IMPACT ASSESSMENT OF SHG LOAN PATTERN USING CLUSTERING TECHNIQUE

Sajeev B. U[1], K. Thankavel[2]

[1]Research Scholar, Center for Research and Development PRIST University, Thanjavoor, T.N., India.
[2]Prof. & Head, Department of Computer Science, Periyar University, Salem, T.N., India.

ABSTRACT

Indian micro-finance sector, dominated by self help groups (SHGs), addresses issues like actualizing equitable gains from the development and fighting poverty. A number of financial institutions provide micro-finance services to the poor through banking and NGOs. Clustering analysis is a key and easy tool in data mining and pattern recognition. We have applied K-Means and Fuzzy C-Means algorithms to study in detail the data's collected from the SHG members of 9 districts in Kerala state through field work and questionnaire. The study reveals that the average range of rate of interest of SHG loans from various government agencies are from 12 to 15 %. Out of total members availing loans, 56% are taking loan from bank. District wise studies on the rate of interest were also carried out. Study on the relationship of education and savings among SHG members' shows that members with higher education shows increased saving habits.

KEYWORDS: *Data mining, Clustering, K-Means, Fuzzy C-Means, self help groups*

I. INTRODUCTION

With the increased and widespread use of technologies, interest in data mining has increased rapidly. Companies are now utilized data mining techniques to exam their database looking for trends, relationships, and outcomes to enhance their overall operations and discover new patterns that may allow them to better serve their customers. Data mining provides numerous benefits to businesses, government, society as well as individual persons [1-5].

For many years, statistics have been used to analyze data in an effort to find correlations, patterns, and dependencies. However, with an increased in technology more and more data are available, which greatly exceed the human capacity to manually analyze them. Before the 1990's, data collected by bankers, credit card companies, department stores and so on have little used. But in recent years, as computational power increases, the idea of data mining has emerged. Data mining is a term used to describe the "process of discovering patterns and trends in large data sets in order to find useful decision-making information." With data mining, the information obtained from the bankers, credit card companies, and department stores can be put to good use. Data mining is a component of a wider process called "knowledge discovery from database". It involves scientists and statisticians, as well as those working in other fields such as machine learning, artificial intelligence, information retrieval and pattern recognition. Before a data set can be mined, it first has to be "cleaned". This cleaning process removes errors, ensures consistency and takes missing values into account. Next, computer algorithms are used to "mine" the clean data looking for unusual patterns. Finally, the patterns are interpreted to produce new knowledge [6-7].

Clustering is very popular descriptive data mining technique that aids describing characteristic of data sets. The goal of clustering is to form group of objects with similar characteristics [8]. Clustering analysis is important, but challenging task in unsupervised learning. Data clustering is a common technique for statistical data analysis and has been used in variety of engineering and scientific

disciplines. The biggest innovation in microfinance in the past five years is the advent of data mining, that is, the analysis of data to inform practical responses to business challenges. In this paper we have discussed about the role of Self Help Groups in India using K-Means and Fuzzy C-Means algorithm for evaluating SHG loan pattern.

1.1 About Self Help Groups

SHG or self help group is a small group of rural poor, who have voluntarily come forward to form a group for improvement of the social and economic status of the members. The core of SHG bank linkage in India has been built around an important aspect of human nature - the feeling of self worth. Over the last ten years, it has come to symbolize an enduring relationship between the financially deprived and the formal financial system, forged through a socially relevant tool known as Self Help Groups (SHGs) [9-10]. An amazingly large number of formal and non-formal bodies have partnered with NABARD (National bank for Agriculture and Rural Development) in this unique process of socio-economic engineering. What had started off in 1992 as a modest pilot testing of linking around 500 SHGs with branches of half a dozen banks across the country with the help of a few NGOs, today involves about 20,000 rural outlets of more than 440 banks, with an advance portfolio of more than Rs.1, 200 crore ($ 240 m.) in micro Finance lending to SHGs. Financial services have reached the doorsteps of over 8 million very poor people, through 500,000 SHGs, hand-held by over 2,000 development partners.

India is fiercely diverse as a nation, and most communities are also diverse in caste, opinion and religion. Indians are also known for their sense of personal independence, which is often translated into indiscipline, whether on the roads, in political assemblies or elsewhere. The SHG system reflects this independence and diversity. It allows people to save and borrow according to their own timetable, not as the bank requires. SHGs can also play a part in a whole range of social, commercial or other activities. They can be vehicles for social and political action as well as for financial intermediation.

A most notable milestone in the SHG movement was when NABARD launched the pilot phase of the SHG Bank Linkage programme in February 1992. This was the first instance of mature SHGs that were directly financed by a commercial bank. The informal thrift and credit groups of poor were recognized as bankable clients. Soon after, the RBI advised commercial banks to consider lending to SHGs as part of their rural credit operations thus creating SHG Bank Linkage [11-13]. The linking of SHGs with the financial sector was good for both sides. The banks were able to tap into a large market, namely the low-income households, transactions costs were low and repayment rates were high. The SHGs were able to scale up their operations with more financing and they had access to more credit products. There are a number of criterias for getting loans for SHG members.

