International Journal of Information System & Technology
Akreditasi No. 36/E/KPT/2019 | Vol. 3, No. 2, (2020), pp. 191-199

Application of K-Means Algorithm for Clustering the Quality of Lecturer Learning at Batam International University

Yefta Christian

Universitas Internasional Batam, Kepulauan Riau, Indonesia

yefta@uib.ac.id

Abstract

The need to measure the quality of learning becomes very significant when the quality of learning is still low. This also applies to the quality of learning at the highest level. The quality of lecturer learning in class is often not in accordance with what it should be, one of the constraints of this low quality is that the quality of lecturer learning is not accurately measured. Data mining is used in this research to map the quality of lecturer learning based on its cluster, namely Clustering, It is hoped that universities can measure the quality of learning from their lecturers accurately. From this research, students with low social studies tend to value lecturers with low grades; conversely students with high social studies tend to judge lecturers with high enough grades. The k-Means algorithm is accurate enough to be used in clustering the quality of lecturer learning based on grades and results of student questionnaires.

Keywords: *Quality of Learning, Lecturers, Data Mining, Clustering, K-Means*

1. Introduction

The Law of the Republic of Indonesia No. 14 of 2005 concerning Teachers and Lecturers outlines various competencies that must be possessed by a teacher and lecturer; competencies that must be possessed by a lecturer are pedagogical competencies, social competencies, and professional competencies. The competence of educators would be very supportive for the quality of learning in the classroom [1]. If the educator lecturer has good competence, the quality of learning in the class will also be good; otherwise, if the competence of the teaching faculty is low, the quality of learning will be low too. Educators must have good competence so that learning can be fun and the material available is easily absorbed so that the quality of learning needed can be obtained [2]. Lecturers with low quality learning will find it difficult to guide their students to achieve learning goals [3].

To learn well about the quality of learning conducted by lecturers in the classroom required measurement, this measurement is part of the evaluation of learning. Evaluation of learning aims to provide information to the community, facilitate decision making and improve the quality of learning itself [4]. The results of other studies conducted an analysis by first providing instruments for mathematics learning for students, then the results were processed and categorized by the EKOP scale [4].

Measurement and processing of data at the present time many use data mining. Data mining is a pretty good tool for measuring and processing data because data mining can process data on a very large scale. Data mining analyzes observational data sets, discovers unexpected relationships and collects data in new ways so that it can help and understand users [5]. Data mining can be used to process data in various fields including processing various types of learning. The world of education has more data to be processed and data mining can be a good means to process data so that it manages education better [6].

ISSN: 2580-7250 IJISTECH

In order to improve the quality of learning in the classroom, more accurate measurements are needed. The data that will be produced to measure the quality of this learning will have different characteristics and must be able to group lecturers according to specified criteria. The Clustering method with K-Means Clustering is a non-hierarchical data clustering method [7]. The results of other studies state that the K-means algorithm has a high degree of accuracy, is effective and is measurable in managing objects [8]. K-Means Clustering also has another advantage that is determining the cluster is not in accordance with the order of the object.

At Batam International University, the quality of learning is taught through questionnaires distributed to students at the end of the learning semester. The results of the questionnaire will be processed by the Academic Development Bureau to improve the quality of instructor learning in the current semester. Measurements were made by calculating the average number of questionnaire results distributed and then lecturers were grouped into two large groups, namely lecturers with good grades and under grades Lecturers with good grades are valued for their performance, while lecturers with good grades will be re-evaluated for their performance. Measurements in this way are still lacking to map the quality of lecturer learning in this study. Measurements will be made to the quality of lecturer learning by grouping and applying data mining with K-Means Clustering.

2. Research Methodology

2.1. Quality of Learning

According to the Kemdikbud Language Center, in a Kamus Besar Bahasa Indonesia the meanings of quality is the level of the good or bad of something [9], besides that quality also means the quality of skills, intelligence, and so on. Based on the same source, the meanings of learning is a process of making learning. In accordance with this research it can be concluded that the quality of learning is the ability of lecturers to make students able to learn something well.

2.2. Clustering

Classification and Clustering are the 2 main methods in Data Mining. Clustering is the process of collecting the same objects. All parts of the data will be grouped into the same group called Clusters [10].

