IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-6, Issue-7, July 2018 E-ISSN: 2347-2693

The Role of Technology and Education in Financial Inclusion – A Data Mining Analysis in a Fuzzy Framework

Binu Thomas

Department of Computer Applications, Marian College/Mahatma Gandhi University, Kerala, India

Corresponding Author: binu.thomas@mariancollege.com, Tel.: +00-919446845140

Available online at: www.ijcseonline.org

Accepted: 22/Jul/2018, Published: 31/Jul/2018

***Abstract*—** Data mining is the exploration and analysis of large data sets, in order to discover meaningful patterns and rules. Fuzzy Logic can be integrated with Data Mining techniques to incorporate human type reasoning in pattern discovery. In this paper we are using Fuzzy Data Mining techniques in a database created from a survey pertaining to Financial Inclusion. The popular Data Mining techniques like clustering and association rule mining are used in a fuzzy framework at various stages of the analysis. Starting from the survey database this paper proceeds through all the steps of Data Mining like pre-processing, attribute selection, segmenting quantitative values, clustering and finally reaching at natural fuzzy association rules. Financial Inclusion is one of the key areas where economists and governments try to concentrate for the eradication of poverty. With this analysis we clearly reach at a conclusion that Education and introduction to Information Technology plays the most significant role in Financial Inclusion.

***Keywords*—** Financial inclusion, Fuzzy Data Mining, Fuzzy Logic, Fuzzy Clustering

I. Introduction

Unrestrained access to public goods and services is the visible outcome of an open and efficient society. The availability of banking and payment services to the entire population without discrimination is the prime objective of this public policy. While in developed countries, the formal financial sector comprising mainly the banking system serves most of the population [1]. In developing countries, a large segment of the society, mainly the low-income group, has little access to financial services, either formal or semi formal. As a result, many people have to necessarily depend either on their own sources or informal sources of finance, which are generally at high cost [2]. It is said that poverty is not merely insufficient income, but rather the absence of wide range of capabilities, including security and ability to participate in economic and political systems. These large numbers of poor are required to be provided with much needed financial assistance in order to sail them out of their conditions of poverty [3]. Importance of financial inclusion arises from the problem of financial exclusion of nearly 3 billion people from the formal financial services across the world [4].

A. Financial inclusion

The term "financial inclusion" has gained importance since the early 2000s, and is a result of findings about financial exclusion and its direct correlation to poverty. Financial Inclusion is delivery of banking services at an affordable cost to vast section of disadvantaged and low income group [5]. It is now a common objective for many central banks among the developing nations. Financial inclusion or inclusive financing is the delivery of financial services at affordable costs to sections of disadvantaged and low income segments of society[6]. A World Bank study had shown that half of the world's population held accounts with formal financial institutions. The study said only nine per cent of the population had taken new loans from a bank, credit union or microfinance institutions. In India, only 35 per cent have formal accounts versus an average of 41 per cent in developing economies [7].

B. Financial inclusion in India

While India has witnessed unprecedented economic growth in recent past, its development has been lopsided with the country trailing on essential social and environmental parameters of development. The approach paper to the Eleventh Plan indicated that the absolute number of poor is estimated to be approximately 300 million in 2004-05. Accordingly, the 11th Five Year Plan has adopted "faster and more Inclusive growth" as the key development paradigm [8]. The importance of this study lies in the fact that India being a socialist, democratic republic, it is imperative on the policies of the government to ensure equitable growth of all

© 2018, IJCSE All Rights Reserved

sections of the economy. With only 34% of population engaged in formal banking, India has, 135 million financially excluded households. Further, the real rate of financial inclusionin India is also very low and about 40% of the bank account holders use their accounts not even once a month. It is universally opined that the resource poor need financial assistance at reasonable costs and that too with uninterrupted pace. However, the economic liberalization policies have always tempted the financial institutions to look for more and more greener pastures of business ignoring the weaker sections of the society. In India, the financially excluded sections comprise largely rural masses comprising marginal farmers, landless labourers, oral lessees, self-employed and unorganized sector enterprises, urban slum dwellers, migrants, ethnic minorities and socially excluded groups, senior citizens and women [9].

C. Aim of the study

There are a variety of reasons for financial exclusion. In remote, hilly and sparsely populated areas with poor infrastructure, physical access itself acts as a deterrent. From the demand side, lack of awareness, low incomes/assets, social exclusion, illiteracy act as barriers. From the supply side, distance from branch, branch timings, cumbersome documentation and procedures, unsuitable products, language, staff attitudes are common reasons for exclusion. All these result in higher transaction cost apart from procedural hassles. On the other hand, the ease of availability of informal credit sources makes these popular even if costlier[10] We selected the remote villages of Idukki districts of Kerala, India for conducting a survey. Idukki is considered as one of the under developed regions in the country. It is evident that high rate of financial exclusion prevail in the district. The Questions were selectively drafted from the above mentioned parameters to reveal the extent of financial inclusion and exclusion in the region and to identify the factors which support financial inclusion. The respondents were also selected carefully to ensure that all factions of the society in the area are coming under the survey. In the first step of the analysis we arrived at the statistical findings of the survey which gives the actual facts and figures of the situation. In the second step we applied the Data Mining techniques on a fuzzy framework to the data set collected from the survey. Through these fuzzy data mining techniques finally we are reaching at certain natural patterns which lead to financial inclusion in rural areas. These patterns are clearly indicating that education and awareness of modern technology plays a vital role in financial inclusion.

Rest of the paper is organized as follows, Section I contains the introduction of the paper, Section II contain the introduction to Data Mining, Section III contain the some introduction to Fuzzy Logic, Section IV contain the is about financial inclusion, section V explain datamining methodology, VI describes Association Rule Mining, Section VII concludes research work with discussions.

II. Data Mining

Data mining is the exploration and analysis of large data sets, in order to discover meaningful patterns and rules. Data Mining or Knowledge discovery refers to a variety of techniques that have developed in the fields of databases, machine learning and pattern recognition [11]. The intent of these techniques is to uncover useful patterns and associations from large databases. The process of finding useful patterns and information from raw data is often known as Knowledge discovery in databases or KDD. Data mining is a particular step in this process involving the application of specific algorithms for extracting patterns (models) from data [12].It is a process of mining a data source for information that is once unaware of prior to the discovery. This spans the entire spectrum for discovering information of which one has absolutely no knowledge before the mining process.

A. Cluster analysis

Cluster analysis is a technique in Data Mining for breaking data down into related components in such a way that patterns and order becomes visible. It aims at sifting through large volumes of data in order to reveal useful information in the form of new relationships, patterns, or clusters, for decision-making by a user. Clusters are natural groupings of data items based on similarity metrics or probability density models. Clustering algorithms maps a new data item into one of several known clusters. In fact cluster analysis has the virtue of strengthening the exposure of patterns and behaviour as more and more data becomes available[13]. A cluster has a center of gravity which is basically the weighted average of the cluster. Membership of a data item in a cluster can be determined by measuring the distance from each cluster center to the data point[14]. The data item is added to a cluster for which this distance is a minimum.

