Enhanced Group Query optimization with Machine Learning Algorithm on spatial data

[1]Mrs. K. MAHALAKSHMI, [2]RAGAVI .P
[1]Assistant Professor, [2]UG Research Scholar
Department of B. Com (Business Analytics)

PSGR Krishnammal College for Women Coimbatore, India

Abstract: In Machine learning algorithm, here we are using proMiSH algorithm to find the Euclidean distance. Grouping the points to find the shortest distance. Machine learning algorithm shows a result by removing unwanted data and shows the best result. Machine learning algorithm much more time and space efficient but returns results whose diameters are within a small approximation ratio of the diameters of the true results. By Group Nearest query, each nearest point matches at least one of the query keywords of the User. Using the dataset, we will be finding out the NEAREST GROUP QUERY it means we are going to fetch the location that the user satisfies all the multiple queries and finds the result by using Euclidean formula.

***Keywords*: Euclidean, Machine Learning and Existing Algorithm.**

I. INTRODUCTION

Machine learning algorithm categorised by Supervised, unsupervised and reinforcement learning. Here we use supervised and unsupervised learnings. Supervised means labelled training data and unsupervised is an unlabelled training data. We use proMiSH algorithm to fine the nearest values. Spatial data, also referred to as geospatial data, may be a term wont to describe any data associated with or containing information a few specific location on the surface. Spatial data is about instances located during a physical space. When spatial information becomes dominant interest, spatial data processing should be applied. Spatial data structures can facilitate spatial mining. Standard data processing algorithms are often modified for spatial data processing, with a considerable a part of pre-processing to require under consideration of spatial information.

II. OBJECTIVE

This research work focused on the following two findings.

1. Using machine learning algorithm to find the nearest group query location.
2. By clustering the data, we are finding the group points.

III. REVIEW OF LITERATURE

Spatial data are data connected to a location, an area on the world. Spatial decision-making exploits the geographic relationships within this data to create the decisions. The joining of spatial and non-spatial data is done by SDSS (i.e.) spatial decision support system and it also analysis and computes the GIS (i.e.) geographic information system and characteristics of problem solution. [5]

The problem faced by the decision support system is based on the spatial and the non-spatial characteristics, with the preceding recording a location's geographical coordinates and spatial relations (i.e., proximity, overlap, containment, distribution pattern). The types of decision problems include; site selection, resource allocation, network routing, location-allocation, and repair coverage. The term SDSS originated with Hopkins & Armstrong.[4]

While mobile technologies facilitate the delivery of SDSS, other developments such as cloud computing have provided the computing capacity to address problems with large amounts of data. Cloud computing is essential as it has a huge volume of data on spatial application to be stored for DSS. [7]