For SHG to get loan from Bank, the SHG should open an account, operate the account regularly, maintain healthy relationship with bank, and the repayment of loan should be regular. The loans initially taken are usually for education, consumption, health, house repair, repaying of old loans. Apart from this, loans are taken for purchase of seeds, fertilizers, development of small business (Petty shops, vegetable vending, flower vending, hotels, saree business, animal husbandry activities, etc).

Sanction of loans to SHGs by banks is based on the quantum of savings mobilized by the SHGs. Loan may be granted by the SHG for various purposes to its members. The bank does not decide the purpose for which the SHG gives loans to its members. A repayment schedules is drawn up with the SHG, and the loan is to be repaid regularly. Small and frequent installments will be better than large installments covering a long period. Problems in repayment of loans by SHGs were quite widespread. Since the amounts involved in these loans at the individual level were not of much significance to the banks, there was a tendency not to take a serious note of irregularities in the repayment schedules of SHGs. However, as the loans to SHGs also had a tendency to slip into the irregular mode more often than not, bankers need to exercise care and caution while dealing with SHGs as they would in case of other borrowers.

These facts were supported by the news came in the Indian national daily **THE HINDU** on Friday, September 11 2010 at Chennai under the heading **"Fall in SHGs' loan repayment rate: NABARD chief"** "The repayment by SHGs is not 100 per cent. It stood at 88 per cent the year before last and is falling further. Two reasons were attributed to it. Firstly, bank managers are not in touch with the SHGs and the loan members are not attending the monthly meetings. Bank managers should visit the

borrowers at least once in three or six months to find out their problems," K.G. Karmakar, Managing Director, NABARD said.

II. BACKGROUND

Earlier SHG data evaluations were done using statistical tools [14]. As research methods, a mix of quantitative and qualitative tools is applied. Through a questionnaire quantitative data are collected. The qualitative information will enable verification of the quantitative findings as well as give more insight into the reasons behind these findings.

The survey has been conducted through structured questionnaires, related to the socio-economic status of SHG members. Since the purpose of the study is to understand the trends within groups, the survey focused on group level information. At the individual level of members and the following information has been collected.

Data was collected regarding the following aspects:

¢ Loan taken and purpose of loan

¢ Savings and credit related activities of the group

¢ Socio-economic composition of the groups

¢ Social issues taken up by the groups

¢ Linkage between the groups and bank

¢ Assets before and after being a member

¢ Literacy and education status of group members

There are 14 districts in Kerala state and this study has been restricted to 9 districts namely Kannur, Calicut, Malappuram, Palakkad, Wyanad, Trichur, Kottayam, Alleppy and Trivandrum. The above mentioned data has been collected from 3500 SHG members with 51 attributes/parameters. Majority members are female. For the better understanding of the financial, utilization of loan, purpose of loan, savings educational and loan repayment status of the SHG members before and after availing the loans has been studied in detailed by applying clustering techniques by means of K- means and Fuzzy C-Means algorithm. Among the various clustering algorithms, K-Means (KM) and Fuzzy C-Means are the most popular methods used in data analysis due to their good computational performances. However, it is well known that KM might converge to a local optimum, and its result depends on the initialization process, which randomly generates the initial clustering.

The main objective of this study is as follows

- To find the rate of interest given by the SHG members to various financial institution
- A study of various financial institution which provide maximum loans to SHG members
- Type of loans availed by the SHG members
- District wise study of rate of interest for loans provided by SHG
- Education status of SHG members in Kerala
- To find the relationship between education and savings

III. MATERIALS AND METHODS

3.1 K-means Algorithm

K-Means [15-18] clustering technique creates a one level partition of data objects. We first chose K initial centroids, where K is a user specified parameter namely number of clusters desired. Each point is then assigned to the closest centroid and each collection of points assigned to a centroid is a cluster. The centroid of each cluster is updated based on the points assigned to the cluster. We repeat the assignment and update the steps until no point changes clusters or equalently until the centroid remains the same. The K-Means algorithm is given below

K- Means algorithm

Require: set of input items, x, in Euclidean space; desired number of clusters, k.

```
1: for 1 ≤ i ≤ k do
2:         kmeans[i]   ← random item from data
3:         centroid[i] ←0
4:         count[i]  ← 0
5: repeat
6:     for all x∈ items do
7:         mindist ←1
8:         for 1 ≤ i ≤ k do
9:         if  || x- Kmeans[i]|| 2 < || x- Kmeans[mindist]|| 2 then
10:                mindist   ← i
11:        cluster[x] ← mindist
12:        centroid [mindist]  ←centroid[mindist]  +  x
13:        count [mindist]  ←  count[mindist   +  1
14:    for 1≤ i ≤ k do
15:        kmeans[i] ← centroid[i]/count[i]
16:        centroid[i] ←  0
17:        count[i] ←0
18: until no item reclassified or repetition count exceeded
19: each x∈ items is now classified by cluster[x]
```

