ANALYSING THE SPATIAL DATA OPTIMISATION WITH MACHINE LEARNING

[1]Mrs. K. MAHALAKSHMI, [2] P. RAAJADHARSHINE

[1]Assistant Professor, [2]UG Research Scholar
[1]Department of B. Com (Business Analytics)
[1]PSGR Krishnammal College for Women, Coimbatore, India.

_____

Abstract: This study is undertaken to analysis and to compare between two algorithms- existing feature algorithm and machine learning algorithms with spatial data. Here Microsoft Visual Studio.Net is used as a tool for better visualization and accuracy of the spatial data. Spatial data is about instances located in a physical space.when spatial information becomes dominant interest, spatial data mining should be applied ,with the results of the two algorithm we will be declaring which algorithm is best when compared.

Index Terms **- clarans algorithm, proMISH algorithm.**

_____

I. INTRODUCTION

The primary objective is to find the nearest location by comparing the algorithms clarans and proMISH algorithms. The location dataset contains longitude and latitude based on Coimbatore spatial datasets. The multi-group queries is used for finding the nearest location, the EUCLIDEAN DISTANCE formula is used on existing feature algorithm and machine learning algorithm to find the nearest group query. The data are clustered to remove all the unwanted data, with the group nearest query each nearest group query matches to at least one of the query keywords of the user, the user has to rank the selected location with respect to the sum of the distance to the nearest interested facilities. As a result the best location can be obtained from the minimized summed distance calculation.

II. RELATED WORKS

The spatial approach to network analysis has principally focused on planar and technical networks from a classic graph theory perspective. Reference to models and methods developed by other disciplines on non-planar networks, such as sociology and physics, is recent, limited, and dispersed. Conversely, the physics literature that developed the popular scale-free and small-world models pays an increasing attention to the spatial dimension of networks. Reviewing how complex network research has been integrated into geography and regional science reveals a high heterogeneity among spatial scientists as well as key directions for increasing their role inside multidisciplinary researches on networks.[1,5]

Multivariate spatial data play an important role in computational science and engineering simulations. The potential features and hidden relationships in multivariate data can assist scientists to gain an in-depth understanding of a scientific process, verify a hypothesis, and further discover a new physical or chemical law. In this paper, we present a comprehensive survey of the state-of-the-art techniques for multivariate spatial data visualization. We first introduce the basic concept and characteristics of multivariate spatial data, and describe three main tasks in multivariate data visualization: feature classification, fusion visualization, and correlation analysis. Finally, we prospect potential research topics for multivariate data visualization according to the current research.[3,2]

An individual is directly or indirectly attached to many smart devices. One can find his or her whereabouts if we monitor the devices they are using by collecting the metadata of the photos posted by them in the social media. Some social media websites have a feature to post the place of their recent past. To provide a simple Android application, this will be using the feature of Geo tagging available with most of the smart phones. We can use this location based data to track the people based on longitude and latitude of Global positioning system(GPS). We can use this to collect the photos posted by individuals and to analyze them to know their present position.[7,9]

Google Analytics will help library web teams determine key performance indicators based on the most used content. The authors highly recommend using "on click" event data through Google Analytics for any library website to track performance. Using event tracking can assist in the development of a composite understanding of library website users. Once library web designers have this data-driven understanding, they can start to improve access by creating more direct pathways. Any library can implement event tracking with Google Analytics, and then choose similar key performance indicators to drive users to the content that they want to access.[2,4]

