1.5% |
| | Combined | 56 | 100.0% | 90 | 100.0% | 66 | 100.0% |

Table 5: Investors buy the share polices for self purposes.

| | | Agree | | Strongly Agree | |
|---------|--------------|-----------|---------|----------------|---------|
| | | Frequency | Percent | Frequency | Percent |
| Cluster | 1 | 3 | 42.9% | 3 | 30.0% |
| | 2 | 2 | 28.6% | 4 | 40.0% |
| | 3 | 2 | 28.6% | 2 | 20.0% |
| | Outlier (-1) | 0 | .0% | 1 | 10.0% |
| | Combined | 7 | 100.0% | 10 | 100.0% |

Table 6: Investors buy the share polices for his/her spouse purposes.

| | | Can't Say | | Strongly Disagree | | Disagree | |
|---------|--------------|-----------|---------|-------------------|---------|-----------|---------|
| | | Frequency | Percent | Frequency | Percent | Frequency | Percent |
| Cluster | 1 | 24 | 30.8% | 27 | 34.6% | 10 | 25.6% |
| | 2 | 46 | 59.0% | 34 | 43.6% | 19 | 48.7% |
| | 3 | 7 | 9.0% | 15 | 19.2% | 9 | 23.1% |
| | Outlier (-1) | 1 | 1.3% | 2 | 2.6% | 1 | 2.6% |
| | Combined | 78 | 100.0% | 78 | 100.0% | 39 | 100.0% |

Table 6: Investors buy the share polices for his/her spouse purposes.

| | | Agree | | Strongly Agree | |
|---------|--------------|-----------|---------|----------------|---------|
| | | Frequency | Percent | Frequency | Percent |
| Cluster | 1 | 5 | 33.3% | 7 | 36.8% |
| | 2 | 9 | 60.0% | 9 | 47.4% |
| | 3 | 0 | .0% | 2 | 10.5% |
| | Outlier (-1) | 1 | 6.7% | 1 | 5.3% |
| | Combined | 15 | 100.0% | 19 | 100.0% |

Table 7: Investors buy the share polices for their children.

| | | Can't Say | | Strongly Disagree | | Disagree | |
|---------|--------------|-----------|---------|-------------------|---------|-----------|---------|
| | | Frequency | Percent | Frequency | Percent | Frequency | Percent |
| Cluster | 1 | 11 | 78.6% | 13 | 86.7% | 47 | 97.9% |
| | 2 | 0 | .0% | 0 | .0% | 0 | .0% |
| | 3 | 0 | .0% | 0 | .0% | 0 | .0% |
| | Outlier (-1) | 3 | 21.4% | 2 | 13.3% | 1 | 2.1% |
| | Combined | 14 | 100.0% | 15 | 100.0% | 48 | 100.0% |

Table 7: Investors buy the share polices for their children.

| | | Agree | | Strongly Agree | |
|---------|--------------|-----------|---------|----------------|---------|
| | | Frequency | Percent | Frequency | Percent |
| Cluster | 1 | 0 | .0% | 2 | 1.7% |
| | 2 | 0 | .0% | 117 | 98.3% |
| | 3 | 33 | 100.0% | 0 | .0% |
| | Outlier (-1) | 0 | .0% | 0 | .0% |
| | Combined | 33 | 100.0% | 119 | 100.0% |

Table 8: Investors buy the share polices for their friends/relatives.

| | | Can't Say | | Strongly Disagree | | Disagree | |
|---------|--------------|-----------|---------|-------------------|---------|-----------|---------|
| | | Frequency | Percent | Frequency | Percent | Frequency | Percent |
| Cluster | 1 | 12 | 24.0% | 21 | 28.0% | 31 | 36.9% |
| | 2 | 31 | 62.0% | 39 | 52.0% | 41 | 48.8% |
| | 3 | 6 | 12.0% | 13 | 17.3% | 11 | 13.1% |
| | Outlier (-1) | 1 | 2.0% | 2 | 2.7% | 1 | 1.2% |
| | Combined | 50 | 100.0% | 75 | 100.0% | 84 | 100.0% |

Table 8: Investors buy the share polices for their friends/relatives.