3.2 Fuzzy C-Mean Algorithm

The Fuzzy C-Means algorithm (FCM) [19-20], which is the best known unsupervised fuzzy clustering algorithm is also used in analyzing the SHG data. However, these FCM algorithms have considerable trouble in a noisy environment and inaccuracy with a large number of different sample sized clusters

It is based on minimization of the following objective function:

$$Jm = \sum_{i=1}^{N} \sum_{j=1}^{C} u_{ij}^{m} \| x_i - C_j \|^2 \quad , 1 \le m < \infty \quad \ldots\ldots\ldots (1)$$

where m is any real number greater than 1, u_{ij} is the degree of membership of x_i in the cluster j, x_i is the ith of d-dimensional measured data, c_j is the d-dimension center of the cluster, and ||*|| is any norm expressing the similarity between any measured data and the center. The algorithm is composed of the following steps.

3.3 Data Cleaning.

As data sets are not perfect, one can expect missing values for some attributes, some errors in

1. *Initialize $U=[u_{ij}]$ matrix, $U^{(0)}$*
2. *At k-step: calculate the centers vectors $C^{(k)}=[c_j]$ with $U^{(k)}$*

$$c_j = \frac{\sum_{i=1}^{N} u_{ij}^{m} \cdot x_i}{\sum_{i=1}^{N} u_{ij}^{m}}$$

3. *Update $U^{(k)}$, $U^{(k+1)}$*

$$u_{ij} = \frac{1}{\sum_{k=1}^{C} \left(\frac{\|x_i - c_j\|}{\|x_i - c_k\|} \right)^{\frac{2}{m-1}}}$$

4. *If $|| U^{(k+1)} - U^{(k)}|| < \varepsilon$ then STOP; otherwise return to step 2.*

transcription or data input, and duplicate entries[21-23]. Dealing with these issues is a topic of major study in itself. Sometimes, a received data set has already been 'cleaned'. Perhaps 'scrubbed' is a better term: missing values are sometimes filled in with average values, or values copied from similar looking records.

3.4. Feature selection

It is a preprocessing method of choosing a subset of features from the original ones. It has proven effective in reducing dimensionality, improving mining efficiency, increasing mining accuracy, and enhancing result comprehensibility[24] .Feature selection methods can broadly fall into the *wrapper* model and the *filter* model [25]. The wrapper model uses the predictive accuracy of a predetermined mining algorithm to determine the goodness of a selected subset. It is computationally expensive for data with a large number of features. The filter model separates feature selection from classifier learning and relies on general characteristics of the training data to select feature subsets that are independent of any mining algorithms. We have chosen filter method for the present study

IV. RESULTS AND DISCUSSION

Surveys were carried out among 3500 SHG members among 9 districts in Kerala. Detailed questionnaires' were prepared. Qualitative information is gathered through semi-structured interviews with SHG members, SHG leaders, federation leaders, Bank officials, moneylenders and government officials. The selected SHG groups were found to be very stable for more than 3 years. From these groups we have collected 3500 objects with 51 attributes. The procedures adopted for clustering include data cleaning. The collected data have been cleaned with the help of domain experts and applying feature selection method. Finally we fixed the data set as 3434 objects with 12 attributes.

The selected attributes are given in the Table I

Table I shows the selected attributes for the study.

Loan amount from SHG(loan I)
Interest rate (%)
Loan period / Month
Loan repay / month
Balance loan in the book
Loan taken other sources(Loan II)
Amount taken
Interest rate (%)
Economic Benefits gained
Savings / month
Assets increased after joining SHG
Savings outside the group

The K-Means and Fuzzy C-Means algorithm discussed in section 3.1 and 3.2are applied for the SHG data collected from 9 districts in Kerala. The K-Mean algorithm is applied for different values of k (number of clusters) to the 3434 members with 12 attributes. The K-mean algorithm has been performed for different values of k and it was found that the best value for k is 2. After analysis the activities and functionalities of two different group members ,one group has been identified as performing group and other one is non performing.

Clusters obtained by k-means algorithm are dominated by the selection of initial seed or centroid. Hence K-Mean algorithm has been performed by selecting different set of initial seed and the result are tabulated in table II

Table II- Results of K-Means for different centroids.