III.   OBJECTIVE OF THE STUDY

The present study is based on the following objective:
- Compare clarans algorithm and proMISH algorithm
- Analysis and declare which algorithm is the best algorithm
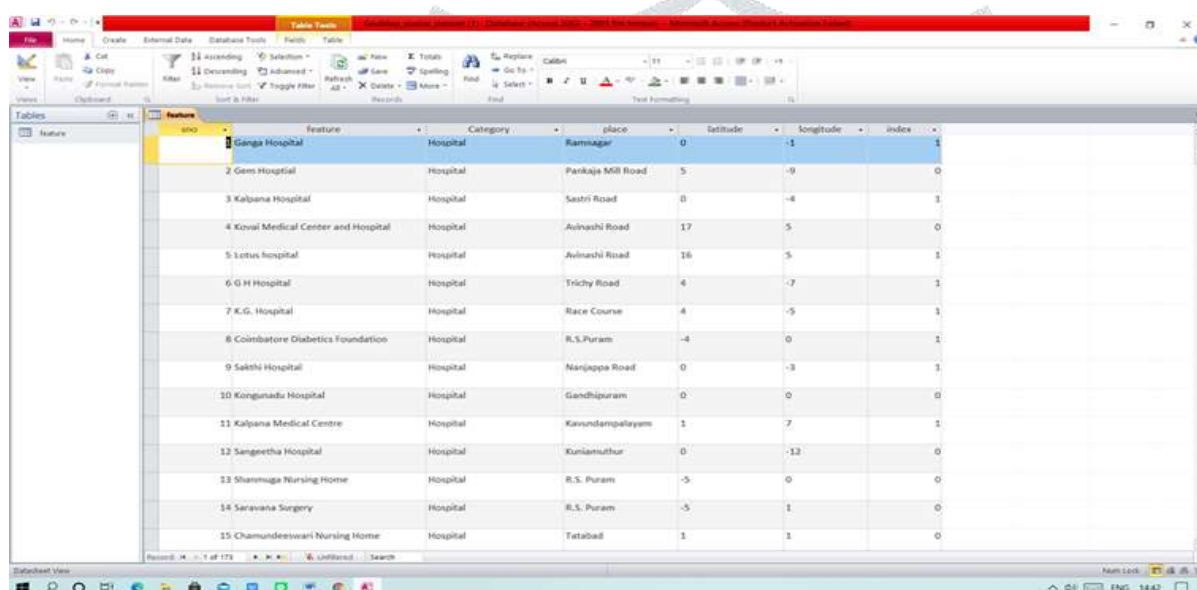
IV.　　　　　　　　　　　　　　　　　　METHODOLOGY

The following are the most used algorithms:
- Clarans algorithm (existing feature)
- proMISH clgorithm (machine learning)

4.1TOOLS AND TECHNIQUES:

Frameworks have become increasingly popular, through them reuse of design as well as code is achieved for object oriented systems. One relatively new framework is the .NET framework from Microsoft. The .NET framework is part of the larger .NET space. It includes the Common Language Runtime, a large number of partially interfaced, partially class-based frameworks packed into assemblies, and a number of tools. .NET is an open platform for enterpriseand web development and its is not bound to a particular programming language. This paper starts with a description of the concept of frameworks. Next we try to cover some of the pieces of .NET framework but due to the extensive size of the .NET not all parts can be covered. The framework perspective of .NET is analysed and we try to focus on the Object Oriented aspects while still covering enough technical parts to let the reader learn about .NET features. We are not trying to paint the .NET features as unique and the only choice on the market, nor do we try to compare .NET as a whole with its competitors. However we can conclude that the .NET framework has advantages over many other frameworks we encountered in the past.[10]

4.2 DATASET:

The datas fetched from coimbatore datas is coined as SPATIAL DATASET. The dataset constitutes attributes such as : FEATURE,CATEGORY,PLACE,LATITUDE,LONGITUDE and INDEX. The latitude and longitude gives the exact geographical location of the place.

4.3 COMPARING THE CLARANS AND proMISH ALGORITHMS:

CLARANS are also called as existing algorithm. In CLARANS, the process of finding k medoids from n objects is viewed abstractly as searching through a certain graph. In the graph, a node is represented by a set of k objects as selected medoids. Two nodes are neighbour if their sets differ by only one object. In each iteration, CLARANS considers a set of randomly chosen neighbour nodes as candidate of new medoidS. We will move to the neighbour node if the neighbour is a better choice for medoids. Otherwise, a local optima is discovered. The entire process is repeated multiple time to find better.[11]

ProMISH algorithm is also called as machine learning algorithm. Machine learning algorithms typically consume and process data to learn the related patterns about individuals, business processes, transactions, events, and so on. Machine learning becomes popular in various application areas, because of its learning capabilities from the past and making intelligent decisions. [3]

While comparing the algorithms- clarans algorithm and proMISH algorithm run the algorithms, the algorithm which takes the lesser time to show the result is considered as the best result. proMISH algorithm declared as the best as it is easy to predict the nearest group query, avoids the duplication of the data and takes the lesser time to show the results.

