

Implementation of K-Means Clustering Algorithm in Mapping the Groups of Graduated or Dropped-out Students in the Management Department of the National University

Muhammad Darwis¹, Liyando Hermawan Hasibuan², Mochammad Firmansyah³, Nur Ahady⁴, Rizka Tiaharyadini⁵

¹Program Studi Magister Ilmu Komputer, Fakultas Teknologi Informasi, Universitas Budi Luhur

²Program Studi Magister Ilmu Komputer, Fakultas Teknologi Informasi, Universitas Budi Luhur

³Program Studi Magister Ilmu Komputer, Fakultas Teknologi Informasi, Universitas Budi Luhur

⁴Program Studi Magister Ilmu Komputer, Fakultas Teknologi Informasi, Universitas Budi Luhur

⁵Program Studi Magister Ilmu Komputer, Fakultas Teknologi Informasi, Universitas Budi Luhur

email: ¹darwis.mawardin@gmail.com, ²liyandohermawan@gmail.com, ³bgolem@gmail.com ⁴nahady18@gmail.com,
⁵rizkatiaharyadini27@gmail.com

Abstract - *The dropout rate at the National University is still high. National Universities must make efforts to anticipate this rate and increase the number of graduates. This study aims to determine the characteristics of students who are likely to graduate or drop out (DO) in the management department of the National University, Jakarta. The study was conducted by implementing the K-Means algorithm, where each data is grouped according to the closest distance to the centroid. Determination of Cluster C1 graduate or C2 drop out is based on the attributes of status of students (active, leave, out and non-active), educational status (graduated or DO), GPA, total credits taken and length of study. To facilitate the clustering process, Orange tools are used that provide K-Means algorithm features. The total data input in this study were 1988 students from various classes. As a result, a pattern or mapping of graduated or DO students was found based on the attributes mentioned earlier. Testing the results of this cluster with the silhouette method, by measuring the distance between cluster members, both C1 and C2, showed good Silhouette value, reaching 85% which indicates that this clustering method can be applied as an effort to overcome the high dropout rate. The management department, National University can use the results of this study to predict the graduation of their students.*

Keywords - data mining; clustering; K-mean algorithm; orange; graduation; mapping;

I. INTRODUCTION

The National University management department is part of the faculty of economics which was founded in 1964, then its status was registered and recognized in 1985 by the Ministry of Education and Culture. Since then, the department has successfully graduated thousands of graduates, who have worked in various sectors in companies, government agencies, universities, and also entrepreneurs. The National University has been recognized as one of the higher education institutions in Indonesia [1].

However, along the way, not a few students have dropped out (DO) in the management department of the National University. This has an impact on the ups and downs of credibility and public trust of the university. To overcome this and maximize the learning process, it is better for the National University to know the patterns and mapping of student groups who have passed or dropped out, based on their academic data. Thus, the National University can make preventive efforts to prevent dropouts in students and maintain their credibility and reputation.

Data mining can provide solutions to the National University through the clustering method. One of the features that can be used in this case is the K-Means Clustering algorithm. With this technique, a pattern or mapping will be produced that can be used by the National University to determine the characteristics of graduated or dropped-out students based on their academic data so far.

The K-Means clustering algorithm has been widely used by various researchers to group objects based on their conditions and characteristics. As long as data is available, this method can be applied in various fields, such as health, education, disaster, military and so on [2], [3], [4], [5], [6], [7], [8], [9], [10], [11]. In the field of education, especially the clustering method at universities, researchers usually try to make patterns or mappings that are related to students, such as interests, academic conditions and students' graduation [12], [13], [14], [15], [16], [17], [18].

The National University has documented the student data well in their system. However, the data is still in raw form so it is difficult to read. To better utilize and maximize the existence of this data, the National University should implement data mining in it. Besides being very helpful, data mining methods are trusted and have been widely used

by various other institutions to extract more value from their data.

Therefore, in this study, the data processing and grouping of students in the Management Department of the National University were carried out using the clustering method. Just as other researchers have done, this study will utilize the K-Means clustering method to do this. The number of datasets processed in this study were 1,988 records. With this method, a pattern and mapping will be generated that can be used to see and predict the likelihood that students will graduate or drop out.

II. RESEARCH METHOD

The research methodology in the clustering process using the K-Means method in mapping the groups of the graduated or dropped-out students at the Management Department of the National University includes several stages: literature review, determination of the data sets and data pre-processing—which consists of data validation, data transformation and data reduction as well as data exploration using Orange application. To test the results of this study, silhouette was applied. The results will show whether or not this clustering method can be used as an effort to overcome the high dropout rate at the National University. The description of the flow and stages of the research that the authors did for this study is as shown in Figure 1.

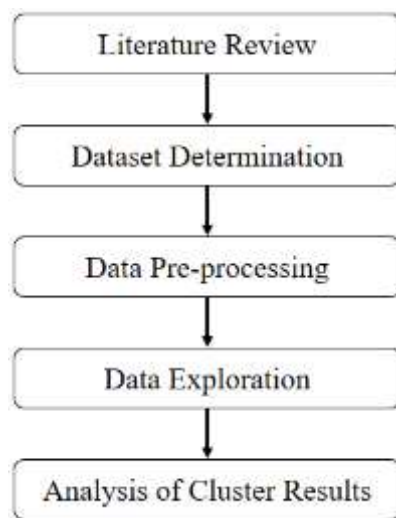


Figure 1. Flow and research stages

A. Literature Review

As literature to support this research, the authors use several trusted scientific journals, proceedings, e-books and websites. These papers contain about clustering, such as the Application of the K-Means Method for Student Clustering Based on Academic Values with a Weka Interface Case Study at the Department of Informatics, UMM Magelang [12], Prediction of Students Academic Execution Using K-Means and K-Medoids Clustering Technique [12] [18], K-Means Cluster Analysis in the Grouping of Students' Capabilities [17] and the Implementation of the K-Means Method in Mapping Student Groups through Lecture Activity Data [15].