| | | Agree | | Strongly Agree | |
|---------|--------------|-----------|---------|----------------|---------|
| | | Frequency | Percent | Frequency | Percent |
| Cluster | 1 | 15 | 44.1% | 18 | 34.0% |
| | 2 | 11 | 32.4% | 32 | 60.4% |
| | 3 | 7 | 20.6% | 1 | 1.9% |
| | Outlier (-1) | 1 | 2.9% | 2 | 3.8% |
| | Combined | 34 | 100.0% | 53 | 100.0% |

Table 9: Investor buys the share directory through broker.

| | | Can't Say | | Strongly Disagree | | Disagree | |
|---------|--------------|-----------|---------|-------------------|---------|-----------|---------|
| | | Frequency | Percent | Frequency | Percent | Frequency | Percent |
| Cluster | 1 | 21 | 38.2% | 14 | 29.8% | 18 | 30.5% |
| | 2 | 27 | 49.1% | 25 | 53.2% | 32 | 54.2% |
| | 3 | 4 | 7.3% | 7 | 14.9% | 8 | 13.6% |
| | Outlier (-1) | 3 | 5.5% | 1 | 2.1% | 1 | 1.7% |
| | Combined | 55 | 100.0% | 47 | 100.0% | 59 | 100.0% |

Table 9: Investor buys the share directory through broker.

| | Agree | | Strongly Agree | |
|--------------|-----------|---------|----------------|---------|
| | Frequency | Percent | Frequency | Percent |
| Cluster 1 | 9 | 20.0% | 11 | 47.8% |
| 2 | 25 | 55.6% | 8 | 34.8% |
| 3 | 11 | 24.4% | 3 | 13.0% |
| Outlier (-1) | 0 | .0% | 1 | 4.3% |
| Combined | 45 | 100.0% | 23 | 100.0% |

Table 10: Investor buys the share directory through an agent.

| | Can't Say | | Strongly Disagree | | Disagree | |
|--------------|-----------|---------|-------------------|---------|-----------|---------|
| | Frequency | Percent | Frequency | Percent | Frequency | Percent |
| Cluster 1 | 11 | 78.6% | 13 | 86.7% | 49 | 94.2% |
| 2 | 0 | .0% | 0 | .0% | 2 | 3.8% |
| 3 | 0 | .0% | 0 | .0% | 0 | .0% |
| Outlier (-1) | 3 | 21.4% | 2 | 13.3% | 1 | 1.9% |
| Combined | 14 | 100.0% | 15 | 100.0% | 52 | 100.0% |

Table 10: Investor buys the share directory through an agent.

| | Agree | | Strongly Agree | |
|--------------|-----------|---------|----------------|---------|
| | Frequency | Percent | Frequency | Percent |
| Cluster 1 | 0 | .0% | 0 | .0% |
| 2 | 0 | .0% | 115 | 100.0% |
| 3 | 33 | 100.0% | 0 | .0% |
| Outlier (-1) | 0 | .0% | 0 | .0% |
| Combined | 33 | 100.0% | 115 | 100.0% |

Table 11: Investor buys the share directory through bank.

| | Can't Say | | Strongly Disagree | | Disagree | |
|--------------|-----------|---------|-------------------|---------|-----------|---------|
| | Frequency | Percent | Frequency | Percent | Frequency | Percent |
| Cluster 1 | 8 | 72.7% | 10 | 83.3% | 49 | 98.0% |
| 2 | 0 | .0% | 0 | .0% | 0 | .0% |
| 3 | 0 | .0% | 0 | .0% | 0 | .0% |
| Outlier (-1) | 3 | 27.3% | 2 | 16.7% | 1 | 2.0% |
| Combined | 11 | 100.0% | 12 | 100.0% | 50 | 100.0% |

Table 11: Investor buys the share directory through bank.

| | Agree | | Strongly Agree | |
|--------------|-----------|---------|----------------|---------|
| | Frequency | Percent | Frequency | Percent |
| Cluster 1 | 0 | .0% | 6 | 4.9% |
| 2 | 0 | .0% | 117 | 95.1% |
| 3 | 33 | 100.0% | 0 | .0% |
| Outlier (-1) | 0 | .0% | 0 | .0% |
| Combined | 33 | 100.0% | 123 | 100.0% |

Table 12: Investors invest in stock market of India to earn more money.

