Article

Student Grouping Based on Final Exam Values of the Courses with the K-Means Classification Method Using KNIME

Sapriani Gustina [1*], **Iwan Purnama** [2], **Windo** [3], **Erlinawaty** [4]

[1] Department of Information Technology, Universitas Proklamasi 45, Yogyakarta; sagustina@up45.ac.id;
[2] Department of Information Technology, Universitas Labuhanbatu, Sumatera; iwanpurnama2014@gmail.com;
[3] Department of Information Technology, Universitas Labuhanbatu, Sumatera; yen.hui@gmail.com;
[4] Department of Information Technology, Universitas Proklamasi 45, Yogyakarta; sagustina@up45.ac.id;

* Corresponding author: sagustina@up45.ac.id

Abstract: In learning, each student must have a different way of education and learning patterns that will impact the results of their learning evaluation at the end of each semester. Furthermore, assessing a student who excels in learning is one of them by looking at the score of the final exam results. Maybe the student can easily get good grades because they have expertise in that field or get good grades because they are diligent in studying. The scores of students final semester exams in several courses will be summarized here to be used as a basis for classifying students into several groups; intelligent, average, and less good students. in this research using AI & K-Means Algorithm

Keywords: K-Means, Classification, Machine Learning, Unsupervised Learning, Students Classification, Online Learning

Citation: Gustina S; Purnama, I; Windo; Erlinawaty, Student Grouping Based on Final Exam Values of the Courses with the K-Means Classification Method Using KNIME. *Iota*, **2021**, ISSN 2774-4353, Vol.01, 02. https://doi.org/10.31763/iota.v1i2.378

Academic Editor : P.D.P Adi

Received : 17 Maret 2021

Accepted : 09 April 2021

Published : 15 May 2021

1. Introduction

Student learning activities are carried out during online learning in different places; most of them have returned to their respective hometowns since the beginning of the online learning period. Some of them experience learning difficulties due to inadequate facilities and infrastructure to achieve good learning outcomes.

The object of research is the students of the Information Technology Department in semester 3; the results of the students Final Examination scores in online learning from several courses are collected for clustering using the K-Means method so that they are expected to be able to group students into three categories, i.e., intelligent, average and less good. The results of clustering are expected to be used as evaluation material for study programs to improve strategies in learning or solve problems experienced by students.

With the value data obtained, then grouping will be carried out using the K-Means classification method, previous research has also been conducted to group students but the measuring indicator is the Student Achievement Index (IP) value or Grade Point Average (GPA) [1], Therefore, in this research, a classification will be carried out based on the final score of the five courses that students undergo in full using the online learning system.

2. Literature Review

In this research, a case study was carried out by taking student grade data from the Information Technology Departement Universitas Proklamasi 45; the data was in the form of a .xls file with the contents of the final scores for each course such as Human and Computer Interaction (IMK), Object-Oriented Programming (PBO), Networking (Jarkom), Operations Research (RO), and Advanced Web Programming (Web) as shown in figure 1. All of these courses are compulsory subjects to be taken by students; in the future, the grades of these courses can be used to formulate student interests, as has been done by previous research [2].

Therefore, then what is classification? It is a method for arranging data systematically or according to some predetermined rules or rules. It can also literally be said that classification is the division of something according to classes [3]. Researchers have developed methods to solve classification cases: Decision Tree, Naïve Bayes, Artificial Neural Network, Statistical Analysis, Genetic Algorithm, Rough Sets, k-Nearest Neighbour, Rule-Based Method, Memory Based Reasoning, Support Vector Machine and others [4].

	A	B	C	D	E	F	G	H
1	NO	N.I.M	NAMA	NILAI MATA KULIAH				
2			MAHASISWA	IMK	PBO	JARKOM	RO	WEB LANJUT
3	1	19450410001	FD	75	80	79	80	78
4	2	19450410002	MZ	70	65	65	80	75
5	3	19450410003	NA	85	87	79	80	80
6	4	19450410004	WO	75	75	85	85	80
7	5	19450410008	EW	85	85	80	80	75
8	6	19450410009	DP	80	90	85	65	80
9	7	19450410014	SS	75	75	65	70	75
10	8	19450410017	DP	80	80	76	85	75
11	9	19450410019	TE	75	75	75	75	75
12	10	19450410021	LM	65	65	65	70	65
13	11	19450410024	MK	80	85	85	80	85
14	12	19450410025	SF	85	90	85	85	90
15	13	19450410026	SU	85	90	85	90	90
16	14	19450410027	PP	75	75	70	70	70
17	15	19450410029	LH	75	75	70	70	65
18	16	19450410031	FA	0	0	0	0	0
19	17	19450410032	AC	78	65	75	80	65
20	18	19450410033	PA	80	80	75	85	75
21	19	19450410036	DF	78	65	70	70	65
22	20	19450410038	SH	75	65	70	85	80