Based on the literature review, it is known that the K-means clustering method in education has been widely used, especially in grouping students, both learning activities and predicting academic scores. The percentage of grouping shows good results, so that it can be applied to solve problems at the university. The focus of this research is to cluster students who are likely to graduate or drop out (DO) using the K-Means method.

B. Dataset Determination

The data set in this study was obtained from the Department of Management, the National University, Jakarta. Furthermore, the data is processed to the data pre-processing stage. The number of datasets processed in this study were 1,988 records or rows of data containing detailed information about the academic data of students majoring in Management at the National University of various generations, from 1990 to 2016. In detail, from the number of student data collected in this study, 56% were male and 44% were female. The number of student data from the class of 1990-1995 was 48% or 960 students, the class of 1996-2000 was 17% or 343 students, the class of 2001-2005 was 20% or 390 students, the class of 2006-2010 was 3% or 65 students, the class of 2010-2016 was 12% or 230 students. For this study, several attributes were used as consideration for creating student clusters, whether they pass or drop out. These attributes are the GPA score, the number of credits taken, length of study, initial status and final status of students. From these attributes, it is hoped that a cluster will be formed that can help the National University to make predictions for students' graduation or drop-out.

C. Data Pre-processing

This stage includes 3 main parts: data validation, data transformation, and data reduction. For information, to facilitate the implementation of this stage, several features in the Orange application are used. The details of the three stages are:

1. Data validation

The data validation stage in this study was carried out to ensure that the training data set was in good condition and there were no more missing values in it. The missing value in question is due to incomplete data, outliers (abnormal data) or data with inconsistent values. From the data validation stage, data will be generated where there are no missing values or normal conditions for the next stage.

2. Data transformation

The next stage is data transformation, which is performed with the aim of maintaining and ensuring the accuracy of the training data set. For this stage, the outlier technique contained in the Orange application is also applied, namely the Outlier Detection Method feature. From this data transformation process, more accurate data will be generated for the next stage.

3. Data reduction

Data reduction is done to take data sampling from the existing training set. The goal is that the selected data sets are really ready to be classified. Of the total 1,988 data records, then filtering was carried out until the remaining 60% of the data or 1,173 records were left. So, from this

stage, a sample data will be generated that is ready to be processed to the next stage.

D. Data Exploration

This stage is one of the important stages in the clustering process, because this is where the data will actually be processed: the value is calculated and the results are analyzed. In this study, data processing and calculation methods were carried out using the features in the Orange application. There are several features that will be used in it, such as the outlier feature, K-Means, select rows and a scatter plot to see the clustering results.

E. Clustering Result Analysis

At this stage, the results of the classification or grouping that have been obtained from calculations using the Orange application are analyzed and reviewed. From this stage, patterns might be as the results of K-Means clustering.

III. RESULTS AND DISCUSSION

Research for the clustering of students majoring in Management at the National University starts from collecting training data sets, designing and implementing data mining. Later, a certain pattern will be generated that can be used to predict the possibility of students graduating or dropping out. These stages include identification of training sets, data preparation, data exploration using Orange and analysis of the results.

A. Data Set

The first stage carried out in this study was the collection and definition of the dataset as shown in Figure 2. In this data, which contains data from students of the Management Department of the National University of various generations from 1990 to 2016, there are 1,988 data records available. The dataset is then uploaded to Orange to facilitate data processing, including data preparation. In this set, it was seen that there were still missing values of 0.3% or around 34 rows or data records.

IDNo	IPK	SKS	Lulus Skori	Status Akademik	Status Studi
1	2.75	144.0	5.70911	1.0	1
2	1.50	80.0	4.99932	2.0	2
3	1.10	144.0	7.00080	1.0	1
4	1.25	144.0	7.00080	1.0	1
5	1.75	95.0	8.99922	2.0	2
6	2.00	144.0	6.00011	1.0	1
7	1.00	144.0	6.00011	1.0	1
8	1.00	144.0	6.00011	1.0	1
9	1.75	144.0	8.99911	1.0	1
10	1.75	75.0	9.99911	1.0	1
11	2.00	144.0	8.00011	2.0	2
12	1.25	62.0	8.99932	2.0	2
13	2.00	144.0	7.00080	1.0	1
14	1.00	144.0	7.00080	1.0	1
15	2.00	144.0	7.00080	1.0	1
16	2.00	144.0	7.00080	1.0	1
17	1.75	80.0	8.99932	2.0	2
18	1.00	144.0	4.99932	1.0	1
19	1.00	144.0	4.99932	1.0	1
20	2.00	144.0	4.99932	1.0	1
21	1.15	144.0	4.99932	1.0	1
22	1.25	45.0	8.99932	2.0	2
23	1.15	144.0	4.99932	1.0	1
24	1.25	144.0	4.99932	1.0	1
25	2.00	144.0	4.99932	1.0	1
26	2.75	144.0	4.99932	1.0	1
27	1.00	144.0	4.99932	1.0	1
28	1.00	144.0	4.99932	1.0	1
29	1.00	50.0	8.99932	2.0	2
30	1.15	144.0	4	1.0	1
31	1.25	144.0	4	1.0	1
32	1.00	144.0	4	1.0	1
33	1.75	144.0	4	1.0	1
34	1.00	144.0	4	1.0	1
35	1.25	70.0	8.99932	2.0	2

Figure 2. Dataset of Management Department students, the National University, Jakarta

From the data table, it can be seen that there are at least 5 attributes that have been selected from the data set and then processed to form a cluster. These attributes include:

- GPA, contains GPA data records for each student and is a numerical attribute. The value is a range of GPA numbers from 0-4 which have been classified based on the Regulation of the Minister of Education and Culture of the Republic of Indonesia Number 49 of 2014 concerning National Higher Education Standards regulating the assessment and cumulative achievement index contained in articles 23 and 24. Assessment reports are in the form of students' success qualifications in taking a course stated in the range of:
 - GPA \leq 4, in good category.
 - GPA \leq 3, in sufficient category.
 - GPA \leq 2, in the poor category.
 - GPA \leq 1, in the very poor category.
- SKS, is a numeric attribute that contains data on the number of credits that have been taken by students. The values are in the range of 1 to 144.
- Duration of study, namely data on the length of study that students have taken until they graduate or drop out and is a numerical attribute.
- Initial Status, is student status data during the lecture process. The authors has classified this data and made it a numerical attribute. The value is 1 for "active" status, 2 for "inactive" status, 3 for "leave" status and 4 for "out" status.
- Final Status is data on the final academic status of students, graduated or dropped out (DO). The authors made it as a numerical attribute with a value of 1 for "pass" status and 2 for "Drop Out" status.

B. Data Pre-processing

This section contains the data preparation stage, which consists of data validation, data transformation and data reduction. The authors use the Orange application in all stages of this preparation to make the data processing easy and effective.

a. Data validation

As previously explained, there are missing values in the dataset used in this study, so it is necessary to pre-process the data first. This is also to improve the validation of existing data. For this reason, the Preprocess feature contained in Orange is used to do this. Furthermore, there is also a Data Table feature that can be used to view data, including data from the validation results. As a result, the missing values in the dataset were eliminated so that the number of data became 1,954 records and could be continued to the next process. These results are as shown in Figure 3.



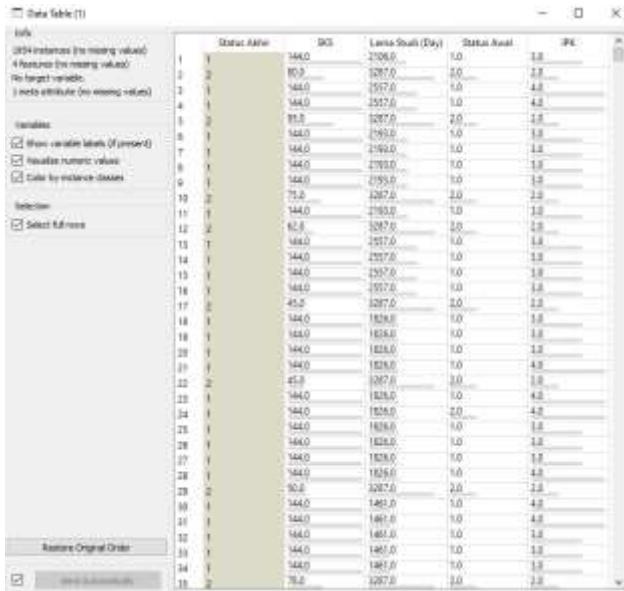


Figure 3. Results of data validation

b. Data transformation

To perform the data transformation process, the Outlier feature in Orange is used. Outlier means there is an unnatural (anomalous) data set in the data set, so that it must be corrected before the next stage. After that, the results are displayed in the form of Data Inliers, which means data that is not normal or all data other than normal data. The description of the Outlier feature as a transformation process in Orange is as shown in Figure 4.

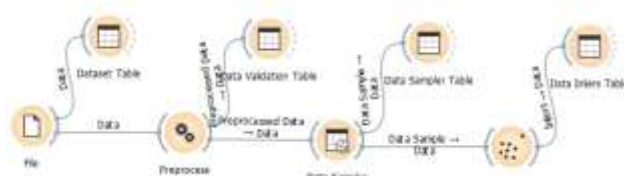


Figure 4. Outlier process in Orange

As additional information, before the dataset is entered into the Orange application, data creations have been carried out. The data creations are the addition of several attributes that are deemed necessary and changing some attributes that contain categorical data into numeric data. This aims to facilitate the classification process with the K-Means feature. These attributes include initial status, final status and length of study as described in the dataset section of this paper.

c. Data reduction

This data reduction is basically a sampling of existing data, especially because there are quite a lot of rows. In addition to streamlining the algorithm, it is also less heavy and doesn't take long to process data. Although several rows of data are reduced at this stage, the quality of the resulting data remains the same, so that it still meets the research requirements.

In the Orange application, to perform this stage, the "Data Sampler" feature can be used which is linked from the Preprocess feature. As previously explained, the data used at this stage is about 60% of the total data or as many as 1,954 records. The technique used is random sampling where the data will be randomly selected. From the "Data Sampler" output, the Outlier feature is then used, so that

later, inlier data is obtained, namely data other than data from outlier data. The data is then processed further so that a total of 592 records are obtained from all existing data. Inliers data obtained as output is as shown in Figure 5. The inliers data will be processed further at a later stage. At this stage, the data is really ready to be processed and explored further.