FLOWCHART:
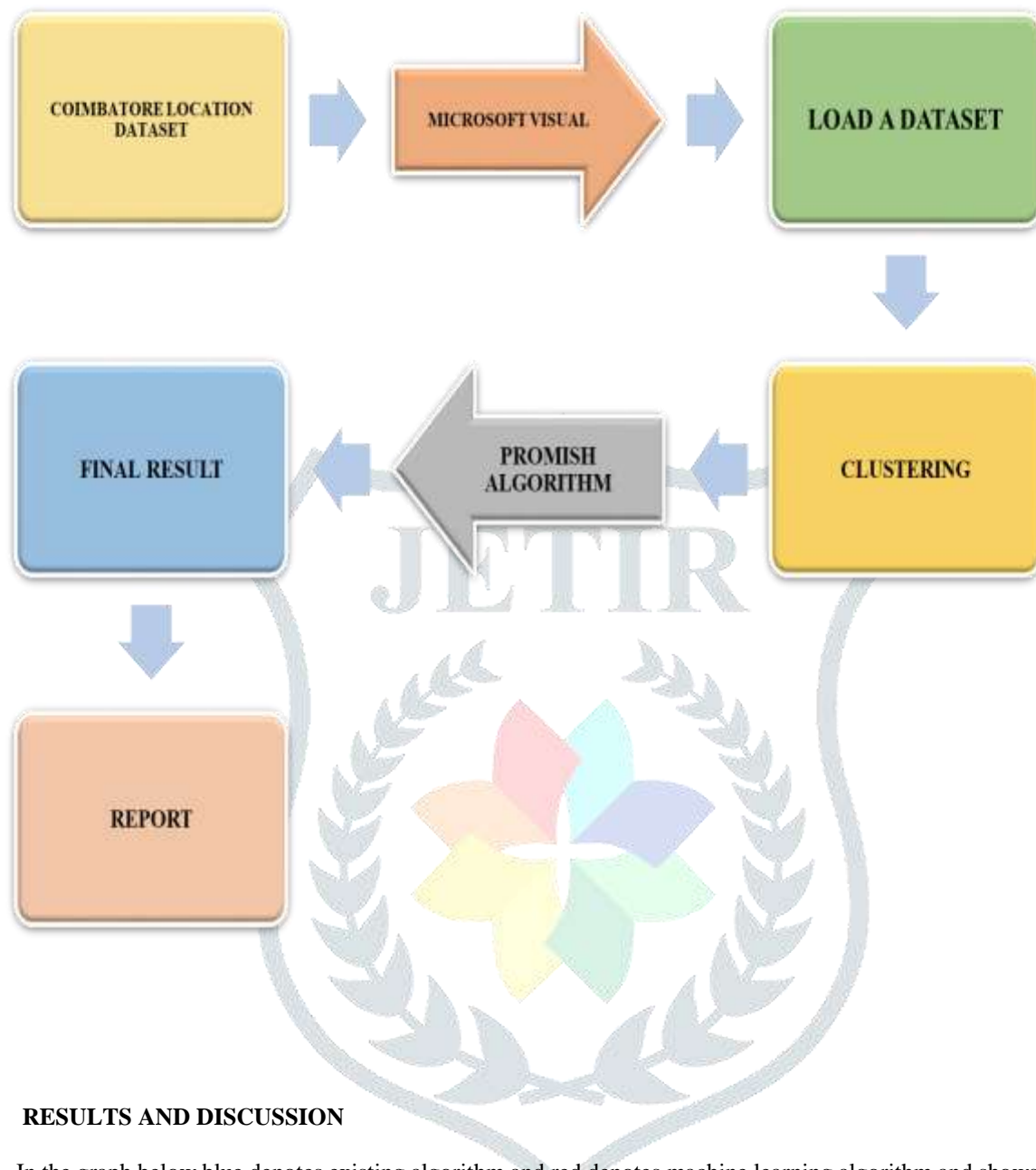
V.   RESULTS AND DISCUSSION

In the graph below blue denotes existing algorithm and red denotes machine learning algorithm and shows the total number of nodes. By comparing the nodes we found machine learning (proMISH) algorithm has given the best result on finding the nearest location. It shows 1342 and 699 node access by executing a Claranss and proMISH algorithm respectively.
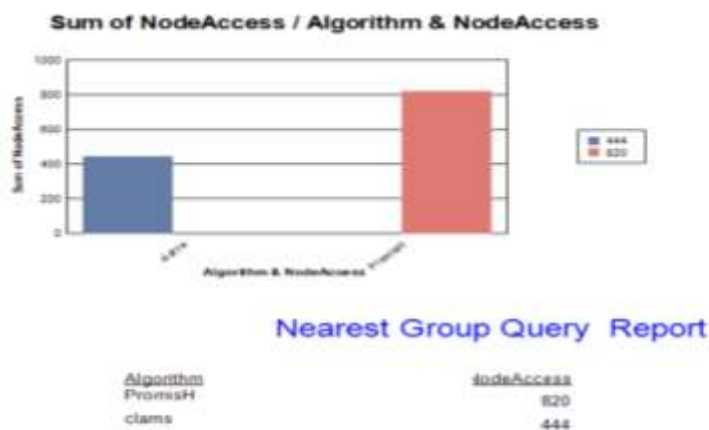
Node Access Result