Figure 1. Data Shown

3. Methods

The K-Means method is one of the methods in Machine Learning which is often called Unsupervised Learning, where data will be grouped without the need for comparisons or data training, K-Means pseudocode, as in figure 2, will be carried out to solve problems from the formation of centroids to forming group groups. or cluster [5], therefore K-Means was chosen because it has proven to be reliable in conducting cluster formation in various cases in various fields such as the classification of measles immunization for children under five per province [6], grouping cities based on human development index in Maluku Province [7], a grouping of outstanding vocational school students [8], a grouping of drug data at Pekanbaru City Hospital [9]. It has also been used to classify students by region of origin to support its marketing strategy [10].

This research will be done using the KNIME application, which has also been done research utilizing this application to classify recommended film and KNIME able to produce good classification results quickly [11]; this research starts from reading the data to be tested in this workflow node one is used, namely Excel Reader, which will function as a data reader that we will test because the file has a .xls format. At first, node one will

be yellow when the data has not been read; it will look like Figure 3. Then we configure node 1 to read the excel file, which we will test, as shown in Figure 4.

The k-means Clustering Algorithm
Input : Data points D, Number of clusters k
Step 1: Initialize k centroids randomly
Step 2: Associate each data point in D with the nearest centroid. This will divide the data points into k clusters.
Step 3: Recalculate the position of centroids.
Repeat steps 2 and 3 until there are no more changes in the membership of the data points
Output : Data points with cluster memberships

Figure 2. K-Means Pseudocode

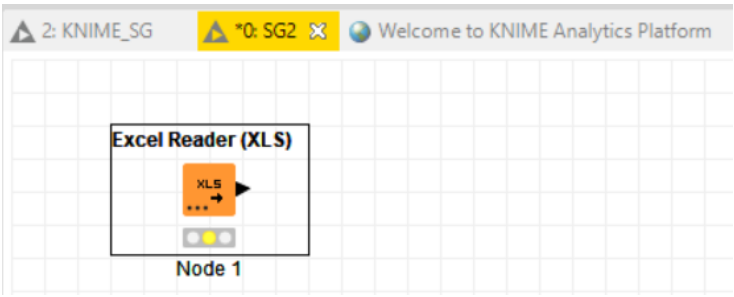

Figure 3. Start Node

Figure 4. Node Configuration

Then after we enter the data, do the execution, and node one will turn green as in figure 5, and for the results of reading the data, it will look like in figure 6.

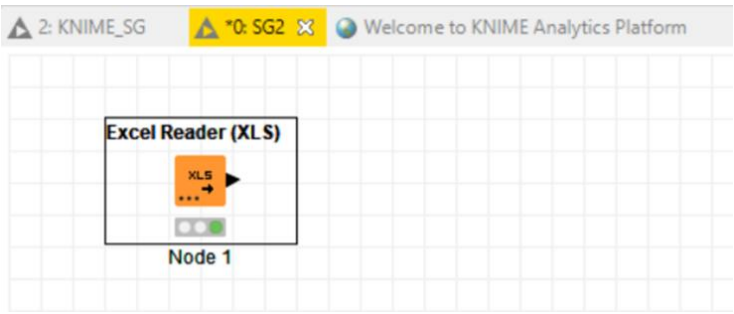

Figure 5. Start Node Ready

To use the K-Means Method, we add another node and place it at node 2 for the K-Means Classification Method; after the excel file on node 1 reads well, the results from node one we connect to node two as in Figure 7 and node two will be yellow before configuration and execution.