2.3. K-Means Algorithm

K-Means algorithm is an algorithm that groups data into several groups, where data in a group has different characteristics from the data in another group [11]. The K-Means algorithm is as follows:

a) Determine the number of clusters to be formed.
b) Generating centroid randomly.
c) Calculate the distance of data to each centroid.
d) Grouping each data based on the closest distance between the data and its centroid.
e) Determine the new centroid position.
f) Calculates the distance of data to each centroid if the new centroid position is not the same as the old centroid position.

2.4. Clustering

This research is focused on knowing and measuring the quality of learning conducted by lecturers in the classroom for 1 semester at Batam International University. The quality of learning in this study was divided into 6 clusters, each divided by 2, namely good clusters, moderate clusters and less clusters. This research is a qualitative study that will answer the research problem with the

research results obtained [12].The method used refers to the Flow of Information in Data Mining [13], as shown below.

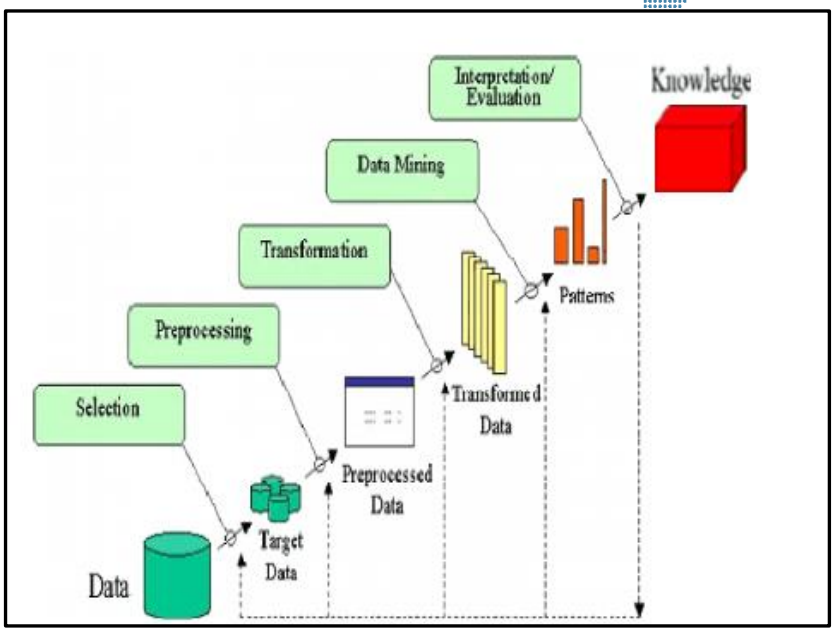

Figure 1. Figure Information Stream on Data Mining Process

The study began by collecting data obtained from the lecturer questionnaire by students. The questionnaire was distributed through the questionnaire website page owned by Batam International University; the questionnaire contained 17 question questions about learning activities in the classroom conducted by the lecturer concerned. The questions given are grouped into 4 groups: 1) the attitude of lecturers in lectures, 2) the atmosphere of lectures, 3) the quality of the material, 4) the ability of lecturers to deliver the material. All the data from the lecturer questionnaire will be used as the target data. The target data is stored on the website and will be exported from a database which will then be converted and normalized using Microsoft Excel. Normalization is done by determining the average of all numerical data targets, the average of all existing data targets will be used to determine the quality status of lecturer learning based on a questionnaire per student.

Normalized target data will be selected and enter the preprocessing stage by checking the validity and reliability of the results of the questionnaire data using the Data Mining Weka tool to avoid blank data and duplicate data. After that in the Data Mining Knime tool data that has been in the form of preprocessed data will be transformed into transformed data. Data transformation is done by separating the value data with the key data. In this study what will be used as key data is the quality of lecturer learning status. This transformed data will be processed in the Data Mining stage. Then the data mining process is carried out by transferring existing data at the K-Means node to then transferred to the next node to be evaluated and interpreted.

3. Results and Discussion

In this research, the students and lecturers in Batam International University will be the object of research. Students will be given a questionnaire in digital form using a questionnaire website owned by Batam International University. Questionnaires given to students are about their opinions while attending lectures in accordance with the lecturers and courses that are followed by these students.

