



Analysis to Determine Barriers to Online Learning Using the K-Means Clustering Algorithm: A Case Study of BBPPMPV BOE Malang

Yuanita Safitri Heryati ^a, Alva Hendi Muhammad ^b, Tonny Hidayat ^c

^{a,b,c} Amikom Yogyakarta University, Jl. Ringroad Utara Yogyakarta 55281, Indonesia

DOI: <https://doi.org/10.55248/gengpi.5.0124.0357>

ABSTRACT

BBPPMPV BOE Malang is a Technical Implementation Unit (UPT) within the Directorate General of Vocational Education, Ministry of Education, Culture, Research and Technology. BBPPMPV BOE has the task of carrying out the development of quality assurance of vocational education according to its field, one of which is by implementing Training or Training for Vocational Teachers (SMK) in the Skills Program at BBPPMPV BOE Malang. Since the Covid19 pandemic, BBPPMPV BOE Malang has also held online training in addition to offline training. The online training or training process is not very effective because the participating teachers cannot focus while doing other tasks in their home area, and internet conditions are not the same between one area and another. To evaluate the implementation of online training at BBPPMPV BOE, researchers need to provide research contributions to determine the level of barriers to online learning, which is divided into two clusters, namely low and high. The conclusion of this research is that using the K-means clustering algorithm can make it easier to cluster online learning barriers among online vocational teacher training participants at BBPPMPV BOE Malang.

Keywords: online learning barriers, k-means clustering algorithm

1. Introduction

Education is one of the factors of national progress and independence. The more advanced a nation's education is, the more advanced and independent that nation will be. Through education, the nation's future generations will shape their qualities. Based on these educational goals, the quality and management of learning in schools or educational institutions needs to be improved. One thing that can be done is to increase the competence of teachers as the spearhead of education. [9]

In the present era where information technology has developed so rapidly and has been used in all aspects of life, learning has also experienced a paradigm shift. Especially when the Covid pandemic occurred. Learning is carried out online and blended, combining technology and virtual face-to-face, with application like Zoom, Google Meet, etc. [5]

Balai Besar Pengembangan Penjaminan Mutu Pendidikan Vokasi Bidang Otomotif dan Elektronika or abbreviated as BBPPMPV BOE Malang is a Technical Implementation Unit (UPT) within the Directorate General of Vocational Education, Ministry of Education, Culture, Research and Technology. BBPPMPV BOE has the task of carrying out the development of quality assurance of vocational education according to its field, one of which is by implementing Training or Training for Vocational Teachers (SMK) in the Skills Program at BBPPMPV BOE Malang. The expertise programs at BBPPMPV BOE Malang are Construction Engineering, Property and Geomatics, Mechanical Engineering, Electrical and Electronics Engineering, Automotive Engineering, Computer and Informatics Engineering and Vocational Education Management (<https://bbppmpvboe.kemdikbud.go.id>).

In order to keep up with changing times and comply with government regulations which during the pandemic urge you to stay at home, BBPPMPV BOE Malang is also holding online vocational teacher training. The training participants are spread from vocational schools in Indonesia which are within the BOE target area. The Computer Engineering and Informatics skills program has target areas in the provinces of East Java, West Nusa Tenggara, East Nusa Tenggara, Maluku and North Maluku. Meanwhile, other Skills Programs are in the provinces of East Java, NTT, NTB, Maluku, North Maluku, South Sulawesi, Gorontalo, North Sulawesi, Central Sulawesi, Southeast Sulawesi, West Sulawesi and Bali. [11]

Data mining is the extraction of information, important or interesting patterns from data contained in large databases. Data mining is often also called Knowledge Discovery in Database (KDD), which is an activity of collecting, processing, and using large data to find rules or patterns that can be stored in a database or other storage media. The term KDD or what is called data knowledge discovery emerged because the main goal of Data Mining is to utilize data in a database by processing it so as to produce new, useful information. [9]

The K-Means clustering analysis algorithm is a research field in analysis and data mining. In this algorithm, the grouping technique is based on similarity of data that does not have any reference (unsupervised). However, it will divide all the data into groups or have the same similarities. Basically, this

algorithm calculates the distance between each data and the data center (centroid) to measure similarity. [6] The K-Means method attempts to group existing data into one group, where the data in one group has different characteristics from the data in other groups. [7]

Research on the k-means analysis clustering algorithm which is a reference includes Grouping Limited Face-to-Face Learning After Covid-19 Vaccination With K-Means [1], this research produces 4 groups of absorption capacity, namely high absorption capacity, standard absorption capacity, low absorption and poor absorption capacity. This research produces new knowledge, namely learning cluster groups which inform students' level of absorption in subjects.

