Jurnal Info Sains : Informatika dan Sains
Volume 15, Number 01, 2025, DOI 10.54209/infosains.v15i01
ESSN 2797-7889 (Online)
https://ejournal.seaninstitute.or.id/index.php/InfoSains

Determining The Amount Of Tuition Fees For New Budidarma Students With Data Mining Using The K-Means Clustering Algorithm

Sinar Sinurat
Department Informatic, Faculty Computer Science and Information Technology, Sisingamangaraja Street No.338 Medan

Article Info
Keywords:
PTS,
PTN,
UKT,
Dynamicsome,
Clustering,
K-Means.

ABSTRACT
One of the variables that determines the quality of education in higher education is the amount of fixed tuition fees that must be paid by students to the academic community, which must be in sync with their parents' income. Although the quality of education can be measured from the consistency of supervision, compliance of teachers and students with the System Operating Procedure (SOP), and the availability of complete teaching and learning infrastructure. In private universities, to determine fixed tuition fees, one method that can be used is to find patterns or information on fixed tuition fees at other private universities in the same area (region) by drawing from a large database, which is data mining. It is very important for a private university to know the patterns of prospective students from the data in the database owned by a campus. This technique is the K-Means Clustering Algorithm. The results of the discussion will describe the amount of affordable tuition fees that will be provided with a list with variations based on the study program chosen by prospective students, where each department is distinguished by the completeness of administration and infrastructure in the study program.

This is an open access article under the CC BY-NC license

Corresponding Author:
Sinar Sinurat
Department Informatic, Faculty Computer Science and Information Technology, Sisingamangaraja Street No.338 Medan
sinurat.sin@gmail.com

INTRODUCTION

Fixed Tuition Fee (PTF) or Single Tuition Fee (STF) based on the Regulation of the Minister of Education and Culture of the Republic of Indonesia Number 39 of 2017 [1], the Single Tuition Fee (PTF) is a fee borne by each student based on their economic ability. The single tuition fee is determined based on the economic ability group of the community. This single tuition fee is determined based on the single tuition fee minus the costs borne by the government. "Single Tuition Fee, hereinafter abbreviated as PTF, is the total operational costs directly related to the student learning process per semester in a study program at a state university." With the implementation of this STF, state universities are not allowed to collect initial fees or other levies from new students who have been accepted at the university, but this is different from private universities. The Dynamicsome algorithm and dynamic programming build association rules, which are modifications of the apriori algorithm that searches for

Jurnal Info Sains : Informatika dan Sains
Volume 15, Number 01, 2025, DOI 10.54209/infosains.v15i01
ESSN 2797-7889 (Online)
https://ejournal.seaninstitute.or.id/index.php/InfoSains

frequent itemsets from transaction data. Frequent itemsets are pairs of items found in transaction data. Because the data sources that will be used are usually electronic data, if applied to manual data (in paper form), it will take a long time to achieve the desired pattern from the results of the dynamic programming algorithm [2].

Objectives And Benefits

The objectives of the study of determining the amount of Fixed Tuition Fee (PTF) for prospective new students at Budidarma University Medan are as follows: 1. Stakeholders are helped to recognize the quality of alumni produced. 2. Prospective new students are helped to choose study programs according to market needs. 3. Helping prospective new students determine their choice of study programs based on the PTF offerings from each existing department with the financial capabilities of prospective students or parents, at least based on the Provincial Minimum Wage (PMW). 4. Obtaining more accurate results of affordable tuition fees according to the financial capabilities of the community in the North Sumatra region 5. Can be used as evaluation material for new student tuition fees in the following academic year without significant differences from year to year. 6. As a reference for determining salary adjustments for teaching staff and the provision of adequate infrastructure from year to year.

The urgency of this research is 1. Following up on the description of the problem in the background so that the use of the K-Means Clustering algorithm can be realized for each determination of Fixed Tuition Fees (FTF) according to the capabilities of the local community. 2. The application of the method can consider the availability of data or tuition fee databases. 3. Frequency pattern analysis to determine the support value and frequency of item sets and form association rules to determine the confidence value of the explored database by combining K-Means Clustering.

Literature Review

Data Mining

Data mining is an activity related to large-scale data mining stored in various storage media. Data mining is often also referred to by other terms, such as "exploring" or "digging/mining data." This term is used to decrypt knowledge about database access. Mining data from several storage media can use statistics, mathematics, artificial intelligence, and so on by recognizing and extracting important information according to user criteria.

Dynamicsome

This algorithm is a derivative of the Apriori Algorithm by searching for the frequency of itemsets based on pairs of items obtained from the mining results. On the other hand, to analyze the relationship between different items in a large set of items, this aims to see the relationship and relationship between these items. The stages of the Dynamicsome algorithm include Step 1: Calculate the itemset from support (transactions from all items). After 2-itemsets are obtained from 1-itemset, whether it was previously minimum support, if it has met the minimum support, 1-itemset will be a high-frequency pattern. Step 2: To get 2-itemsets, a combination of the previous k-itemsets must be carried out; the itemset that meets the minimum support will be selected as the high-frequency pattern of the candidate. Step 3: Determining the value of the k-itemset from the support that has met the minimum support of the k-itemset. Step 4: The next iteration process until there are no more k-itemsets that meet the minimum support. [2]

Dynamic programming

Dynamic programming is a problem-solving technique that extracts or describes solutions with a number of stages so that what is obtained is several decisions that are interconnected, where the decision maximizes its full potential [3]. Dynamic programs are usually carried out because of the use of tables to calculate solutions. Dynamic programs have three characteristics, including 1. There are a number of finite data choices that can be selected. 2. Solutions at each stage are interrelated. 3. Optimization conditions are available.