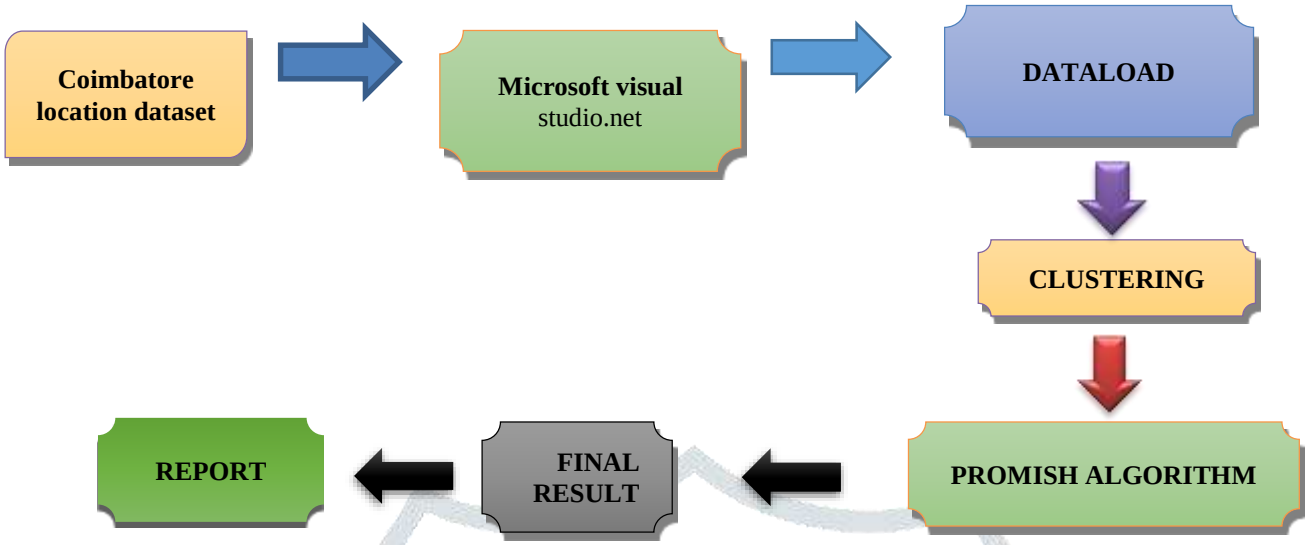
SDSS comprises within the DSS world, because it not only depends upon the reachable technology but also by the available spatial data a part from the organization and outdoors of the SDSS users. The availability of spatial data has greatly improved over the past 30 years, thanks partially to the lowering of knowledge acquisition costs and also thanks to national and transnational geospatial data initiatives. A wider range of knowledge is out there and diverse sources of spatial data are cataloged by Spatial Data Infrastructure (SDI) initiatives, although problems remain in integrating data from different sources at different spatial scales and standards and from different types of organization. [1]

The diversification of the real word, technologies for data collection and processing, database management tools and platforms leads to a enormous amount of contradictory, inconsistent, and incomplete solution. [6] In 2009 The issues of spatial data quality and uncertainty was debated in the International Symposium on Spatial Data Quality which was held in Canada.[2]

The specific data of the organization and the data on geographic properties is joined by the spatial applications; the latter data is usually sourced outside the organization and isn't specific to the organization or the choice problem. The progress of algorithmic, computational, and communication approaches both directly made SDSS technically feasible and indirectly provided the supply of third-party spatial data, which made SDSS applications economically feasible. This is an continuous process, and at all times the data is required for decision making, which makes the decision support system flourish in some fields. [3]

IV. METHODOLOGY

The enhanced Algorithm applied ProMiSH Algorithm and Euclidean distance to find nearest group query.



ProMiSH (Projection and Multi Scale Hashing) that uses random projection and hash-based index structures, and achieves high scalability and speedup. ProMiSH (short for Projection and Multi-Scale Hashing) to enable fast processing for NKS queries. ProMiSH is much more time and space efficient but returns results whose diameters are within a small approximation ratio of the diameters of the true results. [8]

$$d = \sqrt{(x_2 - x_1)^2 + (y_2-y_1)^2}$$

d = distance
(x_1, y_1) = coordinates of the first point
(x_2, y_2) = coordinates of the second point

The Euclidean distance between two points in Euclidean space is the length of a line segment between the two points. It can be calculated from the Cartesian coordinates of the points using the Pythagorean Theorem, therefore occasionally being called the Pythagorean distance. The distance between two objects that are not points is usually defined to be the smallest distance among pairs of points from the two objects. Formulas are known for computing distances between different types of objects, such as the distance from a point to a line. [9]

V.IMPLEMENTATION AND RESULT:

Coimbatore dataset is imported in visual studio software using Microsoft access as a backend tool. Clicking on the data connection select add connection and fetch the path of the data set which is placed on the Microsoft Access. After importing the data we are uploading the additional data to the dataset. And we are finding the nearest group values by using ProMISH Algorithm.