The stages of the research carried out on the KNIME application are shown in the image below:

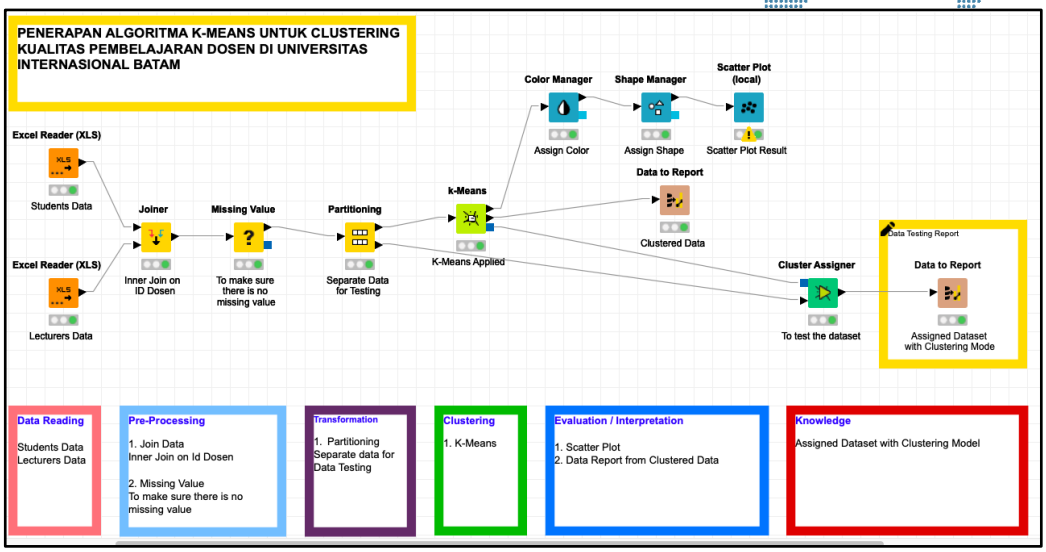

Figure 2. Research Model on KNIME

3.1. Data Selection

The dataset is obtained by applying a query to the information system owned by Batam International University. The dataset used was taken from 9851 questionnaire data obtained from 939 students to assess the quality of learning from 203 lecturers at Batam International University. Below is the dataset:

Table 1. Student Data Table

No	NPM	ID Dosen	ID Kelas	IPS	Jumlah Absensi Mahasiswa	Nilai Kuesioner per Mahasiswa per Kelas
1	1861015	01180001	610000029	4	12	5,00
2	1631037	02170017	310001766	3	12	1,00
3	1861014	02170017	610000039	3	12	5,00
4	1631023	02170023	310001784	3	9	1,00
5	1861014	04180029	610000037	3	8	5,00
6	1861020	04180029	610000037	4	6	5,00
7	1861012	04180029	610000038	3	8	5,00
8	1861014	04180029	610000038	2	8	5,00
9	1651034	06170083	510001337	0	5	1,00
...	...	...	...	...	...	...
9842	1731032	06180037	310001869	3	10	1,00
9843	1861012	07170082	610000061	4	8	5,00
9844	1861014	07170082	610000061	4	8	5,00
9845	1641086	08140050	410002620	2	11	1,00
9846	1641165	08140050	410002638	3	12	1,00
9847	1861020	09150072	610000036	4	6	5,00
9848	1642002	09160042	420002146	4	6	1,00
9849	1611018	12160065	110000776	4	10	1,00
9850	1611035	12160065	110000776	3	11	1,00
9851	1861015	12170130	610000035	4	12	5,00

Table 2. Lecturer Data Table

No	ID Kelas	ID Dosen	Jumlah Absensi Dosen	Nilai Rata - Rata Kuesioner Dosen
1	110000762	10130038	13	5
2	110000774	09170097	6	5
3	110000776	12160065	14	1,5
4	110000835	10130038	12	5
5	110000836	13020005	13	5
6	110000890	10130038	11	5
7	310001869	06180037	13	2,54
8	410002535	08040072	13	2,09
9	410002620	08140050	14	1,8
10	410002623	10170116	14	5
...	...	...	...	...
984	410002793	11170121	14	5
985	410002853	09170104	13	5
986	460000121	08180071	12	2,19
987	460000209	09170093	7	5
988	460000223	08180071	12	2,19
989	460000225	08180071	11	2,19
990	510001337	06170083	14	1
991	510001368	09180092	14	5
992	510001437	09150065	14	2,29
993	510001452	11170121	14	2,65

3.2. Dataset Preprocessing

Pada At this stage the researcher conducted the process of joining data with the ID Dosen field as a key. Data that has been inputted will be checked and cleaned using the missing value node to ensure there are no errors and data emptiness.

