DECISION SUPPORT SYSTEM FOR BANKING ORGANIZATION

A THESIS REPORT

Submitted by

P.MEGALA

Under the guidance of

Dr.R.SHRIRAM

in partial fulfillment for the award of the degree of

MASTER OF PHILOSOPHY

in

COMPUTER SCIENCE

B.S.ABDUR RAHMAN UNIVERSITY

(B.S. ABDUR RAHMAN INSTITUTE OF SCIENCE & TECHNOLOGY)

(Estd. u/s 3 of the UGC Act. 1956)

www.bsauniv ac.in

DECEMBER2012

B.S.ABDUR RAHMAN UNIVERSITY

(B.S. ABDUR RAHMAN INSTITUTE OF SCIENCE & TECHNOLOGY)

(Estd. u/s 3 of the UGC Act. 1956)
www.bsauniv ac.in

BONAFIDE CERTIFICATE

Certified that this thesis report **DECISION SUPPORT SYSTEMFOR BANKING ORGANIZATION**isthe bonafide work of **P.MEGALA(RRN1145212)**who carried out the thesis work under my supervision. Certified further that to the best of my knowledge the work reported herein does not form partof any other thesis report or dissertation on the basis of which a degree or award wasconferred on an earlier occasion on this or any other candidate.

SIGNATURE

Dr. R.SHRIRAM
SUPERVISOR
Professor
Department of CSE
B.S. Abdur Rahman University
Vandalur, Chennai – 600 048

SIGNATURE

Dr. P.SHEIK ABDUL KHADER
HEAD OF THE DEPARTMENT
Professor & Head
Department of Computer Application
B.S. Abdur Rahman University
Vandalur, Chennai – 600 048

ABSTRACT

Decision Support System (SDSS) have been an active area of research for over a decade, addressing the growing data management and analysis needs of spatial applications such as Geographic Information Systems (GIS).In this research, Spatial Decision Support System (SDSS) concentrates on to overcome the difficulties of integrating Geo-graphic information data and non-spatial data in selecting a location.The approach here is to represent the methodology for the development ofeffective spatial decision support environment (SDSS) for branch bank rationalization by integrating spatial and non-spatial data. Bank manager can select the most appropriate city for opening a new branch among the Multi Criteria Decision Making (MCDM). The implementation uses the Visual Basic .Net programming environment and SQL.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

FIGURE NO	TITLE	PAGE NO
3.14(c)	Contribution of Spatial and non-spatial data in Area1	39
3.15	Output of K-mean Algorithm	40
3.16	Output of K-medoids Algorithm	41
4.1	Interaction of SDSS Components	44
5.1	Output of K-mean Algorithm	47
5.2	Cluster values of K-means	48
5.3	Output of K-medoids	49
5.4	Cluster values of K-medoids	50
5.5	Output of CLARANS	51
5.6	Cluster values of CLARANS	51
5.7	Final Results of 4 Areas	52
9.1	Screen Shot of K-mean	58
9.2	Screen shot of location variables	59

B.S. ABDUR RAHMAN UNIVERSITY

SCHOOL OF COMPUTER AND INFORMATION SCIENCES

DEPARTMENT OF COMPUTER APPLICATIONS

21.3.2013

SUBMITTED TO

Dean (Academic Research)

Through HOD/CA

Sub: Submission of Minutes of Viva voce meeting – reg.

The Viva voce examination of M.Phil research scholar **P.MEGALA (RRN: 1145212)** was conducted on 21.3.2013 successfully. The Proceedings and attendance sheet of the M.Phil oral board examination is sent herewith.

24/3/13

Dr. P. Sheik Abdul Khader
HOD/CA

Dr. R. Shriram
Supervisor
Professor / CSE

B.S. ABDUR RAHMAN UNIVERSITY, CHENNAI – 48
DEPARTMENT OF COMPUTER APPLICATIONS

Date: 21.03.2013

PROCEEDINGS OF THE M.PHIL VIVA-VOCE EXAMINATION OF **MS. P. Megala (RRN 1145212)** , HELD AT 10.00 A.M ON 21.03.2013 IN SEMINAR HALL OF DEPARTMENT OF COMPUTER APPLICATIONS, B.S.ABDUR RAHMAN UNIVERSITY, CHENNAI - 600 048.

The open defense viva voce Meeting of **P.Megala (RRN 1145212)**, M.Phil Research Scholar , Department of Computer Applications, of the University was conducted on 21.03.2013 at 10.00 am in the seminar hall of Department of Computer Applications. The following members were present,

1. **Dr. P.Yogesh,** — **- Examiner**
   Associate Professor,
   Department of Information Science and Technology,
   Anna University,
   Chennai – 600 025

2. **Dr.P.AnandhaKumar** — **- Expert Member**
   Professor
   Department of Information Technology
   MIT Campus, Anna University,Chennai – 600044

3. **Dr. P. Sheik Abdul Khader** — **- Member (Ex. Officio)**
   Professor and Head of CA

4. **Dr. R.Shriram** — **- Supervisor & Convener**
   Professor,
   Department of CSE,
   B.S. Abdur Rahman University

P. Megala (RRN 1145212), has presented her research work on **"Decision Support System for Banking Organization"**. This was followed by question from the board members. The Scholar answered the questions to the full satisfaction of board members. The corrections suggested by the examiners have been carried out and incorporated in the thesis before Oral Examination.

Based on the scholar's research work, her presentation and also the clarifications and answers by the scholar to the questions, the board recommends that **P. Megala**, be awarded M.Phil degree in Computer Science.

P. [signature] 21/3/13	P. [signature] 21-3-13	[signature] 21/3/13	[signature] 21/3/13
(Dr.P.AnandhaKumar)	(Dr.P.Yogesh)	(Dr.P.Sheik Abdul Khader)	(Dr.R.Shriram)
Expert Member	**Examiner**	**(Ex. Officio)**	**Supervisor & Convener**

CHAPTER I

INTRODUCTION

1.1 DATA MINING IN SPATIAL DATABASE

Data mining in general is the search for hidden patterns that may exist in large databases. Spatial data mining in particular is the discovery of interesting relationship and characteristics that may exist implicitly in spatial database. Because of huge amounts of spatial data that may be obtained from satellite images, medical equipments, video cameras, etc., Spatial data mining aim to automate such a knowledge discovery process.

That is plays important role

- Extracting Interesting Spatial patterns and features.
- Capturing intrinsic relationships between spatial and non-spatial data.
- Presenting data regularity concisely and at higher conceptual level.
- Helping to reorganize spatial databases to accommodate data semantics as well as to achieve better performance.

Data mining techniques are the result of a long process of research. This evolution began when business data was first stored on computers, continued with improvements in data access, and more recently, generated technologies that allow users to navigate through their data in real time.

Data mining takes this evolutionary process beyond retrospective data access and navigation to prospective and proactive information delivery.

Data mining is ready for application in the business community because it is supported by three technologies that are now sufficiently mature:

- Massive data collection
- Powerful multiprocessor computers
- Data mining algorithms

1.2 CLUSTER ANALYSIS

Cluster analysis groups objects (observations, events) based on the information found in the data describing the objects or their relationships. The goal is that the objects in a group will be similar (or related) to one other and different from (or unrelated to) the Objects in other groups. The greater the similarity (or homogeneity) within a group, and the greater the difference between groups, the “better” or more distinct the clustering.

Clustering can be considered as the most important unsupervised learning problem; so, as every other problem of this kind, it deals with finding a structure in a collection of unlabeled data. A cluster is therefore a collection of objects which are “similar” between them and are “dissimilar” to the objects belonging to other clusters.

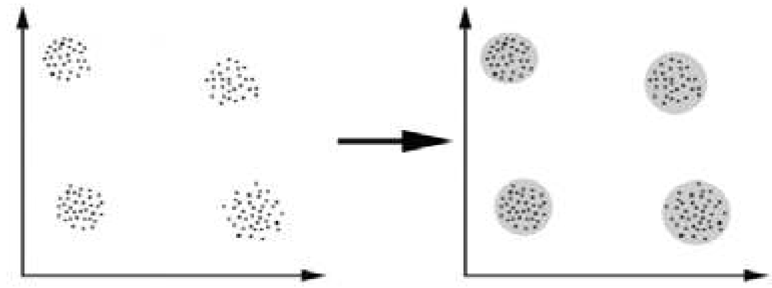


Figure 1.1: A Graphical Example for Clusters

Figure 1.1 shows the data it can easily identify the 4 clusters into which the data can be divided; the similarity criterion is distance: two or more objects belong to the same cluster if they are “close” according to a given distance (in this case geometrical distance).This is called Distance-based clustering.

1.3 DECISION SUPPORT SYSTEM

Decision Supports Systems (DSS) are computer-based information systems designed in such a way that help managers to select one of the many alternative solutions to a problem. It is possible to automate some of the decision making processes in a large, computer-based DSS which is sophisticated and analyze huge amount of information fast. It helps corporate to increase market share, reduce costs, increase profitability and enhance quality. The nature of problem itself plays the main role in a process of decision making. A DSS is an interactive computer based information system with an organized collection of models, people, procedures, software, databases, telecommunication, and devices, which helps decision makers to solve unstructured or semi-structured business problems.

1.3.1 Characteristics of DSS

A Decision Support System (DSS) is an interactive, flexible, and adaptable computer based information system that utilizes decision rules, models, and model base coupled with a comprehensive database and the decision maker’s own insights, leading to specific, implementable decisions in solving problems that would not be amenable to management science models. Thus, a DSS supports complex decision making and increases its effectiveness.

1. Handle large amounts of data like database searches.
2. Obtain and process data from different sources including internal and external data stored on mainframe systems and networks.
3. Provide report and presentation flexibility to suit the decision maker's needs.
4. Have both textual and graphical orientation like charts, trend lines, tables and more.
5. Perform complex, sophisticated analysis and comparisons using advanced software packages.
6. Support optimization, satisfying, and heuristic approaches giving the decision maker a great deal of flexibility in solving simple and complex problems.
7. Perform "what-if" and goal-seeking analysis

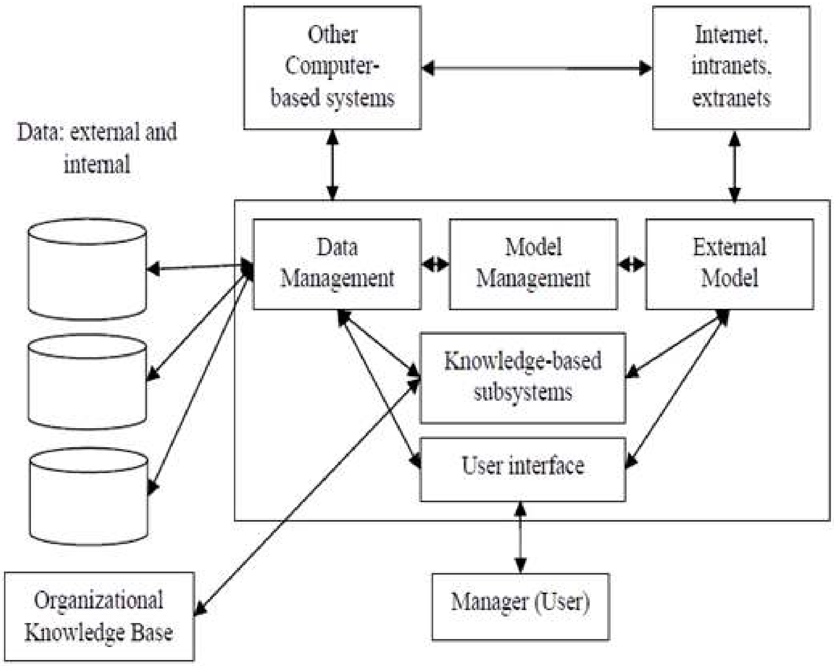


Figure 1.2: A systematic View of DSS

1.3.2 GIS (Geo-graphic information system)

Geographic information system" is a computerized system for the collection, storage, manipulation (analysis), and output of information that is spatially referenced…" GIS is apart from other types of information systems is their reliance on spatial information as their organizing framework and their ability to perform geographic analysis. An added advantage of DSS figure1.2 shows the ability to display information graphically for easy interpretation. The author sees GIS as an information system that models the real word.

1.3.3 Decision support systems

There has been a separate but parallel development of systems to support decision-making in the geographic literatures. Decision support systems (DSS), from management science; have evolved out of Management Information Systems (MIS) which have existed since the 1960s. Decision support systems, from geography, are evolving from Geographic Information

Systems (GIS) which have existed since the early 1960s. Most GIS provide capabilities for map overlay but do not support the analytical and statistical modeling required by many decision-makers (Armstrong and Densham, 1990).

There are six characteristics that distinguish a DSS:

1. They are used to tackle ill or semi-structured spatial problems - these occur when the problem, the decision-makers objectives, or both cannot be fully and coherently specified.

2. They are designed to be easy to use; the often very sophisticated computer technology is accessed through a user-friendly front end.

3. They are designed to enable the user to make full use of all the data and models that are available, so interfacing routines and database management systems are important elements.

4. The user develops a solution procedure using the models as decision aids to generate a series of alternatives.

5. They are designed for flexibility of use and ease of adaptation to the evolving needs of the user.

6. They are developed interactively and recursively to provide a multiple-pass approach which contrasts with the more traditional series approach - involving clearly defined phases through which the system progresses.

Decision support systems represent the role of computer information systems in the decision making process. For the purposes of this research, the term SDSS will include clustering techniques is used for grouping the similar Geo-graphical information data. It will also incorporate understanding of the decision making process involved in solving a specific bank location problem for integration of spatial and non-spatial data fulfilling the characteristics set forth by Geoffrion.

1.4 BANK-BRANCH LOCATION ANALYSIS

Figure 1.3 explain One of the most widely known procedures used for evaluating and selecting a new bank branch office location is the technique suggested by the American Bankers Association (ABA) in their book "A Guide to Selecting Bank Locations". A flow chart of the procedure is displayed in Figure 1.3 .

