PERFORMANCE ANALYSIS OF STUDENTS THROUGH CLUSTERING IN DECISION SUPPORT SYSTEM

M.Sindhuja
Student
Information Technology
KCG College of Technology
Chennai,India
sindhuja1903@gmail.com

S.Gayathri
Student
Information Technology
KCG College of Technology
Chennai,India
gayathrijuly8@gmail.com

R.B.Akshaya
Student
Information Technology
KCG College of Technology
Chennai,India
akshaya110694@gmail.com

I.R.Praveen Joe
Assistant Professor
Information Technology
KCG College of Technology
Chennai,India
praveenjoeir@yahoo.com

Abstract— It is the choice emotionally supportive network for the execution investigation of the understudies. The info information comprises of marks(cgpa) and the behavioral test score, which is gathered by leading Myer's test for the understudies. Bunching is done on the gathered information. Bunch examination or grouping is the errand of collection an arrangement of articles in a manner that questions in the same gathering (called a group) are more comparable (in some sense or another) to one another than to those in different gatherings (groups). This paper manages the most famous thickness based grouping strategy called DBSCAN. As opposed to numerous strategies, it highlights a bunch model called "thickness reachability" and it depends on interfacing focuses inside of certain separation measures. In any case, it just associates focuses that fulfill a thickness measure, characterized as a base number of different items inside of this range. The bunch comprises of all thickness associated objects (which can frame a group of a discretionary shape, as opposed to numerous different techniques) in addition to all protests that are inside of these articles' reach. The bunches are gotten. The qualified understudies and the individuals why should likely be put goes under one group and afterward alternate understudies class ,who are not prone to be put are acquired at the last. At long last the intra-bunch separation among the group is low and the between bunch separation between the group is high. It is the choice making process for the office and the school to give encourage more enhanced preparing for the "Qualified understudies".

Keywords—*Clusteranalysis,Clustering,DBSCAN,density reachability.*

I. Introduction

This paper for the most part spotlights on gathering information from the understudies, preprocessing it and after that bunching is carried on the understudies information in light of the inputs given, which determines the qualified and those understudies why should likely be put and the individuals who donot..An unsupervised grouping calculation is relied upon to cook the prerequisite than a regulated one since it winds up in an unexpected gathering where as the later requests settled class names ahead of time before gathering.

The following part of the paper gives a nitty gritty note on the study made on the way of grouping calculation suitable for the problem.The purposes behind picking DBSCAN calculation are recorded. The working of the calculations are given. A brief rundown of how the calculations are utilized as a part of the reference papers are likewise highlighted. The outline of methodologies and calculations took after for understudies grouping are talked about in this segment. Area 3 depicts the model of the complete issue configuration and its usage. Area 4 compresses the outcomes and the translation of the same. Area 5 finishes up the paper with the future upgrade.

1.1 Clustering of Students

With bunching the gatherings (or groups) depend on the similitudes of information occurrences to one another. No officially characterized yield class is utilized as a part of preparing and the bunching calculation should take in the gathering. Bunching is discovering gatherings of articles such that the items in one gathering will be like each other and unique in relation to the articles in another gathering . Bunching can be viewed as the most vital unsupervised learning strategy. Characterization of grouping calculation. In instructive information mining, bunching has been utilized to aggregate the understudies as indicated by their conduct for e.g. grouping can be utilized to recognize dynamic understudy from non-dynamic understudy as per their execution in exercises. However a suitable decision of an unsupervised calculation and the correct working of components sets considering both useful and non useful attributes would enhance the nature of bunching [1].

1.2 Unsupervised Clustering

The Clustering capacity looks the info information for qualities that much of the time happen .It gathers the data information into groups. The individuals from every bunch have fundamentally the same properties. There are no assumptions of what examples exist inside of the information. Bunching is a revelation process. Information in every subset (in a perfect world) share some regular quality - frequently vicinity as indicated by some characterized separation measure. "Groups" is regularly used to portray totally unrelated sub-sections without a rundown of predefined attributes. In the event that we have predefined marks it is just a characterization. Consequently it is required to perform grouping in an unsupervised way [2].

1.3 Density Based Clustering

The bunching techniques like K-means or Expectation-Maximization are suitable for discovering ellipsoid-formed groups. Be that as it may, for non-curved groups. these techniques experience difficulty finding the genuine bunches, since two focuses from any distinctive groups might be closer than two focuses in the same bunch. The thickness based strategies which we consider in this part can mine such non-raised or shape-based bunches [3].