B. Association rule mining

Among the areas of data mining, the problem of deriving associations from data has received a great deal of attention. The problem was formulated by Agrawal *el al.* in 1993 and is often referred to as the *market-basket problem*[15]. In this problem, we are given a set of items and a large collection of transactions, which are subsets (baskets) of these items. The task is to find relationships between the presences of various items within these baskets.

There are numerous applications of data mining which fit into this framework. The classic example from which the problem gets its name is the supermarket . In this context, the problem is to analyse customer's buying habits by finding associations between the different items that customers place in their shopping baskets. In these and many other areas,

© 2018, IJCSE All Rights Reserved

association rule mining has lead to new insights and new business opportunities[15]. Of course the concept of a market basket needs to be generalized for these applications. For example, a market basket is replaced by the collection of medical services received by a patient during an episode of care, the web pages visited by a user in one browsing session or the set of words or concepts used in a web page. Thus when applying association rule mining to new areas one faces two core questions:

i) What are the “items” and
ii) What are the “market baskets”

The answer of these questions is facilitated if one has an abstract mathematical notion of items and market baskets.

III. UNCERTAINTY AND FUZZY LOGIC

When we examine the features limitations of natural data, two categories of limitations emerge quite naturally – Uncertainty and Vagueness[16]. Vagueness is associated with the difficulty of making sharp and precise distinctions in the world. Uncertainty is a situation in which the choice between two or more alternatives is left unspecified. The modelling of imprecise and qualitative knowledge, as well as handling of uncertainty at various stages is possible through the use of fuzzy sets. Fuzzy logic is capable of supporting, to a reasonable extent, human type reasoning in natural form by allowing partial membership for data items in fuzzy subsets. Integration of fuzzy logic with data mining techniques has become one of the key constituents of soft computing in handling the challenges posed by the massive collection of natural data. Fuzzy logic is a logic of fuzzy sets. A Fuzzy set has, potentially, an infinite range of truth values between one and zero. Propositions in fuzzy logic have a degree of truth, and membership in fuzzy sets can be fully inclusive, fully exclusive, or some degree in between.The fuzzy set is distinct from a crisp set is that it allows the elements to have a degree of membership. The core of a fuzzy set is its membership function: a function which defines the relationship between a value in the sets domain and its degree of membership in the fuzzy set(exp 1). The relationship is functional because it returns a single degree of membership for any value in the domain [17].

$\mu=f(s,x)$ (1)

Here,

μ : is he fuzzy membership value for the element
s : is the fuzzy set
x : is the value from the underlying domain.

Fuzzy sets provide a means of defining a series of overlapping concepts for a model variable since it represent degrees of membership. The values from the complete universe of discourse for a variable can have memberships in more than one fuzzy set [17].

IV. THE SURVEY ON FINANCIAL INCLUSION

In this work we are using the data collected by the Post graduate Department of Commerce in Marian College, Kuttikkanam (www.mariancollege.org)through a field survey. This was conducted personally by small groups of students by visiting villages located in remote and comparatively undeveloped areas of Idukki District The respondents are selected from remote villages of Idukki which is a backward district of Kerala. The approach to the subject of financial inclusion was based on three major premises:

1. The national agenda of securing justice social as enunciated in our Constitution is far from complete even after six decades of planning and policy formulation;

2. Inequalities of income and of opportunities have arisen mainly from imbalances in the process of growth and have resulted in the endemic problems of poverty, unemployment and backwardness persisting in the rural area;

3. Development and growth in their true sense can be achieved only by empowerment of rural masses and by correcting the serious anomaly involved in the fact that agriculture on which 77 per cent of our population depend for their livelihood, accounts for only 13.5 per cent of the national GDP.

In the light of these considerations, the enquiries have been directed to shed some light on local conditions of denial or deprivations that might have contributed to backwardness or lack of development in remote areas and to examine the role of financial inclusion to serve as an effective tool to mitigate such conditions. The statistical findings from the survey are presented first in this chapter. These findings are showing the real situation regarding banking inclusion in the region. After this we are using the data and the findings to apply data mining tools and concepts presented in this thesis so that certain hidden patters pertaining to financial inclusion will be revealed. For the analysis we have converted the questions into a database table. The database table contains 34 fields and its structure is given in table 8.

A. *Statistical findings from the survey*

The statistical analysis of the awareness and usage of the banking services are found and it is presented in the following illustrations. There are 512 respondents for the survey and the main areas of analysis are given below.

B. *Awareness and usage of banking service*

This is to find the awareness of the people about various facilities available with modern banking such as Deposites,

© 2018, IJCSE All Rights Reserved

Credits, ATM, Money transfer etc., The percentage of people who actually use these kind of facilities are also found.

C. Bank Account

This analysis shows the present status of banking inclusion. This simply demonstrates the percentage of people having a bank account and the percentage of people not having any bank account. This information is very important to understand the status of financial inclusion in the region.

D. Details of Bank Account

This is a segmentation of the respondents according to the choice of their banking institution. Mostly we have banks coming under state bank groups and another major player is co-operative banks. Other scheduled banks and other pure private banks are also present in the region but their access to remote areas are very limited**.**

E. Reason for Not having Bank Account

Among the respondents 35.5 percentages are not having accounts in any of the financial institutions. Lack of awareness and accessibility are the main reasons for not having accounts in banks. Even now many of the public sector banks are not having their presence in the remote villages.

F. Type and source of Credit avail

Among the respondents 42% has not availed any credits from formal financial institutions and 58% are having loans from formal banking institutions. Mainly these loans are for the purpose of agriculture, Business, Educations etc,