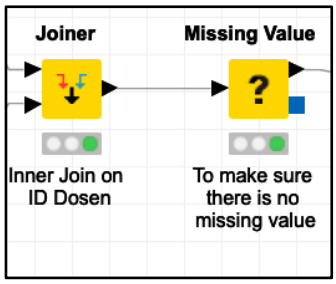

Figure 3. Dataset Preprocessing

The following is a dataset that has been put together and cleaned:

Table 3. Data Table of Missing Value

Table "default" - Rows: 68534 | Spec - Columns: 9 | Properties | Flow Variables

Row ID	NPM	ID Do...	ID Kelas	IPS	Jumla...	Nilai K...	ID Kela...	Jumla...	Nilai R...
Row0_Row...	1861015	01180001	610000029	4	12	5	610000044	8	4.3
Row0_Row...	1861015	01180001	610000029	4	12	5	410002780	10	4.33
Row0_Row...	1861015	01180001	610000029	4	12	5	610000029	14	4.37
Row0_Row...	1861015	01180001	610000029	4	12	5	510001587	14	4.37
Row0_Row...	1861015	01180001	610000029	4	12	5	610000079	14	4.37
Row0_Row...	1861015	01180001	610000029	4	12	5	310001899	12	4.37
Row0_Row...	1861015	01180001	610000029	4	12	5	310001850	11	4.38
Row1_Row77	1631037	02170017	310001766	3	12	1	310001766	14	3.81
Row1_Row...	1631037	02170017	310001766	3	12	1	510001280	14	4.12
Row1_Row...	1631037	02170017	310001766	3	12	1	310001787	14	4.25
Row1_Row...	1631037	02170017	310001766	3	12	1	420002122	14	4.25
Row1_Row...	1631037	02170017	310001766	3	12	1	420002123	14	4.25
Row1_Row...	1631037	02170017	310001766	3	12	1	610000039	14	4.26
Row1_Row...	1631037	02170017	310001766	3	12	1	410003071	13	4.26
Row1_Row...	1631037	02170017	310001766	3	12	1	460000118	14	4.34
Row1_Row...	1631037	02170017	310001766	3	12	1	510001289	14	4.41
Row2_Row77	1861014	02170017	610000039	3	12	5	310001766	14	3.81
Row2_Row...	1861014	02170017	610000039	3	12	5	510001280	14	4.12
Row2_Row...	1861014	02170017	610000039	3	12	5	310001787	14	4.25
Row2_Row...	1861014	02170017	610000039	3	12	5	420002122	14	4.25
Row2_Row...	1861014	02170017	610000039	3	12	5	420002123	14	4.25
Row2_Row...	1861014	02170017	610000039	3	12	5	610000039	14	4.26
Row2_Row...	1861014	02170017	610000039	3	12	5	410003071	13	4.26

3.3. Dataset Transformation

At this stage the researcher conducted the process of separating the dataset with a 9:1 ratio using node partitioning, which consisted of the main dataset at 90% and the testing dataset at 10%. The main dataset will be streamed to the next node for the clustering process, while the testing dataset is used as a dataset at the testing stage.

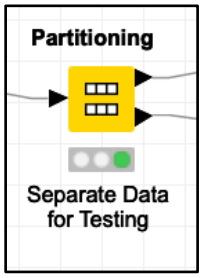

Figure 4. Partitioning Node

3.4. Clustering with k-Means Algorithm

The next stage is clustering using the k-means algorithm. The results obtained at this stage are a dataset that has been clustered according to the number of clusters desired by researchers.

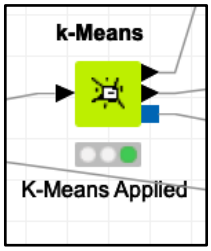

Figure 5. K-Means Node

The next stage is clustering using the k-means algorithm. The results obtained at this stage are a dataset that has been clustered according to the number of clusters desired by researchers.

Figure 6. Results of Clustering Six Cluster

From the test results above it can be seen that the more division of the number of clusters carried out, the results of the existing cluster mapping are increasingly clear. Distribution of datasets of 6 clusters will be used by researchers as a reference for evaluation and conclusion.

