IMPROVING MULTI-FACTOR PERFORMANCE IN E-LEARNING ENVIRONMENT

J. Rajula [1], Mr. S. Baskaran [2].

*[1]Departmentmet of Computer Science, Tamil University, Thanjavur – 613 010, Tamilnadu, India.
*[2]Head, Department of Computer Science, Tamil University, Thanjavur – 613 010, Tamilnadu, India.

Abstract : **Determining the senior high school major is still a dilemma for some junior high school students. The selection of high school majors must be tailored to the interests, talents and academic skills of students so that later students can develop a better competencies, attitudes and academic skills in the new environment. The selection of the appropriate high school major will influence students' interests and abilities in exploring a field of science so that later it will be easier for students to go to university which is expected and in accordance with their current interests and abilities. This will obviously be very beneficial for the student in preparing for his future. Clustering is one technique known in the data mining process. The core concept of clustering is to group a number of data or objects into a group or several groups where each group contains data that has similarities that are very close to other data. There are two types of grouping methods known as hierarchical clustering and partitioning. The hierarchical clustering method consists of several types, namely complete linkage clustering, single linkage clustering, average linkage clustering and centroid linkage clustering. While the partitioning method itself consists of the following types namely k-means clustering and k-means fuzzy clustering.**

In this study, the authors have applied and analyzed the Agglomerative Hierarchical Clustering technique in the data of students of SMP Negeri 2 Purwodadi to classify students based on their respective interests and skills to fit the selection of high school majors. In the implementation, the author uses 5 attributes of pre-processing results which are used as experimental data processing variables. The results of this study succeeded in developing a prototype application that has implemented the Agglomerative Hierarchical Clustering algorithm which is used to visualize data processing so that it can help students determine high school majors. From the various experiments that have been carried out, this application has shown good resultsl.

Keywords—Data Mining, Clustering, Agglomerative Hierarchical Clustering Algorithm, Major Determination in Senior High School

I. INTRODUCTION

Selection of majors by the students according to interests when enrolling in high school has the aim to provide opportunities for students to develop competence attitudes and competence skills of students according with their interests, talents, and academic skills in a group of subjects in science. Determination of the majors will affect the next academic level and will affect the field of science or studies for students who will continue on to the university level so that the selection of incorrect majors could harm students and their future [1].

Placement of students according to ability or capacity is often called the selection of majors by students in senior high school are determined by academic ability that is supported by a factor of interest, because of the characteristics of a science requires the same characteristics of who will be studied. Thus, students who studied a science that correspond to the characteristics of their personality will feel comfortable when studying that science. Student interests can affect the quality of student achievement in a particular field of study. A student that interested in mathematics, for example, will more focus on mathematics than other students who are not interested in mathematics. Because the concentration of intensive attention to the material, students will study harder and try to achieve the achievement [2].

Decision making in choosing majors according with the curriculum in 2013 is determining the majors when students will be enrolled in senior high school by looking at several factors, including the score in progress report when in the junior high school, the score of junior high school national examination, recommendations from counseling teacher /counselor in junior high school. The programs that are available in the senior high school include Science (IPA), Social (IPS), and Linguistics (Bahasa)[1]. In addition, the selection of majors aims to adjust capabilities and interest of the students to the field chosen by the student. Selection of the appropriate majors will increase the interest and provide a sense of comfort to the students in learning. With the same basic capabilities expected in learning activities can proceed smoothly, without any students who have difficulty in learning and increase the interest and student achievement. In contrast, lack of interest in learning as a result of a mistake in choosing majors [2].

Because too many students who will choose the majors, counseling teacher have difficult to perform a calculation to determine the majors because the process of selecting majors still done manually by grouping the student's progress report score based on the relevant courses, and still consider the student's interests and recommendations from counseling teacher, so that way is less effective and less efficient because they have extra work and requires a long time [2]. Clustering is one technique known in the data mining process. The core concept of clustering is to group a number of data or objects into a group or several groups where each group contains data that has similarities that are very close to other data. There are two types of grouping methods known as hierarchical clustering and partitioning. The hierarchical clustering method consists of several types, such as complete linkage clustering, single linkage clustering, average linkage clustering and

centroid linkage clustering. While the partitioning method itself consists of the following types, such as k-means clustering and k-means fuzzy clustering. [3].