Then research on crime rate mapping analysis in Karawang Regency uses the k-means algorithm [2]. Research to analyze crime levels in Karawang Regency using the k-means algorithm was carried out by Fahmi, R, et al, which produced clusters of vulnerable, not vulnerable and very large. This research uses data based on location, number of cases and type of cases in 2019 and 2020.

Application of the K-Means Algorithm to New Vocational High School Students for Major Clustering [3]. This research aims to group new student data according to criteria to form major groups using the k-means clustering algorithm. The results of this grouping were three groups, namely the failed group, the software engineering group and the computer network engineering group.

Application Of Data Mining With The K-Means Clustering Method And Davies Bouldin Index For Grouping IMDB Movies which analyzes and groups films on IMDB based on criteria [4]. This research classifies films from the IMDB database obtained from Kaggle based on rating variables, meta scores, votes, gross.

Implementation of k-means clustering analysis to determine barriers to online learning among case study students: Yapendak Tinjowan Private Middle School which produces low and high clusters based on the level of barriers to online learning [5]. The research results can be used as a reference for evaluating learning for students in the low cluster to add more intensive guidance in learning.

Implementation of the k-means clustering algorithm in determining superior class students which manages centralized assessment data and produces superior student groupings [6]. Researchers developed a web-based information system to group student data into superior classes based on learning outcomes reports.

Based on this, the researcher tried to create research entitled "Analysis to determine barriers to online learning using the K-means Clustering Algorithm: a Case Study of BBPMPV BOE Malang".

The aim of this research is to carry out an analysis to determine barriers to online learning using the k-means clustering algorithm, case study: online vocational teacher training participants at BBPMPV BOE Malang.

2. Method

This type of research is quantitative research with a clustering approach and data processing. Quantitative research is a systematic, planned and structured type of research. Quantitative methods are research methods in which each process is carried out using data in the form of numbers. The stages in this research can be explained by data collection and data analysis. [5]

The nature of this research was carried out independently by taking a quantitative approach in clustering data from training or training participants. The object studied is a dataset taken from the training system database at BBPMPV BOE Malang which will be processed first so that it can be used for clustering.

The data collection method used was by taking data from the training system database at BBPMPV BOE Malang, namely Alfa-T. The data taken is specifically Online Training data which was implemented in 2021 in the Information Technology skills program. The data amounts to around 220 records out of 8000 total data taken from the database.

The data analysis method that will be used is the K-means clustering method. The first step in the data analysis process is to adjust the format in the Excel file so that it can then be processed into the application. The analysis that will be carried out in research with this dataset is as follows:

1. Process the data and assign weights to the specified attributes.
2. Search for groups from the dataset with the K-means clustering algorithm
3. Analyze the results of clustering

2.1 Knowledge Discovery in Database (KDD)

KDD is a method used to obtain knowledge originating from existing databases. The results of the knowledge obtained can be used as a knowledge base used for decision making purposes. In more detail, the KDD process as in the following figure was adopted from [10]:

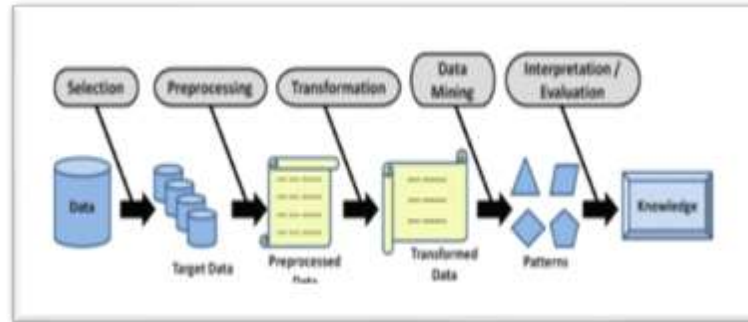


Fig. 1 Process of Knowledge Discovery in Database (KDD)

1. Selection

Selection is used to determine the variables that will be taken so that there are no similarities and unnecessary repetition occurs in data mining processing. The dataset used comes from the training system database at BBPMPV BOE Malang, namely Alfa-T. The data taken is specific to Online Training data carried out in 2021, for training carried out at the Information Technology Program. The data amounts to around 220 records out of 8000 total data taken from the database. The variables from the data that are: Training Class Name, Participant_Name, NIK, NIP, School_Origin, District/City, Province, Age, Gender, Final_Grade.