Number of runs of K- Means algorithm with different seeds(Ri)	Number of patterns in cluster I (C1)	Number of patterns in cluster II (C2)	Seed values

R1	246	3188	Random seed
R2	268©	3166**	Z=10 and Z=100
R3	3166**	268©	Z=10 and Z=3000
R4	145	3289	Z=110 and Z=2000
R5	3166**	268©	Z=300 and Z=3000
R6	3166**	268©	Z=500 and Z=510
R7	268©	3166**	Z=644 and Z=844
R8	268©	3166**	Z=1400 and =1800
R9	3166**	268©	Z=3000 and Z=3010

**, ©, these symbols indicate same number of pattern in the clusters.

Each time when we apply K- means algorithm we obtained two clusters; of which one is performed cluster and other is non- performed cluster. Since the numbers of objects in some clusters are same, which shows their stability; we have selected 5 clusters from 18 clusters for further studies, since the selected clusters have different number of patterns. The following clusters are taken for studies R1C1, R1C2, R3C1, R3C2 and R4C1

By applying K–means algorithm for different centroids, we have obtained the patterns of each parameters like Loan amount from SHG(loan I), Interest rate (%), Loan period / Month, Loan repay / month, Balance loan in the book Loan taken other sources(Loan II), Amount taken, Interest rate (%), Economic Benefits gained , Savings / month, Assets increased after joining SHG, Savings outside the group. Table III shows the patterns obtained for different runs of K-Mean Algorithm.

Table III. Patterns obtained for different runs of K-Mean Algorithm**.**

Patterns	**R1C1**	**R1C2**	**R3C1**	**R3C2**	**R4C1**	**R4C2**
Loan amount from SHG in Rs	28717	20248	20019.35	30725.43	30440.76	20432.31
Interest rate (%)	14	14	14.35218	14.01512	13.87652	14.3457
Loan period per Month in Rs	10	11	10.61592	10.66433	10.44877	10.62724
Loan repay per month in Rs	305	288	288.0054	305.6946	305.5318	288.6744
Balance loan in the book in Rs	6133	4812	4781.629	6381.619	6335.819	4843.49
Loan taken other sources	1	1	0.591282	1.171644	1.089659	0.616601
Amount taken in Rs	89593	3252	3028.747	85146.01	116803	4703.866
Interest rate (%)	13	5	4.81396	13.19781	12.58633	5.154454
Economic Benefits gained in Rs	1574	1414	1410.676	1600.021	1499.722	1422.18
Savings per month in Rs	72	37	36.79659	67.50094	92.41822	36.84646

Assets increased after joining SHG (credit points)	8	7	6.675616	8.190371	8.75899	6.707206
Savings outside the group per month in Rs	216	118	117.4037	214.4172	232.2939	120.2439

To perform a comparative study we have applied Fuzzy C-Means algorithm in the SHG data for different values on 'm'(weight exponent in the fuzzy membership) and the results are tabulated in table IV.

Table IV- Results of Fuzzy C-Means for different values m

Different runs	Number of iterations	No: of members in cluster 1	No: of members in cluster II	m
R1	60	133	3301	1.25
R2	100*	3211	223	1.5
R3	100*	302	3132	1.75
R4	100*	773	2661	2
R5	100*	870	2564	2.25
R6	95	870	2564	2.5
R7	89	2499	935	2.75
R8	71	959	2475	3
R9	77	2453	981	3.25
R10	67	991	2443	3.5
R11	63	2433	1001	3.75
R12	90	1010	2424	4
R13	48	1156	2278	10
R14	46	1166	2268	20
R15	33	1174	2260	30
R16	27	1179	2255	40
R17	3	2164	1270	50

*Indicates maximum number of iterations

We have applied different values for m and a total of 17 runs were performed. For further studies we have selected clusters with maximum number of iterations (100). Selected clusters are R2C1, R2C2, R3C1, R3C2, R4C1, R4C2, R5C1, and R5C2. The table V shows the **p**atterns obtained for different runs of Fuzzy C-Means algorithm with (m=100) with maximum iterations

Table V shows the **p**atterns obtained for different runs of Fuzzy C-Means

Features	R2C1	R2C2	R3C1	R3C2	R4C1	R4C2	R5C1	R5C2
Loan amount from SHG in Rs	19446	32785	35241	18013	46208	12748	11906	45325
Interest rate (%)	14	14	14	14	14	14	14	14
Loan period per Month in Rs	11	11	11	11	11	11	11	11
Loan repay per month in Rs	288	310	309	286	323	278	276	324
Balance loan amount in the book in Rs	4575	7298	8103	4217	10295	3187	3048	9947
Loan taken from other sources	1	1	1	1	1	1	1	1

Amount taken in Rs	3440	91490	71201	2719	27238	3236	3285	22887
Interest rate (%)	5	13	12	5	8	4	4	8
Economic Benefits gained	1402	1599	1682	1364	2170	1152	1104	2191
Savings per month in Rs	36	78	68	36	56	33	33	54
Assets increased after joining SHG	7	8	8	7	8	6	6	10
Savings per month out side the group in Rs	117	226	211	114	184	105	80	179

We have taken the cluster R3C2 (from table III) and R4C1 (from table V) for further studies since the SHG loan is maximum for these clusters. After analyzing this cluster, majority of SHG loan interest lies between 12 to 15% and maximum interest is 25%. In the cluster R3C2 (using K Means) out of all members who have taken loan, 88 % took loan from bank, 9% from society and only 3% depends on money lenders. In R3C2 highest loan amount is Rs 15000 and it was taken from bank. This group is found to have higher savings and deposits. Their balance loan amount is nominal. In the cluster R4C1 (using Fuzzy C-Means), out of all members who have taken loan, 77% of members took loan from bank, 10% from society and 11% from money lenders. In R4C1 the highest loan amount is Rs3,00,000 and it was taken from bank.