II. RELATED WORKS

Fuzzy C-Means (FCM) algorithm can provide recommendations or determination of the optimal majors for the students and also how Genetic algorithms can distribute students into classes based on the available quota. The process of grouping or the determination of these majors are obtained based on the evaluation value of each subject of the majors and based on the evaluation value of each subject related to the majors [4].

Clustering is the process of grouping a collection from the objects that have the same members into classes of the same object. The purpose of clustering is to group a set of data or objects that have a similarity in one cluster and different from the objects in other clusters. The method that used in this research is Agglomerative Hierarchical Clustering method. This method is done by grouping a set of objects (many objects into one piece). Begins by placing any object on clustering and then combine the clustering structure becomes larger clustering until all the objects currently on the same clustering or to fulfill predetermined conditions. This algorithm conducted with a bottom-up approach [5].

III. THEORETICAL BACKGROUND

Data mining is a method by finding a knowledge in a database that is large enough. Data mining is the process of digging and analyzing very large amounts of data to obtain something right, new, useful and ultimately understandable a pattern or patterns in the data [6].

Hierarchical Clustering Method (HCE) is one of the methods used in performing clustering [3]. Clustering that formed from HCE in the form of a hierarchical (tree). In hierarchical clustering, data is not divided into a clustering in one step but in a series of clustering. The division can be done from a cluster that containing all objects to some clustering in the number of n where each value of n contains an object [3]. Hierarchical Clustering method is classified into two types, namely Agglomerative and Divisible Hierarchical Clustering, depending on the composition conducted from the form of bottom-up or top-up [7]. The method used in this study is Agglomerative Hierarchical Clustering Method. This method is done by grouping a set of objects (many objects into one piece).

Begins by placing any object on clustering and then combine the clustering structure becomes larger clustering until all the objects that are in the same clustering or to fulfill predetermined conditions. This algorithm performed with a bottom-up approach [3].

The algorithm used in Hierarchical Clustering Method is Hierarchical Clustering Algorithm. In the algorithm, if given a number N, the object will be clustering with proximity matrix N * N, then [9]:

1) Positioning Figures Beginning with making each object into one clustering. If the available number N, the object to be in clustering, it will form a clustering as many N which contain only one object.

2) Find the nearest of the pair clustering and combined into one clustering. Clustering can be combined into one clustering if it has an object with the closest similarity to the object on the other clustering.

3) Calculate the distance or similarity between the newly formed clustering and other clustering that has existed before.

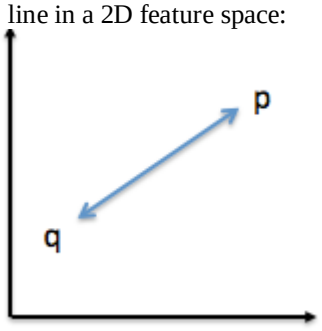
4) Repeat steps 2 and 3 until all objects that performed clustering process currently on one clustering

Agglomerative Hierarchical Clustering is a method by using the strategy design Bottom-Up which begins by placing any object as a cluster itself (atomic cluster) and then combine atomic clusters into a cluster, which is bigger and bigger until finally all the objects fuse into a cluster or the process can also be stopped if it has reached the limits of specific conditions [3].

The measure of distance that is used to combine two objects cluster is Euclidean Distance [3], which can be seen in the following equation:
It is just a distance measure between a pair of samples *p* and *q* in an n-dimensional feature space:

$$d(p,q) = \sqrt{\sum_{i=1}^{n}(p_i - q_i)\,2}$$

For example, picture it as a "straight, connecting" line in a 2D feature space:



The Euclidean is often the "default" distance used in e.g., K-nearest neighbors (classification) or K-means (clustering) to find the "k closest points" of a particular sample point. Another prominent example is hierarchical clustering, agglomerative clustering (complete and single linkage) where you want to find the distance between clusters

Algorithmic steps for Fuzzy c-means clustering

Let X = {x_1, x_2, x_3 ..., x_n} be the set of data points and V = {v_1, v_2, v_3 ..., v_c} be the set of centers.