3.5. Evaluation and Enterpretation

At this stage the clustering data will be streamed to two nodes. The first node is used to map scatter plots from clustered data.

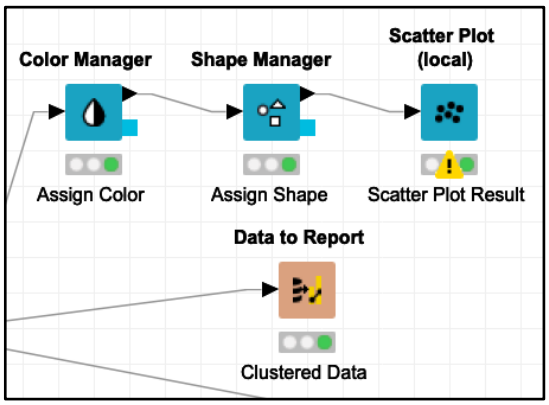

Figure 7. Scatter Plot Mapping Nodes and Data Report

The following are examples of scatter plots that emerged from the results of clustering.

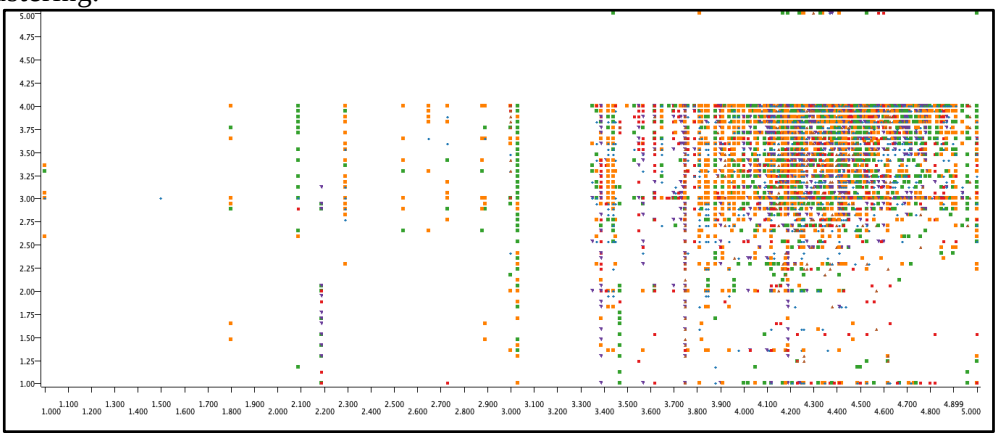

Figure 8. Example of Scatter Plot Six Clustering Results

The second node is used to deduce data from six clusters that have been made. The data generated at this node will be used to draw final conclusions from the dataset being tested from the initial stage to the end. The data below is the data generated as a conclusion of the six clusters that have been made.

Table 4. Clustering Results Table with K-Means Algorithm

Row ID	IPS	Jumlah Absensi Mahasiswa	Nilai Kuesioner per Mahasiswa per Kelas	Jumlah Absensi Dosen	Nilai Rata- Rata Kue...
cluster_0	3.452	9.263	4.34	13.536	4.376
cluster_1	3.213	5.954	4.112	10.404	4.23
cluster_2	3.46	10.853	4.374	7.722	4.339
cluster_3	3.474	10.807	4.338	11.296	4.35
cluster_4	4	11.65	4.48	13.857	4.443
cluster_5	2.636	11.455	4.296	13.817	4.364

3.6. Knowledge

This stage was made to test the initial dataset with the clustering model that was created.

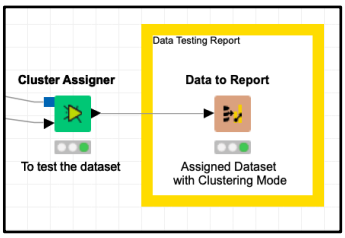

Figure 9. Node Cluster Assigner

The dataset used does not have to be the same dataset as it is currently used, but it can use another dataset with the same table structure to be tested. At present the researcher uses the same dataset with the results as shown below:

Table 5. Cluster Assigner Results Table (Data Testing)