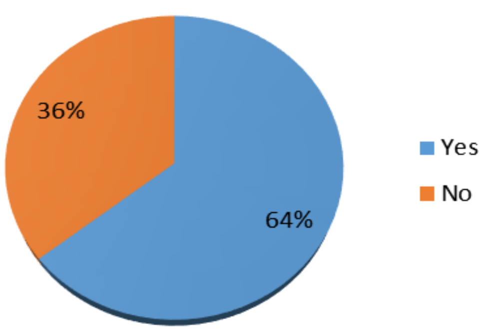

Fig.1 The percentage having Bank Account

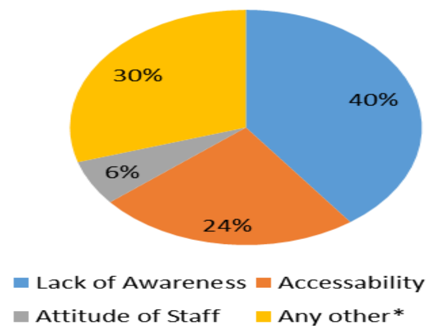

Fig.2 Reasons for not having bank account

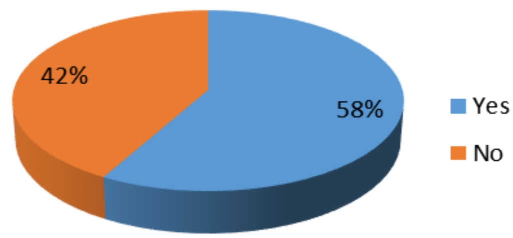

Fig3. Credit loan availed

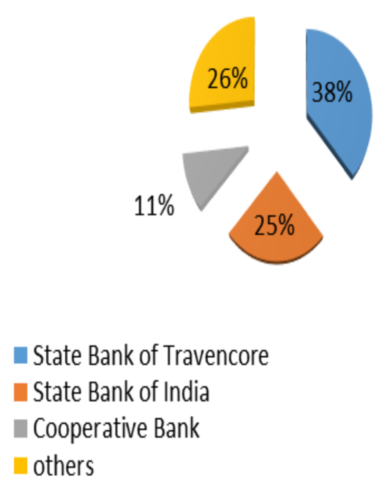

Fig.4 Distribution of bank accounts

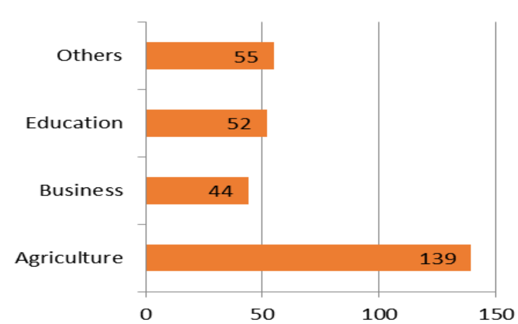

Fig.5 Type of credit availed by respondents

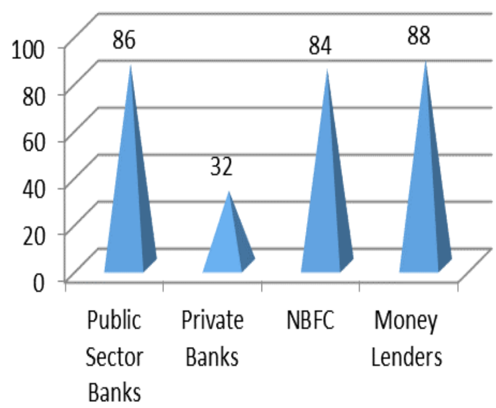

Fig.6 Source of Credit

V. DATA MINING APPROACH TO FINANCIAL INCLUSION

Data mining is a field where computer science and statistics intersect to discover hidden patterns in large data sets. The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use. Aside from the raw analysis step, it involves database and data management aspects, data preprocessing, model and inference considerations, interestingness metrics, complexity considerations, post-processing of discovered structures and visualization[18].

The data mining approach to financial inclusion is to find out the hidden patterns and valuable knowledge from the collection of data. These patterns and acquired knowledge must give light to the factors affecting financial inclusion and exclusion. The patterns identified from such a data mining analysis in this area will be a set of rules revealing the socio-economic and educational conditions which leads to either financial inclusion or exclusion. The rules generated by analyzing the hidden patterns from the facts and figures of the respondents of the survey can be later used to predict the behavior patterns of the people from similar context pertaining to financial activities.

A. Attribute selection.

The first step in any data mining analysis is attribute (feature) selection. It is the process of selecting the attributes from the database which may lead to the discovery of valuable patterns. Feature selection implies not only cardinality reduction, which means imposing an arbitrary or predefined cutoff on the number of attributes that can be considered when building a model, but also the choice of attributes, meaning that either the analyst or the modeling tool actively selects to discards attributes based on their usefulness for analysis. In a data mining analysis one typically wants to remove unnecessary columns because they might degrade the quality of discovered patterns and we may not be able to reach at the required information. Many feature selection methods are available these days and the simplest and popular method is clustering[19,26]. We are using popular k-means method to cluster the database attribute values into groups.

5.2 K-means algorithm

K-means algorithm and its different variations are the most well-known and commonly used partitioning methods. The value 'k' stands for the number of cluster seeds initially provided for the algorithm. This algorithm takes the input parameter 'k' and partitions a set of m objects into k clusters . The technique work by computing the distance between a data point and the cluster center to add an item into one of the clusters so that intra-cluster similarity is high but inter-cluster similarity is low [20] . A common method to find the distance is to calculate to sum of the squared difference as follows and it is known as the Euclidian distance (exp.2).

$$d_k = \sum_n \left\| X_j^k - Ci_j \right\|^2 \qquad (2)$$

where,

d_k : is the distance of the k^{th} data point
n : is the number of attributes in a cluster
X_j^k : is j^{th} value of the k^{th} data point
C_j^i : is the j^{th} value of the i^{th} cluster center

Table 1. K-means clustering algorithm

initialise k=number of clusters
initialise C_j (cluster centers)
Set Cycle variable t=1
Repeat
For i=1 to n Distribute sample points(x_i)into k clusters
For j=1 to k : Calculate $S_j^{(t)}$ for x_i applying (2)
For j=1 to k : Compute new cluster centers by calculating weighted average
t=t+1
Until Cj estimate stabilize

With the definition of the distance of a data point from the cluster centers, the k-means the algorithm is fairly simple. The cluster centers are randomly initialized and we assign data points x_i into clusters based on their minimum distance to cluster centers using Euclidean distance.

5.3 Attribute selection using k-means algorithm

To use clustering as a feature selection tool for our analysis we did sufficient pre-processing with the original data. All the non-numeric field values are converted into

© 2018, IJCSE All Rights Reserved

corresponding values in a numeric point scale. Then we applied k-means algorithm with k value of 4 to segment the respondents into clusters. IBM SPSS Modeler 14.1 is used for the analysis. This Data Mining software tool displays the importance of attributes in clusters. From this information we identified the attributes which are significant in each of these clusters. These attributes are later included into the analysis and the attributes which are found insignificant in the cluster analysis are discarded from further analysis. We found that 20 attributes mentioned in the above table are present in the clusters formed with significant contribution (Fig. seven).