K-means clustering

K-Means Clustering is a process that groups (clusters) by dividing data into several groups. Mac Queen, in his article in 1967, found the K-Means algorithm, where data in one group has the same characteristics as the others but is different from the data in other groups. Steps of the K-Means Clustering algorithm [4] :

1. Selecting k data randomly as the central cluster 2. Calculating the distance between data and the central cluster using Euclidean distance. The distance of all data to each central cluster point is the Euclidean formula, formulated as follows: $D(i,j)=\sqrt{(X_{1i}-X_{qj})^2-+(X_{2i}-X_{2j})^2+...+(X_{ki}-X_{kj})^2}$ where : D(i,j) is the distance of data i to the cluster center j, X_{ki} Is the ith data in the toth data attribute k, X_{kj} Is the central point to j on the kth attribute
2. Data is placed in the nearest cluster, calculated from the center of the cluster
3. The new cluster center is then determined when all data has been assigned to the nearest cluster.
4. The formula for calculating the new cluster center point is : $v=\sum_{i=1}^{n} X_i ; i=1,2,3,...n$ where v is centroid in cluster, X_i is i-th object, n is the number of objects/ number of objects that are members of the cluster.
5. Determine the cluster center and place the data in the cluster repeatedly until the centroid value no longer changes.

RESEARCH METHODS

Basis for determining fixed tuition fees

In determining the Fixed Tuition Fee (FTF) at Budidarma University Medan, a portion of the FTF that will be charged to each prospective new student in each department/study program for diploma programs and undergraduate programs is taken into consideration based on their economic ability or the economic ability of the local (regional) community.

Application of k-means clustering

Determination of FTF at Budidarma University Medan is very suitable and relevant to using K-Means Clustering, including in the division or separation of objects into k separate areas. With the use of K-Means, each object must enter a certain group, but if a stage has entered one group, then it shifts to the next stage, and the object will move to another group. Grouping data using the k-means clustering algorithm is done with the steps of the tree diagram model

Jurnal Info Sains : Informatika dan Sains
Volume 15, Number 01, 2025, DOI 10.54209/infosains.v15i01
ESSN 2797-7889 (Online)
https://ejournal.seaninstitute.or.id/index.php/InfoSains

as in the following diagram : The K-Means Clustering algorithm for its search is 1. Set the number of clusters. 2. Determine the central cluster (centroid). 3. Calculate the distance of the data object to the centroid. 4. Process the data group based on the minimum distance to the centroid. 5. If an object moves to another group or is not the same 6. Then determine step 2 again. 7. The rest of the process is complete.

DISCUSSION AND RESULTS

Registration form prototype

Every prospective new student when registering is required to fill out a form that has been provided in advance by the University of Budidarma Medan, and then each data item filled in the form must provide physical evidence that will be documented by the administration. The administration will receive the registration fee and installments that have been determined by the campus management, and the administration will provide proof of payment to the applicant.

Data transformation based on parental income provisions

In processing data using the KMeans Clustering algorithm, the data of prospective new students to be processed is first carried out through a data transformation process. References in FTF mapping must be considered at Budidarma University Medan. Transformation describes the value of the new FTF in the form of scoring according to the level of socio-economic conditions of parents obtained in previous years. The higher the level of welfare or socio-economic conditions of parents, the higher the score given. Scoring used as a reference for processing this research data is

Table 1.Transformation Of Parent Income data

Father's / Mother's Income	Score
Rp. 0 - < Rp. 2.700.000	1
Rp. 2.700.000 – < Rp. 3.000.000	2
Rp. 3.000.000 – < Rp. 4.000.000	3
Rp. 4.000.000 – < Rp. 5.000.000	4
Rp. 5.000.000 – < Rp. 6.000.000	5
Rp. 6.000.000 – < Rp. 7.000.000	6
Rp. 7.000.000 – < Rp. 8.000.000	7
Rp. 8.000.000 – < Rp. 9.000.000	8
Rp. 9.000.000 – < Rp. 10.000.000	9
Rp. 10.000.000 – < Rp. 12.000.000	10
Rp. 12.000.000 – < Rp. 15.000.000	11
Rp. 15.000.000 – < Rp. 20.000.000	12
> Rp. 20.000.000	13

Table 2. Parental Dependents Data Transformation

Dependents	Score
>=1 orang	5
2 orang	4
3 orang	3
4 orang	2
>= 5 orang	1

Table 3. House Status Data Transformation

Home Ownership	Score
Numpang	1
Sewa Bulanan	2
Sewa Tahunan	3
Rumah Dinas	4
Milik Sendiri	5