Row ID	NPM	ID Do...	ID Kelas	IPS	Jumla...	Nilai K...	ID Kela...	Jumla...	Nilai R...	Cluster
Row2_Row...	1861014	02170017	610000039	3	12	5	510001280	14	4.12	cluster_5
Row2_Row...	1861014	02170017	610000039	3	12	5	420002122	14	4.25	cluster_5
Row5_Row...	1861020	04180029	610000037	4	6	5	420002175	14	4.24	cluster_0
Row7_Row...	1861014	04180029	610000038	2	8	5	310001844	12	4.19	cluster_0
Row7_Row...	1861014	04180029	610000038	2	8	5	310002014	12	4.53	cluster_0
Row8_Row...	1651034	06170083	510001337	0	5	1	510001337	14	1	cluster_1
Row9_Row65	1731099	03170027	310001860	0	6	1	510001326	12	3.73	cluster_1
Row10_Ro...	1741221	06160025	410002575	1	12	1	410002582	14	4.73	cluster_5
Row10_Ro...	1741221	06160025	410002575	1	12	1	420002113	14	4.73	cluster_5
Row10_Ro...	1741221	06160025	410002575	1	12	1	420002116	14	4.73	cluster_5
Row11_Ro...	1742164	04100035	420001869	2	10	1	420001870	10	4.03	cluster_3
Row12_Ro...	1831055	10170116	310001931	2	10	1	610000073	7	4.26	cluster_2
Row14_Ro...	1831107	07170082	310002017	4	7	1	510001425	11	3.62	cluster_1
Row14_Ro...	1831107	07170082	310002017	4	7	1	610000061	10	4.45	cluster_1
Row15_Ro...	1831169	07170082	310002017	4	7	1	310002017	10	3.62	cluster_1
Row15_Ro...	1831169	07170082	310002017	4	7	1	610000084	14	4.09	cluster_0
Row15_Ro...	1831169	07170082	310002017	4	7	1	610000061	10	4.45	cluster_1
Row25_Ro...	1842124	04080032	420002060	4	9	1	420001886	13	4.25	cluster_0
Row26_Ro...	1842124	04080032	420002064	4	11	1	420001817	13	4.32	cluster_4
Row27_Ro...	1846025	08180071	460000117	4	9	1	460000225	11	2.19	cluster_3
Row29_Ro...	1846032	08180071	460000121	4	9	1	460000223	12	2.19	cluster_0
Row30_Ro...	1846039	08180071	460000110	4	3	1	460000121	12	2.19	cluster_1
Row32_Ro...	1846044	08180071	460000117	3	10	1	460000117	13	3.47	cluster_0
Row34_Ro...	1851025	10150105	510001586	2	3	1	420001941	14	4.83	cluster_1
Row37_Ro...	1846017	08180071	460000121	3	7	1.118	460000225	11	2.19	cluster_1
Row40_Ro...	1841082	08040072	410002535	2	9	1.176	310001720	13	4.05	cluster_0
Row43_Row7	1841082	08040072	410002535	2	9	1.176	410002535	13	2.09	cluster_0
Row43_Ro...	1841082	08040072	410002535	2	9	1.176	460000196	13	4.05	cluster_0
Row44_Ro...	1841082	08040072	410002535	2	9	1.176	410003123	13	4.49	cluster_0
Row45_Ro...	1841082	08040072	410002535	2	9	1.176	310002021	14	4.05	cluster_0
Row45_Ro...	1841082	08040072	410002535	2	9	1.176	510001423	13	4.5	cluster_0
Row48_Ro...	1841082	08040072	410002535	2	9	1.176	310001720	13	4.05	cluster_0
Row48_Ro...	1841082	08040072	410002535	2	9	1.176	420001946	13	4.49	cluster_0
Row49_Ro...	1841082	08040072	410002535	2	9	1.176	310001720	13	4.05	cluster_0
Row50_Ro...	1841082	08040072	410002535	2	9	1.176	310002021	14	4.05	cluster_0
Row50_Ro...	1841082	08040072	410002535	2	9	1.176	420001946	13	4.49	cluster_0
Row50_Ro...	1841082	08040072	410002535	2	9	1.176	510001423	13	4.5	cluster_0
Row52_Ro...	1831154	10170116	310001931	4	10	1.235	410002629	14	4.53	cluster_0
Row54_Ro...	1846048	03170028	460000123	4	8	1.235	420002120	13	4.23	cluster_0
Row55_Ro...	1831066	10170116	310001931	2	10	1.294	410002629	14	4.53	cluster_5
Row56_Ro...	1842124	09160042	420001830	3	12	1.294	420001814	10	3.75	cluster_3
Row56_Ro...	1842124	09160042	420001830	3	12	1.294	420002143	6	3.75	cluster_2