5.4 Applying fuzzy clustering to segment quantitative values

A modified fuzzy c-means algorithm is presented for natural data exploration [21] and this is a modification to the classical fuzzy c-means algorithm[22] to overcome many of its limitations. In c-means algorithm the membership of a data point in a cluster depends directly on the sum of distances of the point in other cluster centers. Many limitations of the algorithm which affect the performance arise due this method. Instead, if we consider the sum of distances of data members in a cluster for the calculation of memberships in that cluster, it will improve the performance of the algorithm. This leads to the modified equation for membership calculation. The new membership function for i^{th} data point in j^{th} cluster is given below (Exp 3).

$$\mu_j(x_i) = n * \frac{\left(\frac{1}{d_{ji}}\right)^{\frac{1}{m-1}}}{\sum_{i=1}^{n}\left(\frac{1}{d_{ji}}\right)^{\frac{1}{m-1}}} \qquad (3)$$

where

$\mu_j(x_i)$: is the membership of x_i in the j^{th} cluster
d_{ji} : is the distance of x_i in cluster c_j
m : is the fuzzification parameter
p : is the number of specified clusters

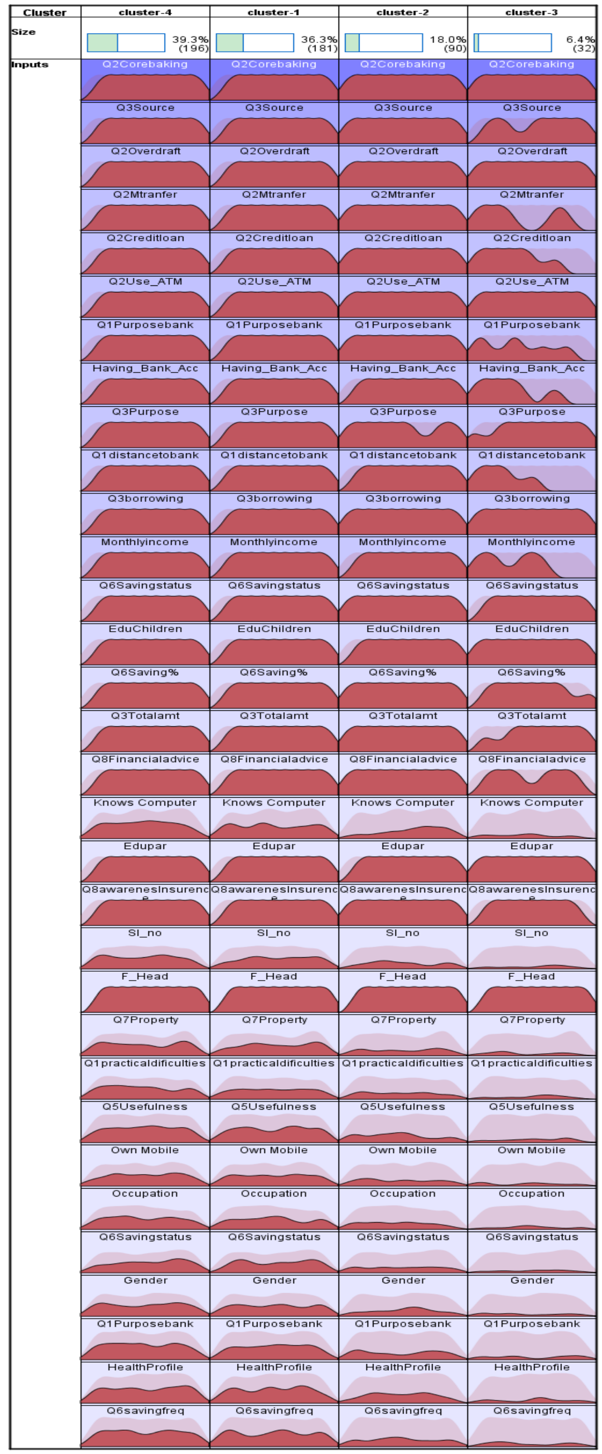

Fig. 7 cluster contributions

Membership values found with the above equation can be used in the same c-means algorithm to produce better results.

Segmenting quantitative attribute values

Among the attributes selected for further analysis, Monthly income, distance to bank and amount borrowed are quantitative attributes. We used the modified fuzzy clustering method[23] to segment these attributes into fuzzy segments like very low, low, medium and high. In the case of education we converted the education into a 20 point scale and the values are segmented using fuzzy clustering to group the values into three fuzzy segments namely low medium and high(Table 2).

For segmenting the users according to their involvement in various banking operations, we selected the fields from the database which directly represent different banking operations. The fields selected are Having bank account, Credit loan, Money transfer, Borrowing, Use ATM, Use core banking, Having deposits. These fields are Boolean in type and their values are converted into 0 and 1 and the sum of these 7 seven fields are found for each respondent. In the next step the sum filed vales are segmented into fuzzy segments named *low* , *medium* and *high* as given in the following table(Table 3).

In the same manner similar to this we also grouped the fields like *having bank account, borrowing, saving, aware insurance,* and *financial advice* to represent their financial inclusion status. The sum obtained from these fields are then mapped into fuzzy linguistic segments *low, medium* and *high.*

Other fields are also processed in a comparable manner and the field values are converted into fuzzy linguistic terms. The following table (Table 4) summarizes the groupings of attributes and the linguistic variables generated from these attributes

VI. FINDING ASSOCIATION AND FUZZY CLASSIFICATION RULES

In Association rule mining, the quantitative attribute values may be converted into Boolean values using intervals to apply conventional association rule mining algorithms for finding relations among the items. These intervals may not be concise and meaningful enough for human users to easily obtain non trivial knowledge from those rules discovered. Fuzzy Clustering is an efficient modeling technique which can be used for segmenting quantitative values into meaningful groups instead of fixed intervals. Membership values of quantitative items in the partitioning clusters are used with weighted fuzzy rule mining techniques to find natural associations[24].

Table 2. Fuzzyfication of Educational qualification

Qualification	20 pt. Scale	Fuzzy segments and Memberships		
		Low	Med.	High
Master Degree	17	0.111	0.500	1.000
Vocational Technical Courses	15	0.143	1.000	0.333
Std 8	8	1.000	0.143	0.100
Vocational Technical Courses	15	0.143	1.000	0.333
Up to degree	15	0.143	1.000	0.333
Master Degree	17	0.111	0.500	1.000
Master Degree	17	0.111	0.500	1.000
Vocational Technical Courses	15	0.143	1.000	0.333
Std 7	7	1.000	0.125	0.091
Std 6	6	0.500	0.111	0.083
Master Degree	17	0.111	0.500	1.000
Master Degree	17	0.111	0.500	1.000
Up to 10 Std	10	0.500	0.200	0.125
Master Degree	17	0.111	0.500	1.000
Master Degree	17	0.111	0.500	1.000
Master Degree	17	0.111	0.500	1.000
Up to 10 Std	10	0.500	0.200	0.125
M Tech	18	0.100	0.333	1.000
Master Degree	17	0.111	0.500	1.000
Vocational Technical Courses	15	0.143	1.000	0.333
Vocational Technical Courses	15	0.143	1.000	0.333
Up to 10 Std	10	0.500	0.200	0.125
Vocational Technical Courses	15	0.143	1.000	0.333
Master Degree	17	0.111	0.500	1.000

A. *Association rules*

Definition: For a given transaction database T, an association rule is an expression of the form $X \Rightarrow Y$, where X and Y are subsets of A and $X \Rightarrow Y$ holds with confidence τ , if τ % of transactions in D that support X also support Y. The rule $X \Rightarrow Y$ (Example Bred $\Rightarrow$ Butter) has support σ in the transaction set T if σ % of transactions in T support $X \cup Y$.