The result of the nearest group query to give the date and time for the report as a result using the existing algorithm (clarans) and machine learning algorithm (proMISH) is found in the below graph
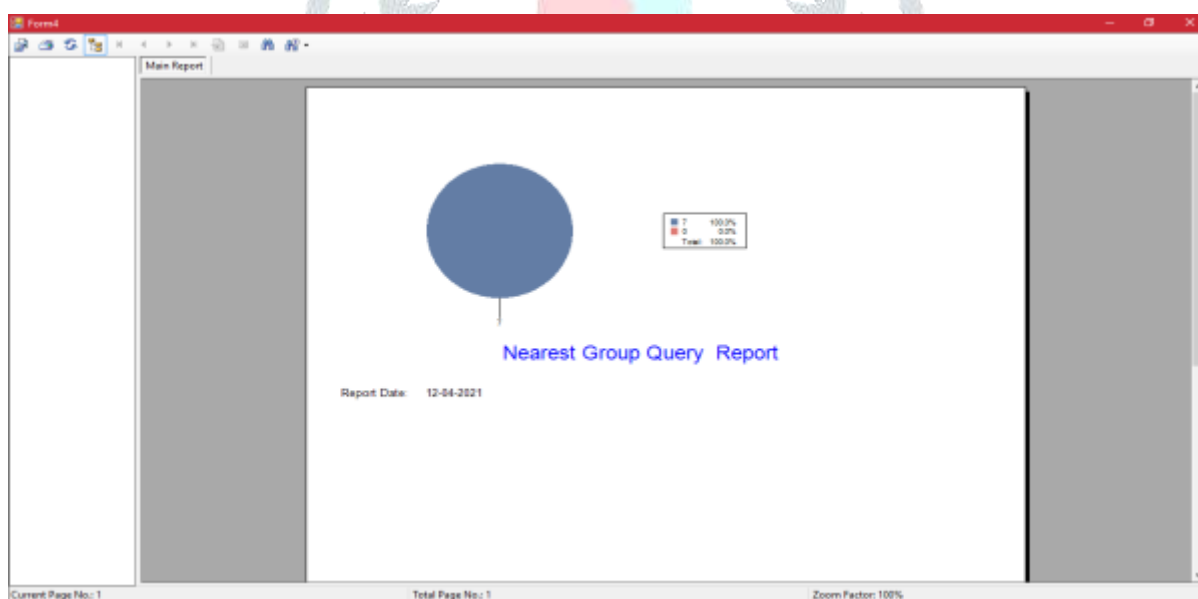
Response Time Graph

VI. CONCLUSION

There has been various methods outlined in this study. These methods have are developed by using two different algorithms: clarans algorithm and proMISH algorithm. The data have summarized and classified this research and compared the two approaches, emphasizing the particular utility of each method and the possible advantages of combining them. By using Machine learning algorithm the clustered dataset is processed to cluster to find the Euclidean distance of the nearest group point. While comparing these two algorithms, Machine learning algorithm is easy to predict the nearest group query in spatial dataset. Its take the lesser time to show the results, avoids the duplication of data hence machine learning algorithm is declared as the best algorithm.

VII.　　REFERENCES

[1]Anderson C, Hellervik A, Lindergren K. **A Spatial networkexplanation for a hierarchy of urban power laws.**Physica A 345(1-2) : 227 – 244 (2005)

[2]Agrawal R, Imieliński T, Swami A. Mining association rules between sets of items in large databases. In: ACM SIGMOD Record. ACM. 1993;22: 207–216

[3]Agrawal R, Gehrke J, Gunopulos D, Raghavan P. Automatic subspace clustering of high dimensional data for data mining applications. In: Proceedings of the 1998 ACM SIGMOD international conference on Management of data. 1998; 94–105

[4]Aha DW, Kibler D, Albert M. Instance-based learning algorithms. Mach Learn. 1991;6(1):37–66.

[5]Ardabili SF, Mosavi A, Ghamisi P, Ferdinand F, Varkonyi-Koczy AR, Reuter U, Rabczuk T, Atkinson PM. Covid-19 outbreak prediction with machine learning. Algorithms. 2020;13(10):249.

[6]Betty, Paul. 2009. **Assessing Homegrown Library Collections: Using Google Analytics to Track Use of Screencasts and Flash-Based Learning Objects.** Journal of Electronic Resources Librarianship 21 (1): 75.doi:10.1080/19411260902858631

[7]Breiman L. Bagging predictors. Mach Learn. 1996;24(2):123–40.

[8]Biswas A, Dutta S, Shan HW, WoodringJ(2013) **An information – aware framework for exploring multivariate data sets.** IEEE Trans Vis Compute Graph 19(12):2683 – 2692

[9]Pan, Bing, John C. Crotts, and Brian Muller. **Developing web-based tourist information tools using google map.** Information and Communication Technologies in Tourism 2007 (2007): 503-512

[10]PatrikTörnros, Lisa Walterfeldt, MälardalenPublished 2004.

[11]R. Ng and J. Han. CLARANS: A Method for Clustering Objects for Spatial Data Mining. IEEE TRANS. KNOWLEDGE AND DATA ENGINEERING, 2002.