Table 4. Land and Building Tax Bill Data Transformation

Land & Building Tax Bill	Score
Rp. 0	1
< Rp. 50.000	2
Rp. 50.000 – Rp. 100.000	3
Rp. 100.000 – Rp. 150.000	4
Rp. 150.000 – Rp. 200.000	5
Rp. 200.000 – Rp. 300.000	6
Rp. 300.000 – Rp. 500.000	7
> Rp. 500.000	8

Jurnal Info Sains : Informatika dan Sains
Volume 15, Number 01, 2025, DOI 10.54209/infosains.v15i01
ESSN 2797-7889 (Online)
https://ejournal.seaninstitute.or.id/index.php/InfoSains

Table 5. Electrically Bill Data Transformation

Electricity Bills	Score
Rp. 0	1
< Rp. 50.000	2
Rp. 50.000 – Rp. 100.000	3
Rp. 100.000 – Rp. 200.000	4
Rp. 200.000 – Rp. 300.000	5
Rp. 300.000 – Rp. 400.000	6
Rp. 400.000 – Rp. 500.000	7
> Rp. 500.000	8

Table 6. Water Bill Data Transfromation

PDAM Bill	Score
Rp. 0	1
< Rp. 25.000	2
Rp. 25.000 – Rp. 40.000	3
Rp.40.000 – Rp. 60.000	4
Rp. 60.000 – Rp. 80.000	5
Rp. 80.000 – Rp. 100.000	6
Rp. 100.000 – Rp. 150.000	7
> Rp. 150.000	8

Table 7. Transformation of Credit Billing Data

HP Credit	Score
Rp0.00	1
< Rp. 40.000	2
Rp. 40.000 – Rp. 60.000	3
Rp. 60.000 – Rp. 80.000	4
Rp. 80.000 – Rp. 100.000	5
Rp. 100.000 – Rp. 150.000	6
Rp. 150.000 – Rp. 250.000	7
> Rp. 80.000	8

Table 8. Television Package Data Transformation

TV Subscription	Score
Tidak Memiliki TV	1
Tidak Langganan/Punya TV	2
Rp. 0 - Rp. 100.000	3
Rp. 100.000 – Rp. 150.000	4
> Rp. 150.000	5

Table 9. Car Text Transformation

Car Tax	Score
Rp. 0	1
Rp. 1 – Rp. 250.000	2
Rp. 250.000 – Rp. 750.000	3
Rp. 750.000 – Rp. 1.250.000	4
Rp. 1.250.000 – Rp. 2.000.000	5
Rp. 2.000.000 – Rp. 2.500.000	6
Rp. 2.500.000 – Rp. 3.000.000	7
> Rp. 3.000.000	8

Table 10. Transformation Motorbike Tax Data

Motorcycle Tax	Score
Rp. 0	1
Rp. 0 – Rp. 100.000	2
Rp. 100.000 – Rp. 200.000	3
Rp. 200.000 – Rp. 300.000	4
Rp. 300.000 – Rp. 400.000	5
Rp. 400.000 – Rp. 500.000	6
Rp. 500.000 – Rp. 600.000	7
> Rp. 600.000	8

After all the data is clustered further with all the transformation tables above, the following sample of prospective new student applicants at Budidarma University is obtained:

Table 11. Overall Data Transformation

Id. Mhs	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12
1	9	2	1	5	2	6	8	8	2	4	3	2
2	5	3	1	5	4	5	7	2	2	8	4	1
3	5	4	6	5	2	4	2	3	2	5	4	1
4	8	3	1	4	2	5	2	7	3	1	5	1
5	8	3	1	4	2	5	2	7	3	1	5	1
6	3	4	1	5	3	8	1	8	2	1	4	1
7	6	2	1	1	2	3	6	7	2	1	4	1
8	7	2	1	5	2	4	5	2	2	1	3	2
9	3	2	1	1	5	8	2	4	2	1	4	1
10	2	5	1	5	2	4	2	2	2	1	4	1
11	5	3	5	5	2	2	2	2	2	1	1	1
12	3	1	1	5	2	6	2	2	2	1	4	1
13	2	2	1	5	3	5	1	5	2	1	1	1
14	2	2	1	3	3	5	2	2	2	1	4	2
15	2	5	1	5	2	4	1	3	2	1	1	1
16	1	3	3	2	1	4	4	2	2	1	3	1
17	1	5	1	5	2	5	2	2	2	1	1	1
18	5	1	5	5	2	1	1	2	2	1	1	1
19	3	1	1	5	2	4	1	3	2	1	3	1
20	2	2	1	5	2	4	2	2	2	1	3	1
21	2	4	2	5	2	2	1	2	2	1	1	1
22	2	3	1	1	2	3	2	2	2	1	5	1
23	2	1	1	5	2	3	2	2	3	1	1	1
24	3	2	1	1	3	3	1	2	2	1	3	1
25	3	1	1	3	1	4	1	4	3	1	1	1
26	3	3	1	1	2	3	2	2	2	1	3	1
27	2	1	1	1	2	4	2	2	1	1	3	1
28	3	3	1	1	3	1	1	3	2	1	1	1
29	3	1	1	1	2	3	1	2	2	1	1	1
30	3	2	1	1	2	3	1	1	2	1	1	1