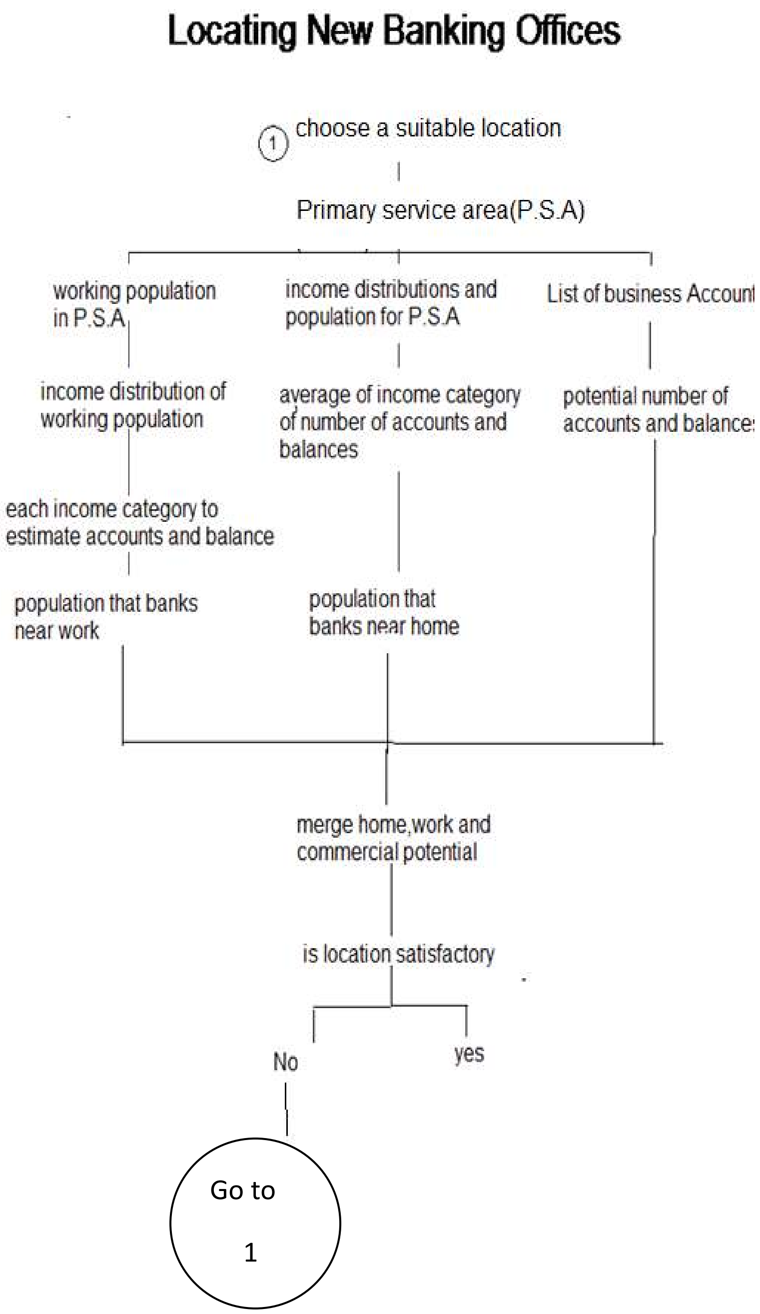


Figure 1.3: Locating New Bank branch

A checklist method is presented, leaving to the discretion of the location analyst the choice of potential facility locations or candidates, the definition of the primary service area, the estimation of the proportion of the highly valuable customer population who bank near home or work, estimates of realizable market growth, and final adjustments to accounts and balances (Lundsten, 2009). The ABA method suggests that analysts evaluate factors that are critical to the success of a branch such as socioeconomic and demographic characteristics of consumers, the level of competition, and consumer expenditure patterns. In addition, there are site-specific factors such as traffic patterns, parking availability, route access, and visibility that are considered by decision-makers.

1.5 THE NEED FOR A DECISION SUPPORT SYSTEM

Making decisions about branch bank closure is a typical strategic planning problem for financial institutions requiring an integrated approach to data management and analysis. The use of the analytical tools of GIS coupled with MCDM to formally create a SDSS is one way of providing this needed integrated approach that can generate valuable information for bank decision-makers. A SDSS is designed to support decision making in an unstructured or semi-structured situation where the problem is complex and not amenable to human analysis alone. SDSS typically use models to allow rapid "what-if" analyses and make use of data stored in a spatial database management system.

Although GIS provide a number of analytical facilities/models to support spatial analysis and information processing, particularly through 'cartographic modeling', it is generally regarded by the user community that this is not sufficient for most business applications.

On the other hand, GIS have good database management facilities that are essential to make a SDSS powerful. Therefore there is a need to augment the power of GIS with the additional analytical capabilities suited to a particular problem. As pointed out above, given that the branch closure problem involves multiple and conflicting factors, a multi-criteria approach is required for the evaluation of branch banks in order to arrive at a rational and justifiable ranking of candidates for closure. MCDM can formally provide this additional analytical capability. The integration of MCDM techniques with GIS is therefore a logical approach to the development of explicitly spatial decision support tool for branch closure analysis and planning.

1.5.1 Advantages

Banking organization plays are very important role for developing our country. GIS software helps banks to integrate spatial and non-spatial data to improving the business needs.

- ✓ GIS software helps to bank managers to identify the valuable customers with needs, demands, and preferences that should be met by any successful bank.
- ✓ We can improve standards of living.
- ✓ Banking organization is looking to position the right products and services to customers,
- ✓ Find new branch/ATM location is, improve workplace safety, or build comprehensive business Continuity plans, GIS helps them achieve these goals.

The chapter 1 discussed about Data Mining techniques, Decision Support System (DSS), location Analysis, and the flow chart of decision support system. The rest of the chapter is organized as follows. Chapter 2 gives a literature Review of Decision support system for bank branch location Analysis. Chapter 3 we evaluate the Proposed Method with respect to both efficiency and effectiveness of Decision support system. Chapter 4 is the result analysis of the proposed work and chapter 5 is the Conclusion of the work.

CHAPTER 2

LITERATURE REVIEW

Densham P.J (1991) has proposed the new framework of Decision support system (SDSS) for banking organization. According to Densham P J emerge the three levels of technology used in the SDSS framework. (1) Lowest level, (2) SDSS generator, (3) Intermediate level. At the lowest level is the SDSS toolbox. This is a set of hardware and software components that can be assembled to built a variety of system modules. Second level of technology is the DSS generator. A generator is a set of mutually compatible hardware and software modules that can be configured easily to produce a specific SDSS. The intermediate level of technology containing a series of mutually compatible modules, it can be rapidly configured to provide a particular set of capabilities.

Densham P J (1995) discussed the characteristics of a Decision support system. SDSS can be characterize as (1) Iterative, (2) Integrative, and (3) Participate. Iterative a set of alternative solution is generated which the decision maker evaluates. The Participation occurs the decision makers plays an active role in defining the problem carrying out the analyses and evaluating the outcomes. Integrative is a value judgments that materially affect the final outcome are made by decision makers. The scope of this paper discussed proposed the new framework and characteristics of SDSS for developing new bank branch location.

Nihan Cinar (2010) have Proposed FAHP and TOPSIS methods are used together for selecting bank branch location. FAHP is utilized for determining the weights of the criteria and TOPSIS method for determining the ranking of the cities. As the weights of the criteria are determined by bank managers from different areas, the result indicates an overall performance ranking. This study indicates that both fuzzy AHP and TOPSIS can be used a decision support system by the organizations in order to make effective decision on the bank branch location selection.

B.U. Sajeev, Thangavel (2012) A new hybrid clustering algorithm of K-Means and Fuzzy C-Means is introduced. In K-Means algorithm we have selected initial seeds randomly. The new hybrid algorithm proposed uses an initialization method to choose suitable initial seed which have high potential to form high quality clusters. This algorithm proposes a method to choose the initial seeds of K-Means algorithm which generates high quality performed and non- performed clusters. Author used clustering analysis for the first time in the socially relevant Self Help Group (SHG) data's.

Avkiran (1995) offers an interdisciplinary and multivariate perspective for an integrated spatial and non-spatial data analysis of bank branch performance. The author's contribution is, therefore, relevant in the sense that he aims to minimize the gap between current branch performance and branch potential based on the spatial location variable. His use of econometric techniques is based on variables that are controllable by bank management. Thus, his study is considerably different than ours, not only in methodological terms, but also because we believe that there are several other variables that fall out the bank's sphere of control that may influence bank branch potential attractiveness.

Dingming Wu (2010) have emerged two different algorithm to reteriveing the top-k spatial keyword quries . (1) Iterate Algorithm computes the top-k results for each sub query separately, which processing a sub query Q, the algorithm maintains a priority queue on the nodes to be visited. The algorithm Utilizes the keyword information to prune the search space. A non-leaf entry is pruned if it does not match all the keywords. The algorithm returns K elements the have the smallest Euclidean distance to the query and contain the query keywords. (2) Group Algorithm to process all sub queries of Q concurrently by employing a shared priority queue to organize the visits to the tree nodes that can contribute to closer results. Group algorithm guarantees that each node in the tree is accessed at most once during query processing. A discussion of the two algorithms having advantage and some disadvantage falls out of the this paper's scope.

Boufounou (2006) employs econometric models to produce a set of equations that predict the main dimensions of branch performance. He argues that external elements should be included in the decision making process, and regards *Volume of Deposits* as the major evaluation criterion measure of the branch performance. Volume of Deposits consider as a Non-Spatial data. He then establishes causal relationships between this measure of performance and the *Number of Rentiers* in the branch trade area, *Branch Age*, *Number of Employees* (associated to the branch's size) and presence of *Night Deposit Facilities* (which represents an exterior attractiveness design feature, according to the author). Finally, he estimates branches' potential attractiveness by comparing each one of the branches' scores with the overall average.

Zhao *et al.* (2008) explore the way in which geographical criteria and a more explicitly spatial approach can be used to identify branches as candidates for closure and to provide decision makers with a more formal approach to branch bank strategy planning. The contribution of these authors seems to be extremely important in the context of the present paper, because despite the fact that financial performance is typically seen as the most important in evaluating 8 branches' viability, they suggest that a number of more general factors should also be considered in assessing branch potential. Besides, their study is partially based on MCDM – Multiple Criteria Decision Making – techniques, which corroborate some of our orientations (for a deeper discussion on MCDM and MCDA, see *e.g.* Roy and Vanderpooten, 1997 and Belton and Stewart, 2002).

Guboin Li (2009) presents the process of spatial query handling. The query handling is divided into two steps (1) filtering , (2) Optimizing. The filtration step a query processor first visits the spatial index. Spatial index store similar description of the spatial objects and to exclude the object which may not meet the query conditions. The optimizing step is the geometric operation for the actual data. The significant proportion of candidates sets do not meet the query conditions, as much as possible to reduce this kind of object into further refinement in order to avoid unnecessary geometric operation. This paper discussed the methods and algorithm to improve the query efficiency.

Paradi and Schaffnit (2004) offer a DEA application where two production models are developed. In one of those models, an environmental factor is introduced with the scope of capturing the level of economic growth in each one of the geographical areas under study. Although this study does not offer much to the potential attractiveness context, it seems to be important in the sense that it tries to align bank managers' judgements with performance measures that support the strategic goals.

CHAPTER 3

PROPOSED SYSTEM

3.1 NON-SPATIAL FACTORS RELATING TO BRANCH CLOSURE

Apart from purely spatial considerations, many other Non-spatial factors and criteria must be considered in planning branch bank networks. There is now a considerable with these as well as performance indicators that typically are considered in making decisions about the closure of individual branches.

Table 3.1 shows the Non-Spatial Data variable for banking organization. The seven variables are playing most important role in the banking sector. The variables are banking population, Population growth rate, Average annual family income, etc..,

Table 3.1: Non-Spatial Data for Banking Organization

Variables	Code
Banking population (age 15 & over)	NS1
Population growth rate	NS2
Average annual family income	NS3
Average age	NS4
Aged population (over 55 years)	NS5
Employed people	NS6
Total number of (small) businesses	NS7
Competitors’ branches	NS8

The group of eight variables relates mainly to the demographic and population characteristics of the bank branches location: the current feasibility of a particular branch, as well as its future potential, is a function of the demographic profile of its trade area and patterns of customer behavior. These variables link to the demand side of a branch and include: total banking population, the growth rate of the population, average age of the population, population aged 55 or more, average annual family income, and the employed population are selected as Non-Spatial data representing the residential demand for banking services.

Demand from the commercial sector is represented by the total number of small businesses found in the trade area and capture the extent competition to branches, the number of branches of competitor banks located within the trade area.

3.2 SPATIAL FACTORS RELATING TO BRANCH CLOSURE

Next consider spatial related data for opening new bank branch. The spatial data of eight location variables relates largely to the spatial supply of a branch location, such as the numbers of small businesses and working population within in 200 meters travel zone, competition in nearby the branches, and accessibility to these. Proximity is defined in terms of various distance zones around each branch. Key criteria considered to affect branch supply capability are the number of small businesses, the working population, and shopping centers within the area, and spatial competition by the number of branches of competitor banks and branches of the particular branch.

Travel time and distance are important measures of how easily customers can reach a branch to obtain financial services and it has been pointed out above that customer patronage is subject to distance-decay effects.

Two criteria are included that measure ease of reaching a particular branch - an index of accessibility and proximity to public transport. Since this group of Sub-Criteria relates to distance zones around individual branches, they are ‘location’ specific. Table3.2 shows the spatial variables and the code

Table 3.2: Spatial Variables (Location Variables)

Variables	Code
Small business in 200m travel zone	S1
Competitor branches in 200m buffer	S2
Working population in 200m travel zone	S3
Branches in 500m buffer	S4
Branch at shopping centre	S5
Shopping centers within 500m buffer	S6
Accessibility index	S7
Proximity to public transport	S8

Of the spatial variables (location variables). After collection of spatial and non-spatial data the research is required into the weights are assigned separately for spatial and non-spatial data based on the criteria. The criteria selected should be evaluated on the basis of the importance of their contribution, validity, and redundancy. A number of methods may be used to evaluate the criteria: in this research, using clustering techniques and K-mean algorithms and find out the distance between the spatial and non-spatial data with the help of distance relation (Euclidian Distance).

3.3 METHODOLOGY

The Decision support system concentrates on to overcome the difficulties of integrating Spatial and Non-Spatial data in the process of selecting a location for new bank branch. The approach here is to represent the methodology for the development of effective spatial decision support environment (SDSS) for branch bank rationalization by integrating spatial and non-spatial data. Bank manager can select the most appropriate city for opening a new branch. Spatial location is playing the important role in the research.

The selection of the location must satisfy a spatial and Non-Spatial data to opening a new branch for a particular bank.

Bank manager select the city to opening a new branch for a bank. Figure 3.1 shows the different areas of Bangalore city. Once selecting the city gives the important to selection of the area (region) for a particular city.

Figure 3.1: Bangalore City Map

Decision maker analyzing spatial and non-spatial data for all the regions of the city. Decision maker first analyzing spatial related data for opening a new bank branch. The spatial data of eight location variables relates largely to the spatial supply of a branch location Table 3.1 shows the eight spatial location variables.

