ISSN 2319-2720

Volume 2, No.2, April - June 2013
International Journal of Computing, Communications and Networking
Available Online at http://warse.org/pdfs/2013/ijccn02222013.pdf

Cluster Identification for Micro & Small Enterprises in North East India: A case study on KVIC sponsored Village Industries in Assam

Nandini Sarma*
Assistant Director of Training,
Assam Administrative Staff College, Guwahati-22.
* e-mail: sarmanandini@yahoo.com

Dr. P.H. Talukdar
Professor, Deptt. of Instrumentation & USIC,
Gauhati University, Guwahati- 14
e-mail: phtalukdar@gmail.com

ABSTRACT

Although the primary occupation in the villages of India (Assam) is agriculture, its share is very low and there is high degree of under employment in Agriculture. So the village economy is leaning towards small industries, having backward linkages with urban areas. There is a growing trend of migration of people from the rural areas to the urban areas. Therefore, people with enterprising mindset need to come forward and undertake initiatives for establishing viable economic enterprises in the rural areas, which would create a number of direct/indirect jobs, apart from opportunities for self employment for the rural entrepreneur themselves.

KVIC (Khadi and Village Industry Commission) India, under Ministry of Micro, Small and Medium Enterprise, is an organization engaged in the task of promoting and developing environment friendly khadi and village industries, with the mission for creating employment opportunities in the rural areas, thereby strengthening the rural economy.

This paper deals with knowledge discovery in databases, for KVIC scheme, namely- REGP of Barpeta, Jorhat, Kamrup and Nagaon districts of Assam (India), for the year 2002-2008. This is to identify and quantify distinct parameters, such as 'pin code', 'industry/product choice' etc., impacting the units setup under KVIC patronage. It is expected that the aforesaid parameters shall unfold vital collective and spatial attributes that not only define the rural industrial setting, but also help in identification of Cluster through data mining, for the purpose of establishing the common facility unit for village industry activity, in the state of Assam.

Key words: Cluster, Data mining, Entrepreneurship, Enterprise, Economy, KVIC, PIN code, RISC, Village industry.

1 INTRODUCTION

India, after her independence, embarked on the path of industrialisation to achieve the much-needed economic growth for development. The government viewed industrial growth as a mode of achieving economic growth in the country. Like many of the developing economies in the post-war period, India opted for rapid industrialisation [11]. Targets were fixed for Industrial investments in the five-year Plans that were the backbone of India's planned development. Apart from medium to big industries, another important segment was encouraged by the planners was the small-scale industry sector. These small-scale and micro enterprises besides being instrumental to the economic development of the country also generated vast employment opportunities.

Economic history of developed countries bears the evidence that industrial development proceeds economic development and micro & small enterprise development proceeds industrial development. Micro and small enterprises play an important role in the economics of both advanced and developing countries. In developing economies these enterprises are very important in the context of employment opportunities, equitable distribution of wealth, balanced regional growth and development of rural and semi urban areas. This sector is considered to be an engine of growth, particularly in developing country like India because of their contribution to employment, national income, enterprise

development and export earnings. As per India Development Report 2008, there are 302 million poor people in India, 27% of the total population. Thus in India micro and small enterprises allow the rural poor including some of the most marginalized and vulnerable stratum such as rural woman, youth and landless to diversify their incomes, create new sources of economic growth and generate additional employment (including self-employment) in rural areas. There is a difference between enterprise development and entrepreneurship development [12]. The most important aspect of development lies in economically empowering the individuals by enhancing their entrepreneurial skills for managing sustainable enterprises [10].

1.1 Definition of Micro, Small and Medium Enterprises

Under Micro, Small and Medium Enterprise Development (MSMED) Act, 2006, concept of **'Industry'** has been widened to that of '**Enterprise**' including both the manufacturing and service sector besides defining the medium enterprises. Under this act the Enterprises have been defined in terms of investment as given in the **Table 1** [22]**:**

Table 1 : Definition of Micro, Small and Medium Enterprises

	Investment in plant and machinery/equipment (excluding land and building)	
	Manufacturing Enterprises	Service Enterprises
Micro	Up to Rs. 25 lakh	Up to Rs. 10 lakh
Small	More than Rs. 25 lakh and up to Rs. 5 Crore	More than Rs. 10 lakh and up to Rs. 5 Crore
Medium	More than Rs. 5 Crore and up to Rs. 10 Crore	More than Rs. 2 Crore and up to Rs. 5 Crore

Source: MSMED Act, 2006

The micro and small enterprises in India comprises of modern and traditional enterprises. A modern segment includes Micro and Small Enterprise (MSE)s under Small Industries Development Organisation (SIDO) and power looms. MSEs under SIDO comprise the residual enterprises that are not covered by any assistance programme of the sector specific statutory bodies viz., the Handicraft, Handloom, Sericulture and Coir. The other segments of traditional enterprises are looked after by Khadi & Village Industry Commission (KVIC) [16]. Thus rural industrialization has been given prominence in the current decade and enterprises under SIDO, KVIC etc. have received more attention [23].