1.4 DBSCAN

DBSCAN (for thickness based spatial grouping of utilizations with commotion) is a thickness based bunching calculation. It is utilizing the idea of "thickness reachability" and "thickness interface ivity "both of which relies on information parameter-size of epsilon neighborhood e and least terms of nearby dissemination of closest neighbors. Here parameter e controls the measure of the area and size of groups. It more often than not begins with a self-assertive beginning stage that has not been gone to DBSCAN calculation is a vital piece of grouping procedure which is principally utilized as a part of logical writing. Thickness is measured by the quantity of articles which are closest the group [4].

II. Related works

2.1 K-means

K-means was proposed by Macqueen ,is a standout amongst the most well known dividing strategies. It segments the dataset into k subsets, and k is as of now characterized. The calculation continues altering the items to the nearest current bunch until no new assignments of articles to groups is made. One Advantage of this calculation is its effortlessness. It additionally has a few downsides in it. It is extremely hard to determine number of groups ahead of time. Sinceit works with squared separations, it is delicate to exceptions. Another disadvantage is centroids and it is not significant in generally problems[5].

2.2Hierarchical Clustering

Progressive grouping calculation for the most part gap or union dataset into a progression of settled segments. The method for the settled allotments can be either base up or beat down. In the base up strategy, grouping is finished with every single item in a solitary bunch and it keeps on bunching the nearby combines of bunches until every one of the articles are as one in one and only bunch. Top-down various leveled grouping, then again, begins with all articles in one bunch and continues isolating bigger groups into littler groups until all items are isolated into single bunch. Both the various leveled strategies demonstrate the most common method for speaking to the bunches, called as dendrogram. Samples of this calculations are ROCK, BIRCH (Balance Iterative Reducing and Clustering utilizing Hierarchies), CURE (Cluster Using Representatives).It handles just numeric records and delicate to information records[5].

III. Algorithm Characteristics

Taking after are the disadvantages saw in the above works which are to be killed at present while picking the bunching calculation. However no specific grouping calculation can be cited as 'best', everything relies on upon the need and application. This is an outcome on the review made on numerical, managed, unsupervised calculations and other swarm based methodologies.

IV. PROPOSED ALGORITHMS

A. DBSCAN ALGORITHM DBSCAN (Density-Based Spatial Clustering of Application with noise) is density based cluster formation algorithm for spatial and non spatial high dimensional data base in the presence of outlier. The working is based on the following definitions, for more detail refer DBSCAN :

Def.1: The -neighborhood of an object p, denoted by N (p), is defined as total number of objects lying in the radius , i.e. N (p) = {q D | dist (p, q) ≤ }.

Def.2: An object p is said to be Core object if | N (p)| ≥ μ (minimum objects).

Def.3: An object p is said to be directly density reachable from an object q with respect to and μ if p N (q) and q is a Core object.

Def.4: An object p is said to density-reachable from an object q if there is a chain of objects p1. . . pn, p1= q, pn= p such that pi+1 is direct density-reachable from pi with respect to and μ.

Def.5: An object p is said to density-connected to an object q with respect to and μ if there is an object o such that both the p and q are density reachable from o with respect to and μ.

Def.6: An object which is lying at the border is not a Core object, but it will be a part of cluster. An object which is not lying in any of the cluster is treated as a noise object.

Def.7: A cluster X is non empty subset of database with respect to and μ, for every p, q: if p X, q is density reachable from p then q X and p is density- connected to q. DBSCAN detects density connected clusters by discovering one of its core object's p and computing all objects which are density-reachable from p. The collection of density reachable objects is performed by iteratively computing directly density reachable objects. DBSCAN checks the -neighborhood of each object p in the database. If N (p) of an object p consists of at least μ objects, i.e., if p is called as the core object, a new cluster X containing all objects of N (p) is created. Then, the - neighborhood of all objects q X, which have not yet been processed, is then checked. If object q is also a core object, the neighbors of q, which are not already assigned to cluster X, are added to X and their - neighborhood is checked in the next step. This procedure is then repeated until no new object can be added to the current cluster X. The DBSCAN algorithm proposed in this work is - 1: Label all the points as core, border, or noise points. 2: Eliminate the noise points. 3: Put an edge between all the core points that are within Eps of each other. 4: Make each group of the connected core points into a single cluster. 5: Assign each of the border point to one of the clusters with its associated core points.