Decision maker clustering spatial data of all the areas using distance (Euclidian Distance) relation. Euclidian Distance calculating distance between the spatial variable and around 200km what are spatial variables are nearly close to the particular posh area.   With the help of clustering techniques the two points are clustered which two spatial variable distance is very close to the particular posh area. Next finding new clustering point with help of clustering techniques and K-means Algorithm.

The bank manager has to select new branch at Bangalore city for a **XYZ** bank. The branch should cover 4 areas. Areas are Bhagalpur (A1), Budigere (A2), kaglipur (A3), and Hosur (A4) of the city to opening a new branch for **XYZ** bank.

Area1 (Bhagalpur) A1

Area (A1) having no. of P attributes. The P attributes represent the posh areas of the Bhagalpur (A1) regions. The attributes donates

Let A1 = {A11, A12, A13, A14…..A1m},

Figure 3.2: Area1 (Bhagalpur)

Figure 3.2 shows posh areas of the Bhagalpur area. Bhagalpur area having 5 no's Posh Areas. The Areas are (P&T colony) P1, (Nathnagar) P2, (Barari) P3, (Tatarpur) P4, and (Sabour) P5. Decision maker separately analyzing spatial and non-spatial variables of P1, P2, P3, P4, P5 Posh Areas. It's denote

Let A1 (P1, P2, P3, P4, P5)

Analyzing Spatial Variables of Posh Area (P&T Colony) P1

Decision maker analyzing eight spatial variables of like Small business in 200m travel zone (S1), Competitor branches in 200m buffer (S2), Working population in 200m travel zone (S3). Table 3.2 shows eight spatial location variables. It denotes

Let P1 = (S1, S2, S3, S4, S5, S6, S7, and S8)

So

A1 ^ (P1, P2, P3, P4, P5) and P1 ^ (S1, S2, S3, S4, S5, S6, S7, S8)

Figure 3.3: P&T Colony (P1) of Bhagalpur

Figure 3.3 shows the no.of posh Area available in (P&T colony). so P1 having 5 small business (S1) in 200m travel zone, 3 competitor branches (S2) in 200m buffer, 62% of working population (S3), 1% same branch in around 200m , Table 3.3 shows the contribution of the spatial variables in Posh area P1.

It denotes

Let S1 = (S11, S12, S13, S14, S15)   S2 = (S21, S22, S23)

S3 = 62%   S4 = (S41)

S5 = (0)   S6 = (S61, S62, S63, S64, S65)

Table 3.3: Contribution of spatial variables in Posh Area (P1)

Spatial Variables	Contribution
Small business in 200m travel zone	5
Competitor branches in 200m buffer	3
Working population in 200m travel zone	62%
branches in 500m buffer	1
branch at shopping centre	2
Shopping centers within 500m buffer	5
Accessibility index	Y
Proximity to public transport	88%

Analyzing the spatial variables on Posh area (P&T colony) P1. After, Finding distance between two spatial variables with help of Distance relation (Euclidian Distance). Randomly select any one of the small business from the spatial variable S1 using distance relation (Euclidian Distance) find out which small business is very close to randomly selected one, with the help of K-mean value clustering the similarity of spatial variables of S1 (Cluster 1).

Next calculate the new K-mean value again find out next which small business in very nearby the new K-mean value the process will continue clustering all the nearby spatial variables on S1.

3.3.1 Distance Relation

Distance relations compare the distance between the attribute and the spatial location variables of each attribute belong to S1 variable. Two objects with a given constant using arithmetic operators such as <, >, =. The distance between two objects is defined as the minimum distance between them (i.e. select all elements inside a radio of 50 km from a "*x*" point). Figure 3.4 shows two examples of this type of relation; in the Figure 3.4 representing how close and how far two spatial variables are each other.

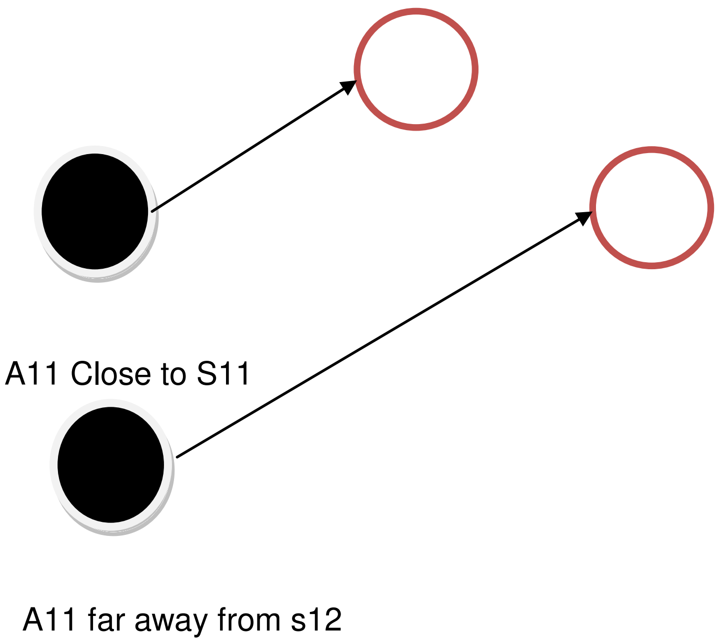


Figure 3.4: Distance Relation between two points

Euclidian Distance

We propose to use Euclidian distance which is defined as the straight line distance between two points.

Euclidian distance is

$$A11\ (S11) = \sqrt{((x1 - x2)^2 + (y1 - y2)^2)}.$$

Apply Euclidian distance to all the region attributes and attributes compare with the eight spatial location variables s1,s2, s3…..s8.

A*11* at (*x1*, *y1*) and

S11 at (*x2*, *y2*) ^ S12 (x2, y2) ^S13(x3, y3)….s1n (xn, yn),

S21 at (*x2*, *y2*) ^ S22 (x2, y2) ^S23(x3, y3)….s2n (xn, yn),

S31 at (*x2*, *y2*) ^ S32 (x2, y2) ^S33(x3, y3)….s3n (xn, yn),

…………..

…………..

Sn1 at (*x2*, *y2*) ^ Sn2 (x2, y2) ^Sn3(x3, y3)….s1n (xn, yn)

Hear A1 is the Bhagalpur and A11, A12, A13……. A1m is the no. of posh areas of the Bhagalpur. Similarly A2, A3, A4 having no. of P attributes. Figure 3.5 represents the four Areas and Posh Areas of the city and Applying Distance relation, Clustering techniques and K-means Algorithms for all A1, A2, A3, and A4 areas and find out which point is very suitable for opening new branch of the **XYZ** bank.

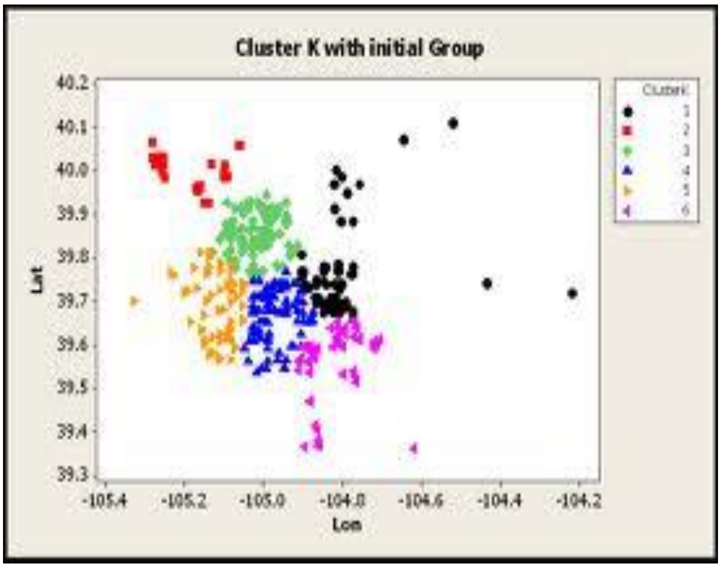


Figure 3.5:Spatial variables of Posh Area1

Posh Area (P&T colony) P1	=	(100,100)
Small business S1	=	S11 (300,100), S12 (500, 50), S13(100,150) S14(50,75)
Competitor Branch S2	=	S21 (300,275),S22(150,75), S23(300,700)
Branch in 500 m S4	=	S41(800,200)
Branch at shopping centre S5	=	S51 (25, 25), S52 (75,400)

Posh Area P1 location point is (100,100). P1 value is constant; it cannot change the location value of P1. All the spatial location variables of small business S11 is (300,100) and S12 (500, 50)…., competitor branch S21 is (300,275) and S22 (150, 75) point out all the spatial location variables posh area P1. With the help of distance relation (Euclidian distance) finding the distance between constant value of (100,100) P1 and the small business of S11,S12,S13, and S14 analyzing which small business is very close to the S1constant value of P1 and cluster the Posh area and very closed S1 variable and finding K-mean value of the Posh area P1 and closed S1 variable and analyzing which S1 variable is nearby K-mean value so the process continue until the similarities of all S1 variables will be covered .

3.3.2 K-means Clustering

The K-means algorithm assigns each point to the cluster whose distance (also called centroid) is the nearest to the region attribute (A11). The center is the average of all the points in the cluster, that is, its coordinates are the arithmetic mean for each dimension separately over all the points in the cluster. For example, a data set has three dimensions and the cluster has two points: X = (x1, x2, x3) and Y = (y1, y2, y3).

Then, the centroid Z becomes

Z = (z1, z2, z3), where z1 = (x1 + y1)/2 and z2 = (x2 + y2)/ 2 and

z3 = (x3 + y3)/2.

The Algorithm steps are:

i. Choose the number of clusters, k.

ii. Randomly generates k clusters and determines the cluster centers, or directly generates k random points as cluster centers.

iii. Assign each point to the nearest cluster center.

iv. Recomputed the new cluster centers.

v. Repeat the two previous steps until some convergence criterion is met (usually that the assignment has not changed).

Based on the K-means Algorithm first randomly select any one of the posh areas from A11, A12, A13 and A14. After selection of the Posh areas from Area A1, checking eight spatial location variables of the particular posh area like how many small business available 200km (S1), Competitor branches in 200m buffer(S2), Working population in 200m travel zone(S3), etc., Table 3.1 shows the eight spatial location variables. Each spatial location variables having no. of n attributes.

The Attributes Denotes

Let $S1 = \{s11, s12, s13……s1n\}$, $S2 = \{s21, s23, s23……s2n\}$,

$S3 = \{s31, s32, s33 ….s3n\}$, $S4 = \{s41, s42, s43 ….s4n\}$,

$S5 = \{s51, s52, s53 ….s5n\}$, $S6 = \{s61, s62, s63 ….s6n\}$,

$S7 = \{s71, s72, s73 ….s7n\}$, $S8 = \{s81, s82, s83 ….s8n\}$,

Each location variables must belong to the region of A1, A2, A3, and A4.so

Let $A1 \wedge (A11,A12,A13,A14) \wedge A11 \wedge (s1,s2,s3,s4,s5,s6,s7,s8) \vee$

$A12 \wedge (s1, s2, s3, s4, s5, s6, s7, s8) \vee$

$A13 \wedge (s1, s2, s3, s4, s5, s6, s7, s8) \vee$

$A14 \wedge (s1, s2, s3, s4, s5, s6, s7, s8) \vee$

Distance Relation (Euclidian distance) between the two spatial variables

P1 ^ {S11, S12, S13, S14, S21, S22, S23, S41, S51, S52}

Distance Relation is

A11 (S11) = $\sqrt{((x1 - x2)^2 + (y1 - y2)^2)}$.

P1 & S11 = (100,100), (300,100) => 200

P1 & S12 = (100,100), (500, 50) => 287

P1 & S13 = (100,100), (100,150) => 50

P1 & S14 = (100,100), (50, 75) => 56

P1 &S21 = (100,100), (300,275) =>265

P1 & S22 = (100,100), (150, 75) =>56

P1 & S23 = (100,100), (300,700) =>632

P1 & S41 = (100,100), (800,200) =>707

P1 & S51 = (100,100), (25, 25) => 106

P2 & S52 = (100,100), (75,400) =>301

Analyzing the distance point with the help of distance relation (Euclidian distance) between the posh area P1 and the all the spatial variables of S1,S2,S3,S4,S5,S6,S7,S8 of the posh area . Clustering techniques clustering the similarity of the spatial variable and posh area. K-means Algorithm continue the process until cover all the similarity of the spatial variable in Area1(Bhagalpur)

3.3.3 Clustering Techniques

Posh area (P&T Colony) of P1 location point is (100,100) and Euclidian distance value of spatial location variables S13 is 50. So the spatial variable

S13 is very nearby to the posh area P1. Next finding the K-mean value of Posh Area P1 and Spatial location variables of S13 value is 75. After that analyzing which S1 spatial variables is very close to K-mean value. Euclidian distance value of spatial location variables S14 value is 78. S14 is the very nearby next spatial variables of K-mean value.

Cluster 1 (Spatial location of S1 variable)

Cluster 1(C1)	=	(P1 & S13)
	=	(100, 50)
K-mean value 1	=	75
Cluster 1 (C1)	=	(P1 & S14)
	=	(100, 56)
k-mean value 2	=	78
Cluster 1	=	(P2 & S22)
	=	(100,56) = 78

Cluster 1 (C1) = {P1, S13, S14, S22}

Cluster 2 (Spatial location variable)

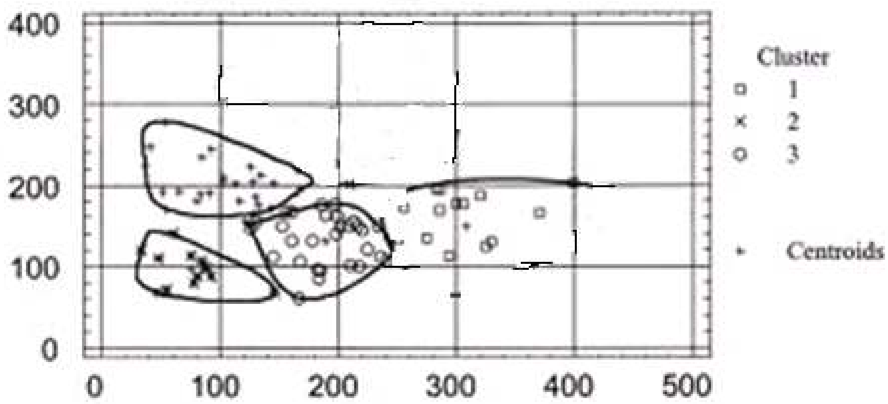
Cluster 2 (C2)	=	(P1) & (S11)
	=	(100), (200)
K-mean value1	=	150
Cluster 2 (C2)	=	(P2 & S21)
	=	(100,265)
k-mean value 2	=	182
Cluster 2 (C2)	=	(P1 & S51)
	=	(100,106)
New K-mean value 2	=	103

Cluster C2 = {P1, S11, S21, S51}

Using K-means algorithm and clustering techniques clustering the spatial variable based on the value of cluster 1, cluster 2. In cluster 1 having the similarity of the s11 variable (Small business), cluster2 having the similarity of the S12 variable (Competitors branch around 200m buffer) and clustering eight spatial variable of area1(Bhagalpur).