1.2 Identification of Village Industry Clusters

In the **Tenth Five Year Plan (2002-2007),** at all India level, it was proposed to set up rural Industrial estates and artisan clusters to provide necessary infrastructure and support services to the village Industries. These common Facility Centers of the Industrial estates would be provided financial assistance by **Khadi and Village Industry Commission (KVIC), National Bank of Agriculture and Rural Development (NABARD**), **Small Industries Development Bank of India** (SIDBI) and **Council for Advancement of People's Action and Rural Technology (CAPART).** All these initiatives are being taken up under **National Programme for Rural Industrialization (NPRI)** through creating rural clusters for Industrialization.

KVIC has been creating employment opportunities for spinners, weavers and other allied artisans [3]. Ghandhiji set up the All India Khadi Board with branches in all provinces in December 1923, for coordinated development of Khadi [13]. Under KVIC there is a Programme for promotion of Village Industry Cluster – **Rural Industry Service Centre (RISC)** scheme, from 2004-05 onwards, for Khadi & Village Industry (KVI) activity with objective to provide Infrastructural support and necessary services to the selected local units with a view to upgrade their production capacity, skill up gradation and market promotion [4].

RISC, provides testing facilities by establishing laboratories for ensuring quality of KVI products; improved machinery/equipment to be utilized as common facilities by nearby units/artisans for enhancing production capacity or value addition of the product; attractive and appropriate packaging facilities and machinery to the local units/artisans for better marketing of their products; training facilities for upgrading artisans' skills in order to increase their earnings and new designs and diversified products in consultation with experts/agencies for value addition of rural manufacturing units [21]. Thus, RISC is the common facility unit which aims to strengthen the rural KVI clusters. All types of KVIs to be covered under RISC except those are in the negative list, of KVIC for all village industries.

Under RISC scheme, financial assistance for establishing smaller projects costing up to Rs.5 lakh each is provided KVI units, while in bigger projects such assistance

is up to Rs. 25 lakh. In case of North Eastern States 90% of project will be provided by KVIC up to a project of Rs. 5 Lakhs, (other areas- 75%). Though Assam is considered as industrially developed compared to other states of North East; in the Indian scenario, Assam shows low industrial development along with micro and small scale enterprises. Assam still remains predominantly an agrarian economy where 53% of her total population depend on agriculture and allied activities [9]. Assam is still one of the industrially backward states [1].

RISC programme implementation (all India) [21]:

- For the purpose of establishing RISC, it may be ensured that the number of village industries/artisans units shall not be less than 25 individual or 5 REGP units.
- Identification of the cluster is the first stage of RISC programme implementation.
- Under the RISC programme, 41 projects were sanctioned in 2004-05 and 62 projects were sanctioned 2005-06, respectively. In 2006-07, 55 projects have been sanctioned upto 31 December 2006. Production increased by 3 to 10 per cent in those institutions which availed the facility for projects.

The number of projects assisted under this programme from 2007-08 to 2010-11 (December 2010) is given in the **Table 2** [22] below:

Table 2 : Number of Projects assisted –RISC

Year	Total No. of KVI Projects Assisted
2007-08	76
2008-09	54
2009-10	18
2010-11*	12

Source: MSME annual report 2010-11 (* up to December 2010)

2 METHODOLOGY

Based on the above background, the present study adopted a methodology wherein both primary and secondary data are analysed. Primary data on KVIC and its rural employment generation schemes- REGP, PMEGP etc. are collected through personal interview at Assam, Guwahati Khadi and Village Industry office/ exhibitions conducted by KVIC Guwahati -as first-hand information. Secondary data are collected from KVIC and KVIB reports/publications, other are from publications/books/websites/ data base etc. **The Secondary data formed the basis of analysis, for the present study.**