PROMISH

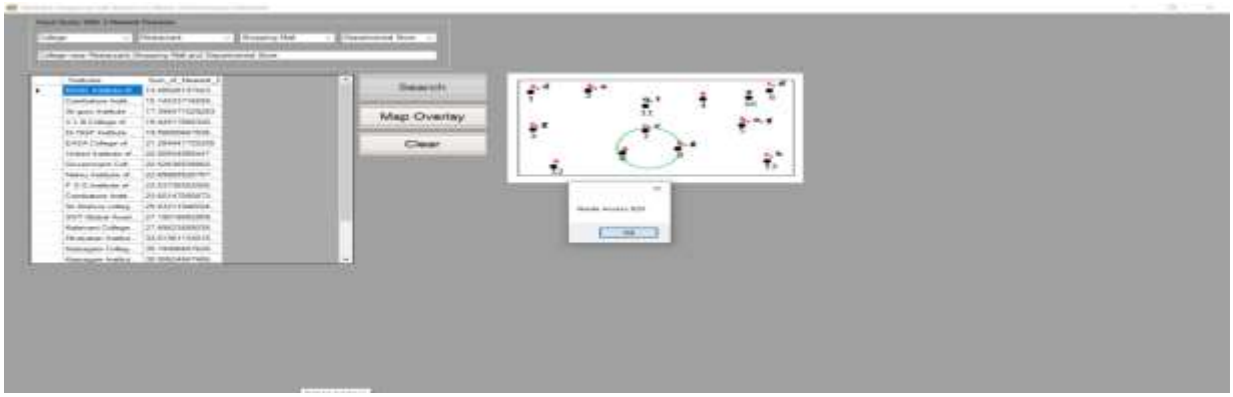


Fig: 1.1 ProMiSH Algorithm Result

In the above fig 1.1 Using Machine learning algorithm we are finding the Euclidean distance by giving different attributes and searching the notes to find which is the best. Here we are using proMiSH algorithm to find the best places.

ProMiSH (Projection and Multi Scale Hashing) that uses random projection and hash-based index structures, and achieves high scalability and speedup. ProMiSH (short for Projection and Multi-Scale Hashing) to enable fast processing for NKS queries. ProMiSH is much more time and space efficient but returns results whose diameters are within a small approximation ratio of the diameters of the true results.[8]

DATA DUPLICATION

QUERY ANALYZER:

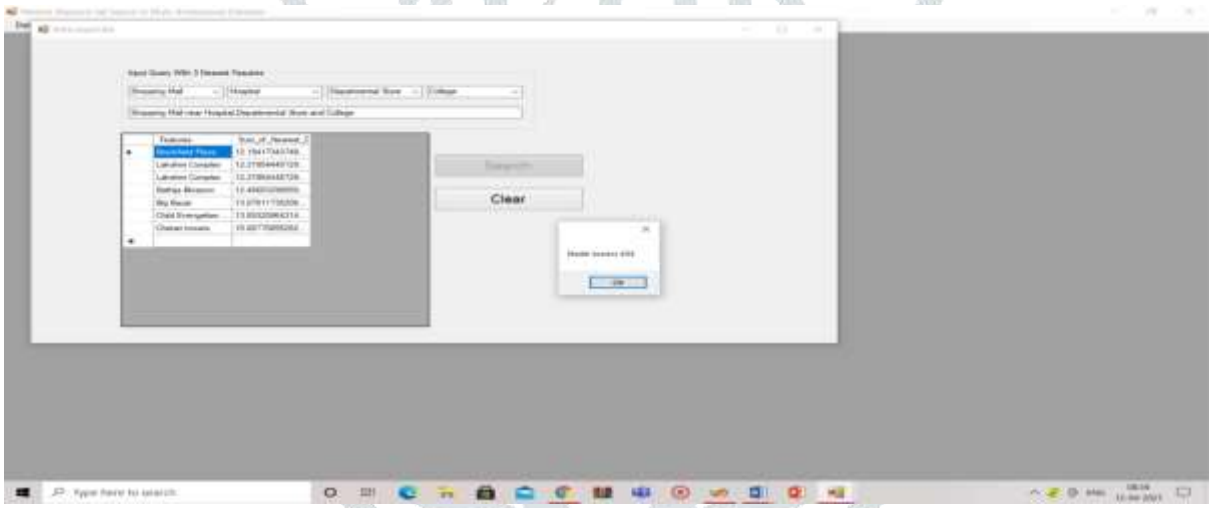


Fig: 1.2 Query Analyzer Result

In the above Fig: 1.2 that from the data set of same attributes it removes a duplication data and showing the result of a node access report. Data duplication is a technique for eliminating duplicate copies of repeating data.

ACCESS REPORT:

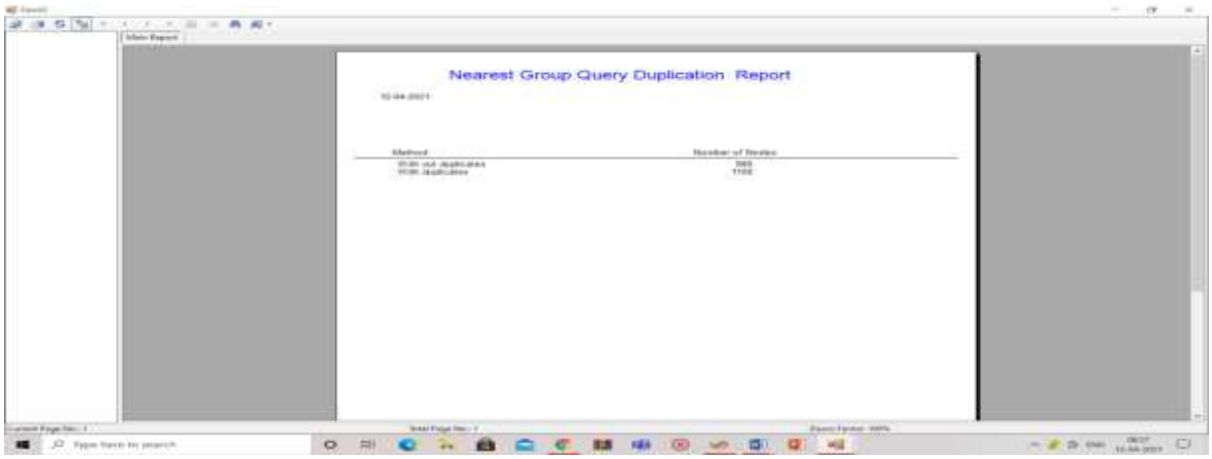


Fig: 1.3(Access Report Result)

In the above Fig: 1.3 it shows the total number nodes with duplication and without duplication. As a result nodes including duplication shows 1100 and by removing duplication we get 880. This shows that the dataset contains of 220 same records were eliminated.