© 2018, IJCSE All Rights Reserved

Table 3. The fuzzy derived attribute Banking Inclusion

Banking Inclusion (BI) fuzzy derived attribute										
Fields related to banking							**Fuzzy Sets for BI**			
Have_Acc	**Credit loan**	**M transfer**	**Borro wing**	**Use ATM**	**banking**	**Deposite**	**Sum**	**low**	**Me dium**	**High**
1	0	1	0	1	1	1	5	0.20	0.50	1.00
1	1	1	0	1	1	0	5	0.20	0.50	1.00
0	1	1	0	1	1	0	4	0.25	1.00	0.50
1	0	0	1	0	0	0	2	0.50	1.00	0.25
1	0	0	1	0	0	0	2	0.50	1.00	0.25
0	0	0	0	0	0	1	1	1.00	0.50	0.20
1	0	1	0	0	1	0	3	0.33	1.00	0.33
1	0	1	1	1	0	0	4	0.25	1.00	0.50
1	0	0	1	0	0	0	2	0.50	1.00	0.25
1	1	0	1	0	1	1	5	0.20	0.50	1.00
1	0	1	1	0	0	1	4	0.25	1.00	0.50
0	1	1	1	1	1	0	5	0.20	0.50	1.00
0	0	1	1	0	0	0	2	0.50	1.00	0.25
1	1	0	0	0	0	0	2	0.50	1.00	0.25
1	1	1	1	1	1	0	6	0.17	0.33	1.00
0	0	0	1	0	0	0	1	1.00	0.50	0.20
0	0	0	0	0	0	1	1	1.00	0.50	0.20
0	0	1	1	0	0	0	2	0.50	1.00	0.25
1	0	0	1	1	0	1	4	0.25	1.00	0.50
0	0	1	0	0	0	1	2	0.50	1.00	0.25
0	0	0	1	0	0	0	1	1.00	0.50	0.20
1	1	1	0	1	1	0	5	0.20	0.50	1.00

Support ($(X \Rightarrow Y) = P(X \cup Y)$) measures how often two items occur together as a percentage of the total transactions. Confidence ($(X \Rightarrow Y) = P(Y \mid X)$) measures how much a particular item is dependent on another. The intuitive meaning of such a rule is that a transaction of the database which contains X tends to contain Y. given a set of transactions, T, the problem of mining association rules is to discover all rules that have support and confidence greater than or equal to the user-specified minimum support and minimum confidence, respectively.

Table 4. Derived fields from the survey questions with fuzzy linguistic variables

Fields	**Derived field**	**Fuzzy Linguistic variables**
Edupar	Edupar	Low, medium, high
EduChildren	EduChildren	Low, medium, high
Monthlyincome	Monthlyincome	low, medium, high
Distancetobank	Distancetobank	Near, medium, far
Mtranfer, Use_ATM, Borrowing Having_Bank_Acc, Core banking, Credit loan, Saving status	Banking inclusion	Low, medium, high
Amount borrowed	Amount borrowed	Low, medium, high
Purpose (Agriculture, Business, Education, Housing, Personal)	Loan accessibility	Low, medium, high
Savingstatus, borrowing, Having_Bank_Acc, awarenesInsurence, Financialadvice	Financial Inclusion	Low,medium,high
Corebaking, Use ATM, Knows Computer, Own mobile	Modern banking	Low,medium,high

B. Fuzzy Association rules

Fuzzy association rules use fuzzy logic to convert numerical attributes to fuzzy attributes, like “Income = High”, thus maintaining the integrity of information conveyed by such numerical attributes. On the other hand, crisp association rules use sharp partitioning to transform numerical attributes to binary ones like “Income = [10000-20000]”, and can

© 2018, IJCSE All Rights Reserved

potentially introduce loss of information due to these sharp ranges. Fuzzy Apriori and its different variations are the only popular fuzzy association rule mining (ARM) algorithms available today[25].

In fuzzy association rule mining, a two-fold pre processing approach is used to handle quantitative attributes. Firstly, quantitative attributes are discredited into different fuzzy linguistic intervals and weights assigned to each linguistic label. A mining algorithm is applied then on the resulting dataset by applying fuzzy weighted support and confidence measures. Let the input data D with transactions t = $\{t1,t2,..,tn\}$ have a set of items $I = \{i_1,i_2,..,i_I\}$ and a set of positive real numbered weights $W=\{ w_1, w_2, w_I \}$ corresponding to each item i. Each i^{th} transaction t_i is some subset of I and a weight w is attached to each item $t_i[i_j]$ (j^{th} item in the "i^{th}" transaction). A pair (i, w) is called a weighted item where '*i' is an* item and '*w' is its weight*[24]

Definition 1. Fuzzy Item Weight is a non-negative real value given to each item ranging [0..1] with some degree of importance, a weight .

Definition 2. Fuzzy Itemset Transaction Weight is the aggregated weight of all the items in the itemset present in a single transaction. Itemset transaction weight for an itemset X can be calculated as follows (expr. 4):

Weight of transaction

$$t_i = \prod_{k=1}^{|x|} {}_{(\forall[t[w]]\in X)} t_i[i_k[w]] \quad (4)$$

Table 5 contains 10(ten)transactions and the weights of items in a transaction. The transaction t_1 involves items A and B with weights *0.6* and *0.9*. The weight of this transaction is calculated as:

TW(B,D)=0.6*0.9=.54

Definition 3. Fuzzy *Weighted Support FWS* is the aggregated sum of itemset transaction weight of all the transactions in which itemset is present, divided by the total number of transactions. It calculated as follows(expr 5)

Fuzzy Weighted support

$$FWS = \frac{\sum_{i=1}^{n} \prod_{k=1}^{|x|} {}_{(\forall[t[w]]\in X)} t_i[i_k[w]]}{n} \quad (5)$$

In the ten(10) transactions items A & B appear together in three(3) transactions. The weighted support for (A,B) is calculated as:

FWS(A,B)= (0.54+0.54+0.54)/10=0.162

Table 5. Transactions and item weights

Trn No	Items				
	A	B	C	D	E
t_1	.6	.9	0	0	0
t_2	0	.9	0	.1	0
t_3	.6	0	0	.1	0
t_4	0	0	.3	0	0
t_5	.6	.9	0	.1	.2
t_6	.6	.9	.3	.1	.2
t_7	0	.9	0	0	.2
t_8	0	0	0	.1	.2
t_9	.6	.9	0	.1	0
t_{10}	0	.9	.3	.1	.2

Definition 4. Fuzzy *Weighted Confidence FWC i*s the ratio of sum of votes satisfying both *X and Y* to the sum of votes satisfying X. It is formulated in the following manner (expr 6).