The Knowledge Discovery in Databases (KDD) is the process of automatic discovery of previously unknown patterns, rules and other regular contents implicitly present in large volume of data [5]. Data stored in databases cannot be used directly- one has to use Report, Query etc. This can be used, if the analysed data is relatively small. For large volume of data one has to go for automation of data analysis tasks and Data Mining (DM) comes into picture [17]. The KDD process consists of the following steps:

2.1 Selection

The data used in the present study is based on secondary data, collected from IIE Guwahati Training Database and the database used by the Indian Institute of Entrepreneurship, Guwahati while conducting EDP trainings for KVIC and many other sector/sponsors (outside Guwahati also). The institute keeps the participants data for all these EDP trainings and has a huge amount of EDP training participant's computerized data. Data mining applied to this huge amount of data, to extract EDP training participants' data for KVIC REGP scheme, for State of Assam. These data are analysed after filtering and normalizing for redundancy and delicacy. EDP Training Participants' data /profiles of First Generation entrepreneurs under REGP schemes of KVICs in Assam were selected for the year 2002-2008.

The study area covered four districts of the State of Assam. **The districts under study comprise Barpeta, Jorhat, Kamrup and Nagaon districts of Assam.** Undivided Kamrup and Nagaon districts are purposively selected considering high number of KVIC REGP EDP training participants- 1637 and 1370 respectively. One lower Assam district, Barpeta and one upper Assam district Jorhat selected with 755 and 669 number of EDP training participants respectively. The four districts together have 4431number of KVIC REGP EDP training participants' data.

2.2 Preprocessing

The data to be used by the process may have incorrect or missing data, which requires preprocessing. Our data are scrutinized for these and erroneous/ missing data are corrected/added/removed, such as -erroneous /missing Industry type etc. Number of attributes was reduced to make it more effective.

2.3 Transformation

Data from different sources are converted into a common format for processing which is called **transformation.** Here some data are transformed into more usable formats. Data reduction and district wise processing is done – for Barpeta, Jorhat, Kamrup and Nagaon districts of Assam.

2.4 Clustering in terms of data mining

Many people use the terms clustering and classification synonymously. Some view clustering as a special type of classification. However, though clustering is similar to classification in grouping of data; unlike classification, the groups are not predefined. Thus cluster and classification are different and we can define clusters as set of like elements, where groupings are not predefined. Clustering divides data into groups called clusters that are meaningful and useful. Elements from different clusters are not alike.

In the present data mining task, clustering is done for finding the possibility of identification of Cluster for the purpose of establishing RISC for KVI activity. Here, out of the EDP training trainee data, we proposed common PIN code number (for location), with similar village industry/product type for Cluster formation. PIN code can be location identification code as:

2.4.1 PIN Code delineation

India has the largest Postal Network in the world with over 1,55,015 Post Offices (as on 31.03.2009), of which 1,39,144 (89.76%) are in the rural areas. At the time of independence, there were 23,344 Post offices, which were primarily in urban areas. Thus, the Post Office network has registered a seven fold growth since independence, with the focus of this expansion primarily in rural areas; and the introduction of PIN code number, acts a unique identification number. On an average **a Post Office serves an area of 21.21 Sq. Km and population of 7175 people** [20].

We want to use PIN code number for identification of locations with similar village industry/product type, for formation of Clusters (from REGP EDP training trainee data). KVIC, Assam can verify them physically, for RISC potential.

Postal Index Number (PIN) or PIN code is a 6 digit code of post Office numbering used by India Post. The PIN was introduced on August 15, 1972. There are 9 PIN regions in the country. The first 8 are geographical regions and the 9th the digit 9 is reserved for the Army Postal Service. The first digit (1 to 8) indicates one of the regions. The first digit of PIN indicates as below:

Table 3: Example of first digit of PIN code

First Digit	Region	States Covered
7	Eastern	West Bengal, Orissa & North Eastern States

The first 2 (out of 6) digits together indicate the sub region or one of the postal circles. The first 2 digits of PIN indicate as below:

Table 4 : Example of first 2 digits of PIN code

First 2 digits of PIN	Circle
78	Assam
79	Other North Eastern States

The first 3 digits together indicate a sorting/revenue district. For our four districts:

Table 5 : Example of first 3 digits of PIN code

First 3 digits of PIN	Revenue District
785	Jorhat, Assam
782	Nagaon, Assam
781	Barpeta & Kamrup, (Earlier undivided Kamrup District)

The last 3 digits refer to the delivery Post Office. So delivery Post office wise, PIN code numbers become different for Kamrup and Barpeta districts.