Jurnal Info Sains : Informatika dan Sains
Volume 15, Number 01, 2025, DOI 10.54209/infosains.v15i01
ESSN 2797-7889 (Online)
https://ejournal.seaninstitute.or.id/index.php/InfoSains

Determining Initial Cluster Centers of Iteration 1

Determining the initial cluster center is done by selecting the data that will be used as the initial centroid, where the FTF magnitude group in the form of a table contains 5 groups, so that the cluster central point (centroid) is selected as many as 5 types of prospective new students as a result of the transformation in Table 12, namely C_0, C_1,C_2, C_3, C_4 with the distribution of member attributes is X1, X2, X3,X4,X5,X6,X7,X8,X9,X10,X11,X12 as seen in the following table :

Table 12. Preliminary Central Data

Sentral Cluster Awal	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12
C0	5	3	5	5	2	2	2	2	2	1	1	1
C1	7	2	1	5	2	4	5	2	2	1	3	2
C2	3	4	1	5	3	8	1	8	2	1	4	1
C3	1	3	3	2	1	4	4	2	2	1	3	1
C4	1	5	1	5	2	5	2	2	2	1	1	1

From table 12, the initial cluster central data C_0 taken from student data with Id_Student 11 that is C1 from Id_student 8, C_1 from Id_student 6, C_3 from Id_student 16 and C_4 from Id_student 17

Calculating the Distance of Data to the Centroid in Iteration-1

With the central cluster, calculate the distance of each data point to the central cluster (centroid) using the Euclidean Distance formula, starting from the distance of prospective student data 1 to the distance of prospective student data 30.

Calculating the distance of student data 1:

$$DC_0=\sqrt{(9-5)^2+(2-3)^2+(1-5)^2+(5-5)^2+(2-2)^2+(6-2)^2+(8-2)^2+(8-2)^2+(2-2)^2+(4-1)^2+(3-1)^2+(2-1)^2}=1161$$
$$DC_1=\sqrt{(9-7)^2+(2-2)^2+(1-1)^2+(5-5)^2+(2-2)^2+(6-4)^2+(8-5)^2+(8-2)^2+(2-2)^2+(4-1)^2+(3-3)^2+(2-2)^2}=787$$
$$DC_2=\sqrt{(9-3)^2+(2-4)^2+(1-1)^2+(5-5)^2+(2-3)^2+(6-8)^2+(8-1)^2+(8-8)^2+(2-2)^2+(4-1)^2+(3-4)^2+(2-1)^2}=1024$$
$$DC_3=\sqrt{(9-1)^2+(2-3)^2+(1-3)^2+(5-2)^2+(2-1)^2+(6-4)^2+(8-1)^2+(8-4)^2+(2-2)^2+(4-1)^2+(3-3)^2+(2-1)^2}=1204$$
$$DC_4=\sqrt{(9-1)^2+(2-5)^2+(1-1)^2+(5-5)^2+(2-2)^2+(6-5)^2+(8-2)^2+(8-2)^2+(2-2)^2+(4-1)^2+(3-1)^2+(2-1)^2}=1264$$

From the distance calculation above, student 1 is the closest to DC1, so student 1 is included in cluster C1. Calculate the distance of student 2's data to the 30th student's data in the same way so that the data in table 13 is obtained and the calculation of determining the distance of the closest result group with the minimum value of prospective student 1 previously where the minimum distance is at DC1, then prospective student 1 enters group C1. And so on until the 30th student so that the following results are obtained :

Jurnal Info Sains : Informatika dan Sains
Volume 15, Number 01, 2025, DOI 10.54209/infosains.v15i01
ESSN 2797-7889 (Online)
https://ejournal.seaninstitute.or.id/index.php/InfoSains

Table 13. Distance Calculationfor All Data

Id Mhs	DC_0	DC_1	DC_2	DC_3	DC_4
1	11,619	7,874	10,247	12,042	12,649
2	10,583	9,849	11,662	9,899	10,344
3	5,657	9,644	9,381	7,616	8,307
4	8,775	7,211	6,403	9,644	9,798
5	8,775	7,211	6,403	9,644	9,798
6	10,198	11,790	0,000	9,165	7,810
7	9,220	8,367	7,937	9,000	10,198
8	8,426	0,000	11,790	10,817	11,136
9	9,747	10,954	6,403	7,141	7,746
10	6,481	10,050	7,483	4,899	3,317
11	0,000	8,426	10,198	6,481	6,708
12	7,000	8,944	7,141	5,568	5,477
13	6,782	10,630	5,657	6,481	4,583
14	7,071	9,849	7,483	4,243	5,000
15	5,916	10,583	7,280	5,745	2,000
16	6,481	10,817	9,165	0,000	5,196
17	6,708	11,136	7,810	5,196	0,000
18	2,449	9,110	11,136	7,483	8,062
19	5,831	9,110	7,211	5,657	5,196
20	5,831	9,539	7,746	4,472	3,873
21	4,472	10,536	9,165	5,477	3,606
22	7,616	10,630	9,055	4,000	6,403
23	5,568	9,899	9,110	5,385	4,690
24	6,633	9,950	9,055	4,899	6,245
25	6,245	9,798	7,681	5,568	5,657
26	6,403	9,592	9,000	3,873	5,657
27	7,348	10,440	9,055	4,000	6,245
28	6,325	10,630	10,000	6,000	6,557
29	6,481	10,149	9,798	5,292	6,403
30	6,325	10,149	10,198	5,099	5,916