Initial cluster

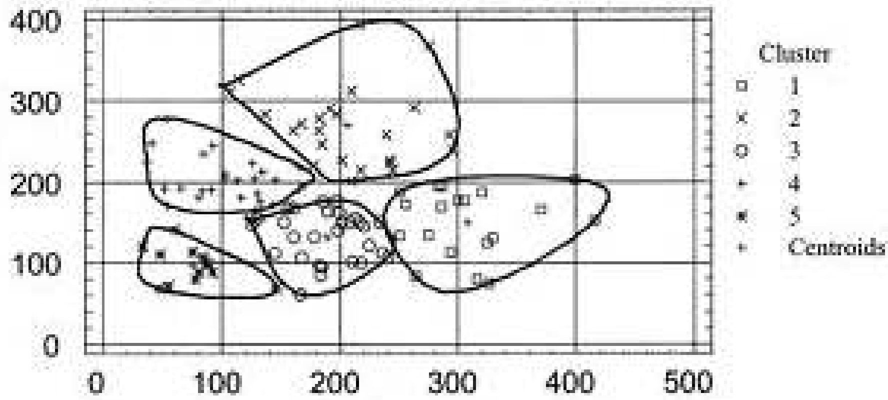
Spatial variable



Location of spatial variable

Final cluster

Spatial variable



Location of spatial variable

Figure 3.6: Result of the K-mean Algorithm

Figure 3.6 shows the result of initial and final cluster of k-mean .Based on the algorithm choosing the centroids point clustering the similarity of the spatial variable. Using distance relation (Euclidian distance) and clustering techniques with K-means algorithm five types of clusters are available in the area1 (Bhagalpur).

3.4 K-MEDOID CLUSTERING

The *k*-medoids algorithm is a clustering algorithm related to the *k*-means algorithm. *K-medoids* algorithms are partition (breaking the dataset up into groups) and attempt to minimize the distance between points labeled to be in a cluster and a point designated as the center of that cluster. In contrast to the *k*-means algorithm, *k*-medoids chooses data points as centers and works with an arbitrary matrix of distances between data points.

K-medoid is a classical partitioning technique of clustering that clusters the data set of *n* objects into *k* clusters known *a priori*. It is more robust to noise and outliers as compared to *k*-means because it minimizes a sum of pair wise dissimilarities instead of a sum of squared Euclidean distances.

A medoid can be defined as the object of a cluster, whose average dissimilarity to all the objects in the cluster is minimal i.e. it is a most centrally located point in the cluster.

k-medoid clustering is the **Partitioning Around Medoids (PAM)**

Algorithm: k-Medoids

Input: The number of clusters k and a database containing n objects.

Output: A set of k clusters that minimizes the sum of the dissimilarities of all the objects to their nearest medoid.

Method: Arbitrarily choose k objects as the initial medoids;

- Repeat
- Assign each remaining object to the cluster with the nearest medoid
- Randomly select a non medoid object, random
- Compute the total cost,S of swapping oj with orandom

- If S < 0 then swap oj with orandom to form the new set of k medoid
- Until no change

Decision maker analyzing eight spatial variables of like Small business in 200m travel zone (S1), Competitor branches in 200m buffer (S2), Working population in 200m travel zone (S3). Table 3.2 shows eight spatial location variables.

It denotes:

Let P1 = (S1, S2, S3, S4, S5, S6, S7, and S8)

So

A1 ^ (P1, P2, P3, P4, P5) and P1 ^ (S1, S2, S3, S4, S5, S6, S7, S8)

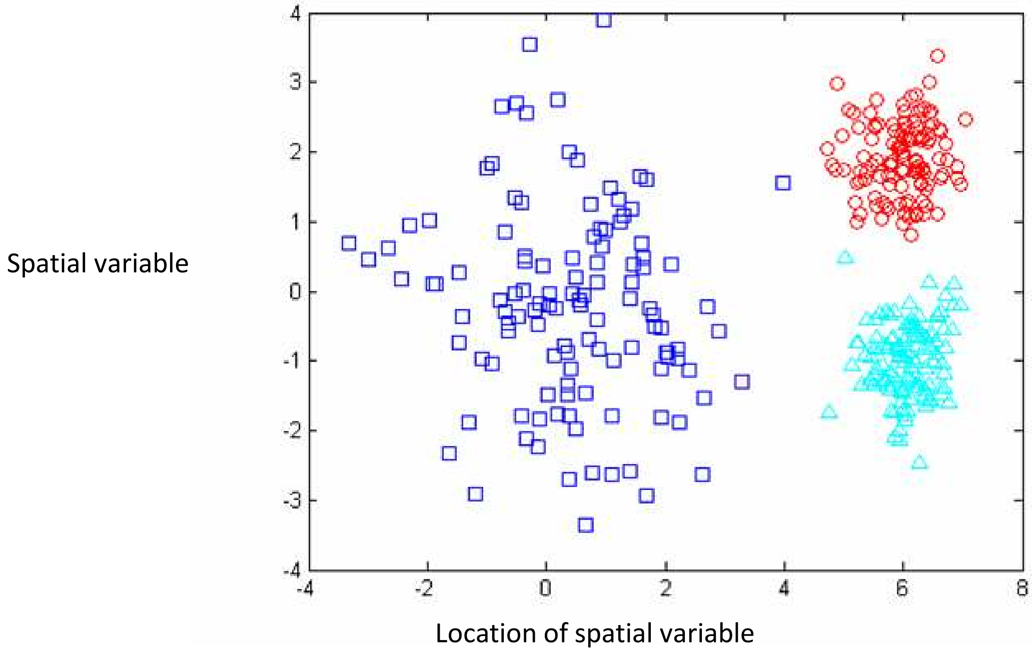


Figure 3.7: spatial variables of Posh Area1

Figure 3.7 shows how many spatial variables are there in the posh area1. The objective of K-medoid Algorithm is to find a non-overlapping set of clusters such that each cluster has a most representative point, i.e., a point that is most centrally located with respect to some measure, e.g., distance. These representative points are called medoids.

After analyzing of spatial variables in Posh area1 finding the distance between the two spatial variables. With the help of distance relation find out which spatial variables are very nearby to the medoids.

Decision maker first clustering the spatial variables of small business (S1) . So randomly select any one the variables from the dataset of S1 variables. The variables called the centre medoid of the cluster. After finding the set of medoids, each object of the data set is assigned to the nearest medoid. It 's clustering the variable in cluster 1. The process will continue until clustering the similarity of all the spatial variables like competitor branches in 200m buffer, branches at shopping centers…, table- 3.2 shows the eight spatial location variables.

Posh Area (P&T colony) P1 having 6 small business (S1) in 200m travel zone, 3 competitor branches (S2) in 200m buffer, 62% of working population (S3), 1% same branch in around 200m , Table 3.3 shows the contribution of the spatial variables in Posh area P1.

It denotes

Let S1 = (S11, S12, S13, S14, S15, S16)

A1 ^ P1 ^ {S11, S12, S13, S14, S15, S16}

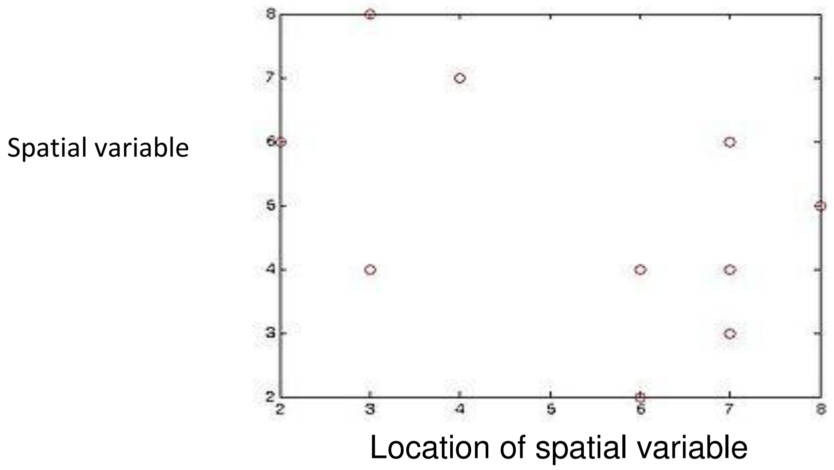
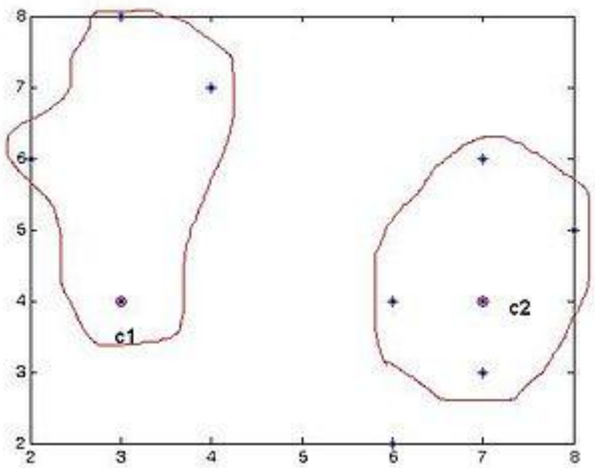


Figure 3.8: Spatial Variables of S1 in Posh Area P1.

Figure 3.8 having no.of spatial variable and randomly select any one the variables from S1 Dataset (S11, S12, S13, S14, S15, S16) the variable is called the representative points of the cluster C1. After finding the center medoid each variable is assigned to the center medoid. Which variable dissimilarity is very minimum to the center of the medoid the variable is assigned to the cluster C1.

Spatial variable



Location of spatial variable

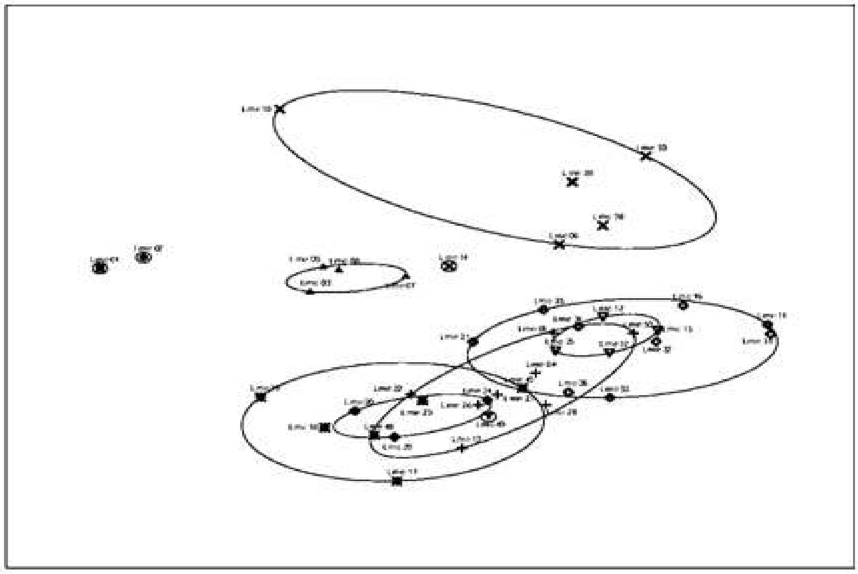
Figure 3.9: clustering of the spatial variable S1

3.4.1 Clustering Techniques and K-medoids Algorithm.

Posh area (P&T Colony) of P1 location point is (100,100) and Euclidian distance value of spatial location variables S13 is 50. In figure 3.9 shows the clustering of the spatial variable S13 is very nearby to the posh area P1. Next randomly select representative point of the each cluster. The representative point minimize the dissimilarity of the all the spatial variables of the posh area (Bhagalpur) P1.representative point value of Posh Area P1 and Spatial location variables of S13 value is 75. After that analyzing which S1 spatial variables is very close to centroids of the cluster 1. Euclidian distance value of spatial location variables S14 value is 78. S14 is next very to the representative point of cluster1.

Final cluster of the K-medoids

Spatial variable



Location of spatial variable

Figure 3.10: Result of the K-medoids Algorithm

In figure 3.10 shows the clustering of the spatial variable S13 is very nearby to the posh areas of area1 using K-medoids algorithm.

3.5 CLARANS

K-medoids partitioning algorithm works effectively for small data sets, but does not scale well for large data sets. To deal with larger data sets, a sampling-based method, called Clara (clustering large applications) can be used. The idea behind CLARA is as follows:

Instead of taking the whole set of data, a small portion of the actual data is chosen as a representative of the data. Medoids are then chosen from this sample partitioning Around Medoids. If the sample is selected in a fairly random manner, it should closely represent the original data set. The representative objects (medoids) chosen will likely be similar to those that would have been chosen from the whole data set. Clara draws multiple samples of the data set, applies PAM on each sample, and returns its best clustering as the output. The effectiveness of CLARA depends on the sample size.

PAM searches for the best k medoids among a given data set, whereas CLARA searches for the best k medoids among the selected sample for the data set. CLARA cannot find the best clustering if any sampled medoid is not among the best k medoids. A k-medoids type algorithm called CLARANS (Clustering Large Applications based upon Randomized Search) was proposed that combines both sampling technique with PAM.

CLARA has a fixed sample with some randomness in each step of the search; CLARANS draws a sample with some randomness in each step of the search. The clustering process can be presented as searching a graph where every node is a potential solution, that is, a set of k medoids.

The clustering obtained after replacing a single medoid is called the neighbor of the current clustering. If a better neighbor is found, CLARANS moves to the neighbor's node and the process starts again; otherwise the current clustering produces a local optimum.