$$\sum_{i=1}^{n} \frac{\prod_{k=1}^{|z|} {}_{(\forall[z[w]]\in z)} t_i[z_k[w]]}{\prod_{k=1}^{|x|} {}_{(\forall[i[w]]\in X)} t_i[x_k[w]]} \quad (6)$$

Fuzzy Weighted confidence of (A,B) is calculated as:

FWC(A,B) = FWS(A,B)/FWS(A)

Fuzzy Weighted support of (A,B) is 0.162 and for A it is 0.3 (A occurs five times (.6+.6+.6+.6+.6)/10=0.3). So the weighted confidence of (A,B) is:
FWC(A,B)=0.162/0.3=0.54

C. Fuzzy associations in financial inclusion

For finding the associations between the new linguistic attributes derived from the survey database, we used the weighted fuzzy association rule mining techniques presented in the preceding section. The fuzzy associations rules generated from the survey database can be later converted into classification rules so that we will be able to derive the facts pertaining to financial inclusion. These rules can be effectively used in the future to address the issue of financial inclusion and to eliminate certain conditions which lead to financial exclusion of common mass in backward areas. To start with the fuzzy association rule mining from the derived data set, we selected the derived attributes *Banking inclusion, Amount borrowed, financial inclusion* and *loan accessibility* as the consequents for the analysis. Then the fuzzy

© 2018, IJCSE All Rights Reserved

associations of other derived fields with these consequents are found.

Consequents											
Banking inclusion			Loan accessibility			Financial Inclusion			Amount borrowed		
low	med	high	low	med	high	low	med	high	low	med	high
0.2	0.5	0.8	0.5	0.9	0.2	0.1	0.3	0.9	0.3	0.4	0.8
0.7	0.3	0.2	0.8	0.4	0.1	0.3	0.7	0.2	0.9	0.4	0.1
0.3	0.6	0.9	0.6	1.0	0.3	0.2	0.4	1.0	0.4	0.5	0.9
0.5	0.9	0.3	0.3	0.5	0.9	0.4	1.0	0.7	0.4	0.9	0.5
0.4	0.8	0.2	0.2	0.4	0.8	0.3	0.9	0.6	0.3	0.8	0.4
0.7	0.3	0.2	0.8	0.4	0.1	0.3	0.7	0.2	0.9	0.4	0.1
0.5	0.9	0.3	0.3	0.5	0.9	0.4	1.0	0.7	0.4	0.9	0.5
0.4	0.8	0.2	0.2	0.4	0.8	0.3	0.9	0.6	0.3	0.8	0.4
0.3	0.6	0.9	0.6	1.0	0.3	0.2	0.4	1.0	0.4	0.5	0.9
0.2	0.5	0.8	0.5	0.9	0.2	0.1	0.3	0.9	0.3	0.4	0.8
0.8	0.4	0.3	0.9	0.5	0.2	0.4	0.8	0.3	0.8	0.5	0.2

Fig. 8 Consequents of fuzzy rules

Anticedents														
Edupar			EduChildren			M_income			Dist_tobank			Modern banking		
low	med	high	low	med	high	low	med	high	near	med	far	low	med	high
0.2	0.5	0.7	0.2	0.6	0.8	0.1	0.3	0.8	0.8	0.4	0.2	0.4	0.3	0.8
0.8	0.3	0.4	0.3	0.7	0.4	0.5	0.7	0.2	0.2	0.5	0.7	0.8	0.4	0.2
0.3	0.6	0.8	0.3	0.7	0.9	0.2	0.4	0.9	0.9	0.5	0.3	0.5	0.4	0.9
0.3	1.0	0.4	0.2	0.4	1.0	0.3	0.6	1.0	0.4	0.7	0.9	0.4	0.8	0.6
0.2	0.9	0.3	0.1	0.3	0.9	0.2	0.5	0.9	0.3	0.6	0.8	0.3	0.7	0.5
0.8	0.3	0.4	0.3	0.7	0.4	0.5	0.7	0.2	0.2	0.5	0.7	0.8	0.4	0.2
0.3	1.0	0.4	0.2	0.4	1.0	0.3	0.6	1.0	0.4	0.7	0.9	0.4	0.8	0.6
0.2	0.9	0.3	0.1	0.3	0.9	0.2	0.5	0.9	0.3	0.6	0.8	0.3	0.7	0.5
0.3	0.6	0.8	0.3	0.7	0.9	0.2	0.4	0.9	0.9	0.5	0.3	0.5	0.4	0.9
0.2	0.5	0.7	0.2	0.6	0.8	0.1	0.3	0.8	0.8	0.4	0.2	0.4	0.3	0.8
0.9	0.4	0.5	0.4	0.8	0.5	0.6	0.8	0.3	0.3	0.6	0.8	0.9	0.5	0.3

Fig 9 The Antecedents of the rules

D. Finding first level fuzzy associations

For finding first level fuzzy associations between the consequents and antecedents, we took the maximum value from the fuzzy sets in a record for each attribute and this fuzzy set which has maximum membership for an attribute is mapped with that attribute (for example, in the first record, if we consider the field *Banking inclusion* [27] the associated fuzzy sets are low, medium and high with values 0.2,0.5 and 0.8 respectively. So we mapped the derived filed Banking inclusion with fuzzy value *high* for that record). This is done for all the records and attributes. If we take the first five record of the derived database we get the following fuzzy mappings (Table 6) .

From this table now we can derive all the first level fuzzy association rules between the consequents and antecedents. In this step one consequent is mapped with one antecedent based on this fuzzy membership value. For example, if we take the first two consequent attributes from the first record, we can derive the first level rules as represented in figures 8 and 9. Similarly the rules can be generated for all the consequents from all the records. The fuzzy support of these rules are found using expression (5) and only the rules with the given support count of 0.20 are retained. In the second stage, the rules with two antecedents are found by using the selected rules from the previous stage. Again the only such rules which satisfy the support count are retained. The rules in this stage will have the following structure

Table 6. First level Association Rules

No	First level rules	Support
1	Edupar=high=> Banking inclusion=high	0.7*0.8=0.56
2	EduChildren=high =>Banking inclusion=high	0.8*0.8=0.64
3	M_income=high =>Banking inclusion=high	0.8*0.8=0.64
4	Dist_tbank=near =>Banking inclusion=high	0.8*0.8 =0.64
5	Modern banking=high =>Banking inclusion=high	0.8*0.8=0.64
6	Edupar=high => Loanaccessibility =medium	0.7*0.9=0.63
7	EduChildren=high =>Loanaccessibility =medium	0.8*0.9=0.72
8	M_income=high =>Loanaccessibility =medium	0.8*0.9=0.72
9	Dist_tbank=near =>Loanaccessibility =medium	0.8*0.9=0.72
10	Modern banking=high) => Loanaccessibility =medium	0.8*0.9=0.72

In the next step , the rules with three and more antecedents those satisfy the support count will be identified in each step until there is no further rule in the next step. Only those rules which satisfy the support and confidence count (A fuzzy confidence of 0.4 is used) are selected as the final rules and these rules can be considered as the classification rules which will provide valuable information to eliminate financial exclusion and to strengthen financial inclusion. The final set of rules are given in table 7.