| | Can't Say | | Strongly Disagree | | Disagree | |
|--------------|-----------|---------|-------------------|---------|-----------|---------|
| | Frequency | Percent | Frequency | Percent | Frequency | Percent |
| Cluster 1 | 9 | 29.0% | 15 | 31.3% | 16 | 25.4% |
| 2 | 13 | 41.9% | 26 | 54.2% | 35 | 55.6% |
| 3 | 8 | 25.8% | 7 | 14.6% | 10 | 15.9% |
| Outlier (-1) | 1 | 3.2% | 0 | .0% | 2 | 3.2% |
| Combined | 31 | 100.0% | 48 | 100.0% | 63 | 100.0% |

Table 12: Investors invest in stock market of India to earn more money.

| | Agree | | Strongly Agree | |
|--------------|-----------|---------|----------------|---------|
| | Frequency | Percent | Frequency | Percent |
| Cluster 1 | 15 | 44.1% | 18 | 34.0% |
| 2 | 11 | 32.4% | 32 | 60.4% |
| 3 | 7 | 20.6% | 1 | 1.9% |
| Outlier (-1) | 1 | 2.9% | 2 | 3.8% |
| Combined | 34 | 100.0% | 53 | 100.0% |

Table 13: Investment inspires the investor for higher standard of living.

| | Can't Say | | Strongly Disagree | | Disagree | |
|--------------|-----------|---------|-------------------|---------|-----------|---------|
| | Frequency | Percent | Frequency | Percent | Frequency | Percent |
| Cluster 1 | 7 | 22.6% | 17 | 35.4% | 22 | 32.4% |
| 2 | 18 | 58.1% | 25 | 52.1% | 32 | 47.1% |
| 3 | 5 | 16.1% | 5 | 10.4% | 12 | 17.6% |
| Outlier (-1) | 1 | 3.2% | 1 | 2.1% | 2 | 2.9% |
| Combined | 31 | 100.0% | 48 | 100.0% | 68 | 100.0% |

Table 13: Investment inspires the investor for higher standard of living.

| | Agree | | Strongly Agree | |
|--------------|-----------|---------|----------------|---------|
| | Frequency | Percent | Frequency | Percent |
| Cluster 1 | 10 | 24.4% | 17 | 41.5% |
| 2 | 23 | 56.1% | 19 | 46.3% |
| 3 | 6 | 14.6% | 5 | 12.2% |
| Outlier (-1) | 2 | 4.9% | 0 | .0% |
| Combined | 41 | 100.0% | 41 | 100.0% |

Table 14: While investing in stock market the most important factor is the speed with which investment grows.

| | Can't Say | | Strongly Disagree | | Disagree | |
|--------------|-----------|---------|-------------------|---------|-----------|---------|
| | Frequency | Percent | Frequency | Percent | Frequency | Percent |
| Cluster 1 | 10 | 76.9% | 12 | 85.7% | 48 | 94.1% |
| 2 | 0 | .0% | 0 | .0% | 2 | 3.9% |
| 3 | 0 | .0% | 0 | .0% | 0 | .0% |
| Outlier (-1) | 3 | 23.1% | 2 | 14.3% | 1 | 2.0% |
| Combined | 13 | 100.0% | 14 | 100.0% | 51 | 100.0% |

Table 14: While investing in stock market the most important factor is the speed with which investment grows.

| | Agree | | Strongly Agree | |
|--------------|-----------|---------|----------------|---------|
| | Frequency | Percent | Frequency | Percent |
| Cluster 1 | 0 | .0% | 3 | 2.5% |
| 2 | 0 | .0% | 115 | 96.6% |
| 3 | 32 | 100.0% | 1 | .8% |
| Outlier (-1) | 0 | .0% | 0 | .0% |
| Combined | 32 | 100.0% | 119 | 100.0% |

4. Within Cluster Percentage & Cluster wise Importance:

As Shown in Figure 1, for categorical variables, a within-cluster percentage plot shows how each categorical variable is split within each cluster. These are generated by selecting "Within cluster percentage chart" under the Plots button. Clusters form the bars, the length of which reflects chi-square, with one plot per variable. For cluster wise importance plots there is a chi-square value on the X axis and cluster number on the Y axis: the longer the horizontal bar for any cluster, the higher the chi-square value, which is a measure of how much the within-cluster distribution of that variable differs from what would be expected based on the whole-sample distribution. In this plot, a dotted blue line represents the critical value line and if a bar is shorter than the critical value line, then the given variable is not important in differentiating that cluster from any of the others. If a variable is a significant differentiator for a given cluster, this means it differentiates that cluster from at least one of the other clusters.