V. Detailed Explanation

The fundamental key thought is that information objects in thick areas are bunched together. The calculation utilizes a settled quality called as edge worth to choose the thick districts. It finds the high thickness districts in space i.e. isolated by low locale thickness. The inconvenience of the calculation is that it catches just certain sorts of commotion when bunches of various densities exist. Not at all like other grouping strategies, it doesn't require the pre-determination of number of bunches. It likewise finds groups of subjective shape in spatial databases with clamor .Variants of DBSCAN are: Incremental DBSCAN goes about as the center calculation of question bunching device. SDBDC (Scalable Density-Based Distributed Clustering strategy, first it by and large takes a shot at every neighborhood site and afterward bunches disseminated objects on worldwide site. Fundamental Definitions: a) Ɛ-neighborhood: The area is separation between two focuses in a bunch. The area in a group is not exactly the limit data esteem, Ɛ. The area inside of a sweep Ɛ of a given item is Ɛ-neighborhood. b) MinPts: It shows the base number of information articles in any bunch. c) Core Object: It alludes Ɛ-neighborhood of an article contains in any event MinPts of items. d) Directly thickness reachable: An information object p is specifically thickness reachable from the information object q if p is inside of the Ɛ-neighborhood of q and q is a center item. e) Density-reachable: An item p is thickness reachable from the article q concerning Ɛ and MinPts if there is a chain of articles p1,p2, . . . ,pn, where p1 = q and pn = q such that pi+1 is straightforwardly thickness reachable from pi as for Ɛ and MinPts.[9] f) Density Connected: An article p is thickness associated with item q as for Ɛ and MinPts if there exists an article o D. g) Density Based Cluster: It is an arrangement of thickness associated objects i.e .the maximal as for thickness reachability. h) Border point: An item p is a fringe point in the event that it is not a center protest but rather thickness reachable from another center article. i) Noise: The items not alloted in any bunch go about as clamor. The calculation fills in as tails: It first checks Ɛ-neighborhood of every point in the space. In the event that the Ɛ-neighborhood of the point p contains more than MinPts, another bunch made in which p goes about as the center article. The calculation repeats and accumulates every one of the items inside Ɛ separation from the center articles. The procedure then ends when there is no new indicate add to any group.

Exception DETECTION DBSCAN manages anomalies (information objects which are diverse with the staying set of the information). The calculation stays away from the commotion or exception to embed into the groups. It's just fit to catch a few sorts of exceptions when distinctive densities of bunches are available. This prompts an enormous loss of critical concealed data as once in a while the exception are specifically noteworthy. Samples are misrepresentation identification, interruption revelation.

Info: D = {t1, t2, t3 … tn}/Set of components MinPts/Number of focuses in bunch Ɛ/Maximum separation for thickness measure Output: K = {K1, K2, K3 … , Kk }/Set of groups Method: k=0;/at first there are no group for i = 1 to n do if ti is not in a bunch, then X = {tj | tj is thickness reachable from t if X is a legitimate group, then k = k+1; Kk = X

DBSCAN manages exceptions (information objects which are distinctive with the staying set of the information). The calculation evades the commotion or exception to embed into the groups. It's just equipped for catching a few sorts of anomalies when diverse densities of groups are available. This prompts a gigantic loss of essential concealed data as some of the time the anomaly are quite compelling. Illustrations are extortion recognition, inter.

VI. STEPS:

1. The input is obtained from the students(Marks, behavioral test)
2. The clustering algorithm is applied and the students clusters are formed.
3. The clusters specifies those students who are likely to be placed and those who do not.
4. Finally the students data is represented in a interpretation tool and the analysis is done.
5. The intra cluster distance among the students in one cluster is less then the inter cluster distance among the clusters are high. The eligible students for placements is also obtained.

VII. SYSTEM ARCHITECTURE:

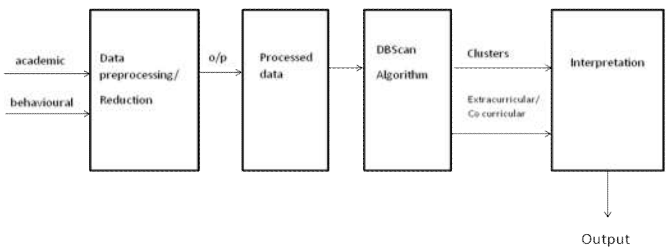

Figure 1

VIII. PERFORMANCE MEASURES:

1.The intra cluster similarity is low.