2. Preprocessing

In preprocessing there are two stages, namely as follows:

- a. Data Cleaning: Eliminate unnecessary data such as handling missing values, noisy data and handling inconsistent and relevant data.
- b. Data Integration: Performed on attributes that identify unique entities.

After the data is obtained, the preprocessing process will then be carried out by cleaning the data. It aims to normalize data, eliminate data inconsistencies, and fix data duplication, or unwanted data. In this research there is still some data that is not needed so that it must be eliminated and cleaned so that the data processing process becomes more leverage, unused data are NIK, gender.

3. Transformation

Changing the data according to the appropriate extension format for data mining processing because some data mining methods require a special format before they can be processed in data mining. If the data used consists of several variables that do not match, it is also necessary to transform these variables. The K-Means Clustering algorithm processes data in numerical form [16]. There is some data that must be adjusted. The following are the results of adjusting the data.

Table 1. Data from Pre-processing Results of Online Training Participants

No	Name	Employee_Status	School_Status	Java_NonJava	Age	Final_Grade
1	Anna Novita	2	2	2	38.00	92.338
2	Fariadin	1	2	1	33.00	28.5
3	Rina Ekawati	1	1	2	40.00	92.75
4	Baltasar Buarlely	2	2	1	46.00	82.013
...						
205	Elfatul Ummah	1	1	2	28.00	83.25
206	Muhamad Saidul Hulam	1	2	2	34.00	87.25

4. Data mining

After the data transformation process, the next step is the data mining process. The data mining process is carried out for patterns from previously selected data using certain algorithms or techniques. The method used in this research is to use k means. The k-means method is a method of grouping data by taking parameters from several k clusters and dividing the data into clusters based on the similarities between the data in a cluster and the differences between clusters. [4]

5. Evaluation/ Interpretation

The last stage is the process of translating the results of the pattern from the data mining process. The resulting pattern of information will then be displayed in the form of data visualization. Presentation of data using visualization makes the data easier to understand. Identify interesting patterns into

the identified knowledge base. At this stage, typical patterns and prediction models are produced which are evaluated to assess whether existing studies have met the desired targets.

2.2 K-Means Clustering Algorithm Process

The K-Means algorithm is an iterative clustering algorithm that partitions a data set into a number of K clusters that have been determined at the beginning. The steps for clustering using the K-Means method are as follows: [1]

1. Determine the k value as the number of clusters you want to form.
2. Initialize k cluster centers (centroids) using the formula [8]:

$$C_i = Min + \frac{(i-1) \cdot (max-min)}{n} + \frac{(max-min)}{2n} \quad (1)$$

3. Calculate the distance of each input data to each centroid using the Euclidean distance formula until the closest distance of each data to the centroid is found. The following is the Euclidian Distance equation:

$$d = \sqrt{\sum_{i=1}^n (x_{2i} - x_{1i})^2} \quad (2)$$

4. Classify each data based on its proximity to the centroid (smallest distance).
5. Compare the total centroid difference and average data for each cluster with the allowable error (error tolerance). If it is smaller, you are done, if not, replace the centroid with the average of the data and repeat step three.

3. Result and Discussion

Clustering in determining online learning barriers for online vocational teacher training participants at BBPPMPV BOE Malang was carried out by manual calculation using Microsoft Excel. The number of clusters used in this research is 2 clusters, namely the low cluster and the high cluster. Training participant data and criteria can be seen in Table 2.