2.4.2 Density-Based Spatial Clustering of Applications with Noise (DBSCAN)

The entire collection of clusters is commonly referred to as a Clustering. We try to see the suitability of clustering technique: density based (DBSCAN) algorithms.

The approach used by DBSCAN is to create clusters with a minimum size and density. This density-based clustering locates regions of high density that are separated from one another by regions of low density. **Density** is defined as minimum number of points within a certain distance of each other. One input parameter, **MinPts,** indicates the minimum number of points in any cluster.

In addition for each point in a cluster there must be another point in the cluster whose distance from it is less than a threshold input value, **Tiv**. The Tiv-neighborhood or neighborhood of appoint is the set of points within a distance of Tiv. This is a density based clustering algorithm that produces a partitional clustering, in which the desired number of clusters, k, is not input but rather is automatically determined by the algorithm itself. Points of low density regions are classified as noise and are omitted; thus DBSCAN does not produce a complete clustering. A density based definition of a cluster is often employed when the clusters are irregular/ unequal and when noise/ outliers are present.

Core point (object):- Point in the interior of density based cluster. A core point (object) is a point which has a neighborhood of user specified minimum density.

Border point (object):- Not a core point (object) but falls within the neighbourhood of a core point.

Noise point (object):- Any point that is neither a core point, nor a border point. It is the set of points (objects) which do not belong to any cluster.

The above points can also be stated as – core points and non-core points only. Non-core points are either border points or noise points. Two border points may not be density reachable from each other. However, a border point is always density reachable from a core point. Non-core noise points are not density reachable from core points. DBSCAN's procedure for finding a cluster – that a cluster is uniquely determined by any of its core points.

In DBSCAN algorithm we find the set of points (objects) in three different categories, such as: classified, unclassified and noise [15]. A cluster-id is associated with each classified objects, indicating the cluster where it is included. A dummy cluster-id may be associated with a noise point (object). The Tiv, neighbourhood for both classified and noise points are planned. The unclassified categories of points do not have any cluster-id and their neighbourhoods are not considered. DBSCAN algorithm converts an unclassified point (object), into a classified or a noise point gradually. A Detail Algorithm [15]:

Algorithm DBSCAN (D, Tiv, MinPts)

Input: Database of objects D
do for all O € D
if O is unclassified
call function **expand_cluster** (O, D, Tiv, MinPts)
end do
Function expand_cluster(O, D, Tiv, MinPts);
Get the Tiv- neighbourhood of O as $N_{Tiv}(O)$
if | $N_{Tiv}(O)$|< MinPts,
mark O as noise
return
else
select a new cluster_id and mark all objects of $N_{Tiv}(O)$
with its cluster_id and put them into candidate-objects.
do while candidate-objects is not empty
select an object from candidate-objects as current-object
delete current-object from candidate –objects
retrieve N_{Tiv}(current-object)
if | N_{Tiv}(current-object)| >=MinPts
select all objects in N_{Tiv}(current-object)
not yet classified or marked as noise,
Mark all of the objects with cluster_id,
Include the unclassified objects into candidate-objects
end do
return.

At every step, the algorithm begins with an unclassified point (object) and a new cluster-id assigned with it. Then the neighbourhood is checked for adequate density. If the density is less than the minimum threshold MinPts, then it is marked as noise point. Else, all the points (objects) that are within the planned neighbourhood are retrieved and put into a list of candidate points (objects). These candidate points may be either unclassified or noise. A neighbourhood cannot contain a classified point (object). If the point is a noise point, the current cluster-id is assigned to it and it is included in the list of candidate points (objects) for which the neighbourhoods are to be planned. The algorithm continues till the list of candidate points (objects) is empty. This way one cluster with own cluster-id is determined. The algorithm repeats this process for other unclassified points (objects) and terminates when all the points (objects) are marked as either classified or noise.

The retrieval of the density-reachable neighbourhood is the important and critical component of the DBSCAN algorithm. This can be accomplished by an R-tree structure, which can be relied for speed and scalability, in its nearest-neighbour search queries. Thus for finding Tiv neighbourhood, search queries are added with the DBSCAN algorithm.

2.4.3 R-tree

R-tree helps in focusing only on relevant parts of the database. Its focusing techniques improve DBSCAN's ability to deal with disk-resident data sets.