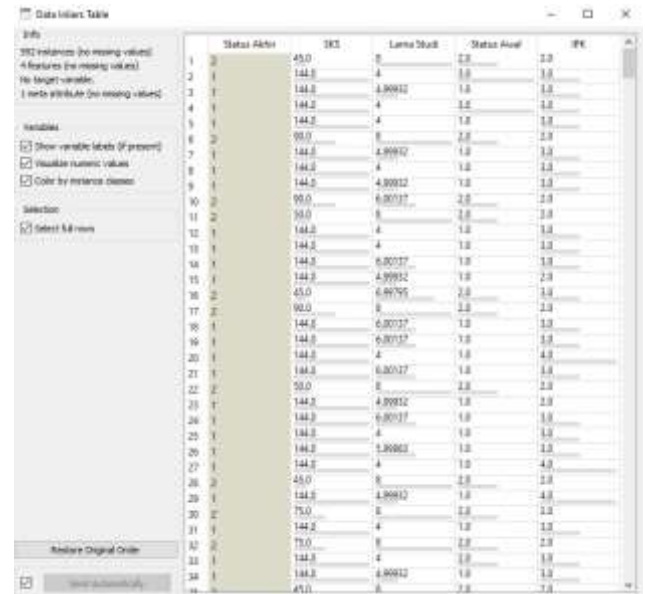


Figure 5. Data sampling inliers

d. Data exploration

In the processing and exploring of the dataset using Orange, the features are emphasized. In addition, the process of identifying the intensity of the relationship between attributes is carried out. At this stage also, a trial with the Univariate Analysis technique was carried out. Actually, at this stage, testing can also be done with the Bivariate Analysis and Multivariate Analysis techniques. However, due to limitations, only Univariate Analysis is used where the properties of each attribute are investigated. The technique is done by applying the "Feature Statistics" feature provided by Orange. The picture is as in Figure 6.



Figure 6. Results of "feature statistics" on Orange

From Figure 6, the results of data processing can generally be read. The data center for each attribute is SKS = 113.43, GPA = 2.87, length of study = 5.75 years, initial status = 1.57 and final status = 1.35. Next is data processing using the K-Means method with the aim of grouping students' data into 2 clusters. To do clustering in Orange, use the "K-Means" feature and then select the attribute row that becomes the center of the cluster with the "Select Row" feature. Figure 7 illustrates the results of clustering against the dataset. Cluster C1 shows groups of students who have



successfully passed while cluster C2 shows students who drop out.

ID	Status	Cluster	Silhouette	SKS	Lama Studi	Status Akhir
1	0	C1	0.99121	45.0	6	2.0
2	0	C1	0.98891	45.0	6	2.0
3	1	C2	0.78829	14.0	4.0000	1.0
4	1	C2	0.74791	40.0	4.0000	1.0
5	1	C2	0.74794	44.0	4	1.0
6	0	C1	0.99999	45.0	6	2.0
7	0	C1	0.99999	45.0	6	2.0
8	0	C1	0.73476	45.0	4.0000	1.0
9	0	C1	0.74820	44.0	4.0000	1.0
10	0	C1	0.74794	44.0	4	1.0
11	0	C1	0.74823	44.0	4.0000	1.0
12	0	C1	0.72389	35.0	6	2.0
13	0	C1	0.74794	44.0	4	1.0
14	0	C1	0.74794	44.0	4	1.0
15	0	C1	0.74794	44.0	4	1.0
16	0	C1	0.74794	44.0	4	1.0
17	0	C1	0.74794	44.0	4	1.0
18	0	C1	0.74794	44.0	4	1.0
19	0	C1	0.74794	44.0	4	1.0
20	0	C1	0.74794	44.0	4	1.0
21	0	C1	0.74794	44.0	4	1.0
22	0	C1	0.74794	44.0	4	1.0
23	0	C1	0.74794	44.0	4	1.0
24	0	C1	0.74794	44.0	4	1.0
25	0	C1	0.74794	44.0	4	1.0
26	0	C1	0.74794	44.0	4	1.0
27	0	C1	0.74794	44.0	4	1.0
28	0	C1	0.74794	44.0	4	1.0
29	0	C1	0.74794	44.0	4	1.0
30	0	C1	0.74794	44.0	4	1.0
31	0	C1	0.74794	44.0	4	1.0
32	0	C1	0.74794	44.0	4	1.0
33	0	C1	0.74794	44.0	4	1.0
34	0	C1	0.74794	44.0	4	1.0
35	0	C1	0.74794	44.0	4	1.0
36	0	C1	0.74794	44.0	4	1.0
37	0	C1	0.74794	44.0	4	1.0
38	0	C1	0.74794	44.0	4	1.0
39	0	C1	0.74794	44.0	4	1.0
40	0	C1	0.74794	44.0	4	1.0

Figure 7. The results of clustering with the K-Means feature

In theory, according to [15], the following are the calculation steps using the K-Means method:

- Determine the number of clusters.
- Allocate data into clusters randomly.
- Calculate the centroid / average of the data in each cluster
- Allocate each data to the nearest centroid / average.
- Return to Step c), if there is still data that has moved to

One alternative to the application of K-Means with several related calculation theories development is Euclidian Distance (L2-Norm). The distance between two points is formulated as follows:

$$d(x, y) = \|x, y\|^2 = \sum_{i=1}^n (x_i - y_i)^2$$

Information:
 d = determinant (Euclidean Distance)
 x = the center of the cluster
 y = data
 n = amount of data
 i = data to-

Furthermore, [15] describes the shortest distance between the centroid and the document to determine the cluster position of a document. For example, document A has the shortest distance to centroid 1 compared to the others, then document A is included in group 1. Recalculate the position of the new centroid for each centroid (C_{i,j}) by taking the average of the documents that enter the initial cluster (G_{i,j}). Iteration is carried out continuously until the group position does not change. The following is the formula for determining the centroid:

$$C_i = \frac{x_1 + x_2 + x_n + x_{...}}{\sum x}$$

Information:
 x1 = the value of the 1st data record

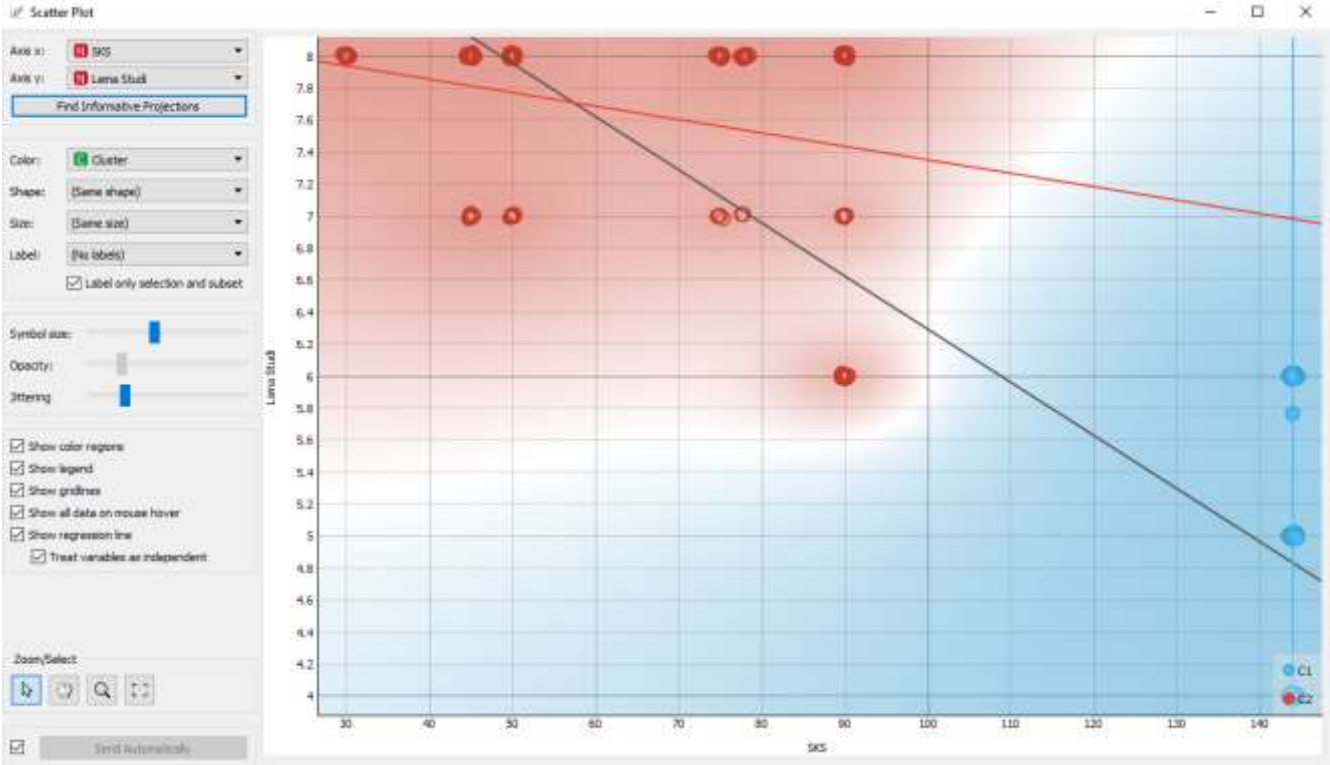


Figure 8. Clustering graphs of graduating and dropout students with the Scatter Plot feature based on credits and length of study

different clusters or if the change in the centroid value is above the specified threshold value or if the value change in the objective function used is above the specified threshold value

x2 = 2nd record data value
 Σx = number of data records



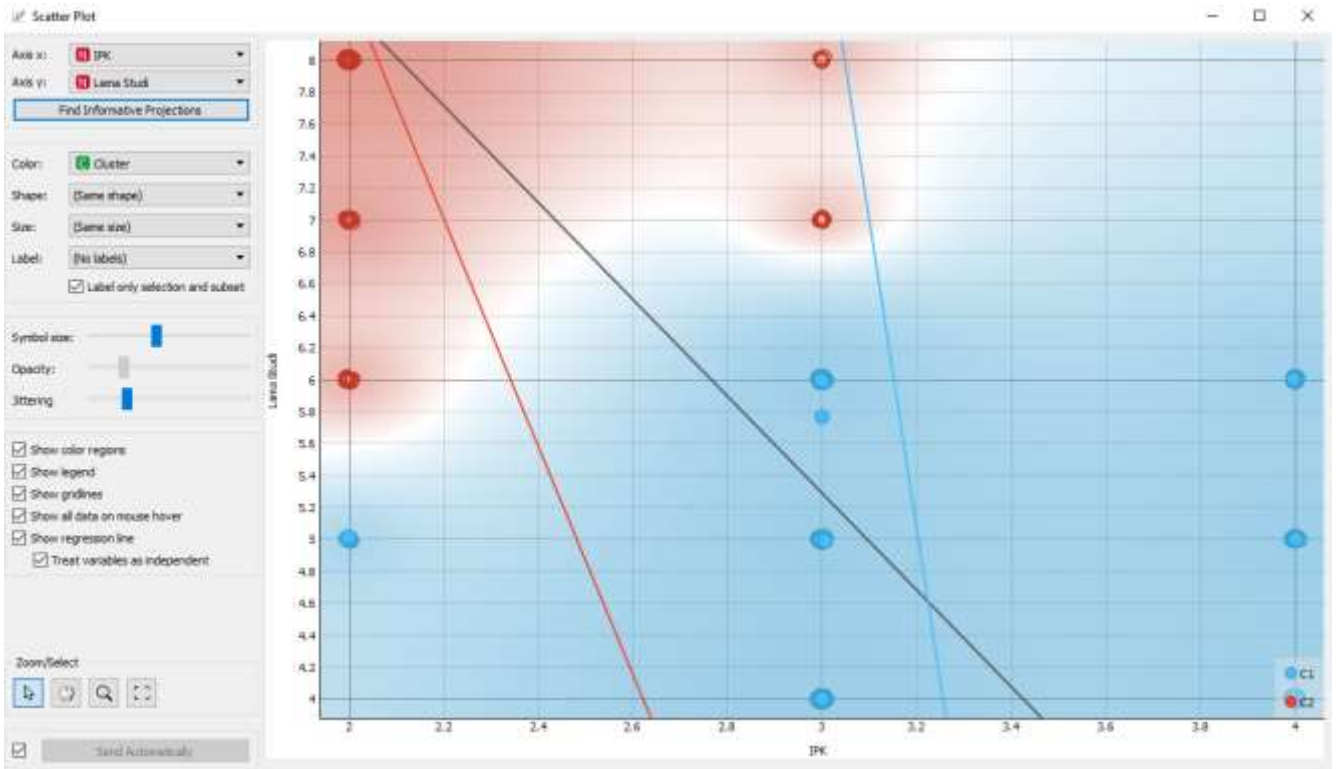


Figure 9. Clustering graphs of graduating and dropout students with the Scatter Plot feature based on GPA and length of study