But in R3C1(from table III- using K Means) 47% members have taken loan from bank, 8% from society and 27% from money lenders and this group shows less savings compared to R3C2. Analysis of R4C2 (from table V-using Fuzzy C-Means) 48% members have taken loan from bank, 9% from society and 29.8% from money lenders. Hence studies on both algorithms explain the same facts and the results are almost same. It shows banks linkage is more important for the smooth functioning of SHG.

Bank linkage can be made more effective by:

1. Providing financial counseling service through face to face interaction.
2. To educate people in rural and urban areas with various financial products available from the financial sector
3. To make the SHG members aware about the advantages of being connected with the formal financial sector

The study reveals that the selected members in the SHG cluster took loans from the following financial institution and the % range interest is shown in the following table VI

Table VI shows financial institutions and the range of interest

No:	Financial institution	Range of interest
1	Bank Loan	10%-16%
2	Society	12%-18%
3	Money lenders(blade)	12%-40%-60%
4	From Other SHG Groups	12%
5	Friends	0-15%

Rate of interest from bank is less compared to the interest taken by the money lenders, so it is necessary that the bank linkage [26-27] with the SHG members should be made effective, so that they can gain maximum benefits. Majority of members are not availing loan facilities, it may be due to the lack of awareness about different types of loans from the standard financial institution or the rules and regulations for getting loans is more difficult. It indicates that the banks or standard financial institution should take necessary steps to provide more loans to SHG members. Our study reveals that out of total members availing loan 56 % is taken from the bank and 22 % are from money lenders, 8 % from society and 13 % from other SHG group.

For ensuring the long term sustainability of SHGs, it is suggested that resource center can be set up in different parts of the country. The SHG - Bank Linkage Programme is now more than 20 years old. To achieve effective linkages with various financial institutions, resource centers can play a vital role. Resource centers can be set up by various stakeholders such as NGOs, banks, Government departments, NABARD at the State/ district level to play an important role in preparing training modules, developing a cadre of trainers, conduct of field studies and in promoting interface between SHG members and service providers. The specific role of Resource Centers could be to :

- Work towards a comprehensive capacity building of SHGs,
- Share innovative ideas and models that can be replicated elsewhere,
- Enhance functional literacy among SHG members,
- Support livelihood interventions among SHG members,
- Facilitate availability of all services to SHG members under one roof.

4.1 SHG Loan interest in each district

SHG members are taking loans from the SHG's with the following rate of interest. Wayanad district is giving minimum rate of interest followed by Mallapuram, Calicut and Trichur. The district wise information regarding the minimum and maximum rate of interest taken and the different sources of loan is shown in table VII and VIII respectively.

Table VII shows the district wise information regarding the minimum and maximum rate of interest

District	**Minimum rate of interest**	**Maximum rate of interest**
Kannur	12%	24%
Calicut	11%	24%
Mallapuram	**11%**	**12%**
Palakad	15%	24%
Wayanad	**9%**	18%
Trichur	11%	24%
Kottayam	12%	24%
Alleppy	12%	24%
Trivandrum	12%	25%

Table VIII District wise information regarding different sources of loan

District	No: of Loan taken	Bank Loan	Society Loan	Blade	other groups	Friends
Kannur	335	117	9	38	62	0
Calicut	179	42	6	8	0	0
Mallapuram	215	85	4	23	29	1
Palakkad	148	20	4	8	31	0
Wayanad	232	128 *	2	29	1	1
Trichur	177	27	24	3	22	2

Kottayam	175	69	2	19	0	0
Alleppy	288	42	14	48	0	0
Trivandrum	552	110	27	74	2	0

District wise analysis of the results shows that maximum bank loans are availed by the SHG groups of Wayanad district, maximum society loans are availed by SHG group of Trivandrum and Trichur.

4.2 Relationship between Education of SHG members and Savings

In India majority of the members of SHGs are illiterate and do not have access to formal Education. Even though this is not true in the case of Kerala state, were formal education is gained by almost all individuals. The handicap of literacy would be a hurdle for achieving many desired results. For example they will be unable to follow the accounts maintained by the group and hence remain ignorant about the amount pooled individually and in the group, and would be unable to draft an application to represent their case. It is therefore essential to provide them education through especially designed modules through distance education that are directly useful as a member of SHG. At this stage they do not need school or university certificate, Diploma or degrees. They need improvement in their professional skills and solving their day-to-day problems in the working and functioning of SHGs. They should be explained the advantage of group based strategies in poverty alleviation, Importance of savings and opening bank account, marketing of products, timely repayment and repeat loaning. It is important to explain that she is not alone and that such problems are being faced universally. Only by self-help they may fight against their misfortune and improve upon the fate of their family and children.