R-tree is helpful for spatial data indexing, special query processing etc. also. Spatial data are data that have a spatial or location component. Spatial data can be viewed as data about objects that themselves are located in a physical space. A spatial object is usually described with both special and non-spatial attributes. Some sort of location type attribute is included. The location attribute could identify a precise point, such as location address, PIN code etc. For these types of data, the non-spatial attributes may be stored in a relational database, while spatial attribute may be stored in some spatial data structure. Each tuple in the relationship represents the spatial object and a link to the spatial data structure is stored in the corresponding position in the non-spatial tuple. Many basic spatial queries can assist in data mining activities [6]. For example:

- A region (pin code etc.) query or range query is a query asking for objects that intersect/match a given region specified in the query.
- A nearest neighbour query asks to find objects, close to an identified object.

These queries can be used to assist in clustering or classification.

In an R-tree cells may actually overlap. Each successive layer in the tree identifies smaller area, location etc. The size of the tree is related to the number of objects. The entire geographic space of Assam is represented by first 2 digits of Pin code '78', which is labeled A, and is shown as the root of the tree in **Figure 1** below.

Algorithm to perform spatial operations using an R-tree, are relatively straight forward and similar to B-trees or binary search trees. To find all the objects that matched/ intersected with a given object, we can search the upper levels of the R-tree to find only those cells that match/ intersect the query; those sub-trees that do not match/ intersect the query can be discarded.

In **Figure 1** R-tree →

As example we may take first

2 digits of pin code as 'A': 78

3 digits as 'B': 782(Nagaon)

Or 'C': 785 (Jorhat)

D,E,F etc.: All 6 digits including last 3 digits of pin code, the delivery post offices

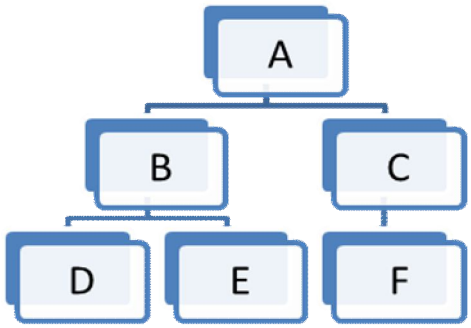

Figure 1

2.4.4 Categorical Clustering CACTUS

Another option of cluster formation can be borrowed from categorical clustering algorithm called CACTUS- Clustering Categorical Data Using Summaries. For this, a database D can be assumed with a set of tuples each having k fields or attributes. Most clustering techniques attempt to cluster the tuples, considering the tuples as primitive object. However, CACTUS attempts to split the

database vertically and then tries to cluster the set of projections of these tuples to only a pair of attributes. The principle is that, a cluster on a set of attributes induces a cluster on a subset of attributes. CACTUS extends this idea to categorical attributes [15]. It tries to find clusters for the subset of attributes, and then clusters for the whole attribute set can be found by intersecting these clusters.

Two important characteristics of this CACTUS algorithm are: First, the algorithm assumes that, the categorical domains have a small number of attribute values, i.e. the domain sizes are small. Second, which is the central idea of CACTUS, the summary information constructed from the data set is sufficient for discovering well-defined clusters. The properties that the summary information can be fitted into main memory and that it can be efficiently constructed in a single scan of the data set, are the key ideas of the algorithm.

In our database attribute Age is numeric attribute; however Pin code can be taken as categorical one. Though the domain of Pin code is numeric, one cannot derive any useful information if one Pin code is greater than another Pin code. In case of Product code, as per KVIC [4], we see that

Product code: 1 to 15 falls under Industry group type:- I;
Product code: 16 to 27 falls under Industry group type:- II;
Product code: 28 to 48 falls under Industry group type:-III;
Product code: 49 to 61 falls under Industry group type:- IV;
Product code: 62to 86 falls under Industry group type:- V;
Product code: 87 to 95 falls under Industry group type:- VI;
Product code: 96 to 115 falls under Industry group type:-VII.

Similar to CACTUS algorithm, one can attempt to split the EDP trainee database vertically and then try to cluster the set of projections of these tuples to only a pair of attributes- Pin code (Pin) and Product code (P code), for generation of clusters for tentative KVI RISC activity.

Going back to CACTUS algorithm – one can consider two attribute values of two different attributes in the earlier mentioned database D. Value a_i of attribute type A and value a_j of attribute type B. There may be tuples where a_i and a_j co-occur. The support of these two values in the database is the proportions of tuples in which they appear together [15]. If this support exceeds a pre-specified value, then these two attribute values a_i and a_j are called strongly connected. D_A denotes domain of the attribute A; s denotes the support value.