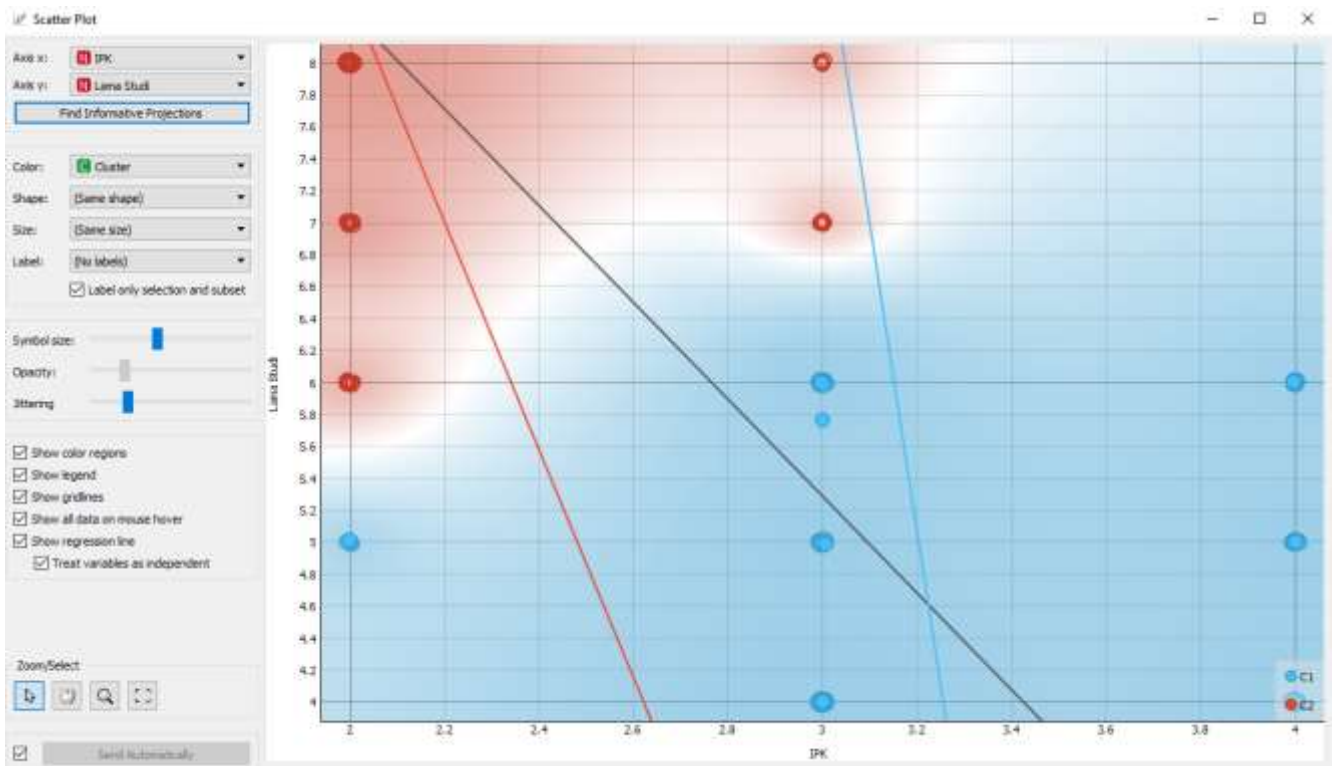


Figure 10. Clustering graphs of of graduating and dropout students with the Scatter Plot feature based on their final status and length of study

By processing data using Orange, it is quite easy to visualize the results of the cluster, namely the "Scatter Plot" feature. The cluster graph was then analyzed and tested. For more details, one of the cluster graphs is shown in Figure 8.

C. Analysis of Classification Results

The cluster graph shown in Figure 8, illustrates the grouping of students in the Department of Management of the National University who have successfully passed or dropped out based on the number of credits taken and the length of study. Cluster C1, which is marked in blue contains a group of students who have successfully passed. Meanwhile, cluster C2 is a group of students who drop out, which is marked in red. In the C2 cluster group, it can be seen that the length of study for DO students is over 6 years with the number of credits taken below 100. Meanwhile, in the C1 cluster, it can be seen that the group of students who successfully passed education for 6 years and under with the full number of credits is 144. This graph shows the irregularity of the academic conditions of dropping out students where they have been recorded and studied for a long time, but still attend lectures with a small number of credits. This is an anomalous condition where the length of the study is inversely proportional to the number of credits taken.

Figure 9 shows the clustering of students who graduated or dropped out (DO) based on the GPA obtained and the length of study. Cluster C2 shows groups of students who drop out and is marked in red, while cluster C1 is a group of students who have successfully completed their studies, which is marked in blue. From this graph, it can be seen that on average students with cluster C1, graduating and completing the study, have a $GPA > 3$ with a length of study ≤ 6 years. Meanwhile, the group of students who dropped out, cluster C2, had a $GPA > 3$ with a length of study > 6 years. This is certainly quite reasonable because the academic conditions of students who drop out usually have a GPA below the average, especially if they have taken courses for more than 4 years.