4. Conclusion

From the results of tests conducted by researchers with a dataset of questionnaires from Batam International University students using the k-Means algorithm it can be concluded that:

a) A low social studies results in students tending to give low ratings to lecturers even though the number of lecturers in the class is quite high.
b) Conversely, IPS with a moderate grade will result in students giving high ratings to lecturers even though the number of lecturers in the class is low.
c) Students with social studies and high attendance rates tend to give high ratings to lecturers who have a high enough class attendance

From all of these conclusions, it can be concluded that the social studies scores tend to be directly proportional to the results of student assessments of lecturers. Students with low social studies tend to value lecturers with low grades; conversely students with high social studies tend to judge lecturers with high enough grades. The k-Means algorithm is accurate enough to be used in clustering the quality of lecturer learning based on grades and results of student questionnaires.

References

[1] V. L. P. Sutrisno and B. T. Siswanto, "Faktor-Faktor Yang Mempengaruhi Hasil Belajar Siswa Pada Pembelajaran Praktik Kelistrikan Otomotif Smk Di Kota Yogyakarta," *J. Pendidik. Vokasi*, vol. 6, no. 1, pp. 111–120, 2016.

[2] T. Nurrita, "Pengembangan Media Pembelajaran Untuk Meningkatkan Hasil Belajar Siswa," *Misykat*, vol. 3, no. 1, pp. 171–187, 2018.

[3] H. Husaini, "Pengaruh Profesional Dosen Terhadap Kualitas Pembelajaran Dosen Agama Islam Di Fakultas Keguruan Ilmu Pendidikan Universitas Pattimura," *PEMBELAJAR J. Ilmu Pendidikan, Keguruan, dan Pembelajaran*, vol. 1, no. 1, pp. 9–16, 2017.

[4] S. Raupu, "Analisis Kualitas Pembelajaran Guru Matematika dengan Menggunakan Model EKOP di SMK Teknologi Tri Tunggal ⍰45 Makassar," *Al-Khwarizmi J. Pendidik. Mat. dan Ilmu Pengetah. Alam*, vol. 4, no. 1, pp. 89–102, 2018.

[5] R. Mythily, A. Banu, and S. Raghunathan, "Clustering models for data stream mining," *Procedia Comput. Sci.*, vol. 46, no. November, pp. 619–626, 2015.

[6] D. Firdaus, "Penggunaan Data Mining dalam Kegiatan Sistem Pembelajaran Berbantuan Komputer," *J. Format*, vol. 6, no. 2, pp. 91–97, 2017.

[7] G. Gustientiedina, M. H. Adiya, and Y. Desnelita, "Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan," *J. Nas. Teknol. dan Sist. Inf.*, vol. 5, no. 1, pp. 17–24, 2019.

[8] A. Bastian *et al.*, "Penerapan Algoritma," *Jsi.Cs.Ui.Ac.Id*, vol. 14, no. 1, pp. 26–32, 2018.

[9] Pusat Bahasa Kemdikbud, "Kamus Besar Bahasa Indonesia (KBBI)," *Kementeri. Pendidik. dan Budaya*, 2016.

[10] D. A. Simovici and C. Djeraba, "Clustering," in *Advanced Information and Knowledge Processing*, 2014.

[11] A. Asroni, H. Fitri, and E. Prasetyo, "Penerapan Metode Clustering dengan Algoritma K-Means pada Pengelompokkan Data Calon Mahasiswa Baru di Universitas Muhammadiyah Yogyakarta (Studi Kasus: Fakultas Kedokteran dan Ilmu Kesehatan, dan Fakultas Ilmu Sosial dan Ilmu Politik)," *Semesta Tek.*, vol. 21, no. 1, pp. 60–64, 2018.

[12] J. W. C. N.-M. H. L.-S. L. H. C. Creswell, *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage Publication, 2013.

[13] R. Helilintar and I. N. Farida, "Penerapan Algoritma K-Means Clustering Untuk Prediksi Prestasi Nilai Akademik Mahasiswa," *J. Sains dan Inform.*, vol. 4, no. 2, pp. 80–87, 2018.

Authors

1st Author
Yefta Christian
Universitas Internasional Batam