VII. CONCLUSION

In this work we started with the database collected for the statistical analysis about financial inclusion situation in Idukki district of India. The findings from the analysis gave a

© 2018, IJCSE All Rights Reserved

clear picture about the statistics of the present situation in this area in the field of financial inclusion and exclusion. Finally we are applying the fuzzy data mining techniques discussed in this paper to the financial inclusion dataset. The concepts presented in the field of clustering and association rule mining are applied here to bring out the hidden patters related with financial exclusion and inclusion. The identified patterns are presented in the form of fuzzy association rules and these rules are quite natural to understand. On closely analyzing the rules generated and the visual illustrations of the patterns, we can easily reach at the conclusion that Education, the introduction to recent Information Technology and modern banking systems and Penetration of the banks in remote areas are the key factors affection financial inclusion. By concentrating more in these areas the situation of financial inclusion can be improved.

- (Edupar=high, EduChildren=high)=> (Banking inclusion=high)

Consequents		Anticedents				
Name	Fuzzy Value	Edupar	Edu Children	M_income	Dist. tobank	Modern
Banking inclusion	high	High	High	-	-	high
Banking inclusion	high	Med	High	-	Near	-
Banking inclusion	med	Low	Med	-	-	-
Banking inclusion	low	-	Low	Low	Far	low
Banking inclusion	high	-	Med	-	Near	High
Loan accessibility	high	High	High	-	-	-
Loan accessibility	med	Med	High	-	Far	
Loan accessibility	med	Med	-	High	-	-
Financial Inclusion	high	-	High	-	Near	high
Financial Inclusion	low	Low	Low	Low	-	-
Financial Inclusion	low	Low	Low	-	Far	-
Financial Inclusion	low	-	Low	Low	Far	
Financial Inclusion	med	Low	High	High	Near	-
Amount borrowed	low	Low	Low	Low		
Amount borrowed	high	Med	High	-	-	high
Amount borrowed	med	Low	Med	-	Mid	-
Amount borrowed	low	Low	Med	Low	Far	-

Fig.10 A portion of the fuzzy rules generated

A. Illustrations

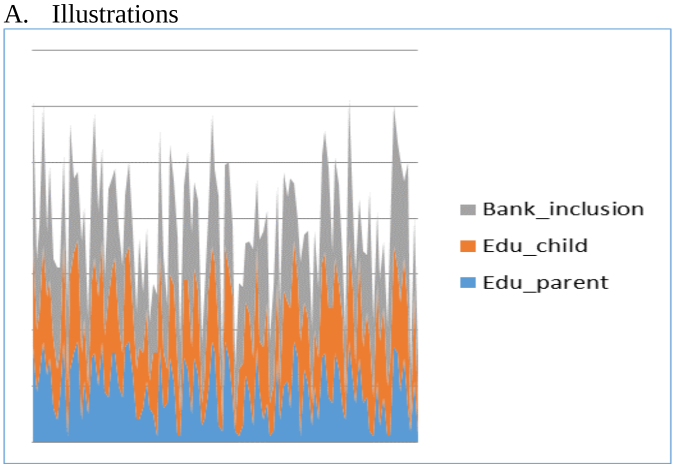

Fig.11 Relationship between education of the parent and child with Banking Inclusion. Always there is a positive correlation between the education and Banking inclusion.

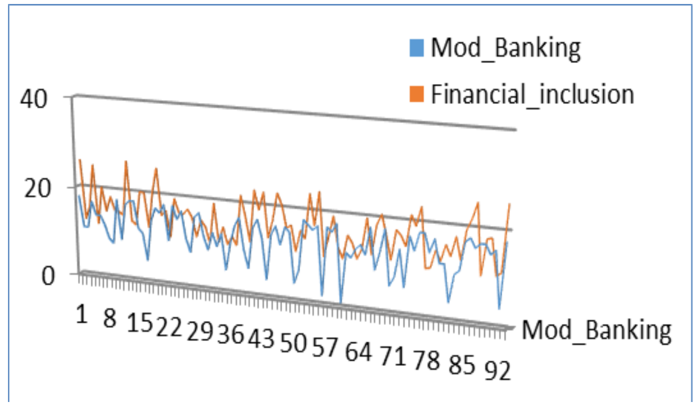

Fig. 12 Relationship between the derived field Modern Banking and financial inclusion. A positive correlation is evident from the graph.

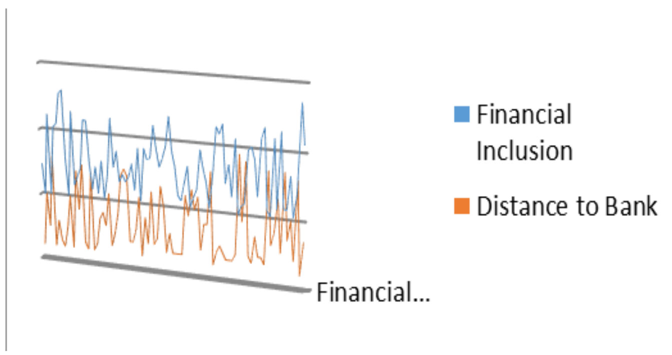

Fig13.Relation between Distance to the Bank and Financial Inclusion. In some cases we can see a positive correlation but it is not very evident

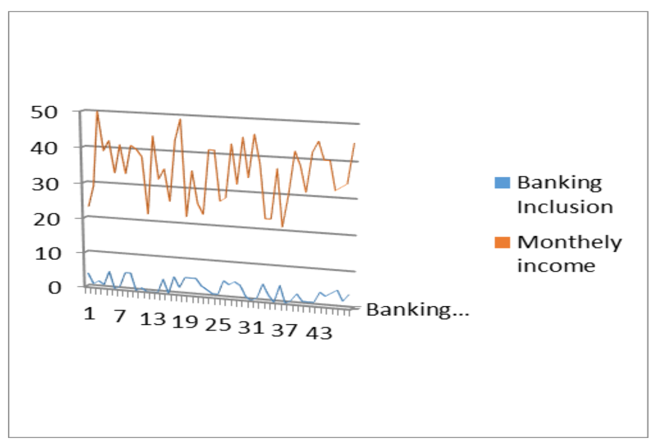

Fig.14 Relation between Distance to the Bank and Banking Inclusion. As it is evident from the illustration there is no significant relation between Banking inclusion and Monthly income