Hence a detailed study was done on the role of education and related saving among SHG members

For the study we have taken 3434 objects with 3 attributes, the attributes taken for the study are educational level, saving/month and saving/month outside the group. K-means and Fuzzy C-Means algorithm was applied in this data. The K-mean algorithm has been performed for different values of k and it was found that the best value for k is 2. Fuzzy C-Means was run with different values of m. Table IX and X shows clusters and patterns obtained by applying K-Means respectively.

Table IX Clusters obtained by applying K-Means in Education and savings

No: of runs of K- Means with different seeds(Ri)	No: of patterns in cluster I (C1)	No: of patterns in cluster II (C2)	Seed values
R1	3091	343	Random seed
R2	3355*	79©	Z=10 and Z=100
R3	3355*	79©	Z=10 and Z=3000
R4	79©	3355*	Z=110 and Z=2000
R5	3355*	79©	Z=300 and Z=3000
R6	3355*	79©	Z=500 and Z=510
R7	79©	3355*	Z=644 and Z=844
R8	79©	3355*	Z=1000 and Z=1500
R9	3355*	79©	Z=1400 and Z=1800
R10	79©	3355*	Z=3000 and Z=3010

**,©, these symbols indicate same number of pattern in the clusters.. Since numbers of members are same in certain clusters, we will consider 6 clusters for our studies

Table X Patterns obtained for different runs of K-Mean Algorithm**.**

	R1C1	R1C2	R3C1	R3C2	R4C1	R4C2
Educational level	1.81	2.081	1.8	2.1	2.1	1.83
Savings per month in Rs	17	230	27.6	527.2	527	27
Savings per month	80	528	107	873	873	107

outside the group in Rs						

Table XI and XII shows clusters and patterns obtained by applying Fuzzy C-Means respectively.

Table XI Clusters obtained by applying Fuzzy C-Means in Education and savings

Different runs	No: of iterations	No: of members in cluster 1	No: of members in cluster II	m
R1	60	376	3058	1.25
R2	100*	3025	409	1.5
R3	100*	516	2918	1.75
R4	100*	2907	527	2
R5	100*	637	2797	2.25
R6	95	2771	663	2.5
R7	89	717	2717	2.75
R8	100	2714	720	3
R9	100	727	2707	3.25
R10	98	2702	732	3.5
R11	81	733	2701	3.75
R12	71	2697	737	4
R13	48	945	2489	10
R14	41	2481	953	20
R15	3	954	2480	30
R16	1	2326	1108	40
R17	60	1114	2320	50

*Indicates maximum number of iterations

We have selected clusters with maximum number of iterations (100). Selected clusters are R2C1, R2C2, R3C1, R3C2, R4C1, R4C2, R5C1, and R5C2.The patterns obtained using Fuzzy C-Means is shown in table

Table XII Patterns obtained for different runs of Fuzzy C-Mean Algorithm

	R2C1	R2C2	R3C1	R3C2	R4C1	R4C2	R5C1	R5C2
Educational level	1.810681	2.053234	2.037019	1.804108	1.798286	2.023985	2.014288	1.792751
Savings per month in Rs	17.03678	172.2002	149.3767	15.28171	13.59459	135.1113	125.8471	12.12893
Savings per month outside the group in Rs	71.65053	462.8756	428.3122	64.67036	59.0495	404.3528	386.7046	54.76946

The analysis clearly depicts that there is relationship a between educational level of SHG members and their savings. The table shows that if educational level is high savings per month within the group and outside the groups are high.

RIC2 has maximum savings of 17 runs of Fuzzy C-Means algorithm. RIC2 shows maximum savings and savings outside the group, so we have taken the clusters RICI and RIC2, which gives the total domain for further studies. The study reveals that the educational level of members is high in the cluster R1C2. This shows that there is a relationship between education and savings. The members with higher education show high savings. The figure 1 shows the relationship between % of members in the clusters with educational level.

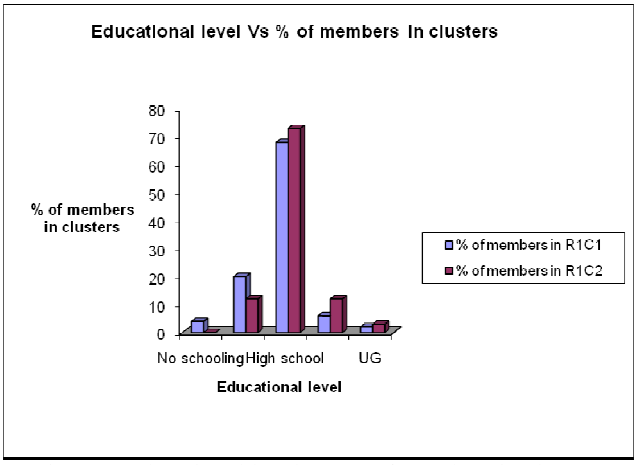

Figure 1 Educational level Vs. % of members in clusters

Further analysis revels that Study of Cluster R1C2, R3C2 and R4C1 (from table X) reveals that the majority of members are in plus 2 educational levels. Figure 2 shows the relationship between educational level and % of members. District wise analysis revealed that SHG members with minimum school educational are from Wayanad and Alleppy.