1) Randomly select *'c'* cluster centers.

2) Calculate the fuzzy membership $'\mu_{ij}'$ using:

$$\mu_{ij} = 1 / \sum_{k=1}^{c} (d_{ij} / d_{ik})^{(2/m-1)}$$

3) Compute the fuzzy centers $'v_j'$ using:

$$v_j = (\sum_{i=1}^{n} (\mu_{ij})^m x_i) / (\sum_{i=1}^{n} (\mu_{ij})^m), \forall j = 1, 2, ...c$$

4) Repeat step 2) and 3) until the minimum *'J'* value is achieved or $||U^{(k+1)} - U^{(k)}|| < \beta$.

where,

'k' is the iteration step.

'β' is the termination criterion between [0, 1].

'U = $(\mu_{ij})_{n*c}$' is the fuzzy membership matrix.

'J' is the objective function.

Advantages

1) Gives best result for overlapped data set and comparatively better then k-means algorithm.
2) Unlike k-means where data point must exclusively belong to one cluster center here data point is assigned membership to each cluster center as a result of which data point may belong to more then one cluster center.

IV. EXPERIMENT AND RESULT

This research uses data that taken from SMP Negeri 2 Purwodadi. Data that provided by the school was presented in hard copy. Student data that provided by SMP Negeri 2 Purwodadi is data about student name and student's national exam score. This dataset consist of 500 records data, with 10 attributes, and with 2 different generations such as: IPA and IPS.

The original student data given by SMPN 2 Purwodadi can't directly use as the experimental data. It is because of the original student data does not have enough predictor variable attributes that become the requirements of supervised learning in this study.Detail of data are shown on table I.

TABLE I. STUDENT RAW DATA

No		NAMA	TUJUAN	JENIS KELAMIN	UN				US					Jurusan
Urut	Pendaf				BI	Bing	Mtk	IPA	Agm	PKN	IPS	BJ	Seni	TIK	
1	305	ABDULLOH ZAKARIA		L	72	60	32,5	55	92	80	82,5	80	81,7	83,3	IPS
2	523	ABDUR ROCHMAN		L	72	52	35	60	88	87	85	87,9	88,5	87	IPS
3	512	ACHMAD TAUFIQ H		L	78	40	50	35	87,7	85	82,6	81,6	80	93	IPA
4	422	AENI APRILIANTI		P	64	76	60	60	93	93,2	83	83,4	87	88	IPA
5	446	BAROYAH		P	88	78	75	82	91,5	87	85	84,3	91,8	85,5	IPS
6	42	DENI KARTIKA SARI		P	60	82	57,5	45	92,5	80	85	85	85	80	IPA
7	486	DERINTAN ARUS SUKATI		P	82	36	42,5	40	90	78	77,5	82,5	80	83,3	IPA
8	517	DETY KOMARIYAH		P	70	70	45	62,5	79,2	76,7	73,3	75,8	76	85	IPS
9	171	DEWI ATIKA		P	84	36	47,5	57,5	84	83	87	82,49	88	87	IPA
10	290	DEWI KHUMAEDAH		P	70	56	35	47,5	84,6	81,7	82,1	82	86,7	83,3	IPA
11	169	EGA UTAMI		P	72	64	40	45	96	96,3	82	84,3	87	91,5	IPA
12	303	FALA AULIA FADHILAH		P	72	60	45	47	92	89	80	83	78	78	IPA
13	427	FAZHIA ULLY DIANTI		P	72	34	42,5	42,5	88	87	77,5	85	80	80	IPA
14	555	FITRIYA		P	64	52	35	42,5	80,9	83,5	80,8	81	90	88,3	IPS
15	469	HIDAYATUS SIBYAN		P	66	48	35	70	94	92,3	92,6	92,2	88	84	IPS
16	350	IIS DAHLIA		P	70	54	42,5	50	90	86	85	72,5	84	83	IPA
17	27	IKA VIDIARTI		P	68	46	52,5	70	86,6	81,5	82,5	87	77	84,6	IPA
18	473	INDAH NAVITASARI		P	72	58	52,5	52,5	90,7	89,1	88	89,3	90	90	IPA
19	324	KARINA SAHRANI		P	77	67	55	69	87	86	90	74	92	78	IPA
20	83	KARYUTI		P	84	52	37	40	80	80	87,5	81,5	80	82	IPS