Output table - 0:1 - Excel Reader (XLS)

File Hilite Navigation View

Table "C:\Users\Asus ZenBook\Downloads\PaperSG\RisetNilai.xlsx [Sheet1]" - Rows: 19 Spec - Columns: 6 Prop

Row ID	S Col0	I Col1	I Col2	I Col3	I Col4	I Col5
Row0	FD	75	80	79	80	78
Row1	MZ	70	65	65	80	75
Row2	NA	85	87	79	80	80
Row3	WO	75	75	85	85	80
Row4	EW	85	85	80	80	75
Row5	DP	80	90	85	65	80
Row6	SS	75	75	65	70	75
Row7	DP	80	80	76	85	75
Row8	TE	75	75	75	75	75
Row9	LM	65	65	65	70	65
Row10	MK	80	85	85	80	85
Row11	SF	85	90	85	85	90
Row12	SU	85	90	85	90	90
Row13	PP	75	75	70	70	70
Row14	LH	75	75	70	70	65
Row15	AC	78	65	75	80	65
Row16	PA	80	80	75	85	75
Row17	DF	78	65	70	70	65
Row18	SH	75	65	70	85	80

Figure 6. Shown Table

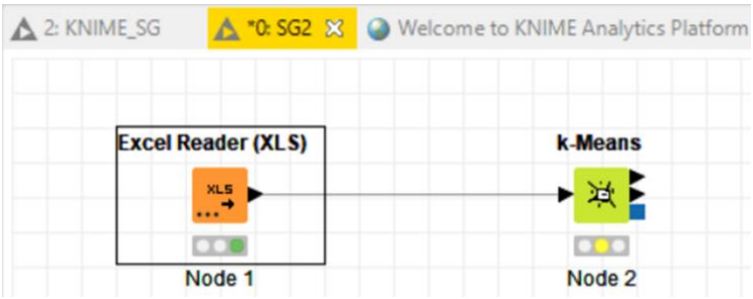

Figure 7. K-Means

What needs to be paid attention to in the configuration of node 2 for the K-Means method? Please pay attention to the clusters we will form; we set them to 3 collections. The formation of the centroid can be arranged based on the initial row of data or random; here, we put it to select the centroid by random and develop the maximum iteration that will occur.

Set the column where the processing will be carried out because the data we have set for processing is only for data on student course values indicated by the code Col, Col2, Col3, Col4, and Col5. All of these configurations will be as in Figure 8.

Finally, when the configuration is complete, then execute and see the color change in the status of node two will turn green as shown in Figure 7, that means the clustering process is running well, and we can see the results as in Figure 9.

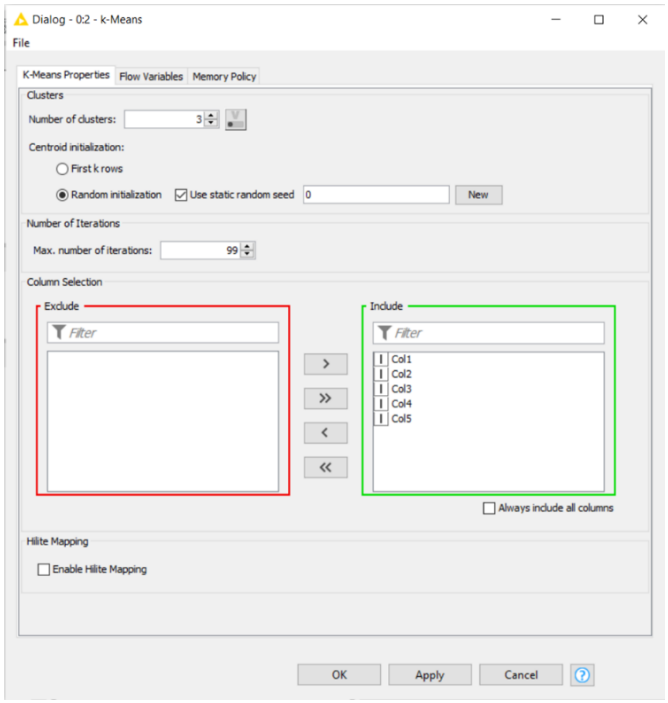

Figure 8. K-Means Configuration

4. Result and Discussion

The Labeled Input in figure 9 appeared, showing clusters that will group the students into 3 clusters. Cluster_0 indicates that if the student is classified as an intelligent student, then Cluster_1 will offer an average class, and Cluster_2 shows a poor student cluster. Clustering can result in the division of smart student groups totaling ten students, students with an intermediate category of 3 students, and as many as six students, including less good students.