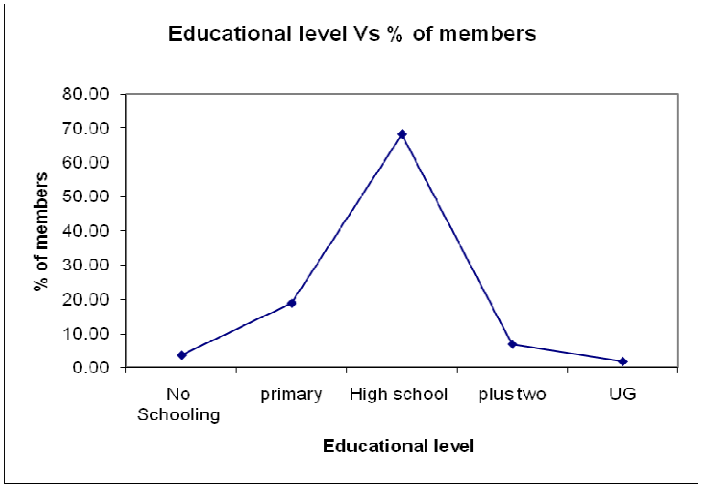

Figure 2 Educational level Vs % of members

So it is necessary that the Government must take effective measures to enroll the members of SHGs in the schemes like Open Schooling. It is observed that open education at present is mainly catering to the needs of elites in the urban areas and it has to make inroads into rural areas where India lives. The Policy planners must think to integrate the economic benefits with education. The economic incentives and effective NGOs participation will definitely make the women empowerment a reality from a distant dream at present.

V. CONCLUSIONS

Surveys were carried out among 3500 SHG members among 9 districts in Kerala with 51 attributes. For this study we have selected 3434 valid data's with 12 attributes. The attributes are loan amount from SHG, interest rate, loan period, loan repay, balance loan in the book, loan taken from other sources, amount taken, interest rate, economic benefit gained, saving, Assets increased after joining SHG, savings outside the group. Data analysis was carried out using K-Means and Fuzzy C-Means algorithm. Studies on both algorithms explain the same facts and the results are almost same.

This study reveals that average range of rate of interest of SHG loans from various government agencies are from 12 to 15 %. But the range of interest from money lenders is from 25-40%. Out of total members availing loan, 56 % is taken from the bank and 22 % are from money lenders, 8 % from

society and 13 % from other SHG group. But majority of members are not availing loan facilities, it may be due to the lack of awareness about different types of loans from the standard financial institution or the rules and regulations for getting loans is more difficult. This can be rectified to a greater extend by providing financial counseling service through face to face interaction, to educate people in rural and urban areas with various financial products available from the financial sector and to make the SHG members aware about the advantages of being connected with the formal financial sector.

District wise studies on the rate of interest of SHG loans showed that minimum rate of interest was given by Wayanad district followed by Mallapuram, Calicut and Thrichur. Maximum bank loans are availed by the SHG members of Wayanad for agricultural purpose. Study on the relationship of education and savings among SHG members shows that members with higher education's shows increased saving habits. 97 % members are literate and out of total members 68.2% are high school educated. It is therefore essential to provide them technical education and financial literacy through especially designed modules through workshops, seminars, open school, distance education that are directly useful as a member of SHG. Only by self-help and training they may fight against their misfortune and improve upon the fate of their family and children. There are other parameters such as loan metrics, socio economic factors and education. This may be considered for further research.