Table 2 - Dataset of online training participants at BBPMPPV BOE Malang

Id	Nik	Nip	Nama	Tanggal_1 ahir	Npsn	Sekol ah	Jenis_kela min	Usia_s kr	Nilai_ak hir	gra de	ket
187 54	332413010880 0001	1980080120100 11020	Nanang Abdullah	01/08/1980	203218 47	SMK N 2 Kenda l	PRIA	42 thn, 07 bln, 11 hari	85.56	B	Baik
188 09	352506220667 0002	1967062220070 11017	Heri Pratikto	22/06/1967	205017 15	SMK N 1 Sidoar jo	PRIA	55 thn, 08 bln, 21 hari	83.24	B	Baik
188 10	351709640271 0002	1971022420001 22002	Erna Kusrini	24/02/1971	205033 72	SMK N 3 Jomba ng	WANITA	51 thn, 12 bln, 17 hari	87.07	B	Baik
188 11	351708160480 0008	1980041620100 11011	Rudi Cahyo Purnomo	16/04/1980	205033 71	SMK N Kudu, Kab. Jomba ng	PRIA	42 thn, 10 bln, 26 hari	76.37	C	Cuk up
188 13	357303261178 0006	1978112620110 11001	Chairur Razikin	26/11/1978	205338 18	SMK N 6 Malan g	PRIA	44 thn, 03 bln, 16 hari	86.64	B	Baik
188 15	352011650272 0001	1972022520031 22001	Yayuk Setya Purwanin gsih	25/02/1972	205341 79	SMK N 1 Madiu n	WANITA	50 thn, 12 bln, 16 hari	71.95	C	Cuk up

Data from online training at BBPMPPV BOE Malang was pre-processed. Pre-processing by selecting the required fields, cleaning unused and null data.

The results of pre-processing (field selection, cleaning of unused data) are shown in table 1. The data has also been filtered only for training in Informatics Engineering skills in 2021, so the data is 206.

The next step is to determine the centroid value, namely by using the formula in (1). Because there are 5 variables used in the data, and here we will determine 2 clusters, namely k1 and k2. k1 is for hampered learning, and k2 is for smooth learning, so look for 5 centroids or center points for each cluster. By determining the max and min values in the data, then entering them into the formula, you will know the centroid value for each variable C1 to C5 for each cluster.

Table 3 - Centroid Calculation Results Using the Formula

	C1	C2	C3	C4	C5
K1 (hampered)	1.25	1.25	1.25	33.75	41.875
K2 (smooth)	1.25	2	2	41.625	91.81075

After knowing the centroid results or center point values, the distance will be searched to find out the clustering results for each data into which cluster. The distance sought is the value of each data to the centroid value. How to determine the distance using the Euclidian Distance formula in formula (2).

The results of the distance calculation can be seen below.

Table 4 - Results of distance calculations and cluster determination

No	Name	Data Distance to Class		Cluster
		K1	K2	
1	Anna Novita	50.65830997	3.739133264	k2
2	Fariadin	13.42164763	63.90386679	k1
3	Rina Ekawati	51.26417487	2.141334995	k2
4	Baltasar Buarlely	41.97986474	10.80273253	k2
5	Dhane Pratignyo Setiyawan	49.68695044	1.260765665	k2
6	Meilida Argitasari	50.17235917	9.656031305	k2
7	Novi Khassifa	52.49167196	9.077495005	k2
....				
205	Elfatul Ummah	41.78086434	16.12419194	k2
206	Muhamad Saidul Hulam	45.38877201	8.888394994	k2

From the calculation results, it was found that for cluster k1 (hampered) there were 5 data, while k2 (smooth) had 201 data.