Similarly, in the EDP trainee database, one can consider one value of attribute Pin code (Pin) and another value of attribute Product code (P code). In this data base there are tuples where these two values co-occur; if they co-occur 5 or more times we denote them as cluster; if there is no co-occurrence then at that run time there is no cluster-cluster may happen after more tuples are added in the database. In this manner another set of the attribute values may be checked for co-occurrence and so on and clustering is done for the whole database.

3 DISTRICT WISE CLUSTERS FOR TENTATIVE VILLAGE INDUSTRY CLUSTER

As described previously, in this data mining task, clustering is done for finding the possibility of identification of Cluster for the purpose of establishing RISC for KVI activity. The proposed cluster would increase the number of extraction from the present level so that that the fruits of development can be achieved [7]. Out of the EDP training trainee data, we proposed common PIN code number for identification of locations, with similar Product/Industry code (type) for Cluster formation. We had separated the data into four selected districts, for district wise identification of cluster. We kept MinPts as 5, so that co-occurrence of same Pin code with same Product code, 5 or more times lead to formation of a cluster.

Identified tentative clusters (district wise) are presented in tabular form. It is to be mentioned here that industry clustering is encouraged in the manufacturing sector. About 70% of total entrepreneurial units set up belong to this sector and about 30% units are in service sector [2].

3.1 Tabular Representation of District wise Cluster (Barpeta)

Table 6: Identified clusters for the district of Barpeta

			BARPETA			
Group No->	2	3	5	6	7	7
Sl No->	21	46	84	87	106	107
Name->	**Cane/Bamboo**	**Leaf plate**	**Fabrication**	**Garments**	**Cycle repair**	**R.C.C**
Pin code	**Colours denote Clusters >= 5 No.**					
781301	-	-	5	5	8	7
781302	5	-	-	-	-	-
781308	-	-	-	5	-	-
781309	9	-	-	-	-	5
781315	-	-	-	6	-	-
781317	-	-	-	-	-	5
781318	-	6	-	-	-	6
781352	7	-	-	-	-	8

From the above **Table 6,** for Barpeta district, we can see that **2nd group** industry product 21-**Cane/Bamboo work**; **3rd group industry** product 46-**Leaf cup making**; **5th group industry product** 84-**Fabrication/Iron grill making**; **6th group industry product** 87-**Tailoring & Readymade Garments**; **7th group industry product** 106,107 has the potential for KVI cluster formation in the respective Pin code as given in the table. As Cane/Bamboo work and Tailoring & Readymade garments are showing more potential, i.e. 3 clusters each and KVIC has already established Cane/Bamboo work related cluster in Barpeta district; Tailoring & Readymade Garments can be considered next KVIC cluster activity. Since service sector (7th group) cluster plan is not there right now, we are not recommending it, though it is showing 5 clusters in 7th group industry 107-Masonary work.

3.2 Tabular Representation of District wise Cluster (Jorhat)

Table 7: Identified Clusters for Jorhat

		JORHAT			
Group No->	3	6	7	7	7
Sl No->	28	87	103	106	113
Name->	**Cereals/spices**	**Garments**	**Hiring gadgets**	**Cycle repair**	**Dhabas**
Pin code	**Colours denote Clusters >= 5 No**				-
785001	9	7	5	-	5
785006	-	-	-	5	-
785104	-	5	-	-	-
785105	-	12	-	-	-
785630	-	8	-	-	-

From the above **Table 7,** for Jorhat district, we can see that **3rd group industry product** 28- Processing, packing and marketing of cereals, pulses, spices, condiments, masala etc. at pin code 785001; **6th group industry product** 87-Tailoring & Preparation of Readymade Garments at pin code 785001, 785104, 785105 and 785630; 7th group industry product 103 at pin code 785001, 106 at pin code 785006, 113 at 785001 has the potential for KVI cluster formation as given in the table. As Tailoring & Preparation of Readymade Garments at pin code 785001, 785104, 785105 and 785630 are showing more potential, i.e. 4 clusters; Tailoring & Readymade Garments can be considered KVI RISC cluster activity. Since service sector (7th group) cluster plan is not there right now we are not recommending it.