Table 14. Data Allocation to Centroids

Id. Mhs	DC0	DC1	DC2	DC3	DC4	Cluser
1	11,619	7,874	10,247	12,042	12,649	C1
2	10,583	9,849	11,662	9,899	10,344	C1
3	5,657	9,644	9,381	7,616	8,307	C0
4	8,775	7,211	6,403	9,644	9,798	C2
5	8,775	7,211	6,403	9,644	9,798	C2
6	10,198	11,790	0,000	9,165	7,810	C2
7	9,220	8,367	7,937	9,000	10,198	C2
8	8,426	0,000	11,790	10,817	11,136	C1
9	9,747	10,954	6,403	7,141	7,746	C2
10	6,481	10,050	7,483	4,899	3,317	C4
11	0,000	8,426	10,198	6,481	6,708	C0
12	7,000	8,944	7,141	5,568	5,477	C4
13	6,782	10,630	5,657	6,481	4,583	C4
14	7,071	9,849	7,483	4,243	5,000	C3
15	5,916	10,583	7,280	5,745	2,000	C4
16	6,481	10,817	9,165	0,000	5,196	C3
17	6,708	11,136	7,810	5,196	0,000	C4
18	2,449	9,110	11,136	7,483	8,062	C0
19	5,831	9,110	7,211	5,657	5,196	C4
20	5,831	9,539	7,746	4,472	3,873	C4
21	4,472	10,536	9,165	5,477	3,606	C4
22	7,616	10,630	9,055	4,000	6,403	C3
23	5,568	9,899	9,110	5,385	4,690	C4
24	6,633	9,950	9,055	4,899	6,245	C3
25	6,245	9,798	7,681	5,568	5,657	C3
26	6,403	9,592	9,000	3,873	5,657	C3
27	7,348	10,440	9,055	4,000	6,245	C3
28	6,325	10,630	10,000	6,000	6,557	C3
29	6,481	10,149	9,798	5,292	6,403	C3
30	6,325	10,149	10,198	5,099	5,916	C3

From the grouping results in Table 4.15, the data grouping in iteration 1 is

Table 15. Data Grouping Results

Kelompok	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12
C0	5	4	6	5	2	4	2	3	2	5	4	1
	5	3	5	5	2	2	2	2	2	1	1	1
	5	1	5	5	2	1	1	2	2	1	1	1
C1	9	2	1	5	2	6	8	8	2	4	3	2
	5	3	1	5	4	5	7	2	2	8	4	1
	7	2	1	5	2	4	5	2	2	1	3	2
C2	8	3	1	4	2	5	2	7	3	1	5	1
	8	3	1	4	2	5	2	7	3	1	5	1
	3	4	1	5	3	8	1	8	2	1	4	1
	6	2	1	1	2	3	6	7	2	1	4	1
	3	2	1	1	5	8	2	4	2	1	4	1
C3	2	2	1	3	3	5	2	2	2	1	4	2
	1	3	3	2	1	4	4	2	2	1	3	1
	2	3	1	1	2	3	2	2	2	1	5	1
	3	2	1	1	3	3	1	2	2	1	3	1
	3	1	1	3	1	4	1	4	3	1	1	1
	3	3	1	1	2	3	2	2	2	1	3	1
	2	1	1	1	2	4	2	2	1	1	3	1
	3	3	1	1	3	1	1	3	2	1	1	1
	3	1	1	1	2	3	1	2	2	1	1	1
	3	2	1	1	2	3	1	1	2	1	1	1
C4	2	5	1	5	2	4	2	2	2	1	4	1
	3	1	1	5	2	6	2	2	2	1	4	1
	2	2	1	5	3	5	1	5	2	1	1	1
	2	5	1	5	2	4	1	3	2	1	1	1
	1	5	1	5	2	5	2	2	2	1	1	1
	3	1	1	5	2	4	1	3	2	1	3	1
	2	2	1	5	2	4	2	2	2	1	3	1
	2	4	2	5	2	2	1	2	2	1	1	1
	2	1	1	5	2	3	2	2	3	1	1	1

Jurnal Info Sains : Informatika dan Sains
Volume 15, Number 01, 2025, DOI 10.54209/infosains.v15i01
ESSN 2797-7889 (Online)
https://ejournal.seaninstitute.or.id/index.php/InfoSains