3.5.1 Clustering Techniques and CLARANS

Area (A1) having no of P attributes. The P attributes represent the posh areas of the Bhagalpur (A1) regions. The attributes donates

Let A1 = {A11, A12, A13, A14…..A1m},

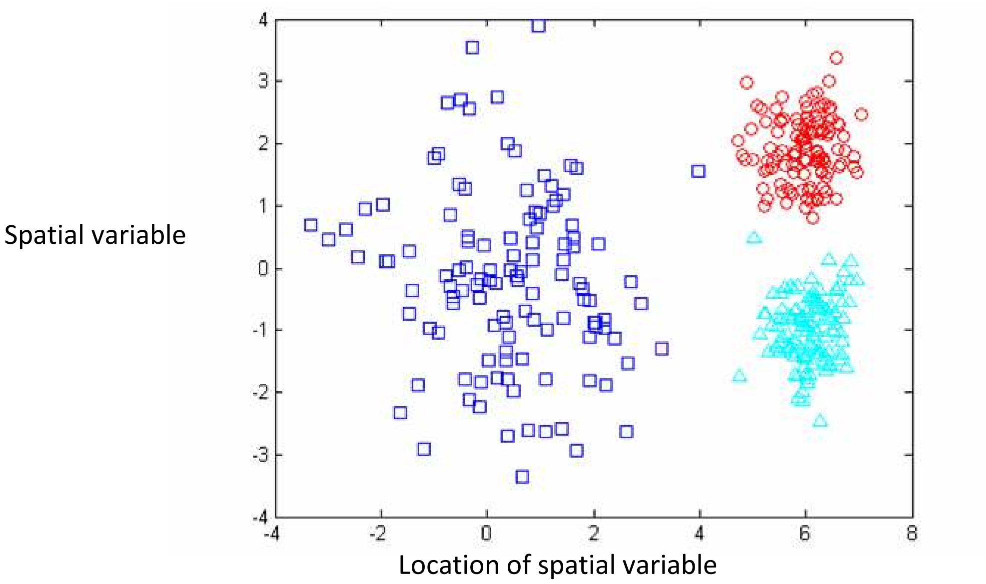


Figure 3.11: Spatial Variables of posh Area (P&T colony)

Figure 3.11 shows the posh area P1 having no. Of spatial variable S1, S2, S3, S4, S5, S6, S7, and S8. Clarans algorithm deals the large database. Clarans algorithm taking sample data from the P1 (P&T colony) variable instead of taking the whole data set. The sample data represent the whole data set of the P1 variable. Randomly select any of the variables from the posh area P1.

It is closely represent the original data set. Clustering the variable which is closely related to the sample data.

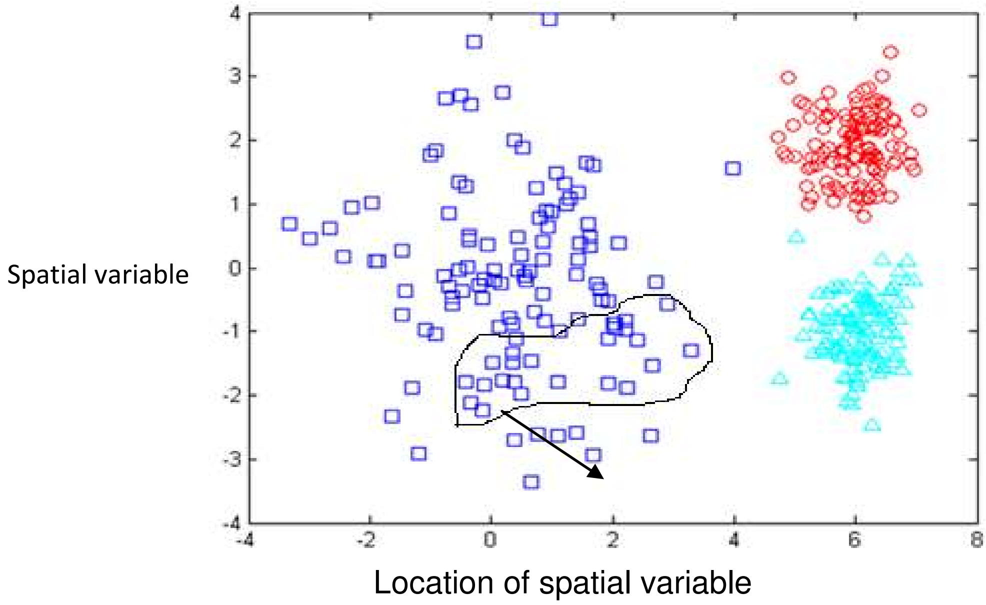


Figure 3.12: Sample dataset of S1 variable

Sample data set of the spatial location variable, applies PAM on sample data, and its returns best clustering as the output. Figure 3.12 the effectiveness of CLARA depends on the sample size. Again the CLARANS algorithm taking the next sample of the data from the P1 variable and applies the PAM on the sample data it's clustering the variable which is similar to the sample data. CLARANS multiple samples of the data set, applies PAM on each sample, and returns its best clustering as the output.

The process will continue cover all the similarity of spatial variable of the Posh area P1 (P&T colony). The effectiveness of CLARA depends on the sample size.

3.6 WEIGHTAGE CALCULATION OF NON-SPATIAL DATA

The Eight Non-spatial data variables are playing most important role in the banking sector. The variables are like banking population, Population growth rate, Average annual family income, etc., Table3.1 shows (Eight Non-spatial data variable). The Non-spatial variables will vary branch to branch.

Although the results of the spatial location variables, SDSS are encouraging, further research is required into the weights assigned to the Non-spatial data. The criteria selected should be evaluated on the basis of the importance of their contribution, validity, and redundancy. Especially into the appropriateness of the 17 criteria used to calculating the weight age of the Non-spatial data. A number of methods may be used to evaluate the criteria: In this research, statistical methods and visualization have been used. It can be expected that the relative importance of the contribution of the criteria will vary from area to area.

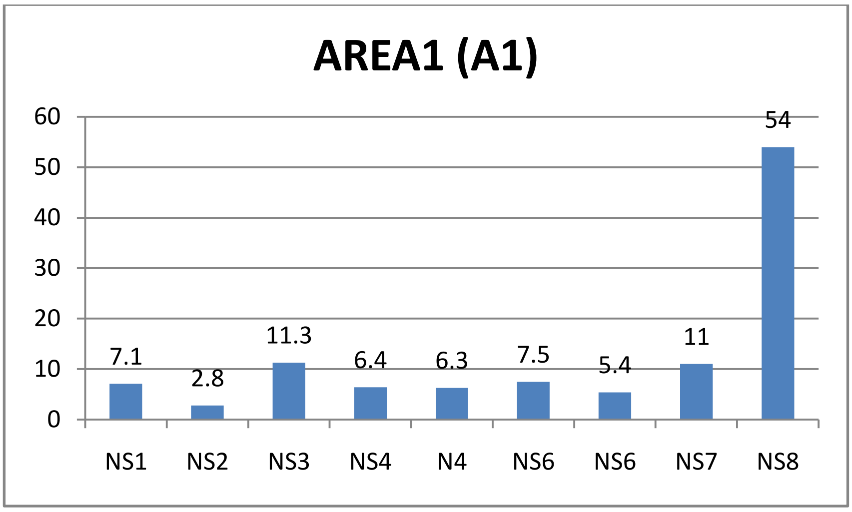
'Analysis by contribution' is a useful way of evaluating the contribution of the variable to make the decision score for each area. An example of the 'Contributions by variable' output is separately calculated for four different areas using a graphical method shown in Table 3.6.1.

TABLE- 3.6.1 (Non-Spatial Data Contribution of Area (A1))

Non-Spatial data variable	Contribution (%)
Banking population (age 15 & over) NS1	7.1
Population growth rate (NS2)	2.8
Average annual family income (NS3)	11.3
Average age (NS4)	6.4
Aged population (over 55 years) NS5	6.3
Employed people (NS6)	7.5
Total number of (small) businesses (NS6)	5.4
Competitors’ branches (NS7)	11
Other variables	54

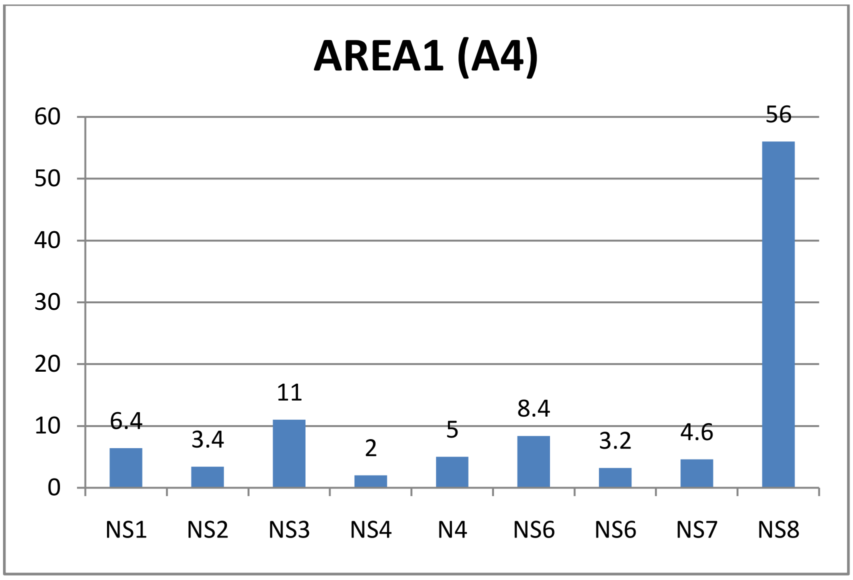
This presents relative contribution made by each of the Non-spatial variable to the total decision score for each area. For example, the variable NS3 (Average annual family income) and NS6 (Employed people) make the biggest contributions to the decision score for the Area1 (A1). Figure 3.13 it shown the contribution of the Non-Spatial data for Area1.

Non-Spatial variable



Values of Non-spatial variable

Figure 3.13(a): Contribution of Non-Spatial Data in Area 1



Values of Non-spatial variable

Figure 3.13(b): Contribution of Non-Spatial Data in Area 4

In any decision situation, however, some variable will make a larger contribution to the Decision scores than others. Determining which variable contribute most to the decision scores and which contribute the least, as well how these variable relate to the spatial location of branches The importance of each variable to the resulting preference ranking is indicated by its overall contribution to the decision score. The percentage total contribution of the seven variables to the decision scores for the four Area. Contribution of the Non-spatial data for four areas is shown in Table , which shows that five variables (NS1,NS3,NS5, and NS6) contribute more than 60% to the total decision scores,

Four criteria (NS2, NS4, and NS7) make a very small contribution (together, less than 3.5%). Another useful method for exploring the contribution made by the criteria in determining the decision scores is to identify those variable that play a very small role in determining the decision score and hence the preference ranking.

TABLE-3.6.2 Contributions of Non-Spatial Data for Four Areas

Non-Spatial Data Attributes	**AREA 1 (%)**	**AREA 2 (%)**	**AREA 3 (%)**	**AREA 4 (%)**
Banking population (age 15 & over)	7.1	6.4	8.9	12
Population growth rate	2.8	3.4	5.1	3
Average annual family income	11.3	11	11	9.6
Average age	4.3	2.0	3.8	1.8
Aged population (over 55 years)	6.3	5	6	9.2
Employed people	7.5	8.4	8.8	7.1
Total number of (small) businesses	5.4	3.2	2.1	1.1
Competitors' branches	1.3	4.6	5	3.1
Other variables	54	56	49.3	53.1

3.7 INTEGRATION OF SPATIAL AND NON-SPATIAL DATA

Integration of Spatial and Non-Spatial data is playing very important role in selection of location for opening new bank branch in the banking sector. Collect the spatial and non-spatial related variables in all the four areas. Next find out posh areas of area1, area 2, area 3 and area4. Using distance relation and clustering techniques identify the right point all the areas. The point should cover 80% of spatial location variables. After that analyze Non-spatial variables of the all the area of the right point. subsequently separately identify how the non-spatial data are performing all the four areas. The four criteria should be followed.

The criteria's are

1. Very strong variables
2. Strong variables
3. Medium variables.
4. Low variables.

The 4 criteria are used all the four area. Table -4 shown Contribution of Non-spatial data in four different areas.

Area 1 having 3 % of very strong variable, 1% of strong variable ,then 2 % medium variable and low variable .

Area 2 performing 2% very strong variable, 2% strong variable, 1% medium variable and low variable is 3%.

Area 3 is 3% very strong variables, strong variable is 2% then Medium variable is 0% and low variable is 2%.

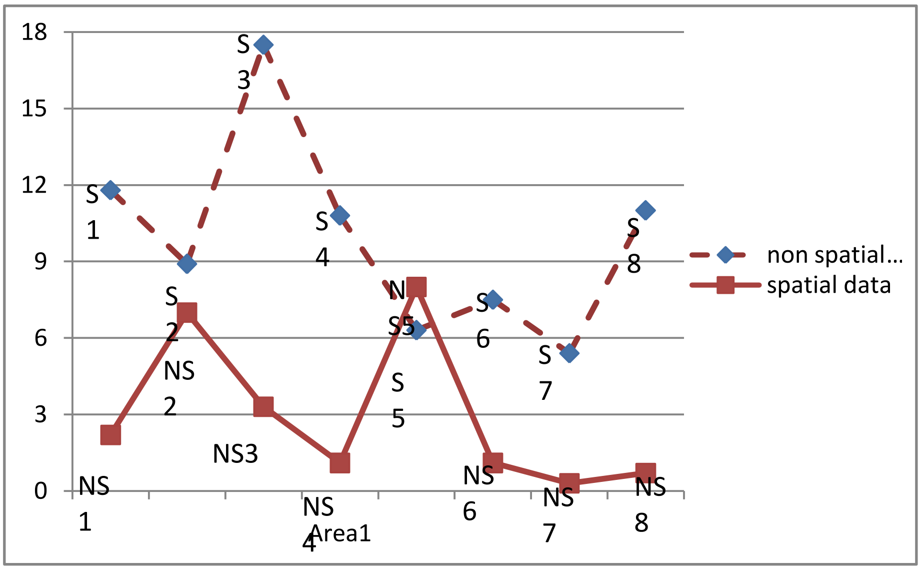
Area 4 playing 4% very strong variables, strong and medium variable is 0% and low variable is 4%.

Table –3.7.1 Variable Performance in 4 Different Areas

Area	Very strong variable (%)	Strong Variable (%)	Medium Variable (%)	Low Variable (%)
Area 1	30	20	28	22
Area 2	10	20	44	26
Area 3	20	30	25	25
Area 4	50	35	10	5

Based on the spatial and non- spatial data decision maker can choose the area to open a new bank branch location. Table 3.7.1 shows Using 3 different algorithm K-means, k-medoids and CLARANS algorithm analyzing the spatial variable of the four areas. After that analyzing the non- spatial variable of the different areas . so the decision system inform to the bank manager which area is very strong in the spatial and non-spatial data.