3.3 Tabular Representation of District wise Cluster (Kamrup)

Table 8 : Identified Clusters for Kamrup

				Kamrup					
Group No->	1	2	3	3	5	5	6	7	7
Sl No->		23	28	32	80	84	87	99	107
Name->	**Brick making**	**Book binding**	**Cereals/spices**	**Rice mill**	**Furniture**	**Fabrication**	**Garments**	**Electric help**	**R.C.C**
Pin code		**Colours denote Clusters >= 5 No.**							
781101	-	-	-	-	-	-	-	7	-
781102	-	-	-	-	-	-	6	-	-
781103	-	-	-	-	-	-	5	12	8
781121	-	-	-	-	-	-	5	-	-
781122	-	-	6	9	-	-	6	7	-
781123	5	-	9	-	-	-	5	-	-
781124	-	-	-	-	-	-	6	-	-
781125	-	-	8	6	-	6	-	-	-
781132	-	-	11	-	-	-	-	-	-
781137	-	-	-	-	-	-	7	-	-
781141	-	-	6	-	-	-	-	-	-
781350	-	-	-	-	5	5	15	-	12
781354	-	5	-	-	-	-	5	-	9
781380	-	-	-	5	8	7	15	-	14
781381	-	-	-	-	-	-	7	-	5
781382	-	-	-	-	-	-	5	-	5

From the above **Table 8,** for Kamrup district, we can see that **1st group industry product Brick making; 2nd group industry product 23-Book binding work; 3rd group industry product 28- Processing/marketing of cereals/spices; 32- Rice mill; 5th group industry product 80-Wood furniture making; 84-Fabrication/Iron grill making; 6th group industry product 87-Tailoring & Readymade Garments;** 7th group industry product 99, 107 has the potential for KVI cluster formation in the respective Pin code as given in the table.

As Processing/marketing of cereals/spices, Rice mill, Fabrication/Iron grill making and Tailoring & Readymade garments are showing more potential, i.e. >=3 clusters each; These can be considered for next KVIC cluster activity. Since service sector (7th group) cluster plan is not there right now we are not recommending it, though it is showing 5 clusters in 7th group industry 99- Electrical work/product and 107-Masonary work.

3.4 Tabular Representation of District wise Cluster (Nagaon)

Table 9: Identified Clusters for Nagaon

				Nagaon					
Group No->	2	3	3	3	5	6	7	7	7
Sl No->	21	28	32	35	80	87	99	103	107
Name->	**Cane/Bamboo**	**Cereals/spices**	**Rice mill**	**Sweet making**	**Furniture**	**Garments**	**Electric help**	**Hiring gadgets**	**R.C.C**
Pin code	**Colours denote Clusters >= 5 No.**								
782001	-	-	-	-	-	6	-	7	-
782002	5	6	-	-	-	7	-	-	-
782124	-	-	-	-	-	7	-	-	-
782125	-	12	-	-	-	-	-	-	-
782136	-	-	-	-	-	-	6	-	-
782425	-	23	-	6	7	22	10	10	-
782426	-	16	-	-	-	9	-	-	-
782446	-	-	8	-	-	-	-	-	5

From the above **Table 9,** for Nagaon district, we can see that **2nd group industry product 21-Cane/Bamboo work; 3rd group industry product 28- Processing, packing and marketing of cereals, pulses, spices, condiments, masala etc.; 32- Rice mill; 35-sweet Making; 5th group industry product 80-Wood furniture making; 6th group industry product 87-Tailoring & Readymade Garments;** 7th group industry product 99, 103, 107 has the potential for KVI cluster formation in the respective Pin code as given in the table. As Processing/marketing of cereals/spices and Tailoring & Readymade garments are showing more potential, i.e. >3 clusters each; these can be considered for next KVIC cluster activity.

4 SOME INFERENCES FROM CLUSTER FINDINGS

Studying the **Tables (6, 7, 8 & 9)** above, for cluster formation, of all the four selected districts, it is observed that:

- 6th group industry product 87-Tailoring & Preparation of Readymade Garments type of manufacturing village industry has come out very prominently in all the four districts under study.

- In the above tables the next visible village industry clusters, in all the three districts except Barpeta, are 3rd group industry product 28- Processing, packing and marketing of cereals, pulses, spices, condiments, masala etc.

Despite its recent economic advances, India's gender balance for entrepreneurship remains among the lowest in the world. Improving this balance is an important step for India's achievement of greater economic growth and gender equality [10]. Above found village industries will be suitable for both male as well as female. However there is significant gender wise difference in choice of village industry [19].

5 CONCLUSION

This paper is an effort for motivating towards application of data mining technology as an aid, in the decision making process for actual village industry cluster identification, in place of or along with traditional processes followed by Entrepreneurship development organizations, such as KVIC etc. Village industries can achieve high levels of competiveness once they work in a cluster environment ensuring complementariness, common activities and institutional stability [14]. Clusters can be upgraded and collective innovations can flow from these efforts. These may be helpful for building cluster-specific skill centers for development of cluster specific labour force, strengthening

the linkages with the local suppliers and facilitating higher level of interactions among the stakeholders of clusters.