Calculating The New Cluster Central to Iteration 2

Calculate new cluster centers for iteration-2. Calculate new cluster centers C_0

$$C_0(X_1)=\frac{(5+5+5)}{3}=5{,}00 \quad C_0(X_2)=\frac{(4+3+1)}{3}=2{,}67 \quad C_0(X_3)=\frac{(6+5+5)}{3}=5{,}33$$

$$C_1(X_1)=\frac{(9+5+7)}{3}=7{,}00 \quad C_1(X_2)=\frac{(2+3+2)}{3}=2{,}33 \quad C_1(X_3)=\frac{(1+1+1)}{3}=1{,}00$$

$$C_2(X_1)=\frac{(8+8+8+3)}{4}=5{,}60 \quad C_2(X_2)=\frac{(3+3+4+2)}{4}=2{,}80 \quad C_2(X_3)=\frac{(1+1+1+1)}{4}=1{,}00$$

$$C_3(X_1)=\frac{(6+2+1+2+3+3+3+2+3+3+3)}{11}=2{,}50 \quad C_3(X_2)=\frac{(2+2+3+3+2+1+3+1+3+1+2)}{11}=2{,}10$$

$$C_3(X_3)=\frac{(1+1+3+1+1+1+1+1+1+1+1)}{11}=1{,}20 \quad C_4(X_1)=\frac{(2+3+2+2+1+3+2+2+2)}{9}=2{,}11$$

$$C_4(X_2)=\frac{(5+1+2+5+5+1+2+4+1)}{9}=2{,}88 \quad C_4(X_3)=\frac{(1+1+1+1+1+1+1+2+1)}{9}=1{,}11$$

And so on up to X12 in the same way as above, then.

Table 16. New Cluster Central Results

Sentral cluster baru	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12
C0	5	2,66	5,33	5	2	2,33	1,66	2,33	2	2,33	2	1
C1	7	2,33	1	5	2,66	5	6,66	4	2	4,33	3,33	1,66
C2	5,6	2,8	1	3	2,8	5,8	2,6	6,6	2,4	1	4,4	1
C3	2,5	2,1	1,2	1,5	2,1	3,3	1,7	2,2	2	1	2,5	1,1
C4	2,11	2,88	1,11	5	2,11	4,11	1,55	2,55	2,11	1	2,11	1

Calculating the Distance of New Cluster Central Data to Iteration 2

Calculating the distance of each prospective student's data to each central cluster (centroid), starting from the distance of student data 1 to the distance of student data 30. Calculation of the distance of prospective student data 1

$$DC_0=\sqrt{\begin{aligned}&(9-5{,}00)^2+(2-2{,}66)^2+(1-5{,}33)^2+(5-5{,}00)^2+(2-2{,}00)^2+(6-2{,}33)^2+(8-1{,}67)^2+\\&(8-2{,}33)^2+(2-2{,}00)^2+(4-2{,}33)^2+(3-2{,}00)^2+(2-1{,}00)^2\end{aligned}}=11{,}21$$

$$DC_1=\sqrt{\begin{aligned}&(9-7{,}00)^2+(2-2{,}33)^2+(1-1{,}00)^2+(5-5{,}00)^2+(2-2{,}66)^2+(6-5{,}00)^2+(8-6{,}67)^2+\\&(8-4{,}33)^2+(2-2{,}00)^2+(4-4{,}33)^2+(3-3{,}33)^2+(2-1{,}67)^2\end{aligned}}=4{,}48$$

$$DC_2=\sqrt{\begin{aligned}&(9-5{,}60)^2+(2-2{,}80)^2+(1-1{,}00)^2+(5-3{,}00)^2+(2-2{,}80)^2+(6-5{,}10)^2+(3-4{,}40)^2+\\&(3-3{,}44)^2+(2-2{,}00)^2+(4-4{,}00)^2+(3-3{,}00)^2+(2-1{,}00)^2\end{aligned}}=7{,}75$$

$$DC_3=\sqrt{\begin{aligned}&(9-2{,}50)^2+(2-2{,}10)^2+(1-1{,}20)^2+(5-1{,}50)^2+(2-2{,}10)^2+(6-3{,}30)^2+(8-1{,}70)^2+\\&(8-2{,}20)^2+(2-2{,}00)^2+(4-1{,}00)^2+(3-2{,}50)^2+(2-1{,}10)^2\end{aligned}}=12{,}05$$

$$DC_4=\sqrt{\begin{aligned}&(9-2{,}10)^2+(2-2{,}80)^2+(1-1{,}11)^2+(5-5{,}00)^2+(2-2{,}11)^2+(6-4{,}11)^2+(8-1{,}55)^2+\\&(8-2{,}55)^2+(2-2{,}11)^2+(4-1{,}00)^2+(3-2{,}11)^2+(2-1{,}00)^2\end{aligned}}=11{,}56$$

Jurnal Info Sains : Informatika dan Sains
Volume 15, Number 01, 2025, DOI 10.54209/infosains.v15i01
ESSN 2797-7889 (Online)
https://ejournal.seaninstitute.or.id/index.php/InfoSains

Calculating the Distance of New Cluster Central Data to Iteration 3

The closest distance to the minimum value of prospective student 1 is obtained before the distance calculation (DC0, DC1, DC2, DC3, DC4), where the minimum distance is at DC1; then student 1 enters group C1. Thus, the next process is up to 30 prospective students. After grouping all data, the results are in table 4.19. Then compare the results of the iteration-2 grouping. From the comparison of the iteration-2 grouping with iteration-1, the group and members of the cluster group C0, C1, C2, C3, and C4, the results are the same. Thus the K-Means Clustering algorithm process is stopped at iteration 2 because there are no more changes/transfers of object group members between iteration 2 and iteration 1.