X - axis



Y - axis

Figure 3.14(a): Integration of Spatial and Non-Spatial data Area1

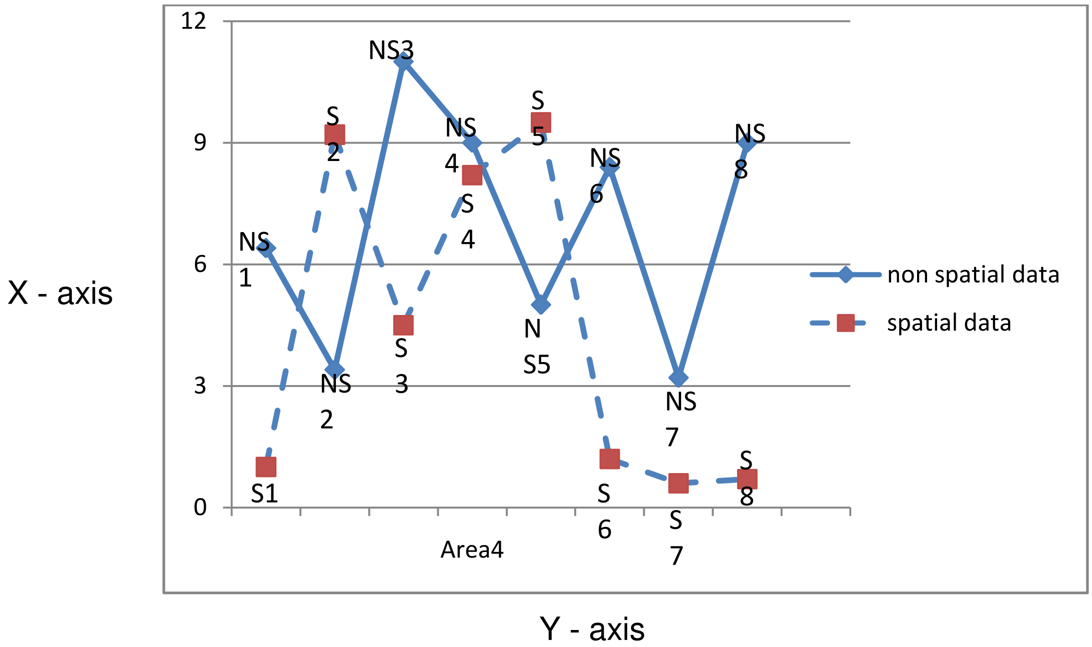


Figure 3.14(b): Integration of Spatial and Non-Spatial data Area4

Analyzing of spatial and non-spatial data of the four different areas. Figure 3.14(a) shows the spatial and non-spatial data of the area 1. Area 1 is not very strong in the area1 (Bhagalpur). Area 2 (Budigere) is very strong in spatial data, but medium variable in non-spatial data. Area 3 (kaglipur) is medium in spatial data and strong in non-spatial data. Figure 3.14(b) shows the area 4 (Hosur). Area4 has very strong in spatial and non-spatial data. So the decision making system suggest the decision maker can open a new bank branch in Area 4 (Hosur)

CHAPTER 4

IMPLEMENTATION OF PROPOSED METHOD

Decision support systems represent the role of computer information systems in the decision making process. For the purposes of this research, the term SDSS will include clustering techniques is used for grouping the similar Geo-graphical information data. It will also incorporate understanding of the decision making process involved in solving a specific bank location problem for integration of spatial and non-spatial data.

Conceptually, a SDSS can be thought of as providing an integrated set of flexible capabilities; the implementation of such a system can be achieved using a set of linked software modules. There are several components that form the core of a SDSS, Typically, a SDSS contains a spatial information system (GIS) integrated with a modeling system. Specifically, the system includes a geo-referenced database, analytical tools, and display and reporting capabilities.

Within a SDSS there are three interface components, a

- User Interface
- System Interface.
- Model base

USER INTERFACE

The user interface is often the most important component in a system's development and perceived success. It provides users with access to the databases, modelbase (analytical routines), and graphical and report generation. System designers may provide different levels of access for different levels of users. Bank management, for example, may require access to all databases and analytical routines, while clerical personnel have restricted access. The user interface also enables users to generate output (maps, charts, etc.) directly from the database management system (DBMS) or from the results of analyses which may be stored in the DBMS.

SYSTEM INTERFACE

The system interface expedites the transfer of data between the DBMS and the model base and contains routines invoked automatically during SDSS execution. A model base, or a core of analytical routines, is integrated into the SDSS through the user and system interfaces.

MODEL BASE

The model base gives decision makers access to a variety of models and assist them in decision making. The model base can include the model base management software (MBMS) that coordinates the use of models in a DSS. This component can be connected to external storage of data.

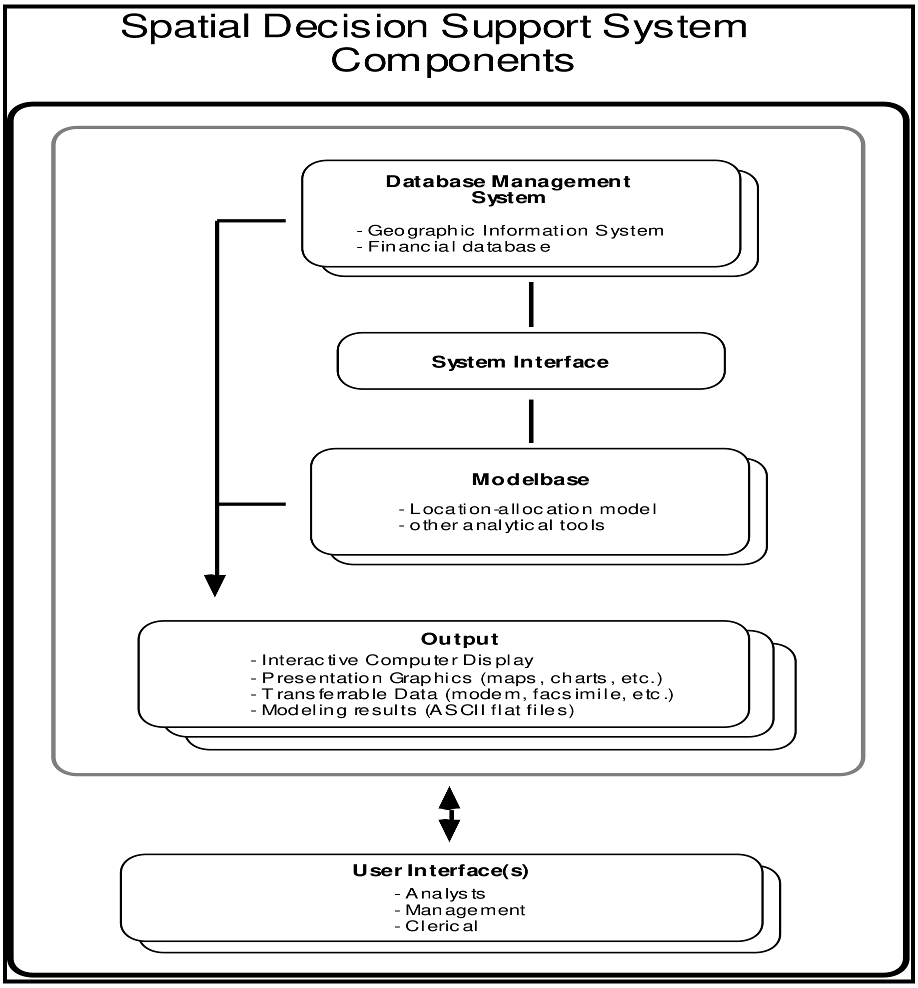


Figure 4.1: Interaction of SDSS Components.

4.1 THE IMPLEMENTATION ENVIRONMENT

The following environment is used for implementation. VB.Net is used to on server side scripting language.

Language	:	VB.Net
Data base	:	MYSQL (5.0)
Hardware	:	Processor i5 , 4 GB RAM ,
Algorithm	:	clustering techniques and K-means, K-medoids and CLARANS Algorithm

Advantages of VB.NET

VB.NET is totally object oriented. VB.NET provides managed code execution that runs under the Common Language Runtime (CLR), resulting in robust, stable and secure applications.

The .NET framework comes with ADO.NET, which follows the disconnected paradigm, i.e. once the required records are fetched the connection no longer exists. It also retrieves the records that are expected to be accessed in the immediate future. This enhances Scalability of the application to a great extent.

VB.NET uses XML to transfer data between the various layers in the DNA Architecture i.e. data are passed as simple text strings.

Error handling has changed in VB.NET. A new Try-Catch-Finally block has been introduced to handle errors and exceptions as a unit, allowing appropriate action to be taken at the place the error occurred thus discouraging the use of ON ERROR GOTO statement.

MYSQL

MYSQL has proved itself to be a fast, reliable and cost effective.

MYSQL is an open source database system which means that anyone can use it for free.

A lot of people around the globe are continuously developing new modules for integration with MYSQL.

This relational database system is free so it reduces the cost of overall database solution for small businesses and companies.

MYSQL as a relational database is secure as all access passwords are stored in an encrypted format restricting any unauthorized access to the system. It also encrypts the transactions so eavesdroppers and data harvest tools cannot replicate or regenerate the database transactions once they are processed.

CHAPTER 5

RESULT ANALYSIS

5.1 CLUSTERING TECHNIQUES AND K-MEAN ALGORITHM

In this study, the k-Means algorithm is explained with an example first, followed by k-Medoids and Clearns algorithm. The experimental results are discussed for the K-Means algorithm. Figure 4.1 shows the number of clusters and data points is given by the user during the execution of the program. The number of data points is 1000 and the number of clusters given by the user is 10 (k = 10). The algorithm is repeated 1000 times to get efficient output. The cluster centers (centroids) are calculated for each cluster by its mean value and clusters are formed depending upon the distance between data points. For different input data points, the algorithm gives different types of outputs.

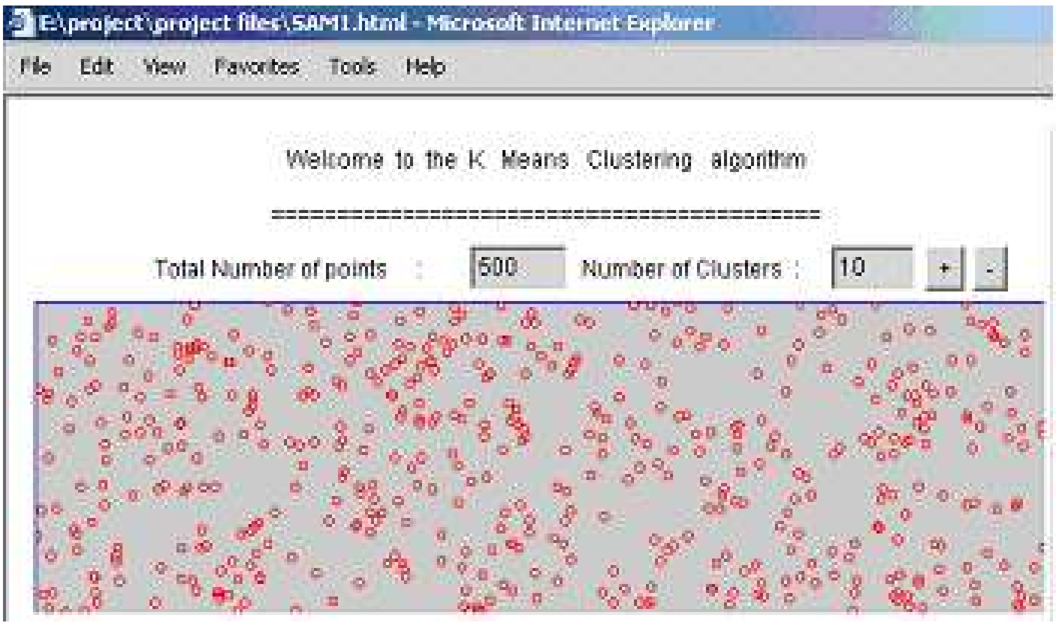
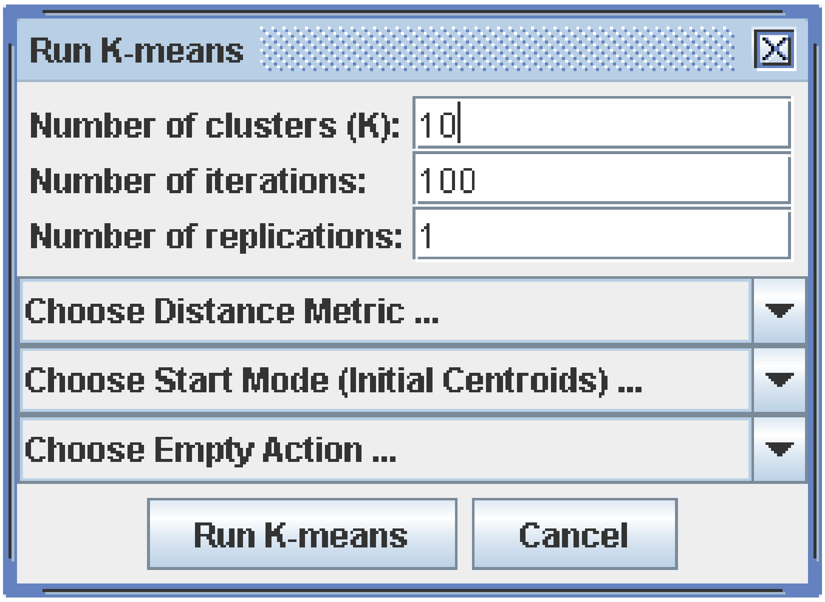


Figure 5.1: output of K-mean Algorithm



Number of random data points -> 500

Number of Clusters -> 10

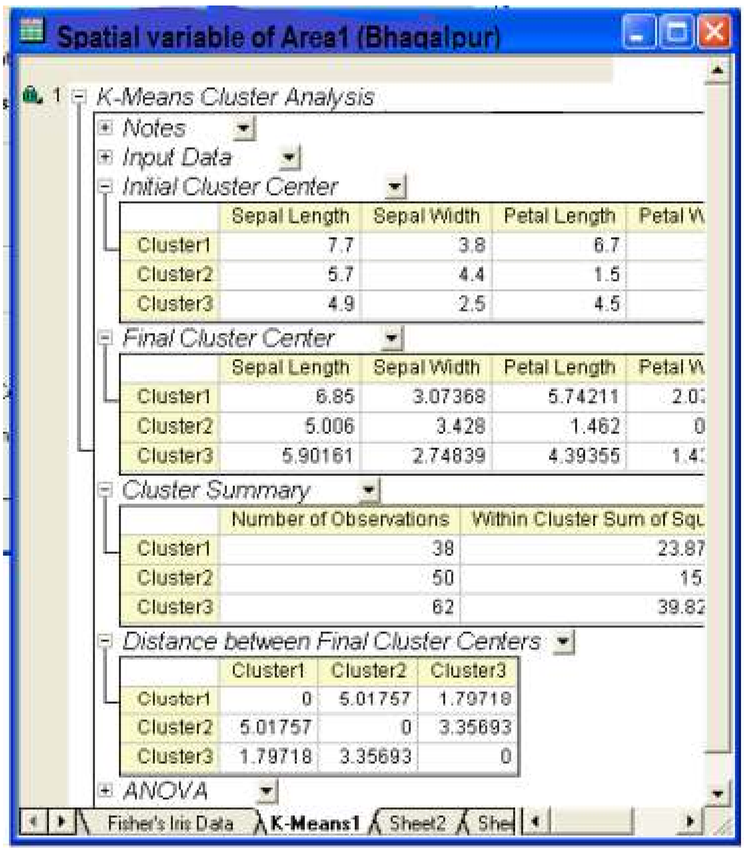


Figure 5.2: Cluster values of K-means

5.2 CLUSTERING TECHNIQUES AND K- MEDIODS ALGORITHM

Hundred random data points are input to this algorithm. The number of clusters and data points given by the user. The k-medoids algorithm is repeated for thousand times to get efficient output. Figure 4.3 shows *K-medoids* algorithms to minimize the distance between points labeled to be in a cluster and a point designated as the center of that cluster. In contrast to the *k*-means algorithm, *k*-medoids chooses data points as centers and works with an arbitrary matrix of distances between data points.