The cluster graphic in Figure 10 emphasizes the relationship between the number of credits a student has taken or the GPA a student has with the length of study he/she has taken. The graph depicts students whose status has finally passed, cluster C1 is marked in blue and groups of students who drop out are marked in red. In line with the pattern in Figure 8 and Figure 9, in this graph it can be seen that the group of students who successfully graduated, cluster C1, has a length of study period under 6 years. Meanwhile, cluster C2, students who drop out (DO), have a study period of more than 6 years.

This study shows that students who have taken education above the normal time, which is 4 years or more but are still recorded taking a small number of credits, it is necessary to take preventive measures so they will not drop out. The university should have taken anticipatory steps, so that the students concerned can complete their studies. Furthermore, students who have a GPA below 3 and have taken education above the normal time, 4 years or below, need to be given special attention as a preventive measure.

There are several other insights that can be taken from research that applies clustering with the K-Means method based on Figure 7, including:

- a) Students who have taken $SKS > 90$ but have $GPA > 2$, need to be given special attention. These students need to be supported in order to increase their achievement index, especially if the students have taken

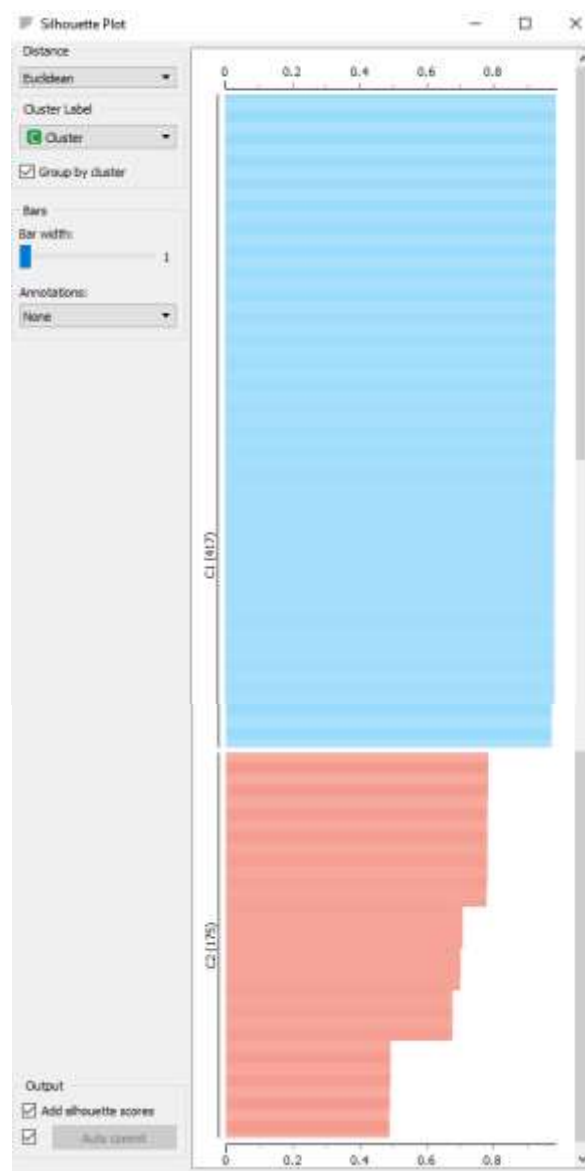


Figure 11. Cluster category

- education above the normal time, 4 years.
- b) Students whose status was initially inactive and had just taken the number of SKS ≤ 90 , need special attention. Do not let it drag on and become a drop out (DO) in the final status, especially if the students have taken education above the normal time, 4 years. The university needs to be pro-active in communicating with these students.
- c) Students whose $GPA > 3$, and have taken $SKS > 90$, but the length of study is above normal, 4 years, need to be given special attention, especially if they have been recorded as inactive or on leave.

To test the results of the clusters produced in this study, the Silhouette Plot feature which is also included in the Orange application is used. According to [19], the Silhouette Coefficient is used to see the quality and strength of the cluster, how well an object is placed in a cluster. This method is a combination of the cohesion and separation method. The stages of calculating the Silhouette Coefficient are as follows:

- a) Calculate the average distance from a document for example i with all other documents that are in the same

cluster:

$$a(i) = \frac{1}{|A|-1} \sum_{j \in A, j \neq i} d(i, j)$$

where j is another document in one cluster A and d(i, j) [2]
is the distance between document i and j.

- b) Calculate the average distance from document i to all [3]
documents in other clusters, and take the smallest value.

$$d(i, C) = \frac{1}{|A|} \sum_{j \in C} d(i, j)$$

where d(i, C) is the average distance of document i with [4]
all objects in other clusters C where $A \neq C$.

- c) The Silhouette Coefficient value is: [5]

$$b(i) = \min_{C \neq A} d(i, j)$$

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))}$$

The results of the Silhouette Coefficient shown by the [7]
application of Orange in this study are as shown in Figure [8]
11. The silhouette method average value approach is used [9]
to estimate the quality of the clusters formed. The higher [10]
the average value is, the better it is. Based on the graph in [11]
Figure 11, it can be seen that the optimal cluster formed at [12]
C1, k = 1, with an average Silhouette value of more than 0.85 [13]
or 85%. All graduate student members, cluster C1, showed [14]
average grades. In cluster C2, k = 2, the Silhouette values [15]
vary between 0.5 - 0.8. However, the value is quite good. [16]
This shows that the clusters formed in this study have been [17]
tested and can be used by the Management Department of [18]
the National University to predict the likelihood that [19]
students will graduate well or experience drop out (DO).

rsitas

IV. CONCLUSION

The K-means algorithm can be used to cluster and [20]
classify students who graduate or drop out (DO). The [21]
results of the clustering are used as predictions to determine [22]
the possibility of students' academic conditions that can [23]
pass or experience dropouts. Insights obtained from the [24]
results of this calculation can be used by the Department of [25]
Management of the National University to take preventive [26]
steps and ensure that students do not drop out. This K- [27]
means method is easily implemented in Orange because it [28]
has proven features, so it is quite fast and accurate.