Table 17. Data Distance Calculation Results

Id. Mhs	DC0	DC1	DC2	DC3	DC4
1	11,210	4,865	7,754	12,052	11,568
2	9,730	4,967	9,814	10,259	9,760
3	4,082	7,572	8,020	8,040	7,313
4	8,347	7,000	3,053	8,405	8,099
5	8,347	7,000	3,053	8,405	8,099
6	9,798	9,381	4,682	8,686	7,175
7	9,110	6,557	5,032	7,513	8,733
8	6,403	4,435	6,366	6,711	6,221
9	9,434	8,963	5,340	6,053	6,850
10	6,481	8,446	6,893	4,883	2,931
11	1,826	8,524	8,290	6,151	5,450
12	6,708	7,550	6,043	4,841	3,469
13	6,831	8,679	5,926	5,181	3,151
14	6,880	8,165	6,093	3,072	3,388
15	6,298	9,147	7,275	4,984	2,502
16	6,532	8,718	7,611	3,720	4,706
17	7,095	9,292	8,032	5,314	2,874
18	2,828	9,557	9,095	6,499	6,212
19	5,657	8,000	5,960	3,955	2,388
20	5,831	8,000	6,717	3,666	1,466
21	5,033	9,557	8,362	4,609	2,893
22	7,439	9,274	6,893	2,800	5,114
23	5,916	8,660	7,882	4,152	2,715
24	6,633	9,165	6,717	1,497	4,587
25	6,351	8,699	6,142	3,527	3,747
26	6,506	8,622	6,551	1,356	4,401
27	7,303	9,092	7,037	1,908	4,706
28	6,733	9,933	7,818	3,292	5,389
29	6,733	9,557	7,611	2,154	4,846
30	6,683	9,730	8,094	2,200	4,776

Table 18. Data Distance Grouping Results

Id. Mhs	DC0	DC1	DC2	DC3	DC4	DC4	Cluser
1	11,210	4,865	7,754	12,052	11,568	11,568	C1
2	9,730	4,967	9,814	10,259	9,760	9,760	C1
3	4,082	7,572	8,020	8,040	7,313	7,313	C0
4	8,347	7,000	3,053	8,405	8,099	8,099	C2
5	8,347	7,000	3,053	8,405	8,099	8,099	C2
6	9,798	9,381	4,682	8,686	7,175	7,175	C2
7	9,110	6,557	5,032	7,513	8,733	8,733	C2
8	6,403	4,435	6,366	6,711	6,221	6,221	C1
9	9,434	8,963	5,340	6,053	6,850	6,850	C2
10	6,481	8,446	6,893	4,883	2,931	2,931	C4
11	1,826	8,524	8,290	6,151	5,450	5,450	C0
12	6,708	7,550	6,043	4,841	3,469	3,469	C4
13	6,831	8,679	5,926	5,181	3,151	3,151	C4
14	6,880	8,165	6,093	3,072	3,388	3,388	C3
15	6,298	9,147	7,275	4,984	2,502	2,502	C4
16	6,532	8,718	7,611	3,720	4,706	4,706	C3
17	7,095	9,292	8,032	5,314	2,874	2,874	C4
18	2,828	9,557	9,095	6,499	6,212	6,212	C0
19	5,657	8,000	5,960	3,955	2,388	2,388	C4
20	5,831	8,000	6,717	3,666	1,466	1,466	C4
21	5,033	9,557	8,362	4,609	2,893	2,893	C4
22	7,439	9,274	6,893	2,800	5,114	5,114	C3
23	5,916	8,660	7,882	4,152	2,715	2,715	C4
24	6,633	9,165	6,717	1,497	4,587	4,587	C3
25	6,351	8,699	6,142	3,527	3,747	3,747	C3
26	6,506	8,622	6,551	1,356	4,401	4,401	C3
27	7,303	9,092	7,037	1,908	4,706	4,706	C3
28	6,733	9,933	7,818	3,292	5,389	5,389	C3
29	6,733	9,557	7,611	2,154	4,846	4,846	C3
30	6,683	9,730	8,094	2,200	4,776	4,776	C3

Evaluation Of Calculation Results

Based on the stages of the K-Means Clustering process that have been carried out, it is known that there are 5 groups of FTF amounts for prospective new students consisting of

1. Cluster C0 is in the FTF category V cluster, where the cost value is 4,000,000.00, namely students 3, 11, and 3,11,18 total 3 people.

2. Cluster C1 is in the FTF category VI cluster, where the cost value is 3,000,000.00, namely students 1, 2, and 1,2,8 total 3 people.
3. Cluster C2 is in the FTF category III cluster, where the cost value is 2,000,000.00, namely students 4, 5, 6, 7, and 9, for a total of 5 people.
4. Cluster C3 is in the FTF category II cluster, where the cost value is 1,000,000.00, namely students 14, 16, 22, 24, 25, 26, 27, 28, 29, and 30, for a total of 10 people.
5. Cluster C4 is in the FTF category I cluster, where the cost value is 500,000.00, namely students 10, 12, 13, 15, 17, 19, 20, 21, and 23, for a total of 9 people.