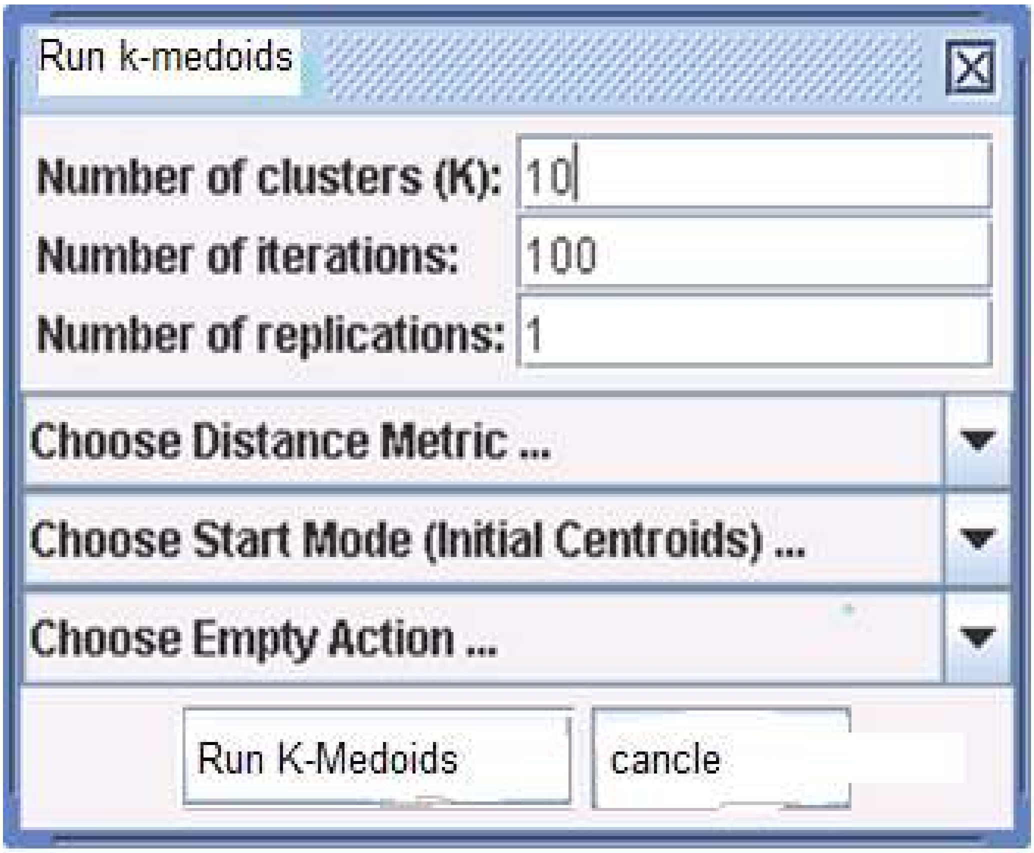


Figure 5.3: Output of K-medoids

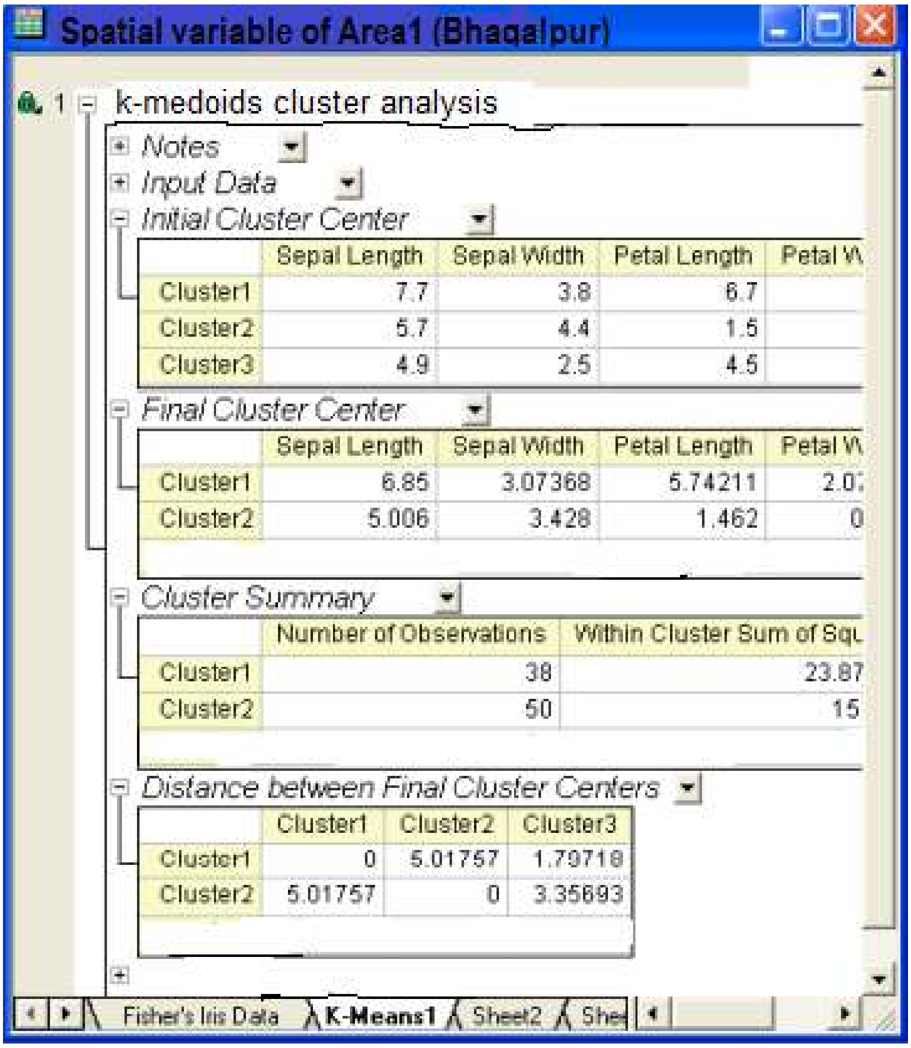


Figure 5.4: Cluster values of K-medoids

Input data and no. Of clusters are given by the user. The K-medoids algorithm randomly choosing the initial cluster form the data points and the initial cluster is the representative point of the each cluster. In figure 4.4 the cluster performed based on the principle of minimizing the sum of the dissimilarities between each object and its corresponding reference point. K-Mediods clustering algorithms is to find no.of clusters in n objects by first arbitrarily finding a representative object (the Medoids) for each cluster. Each remaining object is clustered with the Medoid to which it is the most similar.

5.3 CLUSTERING TECHNIQUES AND CLARANS

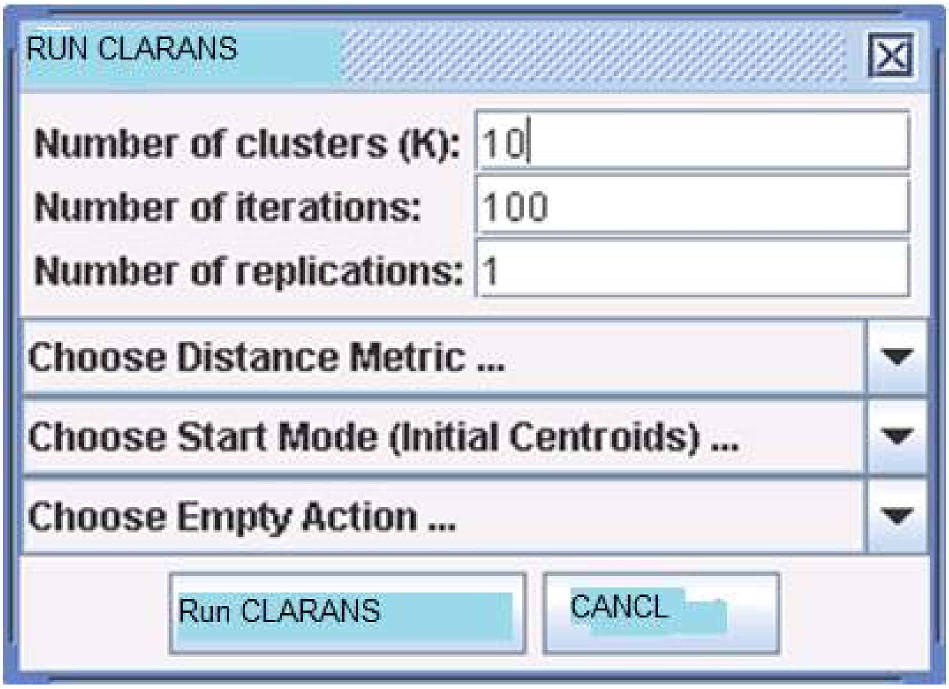


Figure 5.5: Output of CLARANS

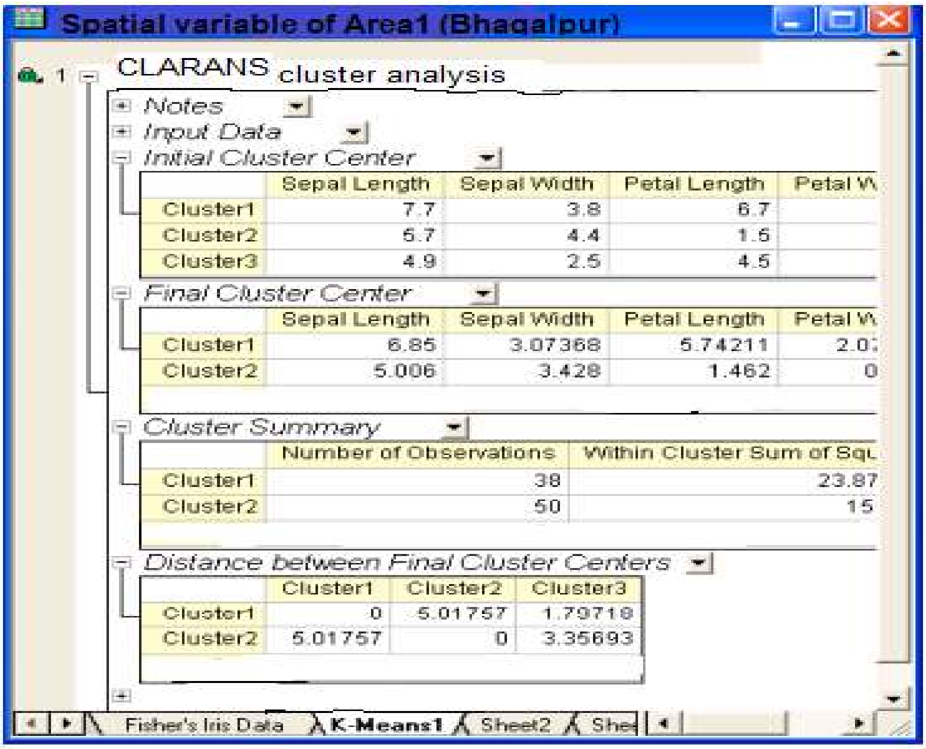


Figure 5.6: Cluster values of CLARANS

5.4 FINAL RESULT ANALYSIS OF FOUR AREAS

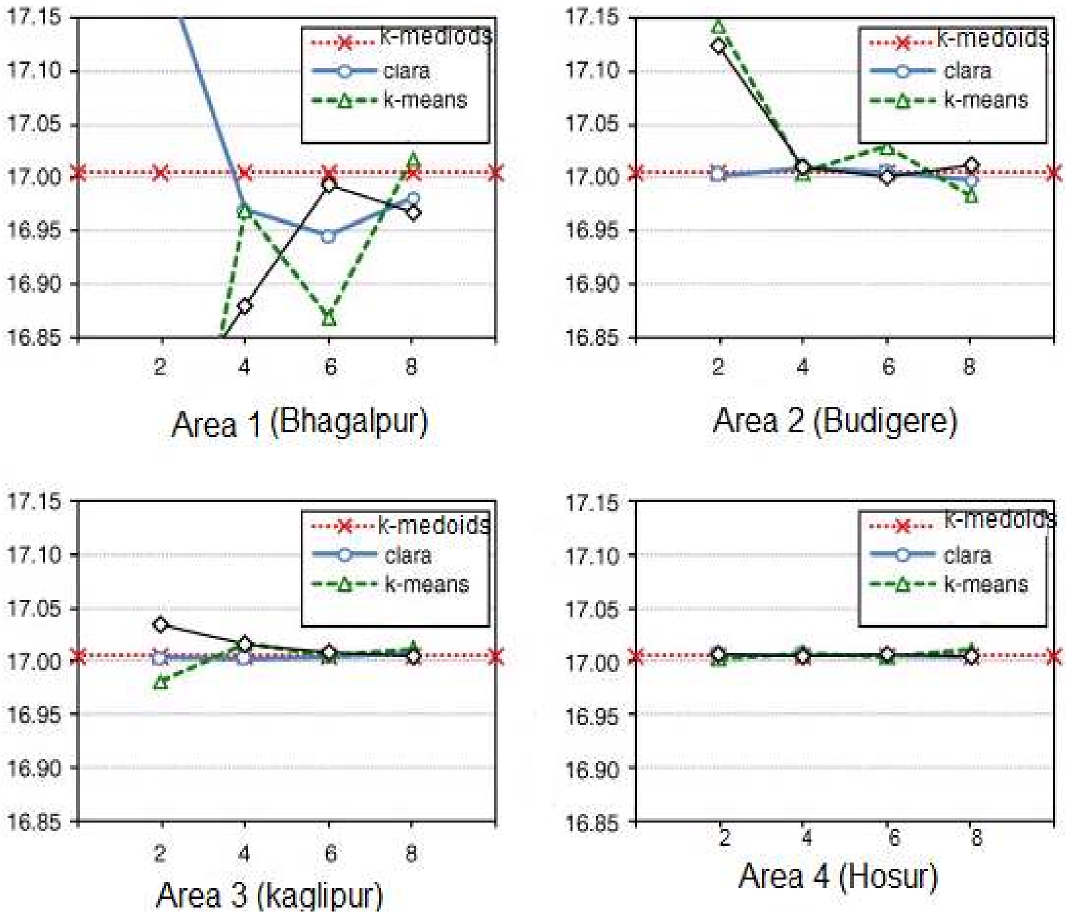


Figure 5.7: Final Results of 4 Areas

Decision maker analyzing spatial location for opening new bank branch for four different areas. Figure 4.7 represents three different algorithms k-means, k- medoids and Clarans are performing to find out the correct location for opening a new branch. In Area 1 the three algorithms are performing in different ways. So spatial and non-spatial data are not very strong in area 1.similarly area2 and area3. But area4 the three algorithms are performing in some way so the spatial and non-spatial data are very strong in area4.So area 4 is the correct location point to opening a new bank branch.