RESPONSE REPORT

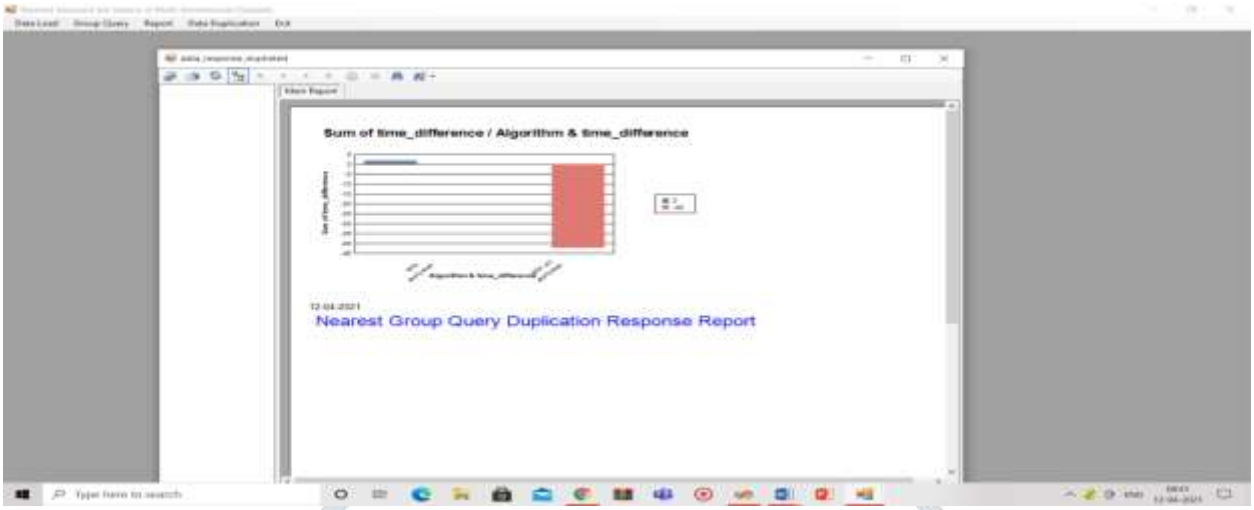


Fig: 1.4 (Response Report Result)

In an above Fig: 1.4 it shows the result sum of time difference of both Clarns and proMiSH algorithm. For finding which takes a less time to find a result. And plotting the graph of both duplication and without duplication of sum of time difference and algorithm &time difference.

VI. CONCLUSION:

Through our project we are trying to offer a better way of serving the users by introducing search options that will be able to answer multiple queries. Our search option will provide results on cluster basis with provide results on a cluster basis with relative distance between them. This will help the user in choosing better locations with accurate results.

REFERENCES:

[1] Dessers. E, G. Vancauwenberghe, D. Vandenbroucke, J. Crompvoets, G. Van Hootegem. **Analysing spatial data performance in inter-organisational processes.** International Journal of Digital Earth, 8 (2015), pp. 403-420

[2] Devillers R., Bedard Y., Fisher P., Stein A., Chrisman N., Shi W., **Thirty years of research on spatial data quality: achievements, failures, and opportunities**. Transaction in GIS, 2010, 14(4), 387-400 5

[3] Goodchild .M.F Citizens as sensors: the world of volunteered geography GeoJournal, 69 (2007), pp. 211-221

[4] Hopkins. L. D, M.P. Armstrong. **Analytic and cartographic data storage: a two-tiered approach to spatial decision support systems**. Proceedings, Seventh International Symposium on Computer-assisted Cartography, American Congress on Surveying and Mapping, Washington, DC (1985)

[5] Jankowski. P. **Spatial decision support systems**. K.K. Kemp (Ed.), Encyclopedia of Geographic Information Science, SAGE Publications, Inc., Thousand Oaks, California (2008)

http://sk.sagepub.com/reference/geoinfoscience

[6] Tong X., Xie H., Liu S., Jin Y., Shi W., Wang J., Pei T., GeY., Zhu, Ch., **Uncertainty of Spatial Information and Spatial Analysis**. In: The Geographical Sciences During 1986—2015, Springer Geography. Springer, Singapore, 2016, 511-522 4

[7] Yue. P, H. Zhou, J. Gong, L. Hu. **Geoprocessing in cloud computing platforms–a comparative analysis**. International Journal of Digital Earth, 6 (2013), pp. 404-425

[8] Shilpa Thakare,Prof.Shinde Jayashri, Novel Method Fot NKS Dearch in Multidimensional Dataset Using Advance Promish & Rankinh Function, International journal for Research in Engineering Application & Managenent(IJREAM), ISSN:2494-9150 Vol-02,Issue 08, Nov 2016

[9]https://en.m.wikipedia.org/wiki/Euclidean_distance#:~:text=In%20mathematics%2C%20the%20Euclidean%20distance,being%20called%20the%20Pythagorean%20distance.