Based on the application of the K-Means Clustering algorithm, it is able to group the FTF value of prospective new students at Budidarma University Medan into 5 groups based on the level of socio-economic conditions of parents; thus, it can be used as a supporting model in decision-making regarding the provisions of the FTF value of prospective new students.

CONCLUSSIONS

Based on the overall results of the discussion above, the conclusions are as follows: Based on the perspective of a decision support system where a number of criteria for consideration in determining FTF each year always adjust the regeneration of human resources, the facilities needed are adjusted into 5 groups based on the level of socio-economic conditions of the parents. Based on the calculation results with K-Means Clustering in the discussion, it is adjusted into 5 groups based on the level of socio-economic conditions of the parents.

RECOMMENDATIONS

Based on the analysis during the research phase, several suggestions are needed for the development of this research in the future, including: Public feedback is needed regarding the amount of FTF through questionnaires or face-to-face meetings with parents so that both parties accept each other. The calculation process using K-Means Clustering seems to need to consider the behavior of prospective students, which allows for certain funding projections in the form of social assistance from various parties.

REFERENCES

[1]. BN 2017/ NO 779; KEMENRISTEKDIKTI.GO.ID : 7 HLM, "Peraturan Menteri Riset, Teknologi dan Pendidikan Tinggi RI nomor 39 Tahun 2017, Tentang Biaya Uang Kuliah Tunggal dan Uang Kuliah Tunggal Pada Perguruan Tinggi Negeri di Lingkungan Kementerian Riset, Riset dan Teknologi dan Pendidikan Tinggi"
[2]. Prasetyani R., Djatna T., "Data Mining Dengan Algoritma Dynamicsome Untuk Penentuan Pengiriman Stock Yang Belum Dikirim Pupuk Subsidi", Journal Of Informatic And Advanced Computing (JIAC), Vol.3, No.1 Mei 2022
[3]. Kamal A., "Analisis Perbandingan Algoritma Dynamic Programming Dengan Pendekatan Forward Dan Backward Melalui Hasil Studi Kasus Distribusi Produk Air Minum Kemasan Galon Di Depot Air Minum Isi Ulang Banyu Belik, Purwokerto", 2017
[4]. Kurniawan H., Defit S., Sumijan, "*Data Mining* Menggunakan Metode *K-Means*

Clustering Untuk Menentukan Besaran Uang Kuliah Tunggal", Journal of Applied Computer Science and Technology (Jacost), Vol.1 ISSN 2723-1453,Tahun 2020

[5]. Triase, Syamsudin, "Implementasi Data Mining Dalam Mengklasifikasikan UKT (Uang Kuliah Tunggal) Pada UIN Sumatera Utara Medan", Jurnal Teknologi Informasi) Vol.4, No.2, Desember 2020, P-ISSN 2580-7927 E-ISSN 2615-2738

[6]. Karim B., Sentinuwo S.R., Sambul A.M., "Penentuan Besaran Uang Kuliah Tunggal untuk Mahasiswa Baru di Universitas Sam Ratulangi Menggunakan *Data Mining*", E-Journal Teknik Informatika Vol 11, No.1 (2017) ISSN: 2301-8364

[7]. Undang-Undang Republik Indonesia Nomor 12 Tahun 2012 Tentang Pendidikan Tinggi

[8]. Peraturan Menteri Pendidikan Dan Kebudayaan Republik Indonesia Nomor 55 Tahun 2013 Tentang Biaya Kuliah Tunggal Dan Uang Kuliah Tunggal Pada Perguruan Tinggi Negeri Di Lingkungan Kementerian Pendidikan Dan Kebudayaan

[9]. Rahmatullah M.F., "Disain dan Analisis Algoritma Pemrograman Dinamis Model Tree Pada Permasalahan URI OnLIne Judge 2911 Ink Colors, Institut Teknologi Surabaya, Tahun 2020

[10]. Prasetiyowati M.I., Wicaksana A., "Implementasi Algoritma Dynamic Programming Untuk Multiple Constraint Knapsack Problem (Studi Kasus : Pemilihan Media Promosi di UMN", Seminar Nasional Aplikasi Teknologi Informasi (SNATI) ISSN : 1907:5022 1 Juni 2013

[11]. Risnawati, Rahmadani, "Implementasi Metode K-Means Clustering Tunggakan Uang Kuliah Pada Stmik Royal Kisaran", Journal of Science and Social Research ISSN 2615 – 4307 Feb 2022, V (1): 118 – 124

[12]. Isnaeni S., Supriyono, Arini F.Y., "Implementasi Algoritma Pemrograman Dinamik Untuk Penyelesaian Persoalan Knapsack Dalam Penentuan Keuntungan Optimal Penjualan Barang" Unnes Journal of Mathematics, ISSN 2252-6943 UJM 3 (2) (2014)

[13]. Chairun N., "*Data Mining* Prediksi Minat Calon Mahasiswa Memilih Perguruan Tinggi Menggunakan Algoritma C4.5" Jurnal Manajemen Informatika (JAMIKA), Volume 11 Nomor 2 Edisi Oktober 2021, E ISSN: 2655-6960 | P ISSN: 2088-4125