This data mining technology can be useful not only for cluster identification, but also for finding trends in area wise selection of village industry type etc. The attractiveness of this data mining is that –most of the required knowledge can be extracted from the existing databases.

Furthermore instead of only selected area, overall computerization may improve efficiency of KVIC like organizations. [18] Demonstrates the use of an approach for the development of performance measurement system at one end and a workable framework of balanced scorecard at other end for a real case company KVIC.

REFERENCES

[1] Bujar Baruah, P. (2012), "Impact of Micro-Finance on Poverty: A Study of Twenty Self-Help Groups In Nalbari District, Assam", NIRD Journal of Rural development Vol.31 No. (2), pp 223-234.

[2] Baruah,S.B. (2005),"A study on Entrepreneurship Development efforts and Needs Assessment for Entrepreneurial Growth in North East", Research publication of IIE, Guwahati.

[3] Busenna P. and Reddy, A.A.(2011), "Khadi and Village Industry: A Case Study of Khadi Institutions in India", NIRD Journal of Rural Development Vol.30 No.(3), pp 273-289.

[4] CKVI, (2005), "CKVI Activities & Programmes", publication of Commission for Khadi & Village Industry Assam State Unit.

[5] Devedzic, V., (2001) "Knowledge Discovery and DM in Databases", FON- School of Business Administration, University of Belgrade, Yugoslavia.

[6] Dunham, M. H., (2009) "Data Mining Introductory and Advanced topics", Pearson Education.

[7] Dutta, G. K., (2003) "A study of development of Bee-keeping Clusters at Nalbari and Jorhat district of Assam", Research publication of IIE, Guwahati.

[8] Ghani,E., Kerr,W.R. and O'Connell, S. (2011), "Local Industrial Structures and Female Entrepreneurship in India", Harvard Business School working paper No: 12-036.

[9] Government of Assam (2006), "Statistical Hand Book of Assam, 2006" Directorate of Economics and Statistics, Government of Assam, Guwahati.

[10] Inayatullah, J. and Birley, S.(1997), "The Orangi Pilot Project: The evaluation of a Micro enterprise Credit Institution," Entreprenuership & Regional Development, 9(4), pp 299-318

[11] Lall, S.(1997), "Learning from the Asian Tigers: Studies in Technology and Industrial policy". London: Macmillan Press Ltd.

[12] Mali, D.D. (1999), "Entrepreneurship Development in North East", Mittal Publications, New Delhi.

[13] Mandal, R. K. (2007) "Khadi & Village Industries in North East India with special reference to Arunachal Pradesh: Retrospect and Prospect", Research center for Social Sciences, Mumbai, India journal of Global Economy, Vol. 3, Iss. 2 (April), pp- 88-99.

[14] Nagayya, D. and Rao, P.T.,(2013), "Small and Medium Enterprises in the Era of Globalisation", NIRD Journal of Rural Development Vol.32 No.(1), pp 1-17.

[15] Pujari, A.K., (2006) "Data Mining Techniques" Universities Press.

[16] Rahman, M. and Day, N.B. (2010), "Micro and Small Enterprises in NE India- problems & prospects", EBH Publishers (India).

[17] Tan, P., Steinback, M., and Kumar, V. (2009), "Introduction to Data Mining", Pearson Education.

[18] Thakkar, J., Deshmukh, S.G., Gupta, A.D. and Sankar, R., (2006) "Development of balanced scorecard-An integrated approach of interpretive Structural Modeling (ISM) and Analytical Network Process (APN)", International Journal of Productivity and Performance Management, Vol. 56 Iss:1, pp 25-59.

[19] Sarma, N. and Talukdar, P.H. (2012), "Environment Friendly Village Industry: A Case Study On KVIC Sponsored Rural Induastries In Assam", Paper presented at National Seminar on 57th Annual Technical Session of Assam Science Society, held at Gauhati University, on 16th March, 2012 .

[20] http://www.indiapost.gov.in/ - Postal Department India, web link.

[21] http://www.kvic.org.in/ - KVIC web link.

[22] http://www.msme.gov.in /- Ministry of Micro, Small & Medium Enterprises.

[23] http://www.nird.org.in/ - National Institute of Rural Development web link.