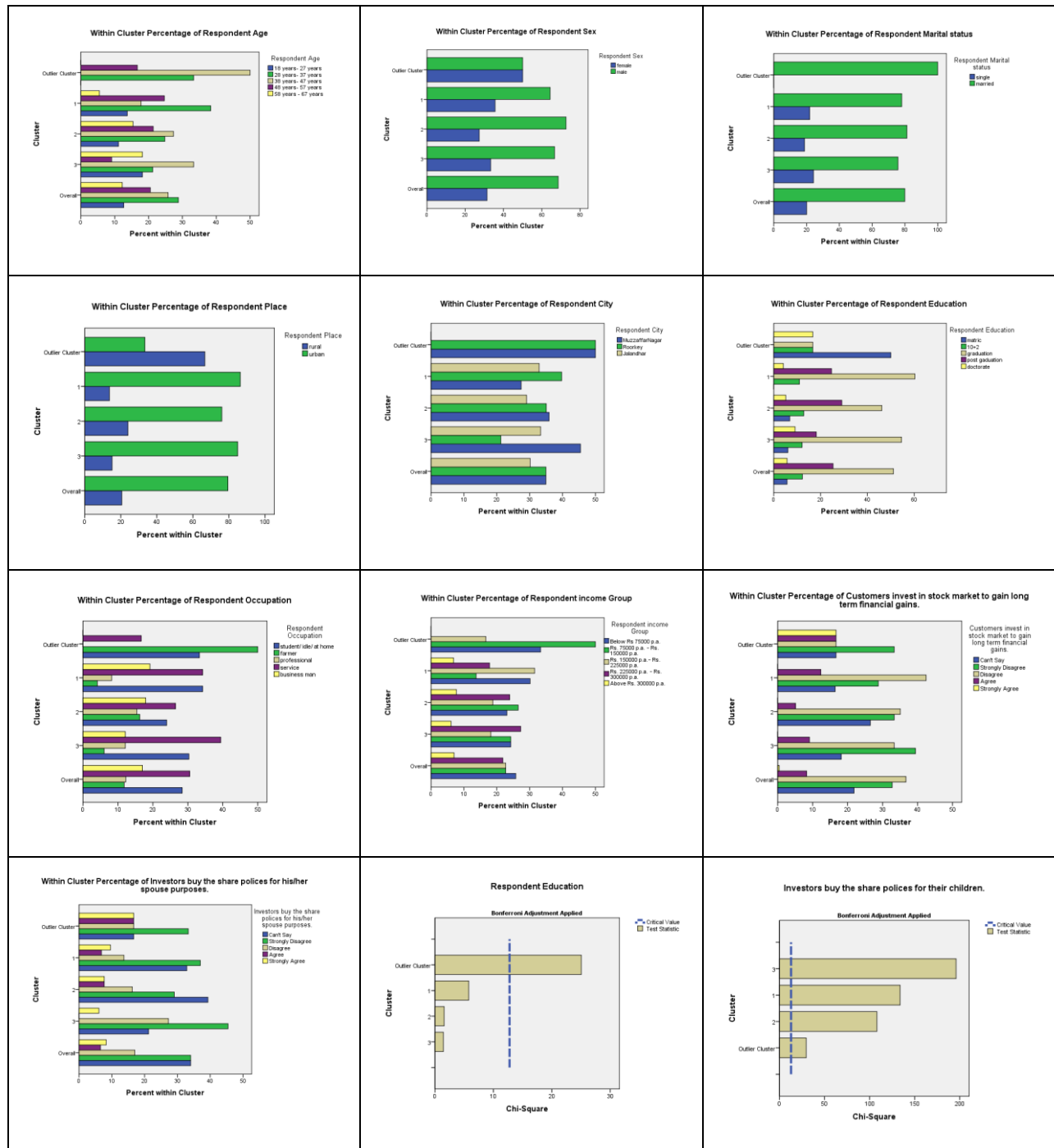


Figure 1: Cluster wise and within cluster importance

5. Formation of Clusters:

After executing the procedure of two-step clustering and result drawn as above in table 1, 2 and figure 1 it is concluded that three clusters are framed as under:

Cluster 1: Male under graduate respondents belonging to urban area whose age is 48 year to 57 year actively involved in the share trading by sitting at home belonging to annual income of Rs. 75000 to 150000 are agreed that most of the investments in share market are made for the purpose of financial gains. Also they are strongly agree that the purpose of buying share may be for the security of their spouse or relatives. People usually buys share through broker or agent to earn more money. This may also rise up their standard of living.

Cluster 2: Married female post graduate respondents belonging to urban area whose age is 58 year to 67 year actively involved in the share trading by profession belonging to annual income of Rs. 75000 to 150000 are strongly agree that most of the investments in share market are made for the purpose of financial gains. Also they are strongly disagree that the purpose of buying share may be for the security of their spouse or relatives. They believe that investment is done to secure the future of their children. These people like to do their share transaction directly with the company. They do not want to deal through an agent at all. If necessary these respondents like to utilize the services of bank to have secure investments. Their ultimate idea to work with share is to earn more money. They also believe that this helps them in improving their standard of living.

Cluster 3: Some of the unmarried male respondents belonging to urban area whose age is 18 year to 27 year who are educated are actively involved in the share trading by profession belonging to annual income of up to Rs. 75000 are strongly agree that most of the investments in share market are made to earn more money. They are strongly disagree that the purpose of buying share may be for the security of their spouse or relatives. They believe that investment is done to secure their future. These people like to do their share transaction through share brokers. They do not want to deal through an agent at all. They also believe that share market is one which gives strengthen to their investment with higher rate of returns.

6. Conclusion:

The research performed on the objective around to central theme concludes that there are three clusters which are very important for the customers of stock market. In these three clusters, it can be stated that the customers of different ages and from different localities have their own opinions on the investments in the stock markets. Most of them believe that the major objective of the investment is in regards of their future. From these, we form up the clusters of the customers as per their investment behaviours.

7. References:

- [1] K. Senthamarai Kannan, P. Sailapathi Sekar, M. Mohamed Sathik and P. Arumugam “Financial Stock Market Forecast using Data Mining Techniques,” in proceedings of International Multi Conference of Engineers and Computer Scientist (IMECS2010), Vol. 1, March 17-19, 2010.
- [2] D. Venugopal Setty, T.M.Rangaswamy and K.N.Subramanya, “A Review on Data Mining Applications to the Performance of Stock Marketing”, International Journal of Computer Applications, Vol. 1, No. 3, pp. 33-43, 2010.
- [3] M S Nagaraja , “Short Term Electrical Load Forecasting using Clustered Data,” IE (I) Journal–EL, Vol. 89, pp.28-30, 2008.
- [4] Madhurima Deb & Kavita Chavali, “A Study on the Significance of E-Trust and E-Loyalty in Online Banking,” AIMS International Journal, Vol. 3, No.3, pp. 241-257, 2009.
- [5] Ms. T.R. Rajeswari a & Prof. V.E. Rama Moorthy, “An Empirical Study on Factors Influencing the Mutual Fund/Scheme Selection by Retail Investors,” downloadable from www.utiiicm.com/Cmc/PDFs/2001/rajeswari.pdf .
- [6] Sudheep Elayidom, Dr. Sumam Mary Idikkula, and Joseph Alexander, “Applying Data mining using Statistical Techniques for Career Selection,” International Journal of Recent Trends in Engineering, Vol. 1, No. 1, pp. 446-449, May 2009.
- [7] Z. Ismail, A. Yahya and A. Shabri, “Forecasting Gold Prices Using Multiple Linear Regression Method,” American Journal of Applied Sciences, Vol. 6, No. 8, 1509-1514, 2009.
- [8] Chiu, T., Fang, D., Chen, J., Wang, Y., and Jeris, C. (2001). A Robust and Scalable Clustering Algorithm for Mixed Type Attributes in Large Database Environment. Proceedings of the seventh ACM SIGKDD international conference on knowledge discovery and data mining, 263.