2.The inter cluster similarity is high.

IX. RESULTS:

Figure 2. Clustering of students

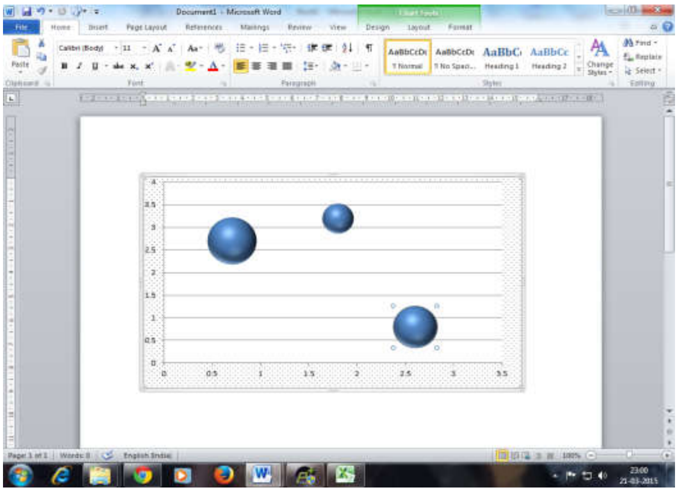

Figure 3 CLUSTERS

X. CONCLUSION & FUTURE WORK

In this paper, we have concentrated how information mining can be connected to instructive frameworks. It demonstrates how helpful information mining can be in advanced education, especially to enhance understudies' execution. We utilized understudies' information from the database of conclusive year understudies' for Information Technology UG course. We gathered all accessible information including their execution at college examination in different subjects. We connected information mining procedures to find learning. Likewise we bunched the understudy into gathering utilizing DBSCAN-grouping calculation. At long last, boisterous information were additionally distinguished. Every one of this learning can be utilized to enhance the execution of understudy. For future work, an approach to sum up the study to more assorted courses to get more exact results should be created. Additionally, examinations should be possible utilizing more information mining strategies, for example, neural nets, hereditary calculations, k-closest Neighbor, Naive Bayes, bolster vector machines and others. At last, the utilized preprocessed and information mining calculations could be inserted into elearning framework with the goal that one utilizing the framework can be profit by the information mining systems.

REFERENCES

[1]Shannaq,B.,Rafael,Y.and Alexandro,V. ”Student Relationship in higher education using data mining Techniques”,Gobal Journel of computer science and technology,vol 10,no.11,pp.54-59.

[2]N.V.Anand Kumar and G.V.Uma,”Improving Academic performance of students by applying Data mining Technique,”European journal of scientific research,vol.34(4),2009.

[3]Edin Osmanbegović,Mirza Suljić,”DATA MINING APPROACH FOR PREDICTING STUDENT PERFORMANCE”, Economic Review – Journal of Economics and Business, Vol. X, Issue 1, May 2012.

[4] K.Shanmuga Priya” Improving the student’s performance using Educational data mining”, International Journal of Advanced Networking and Application, Vol.4,pp-1680- 1685 (2013)

[5]Fhim A.M.Salem A.M.Torrkey F.A. and Ramadan M.A.,”An efficient enchanced K-means Clustering Algorithm,”pp 1626-1633,2006

[6]. Md. Hedayetul Islam Shovon , Mahfuza Haque , “Prediction of Student Academic Performance by an Application of K-Means Clustering Algorithm”, International Journal of Advanced Research in Computer Science and Software Engineering Research Volume 2, Issue 7, July 2012 ISSN: 2277 128X

[7].Mohammed M.Abu Tair,Alaa M.El-Hales,”Mining ,Eductaion Data to improve Student’s Performance:A Case Study, International Journal of Information and Communication Technology Research”, Volume 2,No 2, February 2012

[8].Ayesha.S,Mustafa.T, Sattar.A and Khan.I, “Data Mining Model For Higher Education System", European Journal of Scientific Research”, vol.43,no1,pp24-29.

[9]. Baradwaj.B, Pal.S, “Mining Educational Data to Analyze Student’s Performance”, International Journal of Advanced Computer Science and Applications, vol.2, no.6,pp.63-69

[10]. Han.J and Kamber.M, “Data Moining: Concepts and Techniques”, The Morgan Kauffman Series in Data Management Systems,2nd edition