CHAPTER 6

CONCLUSION

Decision support system (SDSS) aims to help Identifying location for opening a new bank branch. Through the spatial database it is possible gathering the needed data for spatial analysis, storing the order of importance of spatial variables, and viewing a combination of the different data, with the help of k-means, k-medoids and CLARANS algorithms find out distance between the two spatial variables and using clustering techniques cluster the variables which is most similar to one another. The non-spatial data are stored in separate database and weightage is calculated based on important of the variables. Finally integrating spatial and non-spatial data. The location having very strong in spatial and non-spatial data so the bank manager can select the location for opening new bank branch. At the heart of our framework lies a mapping of pairs of spatial database and non-spatial database to identify the distance-score space.

CHAPTER 7

SCOPE FOR FUTURE WORK

Clustering Algorithm with Decision support system has presented a novel and interesting problem in the fields of data mining. In the discussion part of the previous chapter, showed that this problem in decision support system and that much work can be done to improve the application of Decision support system. There are other interesting constraints that require changing the nature and the emphasis of the clustering algorithms.

7.1 REAL WORLD DATA

Aware that the Clustering Algorithm is currently tested with synthetic data, not Real world data. Due to the limitation of time, leave this task to the future study.

7.2 OTHER CLUSTERING ALGORITHMS

In this research using there are three major categories for spatial-clustering Algorithms. Algorithms designed to solve the problem of decision support system. Clustering algorithms as their foundation. For example, the research could develop the decision support system for bank branch location analysis. These three-algorithm families might Serve various purposes and have different strengths. It will be interesting to observe future research of the decision support problem.

7.3 OTHER CONSTRAINT-CLUSTERING PROBLEMS

There remain many other constraint-clustering problems to be tackled. In real-world. The clustering task needs to comply with given rules and conditions and here summarize the constraint-clustering problem into five categories:

1. Constraint on object type

Object constraints are rules and limitations applied to data objects that are being clustered. For example, one could cluster the locations of all location of the nearby the bank branch in Bhagalpur which are priced higher than 1 million dollars. The result would be useful for high-end commercial product promotion.

2. Constraint involving physical obstacles

Spatial constraint is defined as the location probtem in this thesis. Most of the clustering algorithms today have ignored the fact that physical obstacles exist in the real world. The application of this clustering method would be useful for identifying the best location for any type of resource centre.

3. Constraint on single cluster

Single cluster constraint refers to the rules and conditions applied to each of the resulting individual clusters. These can include the size, the area or other cluster characteristics. One example could be clustering the location and income of all Bhagalpur residents so that each of the resulting clusters yield combine early income of the family.

4. Constraints between cluster

Rules can be imposed on the relationship between any two or more resulting clusters. An example would Bhagalpur to compose clusters having a minimum distance from each other but still maintaining a certain level of density.

In this thesis, to solve the real-world problem, it must cater the clustering Algorithm to accommodate the given rules and conditions in order to achieve the optimal performance. Believe that further research in Constraint-clustering will enhance the applicability of existing spatial-clustering algorithms.

REFERENCES

[1] Guobin Li, “Research on Optimized Spatial Data Query Algorithm in the Spatial Database”, IEEE International Conference on Computer Science, 2009.

[2] Nihan Cinar, “A Decision Support Model for Bank Branch LocationSelection’’, International Journal of Human and Social Sciences, pp. 5-13, 2010.

[3] Sajeev B, Thangavel K , “Impact Analysis of Financial Inclusion through SHG Bank Linkage using Clustering Techniques”, International Journal of Research and Reviews in Computer Science IJRRCS, pp. 134-151, 2010.

[4] Dingming Wu, “Joint Top-K Spatial Keyword Query Processing”, IEEE International Conference 2008.

[5] Cebi.F, “A Decision Support Model for Location selection: Bank branch“ Management of Engineering & Technology, pp. 1069 – 1074, 2009.

[6] Shan Gao, “Flexible Support for Spatial Decision-Making”, IEEE 37th International Conference on system Science, 2004.

[7] Joao B, “Efficient processing of Top-k Spatial Preference Queries”, The 37th International Conference on Very large Databases, vol-4, 2011.

[8] Kido H, Yanagisawa Y, “ An anonymous communication technique using dummies for location-based services”, In *IEEE Conference on Pervasive Services*, pp. 88–97, 2005.

[9] Cong G, Jensen C, ”Efficient retrieval of the top-k most relevant spatial web objects”, In *Very large Database*, pp. 337–348, 2009.

[10] Burdziej J, “Multi-criteria spatial analysis of land accessibility for seismic operations”, *Annals of Geometrics*, pp. 23-32, 2009.

[11] Hong M, Koch C, “Rule-based multi-query optimization”, In *Extending Database Technology*, pp. 120–131, 2009.

[12] Marc V., et al “Distributed Ranking Methods for Geographic Information Retrieval”, 20th European Workshop on Computational Geometry, March 25-26, 2010.

[13] Armstrong M, Densham P, (1990) "Database Organization Alternatives for Decision support systems", International Journal of Geographic Information Systems 4. pp. 3-20, 2008.

[14] Samet H, Sankaranarayanan J, “Scalable network distance browsing in spatial databases”, In *Spatial Interest Group on Management Of Data*, pp. 43–54, 2008.

[15] Zhang D, Mondal A, ”Keyword search in spatial databases towards searching by document”, In *International Conference in Database Engineering*, pp. 688–699, 2009.

[16] Hariharan R, “Processing spatial-keyword queries in geographic information retrieval (GIR) systems”, In *SSDBM*, pp-16, 2007.

[17] Kaufman L, and Rousseeuw P.J,” *Finding groups in data: an introduction to cluster analysis”,* New York: Wiley, 1990.

[18] Agrawal R, “Database Mining: A Performance Perspective”, IEEE Transactions on Knowledge and Data Engineering, Vol. 5, No. 6, pp. 914-925, 2004.

[19] Armstrong M, “Database Organization Strategies for Decision support systems”, International Journal of Geographical Information Systems, Vol.4 (1), pp. 3-20, 2008.

[20] Boufounoua P, ” Evaluating bank branch location and Performance”, *European Journal of Operational Research Vol* -8, pp. 389-402, 2010.

APPENDICES

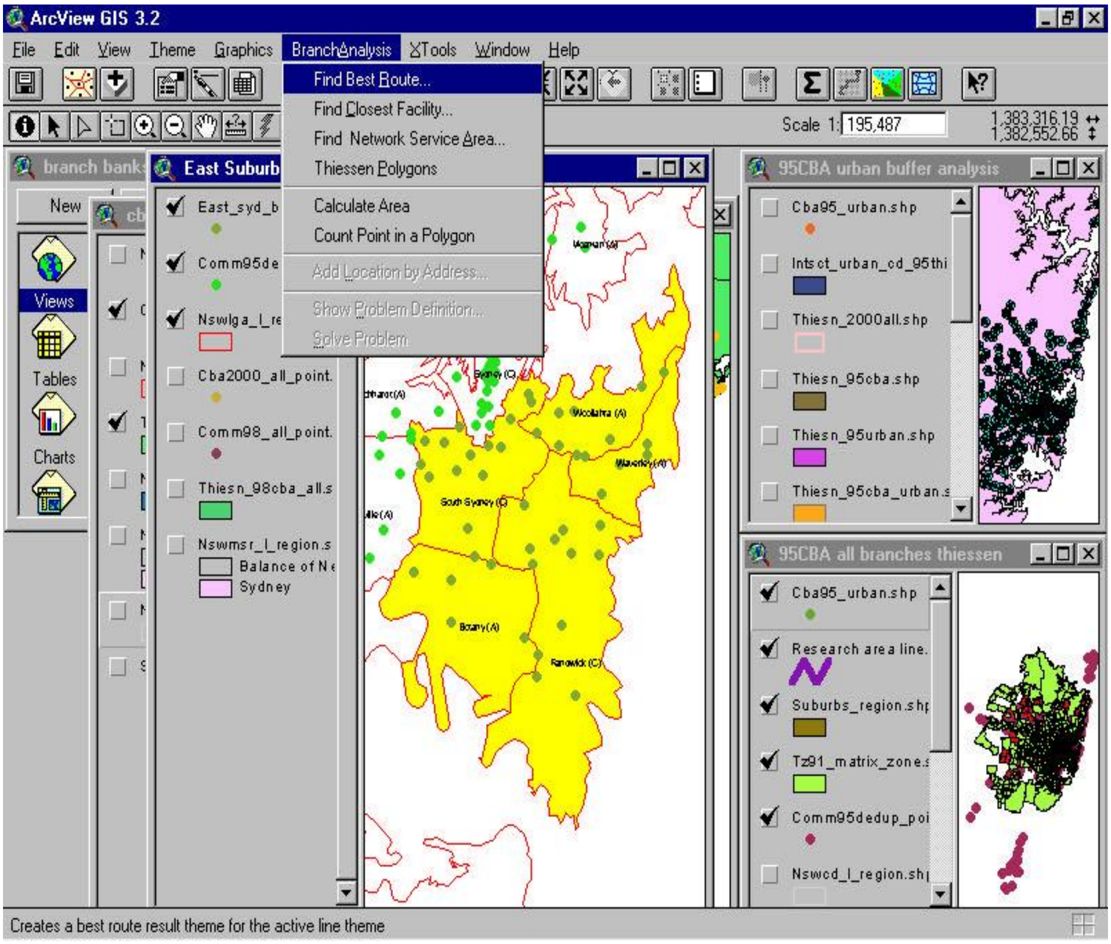


Figure 9.1: Screen Shot of K-mean

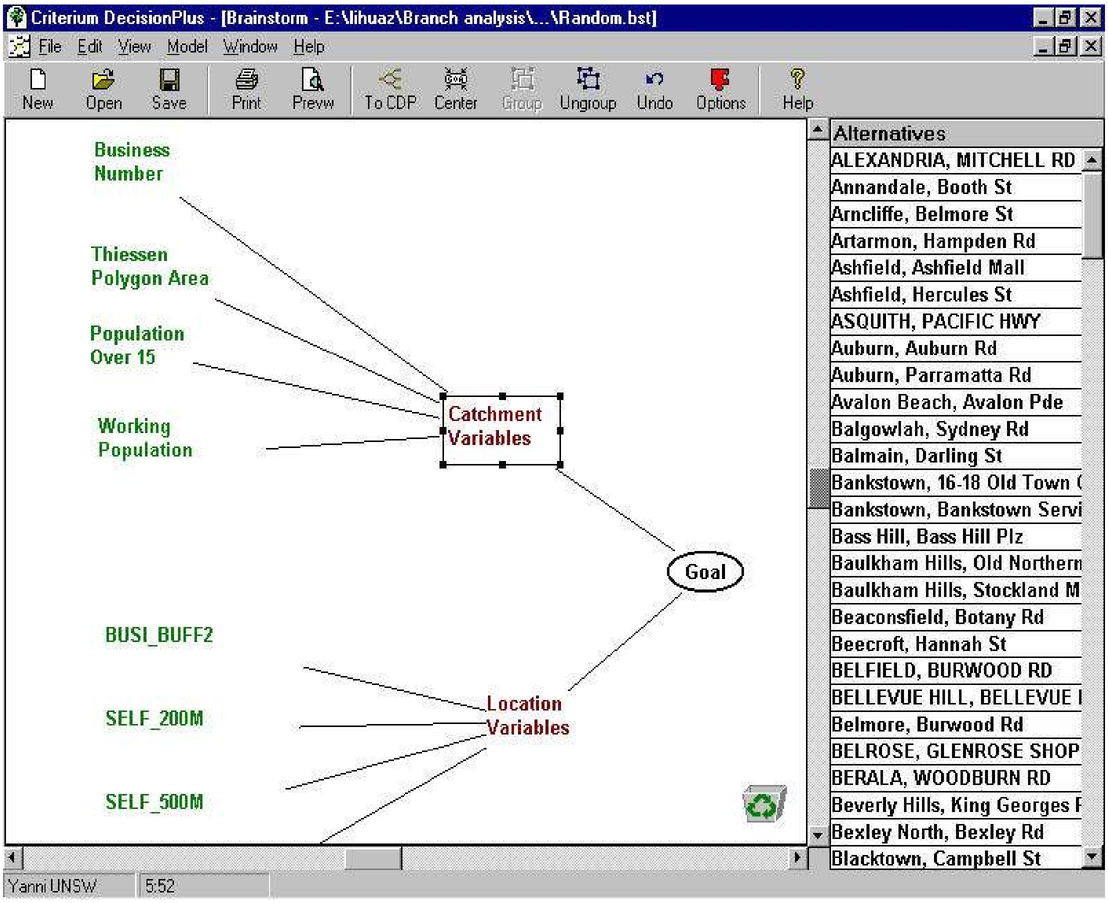


Figure 9.2: Screen shot of location variables

TECHNICAL BIOGRAPHY

Mrs. Megala P. (RRN-1145212) was born on 20th July 1984, in Salem (dt), Tamil Nadu. She did her schooling in Govt.Hr.Sec.School and secured 84% in the Higher Secondary Examination. She received BSc degree in Computer Science from Maharaja College for women in the year 2004 from Bharathiar University. She had the Master's degree in MCA from PKR College of Arts and Science in the year 2008 from Bharathiar University. She is currently pursuing her M.Phil degree in Decision support system in the Department of Computer Science of B.S Abdur Rahman University. Her area of interests is Data mining in clustering techniques. The e-mail ID is: megala.salem@gmail.com and the contact number is: 8148745462.