The results of the silhouette test in this study reached [29]
85%, which indicates that the clustering method to [30]
determine the characteristics of students who are likely to [31]
graduate or drop out (DO) in the management department [32]
of the National University, Jakarta, can be applied as an [33]
effort to overcome the high dropout rate. With the results [34]
of clustering like this, for example, it makes it easier for [35]
universities to control and observe the academic conditions [36]
of their students. Finally, the campus can improve [37]
academic services and graduate students with good [38]
academic conditions.

REFERENCES

- [1] Universitas Nasional, "Sejarah Perkembangan [39]
Fakultas Ekonomi Universitas Nasional,"

Universitas Nasional, 2020. [Online]. Available: [40]
<http://fe.unas.ac.id/profil-fakultas/sejarah-perkembangan-fakultas/>.

- W. M. P. Duhita, "Clustering Menggunakan [41]
Metode K-Means Untuk Menentukan Status Gizi [42]
Balita," vol. 15, no. 2, 2015.
M. H. Adiya and Y. Desnelita, "Jurnal Nasional [43]
Teknologi dan Sistem Informasi Penerapan [44]
Algoritma K-Means Untuk Clustering Data Obat- [45]
Obatan Pada RSUD Pekanbaru," vol. 01, pp. 17- [46]
24, 2019.
A. V. Vidhyapeetham, "Crime Analysis and [47]
Prediction using Optimized K-Means Algorithm 1 [48]
1," no. Iccmc, pp. 915-918, 2020.
P. Manivannan, "Dengue Fever Prediction Using [49]
K-Means Clustering Algorithm," pp. 1-5, 2017.
S. Kaur, "Disease Prediction using Hybrid K- [50]
means and Support Vector Machine," 2016.
X. Chen and P. Miao, "Image Segmentation [51]
Algorithm Based on Particle Swarm Optimization [52]
with K-means Optimization," 2019 *IEEE Int. Conf. [53]
Power, Intell. Comput. Syst.*, pp. 156-159, 2019.
R. Alzu, "Medical Image Segmentation via [54]
Optimized K-Means," pp. 959-962, 2017.
F. Natalia, R. I. Desanti, and F. V. Ferdinand, [55]
"Prediction and Visualization of Flood Occurrences [56]
in Tangerang using K-Medoids, DBScan, and X- [57]
Means Clustering Algorithms," pp. 43-47, 2019.
A. Suresh, "Prediction of major crop yields of [58]
Tamilnadu using K-means and Modified KNN," [59]
2018 *3rd Int. Conf. Commun. Electron. Syst.*, no. [60]
Icces, pp. 88-93, 2018.
E. Hot and V. Popovi, "Soil data clustering by [61]
using K-means and fuzzy K-means algorithm," pp. [62]
890-893, 2015.
ASRONI and R. ADRIAN, "Penerapan Metode K- [63]
Means Untuk Clustering Mahasiswa Berdasarkan [64]
Nilai Akademik Dengan Weka Interface Studi [65]
Kasus Pada Jurusan Teknik Informatika UMM [66]
Magelang," *J. Ilm. SEMESTA Tek.*, vol. 18, no. 1, [67]
pp. 76-82, 2015.
F. Nur, R. Fauzan, J. Aziz, B. D. Setiawan, and I. [68]
Arwani, "Implementasi Algoritma K-Means untuk [69]
Klasterisasi Kinerja Akademik Mahasiswa," vol. 2, [70]
no. 6, pp. 2243-2251, 2018.
H. Priyatman, F. Sajid, and D. Haldivany, [71]
"Klasterisasi Menggunakan Algoritma K-Means [72]
Clustering untuk Memprediksi Waktu Kelulusan," [73]
vol. 5, no. 1, pp. 62-66, 2019.
A. Fadlil, M. Teknik, I. Universitas, A. Dahlan, T. [74]
Elektro, and U. Ahmad, "Implementasi Metode K- [75]
Means Dalam Pemetaan Kelompok Mahasiswa [76]
Melalui Data Aktivitas Kuliah," vol. 3, no. 1, pp. [77]
22-31, 2018.
D. W. Widodo, "Implementasi Algoritma K-Means [78]
Clustering Untuk Mengetahui Bidang Skripsi [79]
Mahasiswa Multimedia Pendidikan Teknik [80]
Informatika Dan Komputer Universitas Negeri [81]
Jakarta," *J. Pinter*, vol. 1, no. 2, pp. 157-166, 2017.
B. Poerwanto and R. Fa'rifah, "Analisis cluster k- [82]
means dalam pengelompokan kemampuan [83]
mahasiswa," *J. Sci. Pinisi*, vol. 2, no. 2, pp. 92-96,



- 2015.
- [18] S. Agrawal, "Prediction of Students Academic Execution Using K-Means and K-Medoids Clustering Technique," *2018 2nd Int. Conf. Trends Electron. Informatics*, no. Icoei, pp. 1308–1315, 2018.
- [19] F. M. IZZADIN, "Optimasi Jumlah Cluster K-Means Dengan Metode Elbow dan Silhouette Untuk Pengelompokan Luas Panen Palawija Kabupaten Magelang Pada Tahun 2017," *Medium The Startup*, 2019. [Online]. Available: <https://medium.com/@16611050/optimasi-jumlah-cluster-k-means-dengan-metode-elbow-dan-silhouette-untuk-pengelompokan-luas-panen-200131515c4f>.