Labeled input - 0:2 - k-Means

File Hilite Navigation View

Table "C:\Users\Asus ZenBook\Downloads\PaperSG\RisetNilai.xlsx [Sheet1]" - Rows: 19 | Spec - Columns: 7 | Properties | Flow Variables

Row ID	Col0	Col1	Col2	Col3	Col4	Col5	Cluster
Row0	FD	75	80	79	80	78	cluster_0
Row1	MZ	70	65	65	80	75	cluster_1
Row2	NA	85	87	79	80	80	cluster_0
Row3	WO	75	75	85	85	80	cluster_0
Row4	EW	85	85	80	80	75	cluster_0
Row5	DP	80	90	85	65	80	cluster_0
Row6	SS	75	75	65	70	75	cluster_2
Row7	DP	80	80	76	85	75	cluster_0
Row8	TE	75	75	75	75	75	cluster_2
Row9	LM	65	65	65	70	65	cluster_2
Row10	MK	80	85	85	80	85	cluster_0
Row11	SF	85	90	85	85	90	cluster_0
Row12	SU	85	90	85	90	90	cluster_0
Row13	PP	75	75	70	70	70	cluster_2
Row14	LH	75	75	70	70	65	cluster_2
Row15	AC	78	65	75	80	65	cluster_1
Row16	PA	80	80	75	85	75	cluster_0
Row17	DF	78	65	70	70	65	cluster_2
Row18	SH	75	65	70	85	80	cluster_1

Figure 9. K-Means Cluster

5. Conclusions

With existing data, testing using the KNIME application using the K-Means classification method can produce student classifications based on the final grades obtained. And also, the use of the KNIME application is efficient. Therefore, it is expected to help solve other problems quickly.

Author Contributions: Conceptualization; Sapriani Gustina (SG), Iwan Purnama (IP), Windo (W), Erlinawaty (E), methodology; SG, IP, W, E, software; SG, IP, W, E, validation; SG, IP, W, E, formal analysis; SG, IP, W, E, investigation; SG, IP, W, E, data curation; SG, IP, W, E; writing—original draft preparation; SG, IP, W, E, writing—review and editing; SG, IP, W, E, visualization; SG, IP, W, E, supervision; SG, IP, W, E, project administration; SG, IP, W, E, funding acquisition; SG, IP, W, E, have read and agreed to the published version of the manuscript.

Acknowledgments: The authors would like to be the head of the Department of Information Technology Universitas Proklamasi 45 and the research team who helped provide data for composing this research.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. R. Z. Alberto, W. K. Sari, and A. Primanita, "Pengelompokkan Performa Akademik MahasiswaBerdasarkan Indeks Prestasi Menggunakan K-Means Clustering," J. Online Inform., pp. 99–104, 2017.
2. S. Anraeni, "Penentuan Peminatan Mata Kuliah Mahasiswa Program Studi Program Studi Teknik Informatika," no. December 2014, 2017.
3. I. D. Lestari, "Klasifikasi Online dan Google," pp. 83–94, 2016.
4. U. Penentuan, J. Pada, and S. Sma, "Perbandingan Analisis Klasifikasi Antara Decision Tree Dan Support Vector Machine Multiclass Untuk Penentuan Jurusan Pada Siswa Sma," J. Gaussian, vol. 4, no. 4, pp. 1007–1016, 2015.
5. D. Singh and C. K. Reddy, "A survey on platforms for big data analytics," J. Big Data, vol. 2, no. 1, 2015, DOI: 10.1186/s40537-014-0008-6.
6. R. W. Sari, A. Wanto, and A. P. Windarto, "Implementasi Rapidminer Dengan Metode K-Means (Study Kasus: Imunisasi Campak Pada Balita Berdasarkan Provinsi)," KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer), vol. 2, no. 1, 2018, doi: 10.30865/komik.v2i1.930.
7. M. W. Talakua, Z. A. Leleury, and A. W. Talluta, "A cluster Analysis By Using K-Means Method for Grouping of District/City in Maluku Province Industrial Based on Indicators of Maluku Development Index in 2014," Barekeng J. Ilmu Mat. dan Terap., vol. 11, no. 2, pp. 119–128, 2017.
8. M. L. Sibuea and A. Safta, "Pemetaan Siswa Berprestasi Menggunakan Metode K-Means Clustring," Jurteksi, vol. 4, no. 1, pp. 85–92, 2017, doi: 10.33330/jurteksi.v4i1.28.
9. 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, doi: 10.25077/teknosi.v5i1.2019.17-24.
10. T. Suprawoto, "Klasifikasi Data Mahasiswa Menggunakan Metode K-Means Untuk Menunjang Pemilihan Strategi Pemasaran," JIKO (Jurnal Inform. dan Komputer), vol. 1, no. 1, pp. 12–18, 2016, doi: 10.26798/jiko.2016.v1i1.9.
11. U. Indriani, S. Informasi, U. Potensi, and U. Medan, "Pendekatan K-Means Clustering Terhadap," vol. 2, no. 1, pp. 84–88, 2018.