© 2018, IJCSE All Rights Reserved

Table 8. The fields of the Database

No	Fields	Remark
1	**Sl_no**	Serial number
2	Gender	Gender of Respodant (String)
3	F_Head	Head of the family (Boolean)
4	Edupar	Parent's education (string)
5	EduChildren	Education of the Children (string)
6	Occupation	Parent' ocupation (string)
7	Having_Bank_ Acc	Head having bank Acc. (Boolean)
8	HealthProfile	Any health problems (string)
9	Monthlyincom e	Monthly income (string)
10	Nameofbank	Name of of bank (string)
11	Distancetoban k	Distance to bank (string)
12	Practicaldificul ties	Difficulties in banking (string)
13	Purposebank	Purpose of banking (string)
14	Creditloan	Availing credit loan **(string)**
15	Mtranfer	Availing Money transfer (string)
16	Use_ATM	Using ATM facilities (string)
17	Overdraft	Using Overdraft facilities (string)

No	Fields	Remark
18	Corebaking	Using Core banking (string)
19	Borrowing	Borrowed money from bank (string)
20	Purpose	Purpose of borrowing (String)
21	Source	Source of borrowing(string)
22	Totalamt	Amount borrowed (string)
23	Guarantee	Guarantee for borrowing
24	Nothavingbankaccoun t	Reason for not having acc.(string)
25	Usefulness	Usefulness of banking (string)
26	Savingstatus	Status of saving (string)
27	Savingfreq	Frequency of saving (string)
28	Percentageofsaving	Percentage of saving (string)
29	P_value	Value of the property (string)
30	AwarenesInsurence	Awareness of insurance (string)
31	Financialadvice	Seeking financial advice (boolean)
32	Awpostoffice	Post office savings (boolean)
33	Mobilephone	Having mobile phone (boolean)
34	Cumputer	Knows computer (string)

VIII. REFERENCES

[1]. ADB, (2007): "Low-Income Households' Access to Financial Services", InternationalExperience, Measures for Improvement and the Future; Asian Development Bank

[2]. Kempson, E. (2006): "Policy Level Response to Financial Exclusion in DevelopedEconomies: Lessons for Developing Countries", Paper for Access to Finance: Building Inclusive Financial Systems, World Bank, Washington, May

[3]. Sen, Amartya, (2000): '*Development as Freedom*', Anchor Books, New York, 2000.

[4]. Peachy, S. and A. Roe, (2004): "Access to Finance - What Does it Mean and How" Applications in the Social Sciences, SAGE Publications, London.

[5]. Kempson, E. (2006): "Policy Level Response to Financial Exclusion in DevelopedEconomies: Lessons for Developing Countries", Paper for Access to Finance: Building Inclusive Financial Systems, World Bank, Washington, May

[6]. H.M. Treasury, (2007): "Financial Inclusion: the Way Forward", HM Treasury, UK, March

[7]. Mohan, R. (2006): 'Agricultural Credit in India: Status, Issues and Future Agenda', *Economic and Political Weekly* (March), pp.1013-23

[8]. Towards Faster and More Inclusive Growth The approach paper to the Eleventh Plan 2006, retrieved from, http://mhrd.gov.in/sites/upload_files/mhrd/files/apppap_11_1.pdf

[9]. Report of the Committee on Financial Inclusion in India (Chairman: C. Rangarajan)(2008), Government of India

[10]. Reasons for financial exclusion - *Reserve Bank of* India*(2007)www.rbi.org.in › Speeches*

[11]. . W. H. Inmon.: The data warehouse and data mining. Commun, ACM, vol. 39, pp. 49–50 (1996)

[12]. Pavel Berkhin, "Survey of Clustering Data Mining Techniques", *http://citeseer.ist.psu.edu/berkhin02survey.html*

[13]. U. Fayyad and R. Uthurusamy, "Data mining and knowledge discovery in databases", *Commn. ACM, vol. 39,* pp. 24–27, 1996.

[14]. Pujari, A. K., 2001, *Data Mining Techniques*, University Press, Hyderabad, India

[15]. K. Pal and P. Mitra(January 2002).: Data Mining in Soft Computing Framework: A Survey. IEEE transactions on neural networks, vol. 13, no. 1 (January 2002)

[16]. G. J Klir, T A. Folger, *Fuzzy Sets, Uncertainty and Information,* Prentice Hall,1988

[17]. Wai-Ho Au, Keith C.C. Chan, "Classification with Degree of Membership: A Fuzzy Approach", *Proceedings IEEE International Conference on Data Mining (ICDM),* 2001

[18]. U.M. Fayyad; G. Piatetsky-Shapiro; P. Smyth, 1996, Advances in knowledge discovery and data mining",MIT Press

[19]. M.S. Aldenderfer; R.K. Blash_eld, 1984, \Cluster analysis", volume 44 of Quantitative Applications in the Social Sciences, SAGE Publications, London.

[20]. Hegland, M. 2003. *Data Mining – Challenges, Models, Methods and Algorithms.* May 8.Scientific Literature Digital Library, citeseer.ist.psu.edu/context/68861/0.

[21]. G Raju, B Thomas , Fuzzy clustering methods in data mining: A comparative case analysis Proceeedings of International Conference on Advanced Computer Theory and Engineering, 2008. ICACTE '08. ,2008 pp 489 – 493

[22]. E. Cox.: Fuzzy Modeling and Genetic Algorithms for Data Mining and Exploration. Elsevier, (2005)

[23]. Binu Thomas, G. Raju, and W. Sonam. "A modified fuzzy c-means algorithm for natural data exploration." *World Academy of Science, Engineering and Technology,* Volume 49, 2009 pp 478-481

[24]. M Muyeba, MS Khan, F Coenen,(2010) citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.157.9281&rep=rep1&type=pdf

[25]. Shu Yue, J , Tsang, E. ; Yeung, D. Mining fuzzy association rules with weighted items Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics ,2000 , pp 1906 - 1911 vol.3

[26]. Patel, Mohnish, Aasif Hasan, and Sushil Kumar. "A survey: Preventing discovering association rules for large data base." *International Journal of Scientific Research in Computer Science and Engineering* 1.2 (2013): 30-32.

[27]. Ghuse, Namrata, Pranali Pawar, and Amol Potgantwar. "An Improved Approch For Fraud Detection In Health Insurance Using Data Mining Techniques." *health* 11.12 (2017): 13.

Authors Profile

Dr. Binu Thomas obtained PhD from Mahatma Gandhi University in 2014. He is currently working at Marian College Kuttikkanam as Associate Professor in Department of Computational Sciences, Department of Computer Applications. He has published more than 30 research papers in reputed international journals including Thomson Reuters (SCI & Web of Science) and conferences including IEEE and it's also available online. He has 116 Google Scholar Citations in his credit. His main research work focuses on Educational technologies, Big Data Analytics, Data Mining, IoT and Educational analytics. He has 20 years of international teaching experience

© 2018, IJCSE All Rights Reserved