REFERENCES

[1]. Hand, D., Mannila H., Smith P, (2001). "Principles of Data Mining", MIT Press,Cambridge: MA,.
[2]. Fayyad, U. M., Piatetsky-Shapiro, G., Smyth,P., & Uthurusamy, R. (Eds.) (1996). "Advances in knowledge discovery and data mining". Boston: AAAI/MIT Press.
[3]. Fayyad, V., G. G. Grinstein, A. Wierse, (2002). "Information Visualization in Data Mining and Knowledge Discovery", Morgan Kaufmann, San Diego: CA,.
[4]. Andrew Kusiak *, Matthew Smith, (2007) "Data mining in design of products and production systems"; Annual Reviews in Control 31 147–156.
[5]. Witten, I. H., & Frank, E. (2005). "Data mining: Practical machine learning tools and techniques" New York: Elsevier.
[6]. Rundensteiner, E. (ed.): (1999) "Special Issue on Data Transformation". IEEE Techn. Bull. Data Engineering 22 (1).
[7]. Parent, C, Spaccapietra, S, (1998). "Issues and Approaches of Database Integration". Comm. ACM 41(5):166-178.
[8]. Rachsuda Jiamthapthaksin, Christoph F. Eick, and Vadeerat Rinsurongkawong, (2009). "An Architecture and Algorithms for Multi-Run Clustering", IEEE, 978-1-4244-2765-9/09.
[9]. Chakrabarti,R.(2004)."The Indian Microfinance Experience – Accomplishments and Challenges" www.microfinacegateway.org.
[10]. Wilson,K.(2002)."The new microfinance-an essay on the self-help group movement in India". Journal of Microfinance, Vol. 4, No. 2, pp.21–245.
[11]. Bansal,H.(1998), "Self-Help Group- NGOs-Bank Linkage Programmes in India: A Case Study", M.S University of Baroda, (www.prism.gatech.edu)
[12]. Wilson Kim,(2002), "The Role of Self Help GroupBank Linkage Programme in Preventing Rural Emergencies in India" NABARD,(see www.nabard.org)
[13]. Puhazhendhi,V,& Badaty K.C,(2002),"SHGBank Linkage Programme-An Impact Evaluation". NABARD, (www.nabard.org)
[14]. B. Narayana swamy, K. Narayana Gowda and G. N. Nagaraj (2007). "Performance of Self Help Groups of Karnataka in Farm Activities"; Karnataka J. Agric. Sci.,20(1):85 - 88.
[15]. J. Pena, J. Lozano, and P. Larranaga, (1999) "An Empirical Comparison Of Four Initialization Methods For The K-Means Algorithm," Pattern Recognition Letters, Vol. 20 No. 10, pp. 1027-10.
[16]. Madhu Yedla, Srinivasa Rao Pathkota, T.M.Srinivasa,(2010)."Enhancing K-means clustering Algorithm with Improved Initial Center". International Journal of computer science and information technologies, Vol.1, pp 121-125.
[17]. Ohn mar san, Van-nam huynh, Voshiteru nakamori, (2004)."An alternative extension of the K-Means algorithm for clustering categorical data". Int. J. Appl. Math. Computer Science Vol. 14, No. 2, pp 241–247.

[18]. J. Macqueen, (1967) "Some Methods For Classification And Analysis Of Multivariate Observations," In proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, pp. 281-297.
[19]. M.S. Yang, C.H. Ko, 1997. On cluster-wise fuzzy regression analysis, IEEE Trans. Systems, Man, Cybern. 27, 1–13.
[20]. Kuo-Lung Wu, Miin-Shen Yang; 2002. Alternative C-Means clustering algorithms, Pattern Recognition 35, 2267 – 2278.
[21]. H. Galhardas, D. Florescu, D. Shasha, E. Simon, and C. Saita, (2001). "Declarative Data Cleaning: Language, Model, and Algorithms," roc. 2001 Very Large Data Bases (VLDB) Conf.
[22]. M. Hernandez and S. Stolfo, (1995). "The Merge/Purge Problem for Large Databases," Proc. ACM SIGMOD Int'l Conf. Management of Data, pp. 127-138.
[23]. M.L. Lee, T.W. Ling, and W.L. Low, (2000), "Intelliclean: A Knowledge- Based Intelligent Data Cleaner," Proc. Sixth ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining.
[24]. L. Yu and H. Liu. (2003) "Feature selection for high-dimensional data: a fast correlation-based filter solution". In: Proceedings of the 20th International Conferences on Machine Learning, Washington DC.
[25]. R. Kohavi and G. John. (1997). "Wrappers for feature subset selection. Artificial Intelligence," Vol. (1-2), pp. 273-324.
[26]. B. Kumar (2005). "Impact of Microfinance through SHG-Bank Linkage in India: A Micro Study" Vilakshan, XIMB Journal of Management, July 9.
[27]. Mahendra Varman P; (2005)."Impact of Self-Help Groups on Formal Banking Habits" Economic and Political Weekly April 23, pp 1705-13.

AUTHOR BIOGRAPHIES

SAJEEV B.U is pursuing Ph. D. in Computer Science and Engineering under the guidance of Dr.Thangavel K, from Center for Research and Development PRIST University, Thanjavoor, Tamil Nadu, India. He received his Masters degree in Mathematics from Calicut University in 1984, M.C.A. from Mahatma Gandhi University and M.Tech in Computer Science from Allahabad Agricultural Institute- Deemed University, Allahabad in 2006. Currently he is working as HOD, Department of Computer Applications at KVM, CE & IT, Cherthala, Kerala, India. His research interest includes Data Mining, Clustering and Pattern Recognition.

THANGAVEL KUTTIANNAN received the Master of Science from Department of Mathematics, Bharathidasan University in 1986, and Master of Computer Applications Degree from Madurai Kamaraj University, India in 2001. He obtained his Ph. D. Degree from the Department of Mathematics, Gandhigram Rural University in 1999. He worked as Reader in the Department of Mathematics, Gandhigram Rural University, up to 2006. Currently he is working as Professor and Head, Department of Computer Science, Periyar University, Salem, Tamilnadu, India. His areas of interest include medical image processing, artificial intelligence, neural network, fuzzy logic, data mining, pattern recognition and mobile computing. He is the recipient of Tamilnadu Scientist Award for the year 2009.