To perform experimental testing in this study, there are some differences in data processing/preprocessing phase. It means the data preparation need to be performed in the very beginning of the experimental process before doing a testing. There are some steps taken to get the original student data become an example set used in this study. Those are data selection and data reduction.

From these regulation, cleaning the data in student data from unselected student data record. After the selection process, not all student data is used because of the limitation variables from some students. The student that does not have a valid status information in predictor variable must be deleted. Deleting some attributes in customer data that does not have an integrity for retrieval information. Those attributes are No Urut, No Pendaf, Nama, Tujuan Sekolah and Jenis Kelamin. Those attributes can be deleted because it does not give any information in data mining process. Moreover, those attributes are used only as an information to the school, which is SMP Negeri 2 Purwodadi, about the student. After data reduction process, student data used for this study is ready to use. This data also called as example set. It consists of 290 records of student data with 5 attributes. Those attributes are UN BI, UN
Bing, UN MTK, UN IPA, and Jurusan.

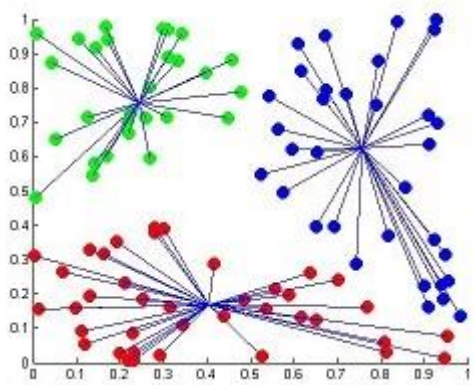


Fig 1 Result of Fuzzy c-means clustering

Figure 1 shows that the Final Result of Clustering that contain final result of whole the calculation process based on the data set that have been uploaded in the form upload. The data here will be divided into 2 types. First 2 clusters and then 3 clusters. So, in the table of clusters will show the detail which data will follow the group.

V. EXPERIMENT AND RESULT

In this study, the implementation of Agglomerative Hierarchical Clustering Algorithm for clustering prediction to determine the major in senior high school get the fair result. The best testing result shown 277 students on cluster 1 and 13 students on cluster 2 from 290 record student data. Agglomerative Hierarchical Clustering Algorithm is the one method that can be applied to determine the major in senior high school that is used to help SMPN 2 Purwodadi student

REFERENCES

[1] M. Nur, "Peminatan pada Pendidikan Menengah," Jakarta : Menteri Pendidikan dan Kebudayaan Republik Indonesia, 2014.
[2] Bahar, "Penentuan Jurusan Sekolah Menengah Atas dengan Algoritma Fuzzy C-Means," Thesis Program Pasca Sarjana Udinus, Semarang, 2011.
[3] J. Han and K. Micheline, Data Mining Concepts and Technique, Morgan Kaufmann, 2001.
[4] Fajar Akbar, "Clustering Data Nilai Siswa SMA Untuk Penentuan Jurusan Menggunakan Algoritma Fuzzy C-Means," 2015.
[5] Marjiyono, " Penerapan Algoritma AHC Algorithm Dalam Aplikasi Pembagian Kelas Siswa Baru," 2015.
[6] J. Han and M. Kamber, Data Mining: Concepts and Techniques Second Edition, San Francisco: Morgan Kaufmann Publishers, 2006.
[7] R. Margo dan E. Ari Eka, Penerapan Algoritma Clustering Untuk Mengelompokkan Ketertarikan Siswa Berdasarkan Aktivitas di Model Pembelajaran Elektronik, 2014. 2018 International Seminar on Application for Technology of Information and Communication (iSemantic) 228