Table 5 – Clustering Result

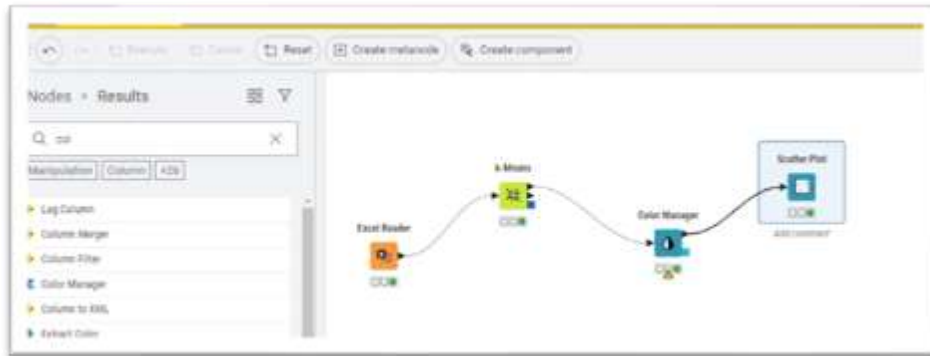
No	Name	Age	Final_Score	Data Distance to Class		Cluster
				K1	K2	
1	Anna Novita	38.00	92.338	50.65830997	3.739133264	k2
2	Fariadin	33.00	28.5	13.42164763	63.90386679	k1
3	Rina Ekawati	40.00	92.75	51.26417487	2.141334995	k2
4	Baltasar Buarlely	46.00	82.013	41.97986474	10.80273253	k2
5	Dhane Pratignyo Setiyawan	41.00	91.013	49.68695044	1.260765665	k2
6	Meilida Argitasari	32.00	92	50.17235917	9.656031305	k2
7	Novi Khassifa	33.00	94.35	52.49167196	9.077495005	k2
....						
205	Elfatul Ummah	28.00	83.25	41.78086434	16.12419194	k2
206	Muhamad Saidul Hulam	34.00	87.25	45.38877201	8.888394994	k2

From the results of clustering with the first iteration, the error tolerance was checked by reducing the average data in clusters 1 and 2 with centroids. It was found that the results of reducing the cluster average with centroids were smaller than the error tolerance, so the iteration only had to be done once. The way to calculate error tolerance is by using the formula: $\Delta * (\max - \min)$. Error tolerance is sought for each variable used in clustering. [1]

Table 6 - Comparison results of cluster mean minus centroid with error tolerance.

	Centroid Mean				
	RV1 - C1	RV2 - C2	RV3 - C3	RV4 - C4	RV5 - C5
	-0.05	0.35	0.55	-0.35	0
	0.088308458	-0.472636816	-0.069651741	-3.779228856	-3.105068408
Error Tolerance	0.01	0.01	0.01	0.31	0.775

The next implementation carried out was clustering with the KNime application. The following are the nodes used in the KNime application to perform clustering with KNime.

**Fig. 2 Workflow and Nodes used in the clustering process in KNime****Fig. 3 Scatter Plot of Clustering results**

There are differences in the results of clustering carried out in Excel and the KNime application. This happens because the centroid determination is carried out randomly in KNime, whereas in Excel a detailed calculation is carried out using the formulas listed in formulas (1) and (2).

5. Conclusion

The conclusion obtained from this research is that the implementation of k-means clustering can determine barriers to online training at BBPMPPV BOE Malang. The cluster results consist of two, namely k1 (low or hampered cluster) for participants who experience obstacles in learning online training, as many as 5 people and k2 (high or smooth cluster) namely for participants who do not experience learning obstacles in online training as many as 201 people. And there are differences in the calculation results with k-means clustering on Ms. Excel and the KNime Application, because Excel determines the centroid in Ms. Excel uses formula calculations, whereas in the KNime application it is determined randomly.

The suggestion that researchers can conclude is that for online learning, there are variables that influence the quality of learning, but do not really influence learning outcomes. What plays a more important role in online learning is the commitment of participants in participating in online training and learning processes in the midst of their respective routine activities. The training team at BBPMPPV BOE can screen training participants, with teachers who really have the will and commitment to take part in online training.

References

- Lili Kartikawati, Kusri, Emha Taufiq Lutfi (2021). Algoritma K-Means Pada Pengelompokan Pembelajaran Tatap Muka Terbatas Sesudah Vaksinasi Covid-19. *Jurnal Eksplorasi Informatika* [1]
- Resti Noor Fahmi, Mohamad Jajuli, Nina Sulistyawati (2021). Mapping Analysis Of Criminality Level In Karawang Using K-Means Algorith. *Journal of Information Technology and Computer Science (INTECOMS)* [2]
- Fauziah Nur, Prof. M. Zarlis, Dr. Benny Benyamin Nasution (2017). Penerapan Algoritma K-Means Pada Siswa Baru Sekolah Menengah Kejuruan Untuk Clustering Jurusan. *Jurnal Nasional Informatika dan Teknologi Jaringan*. [3]
- Ilham Firman Ashari, Romantika Banjarnahor, Dede Rodhatul Farida (2022). Application of Data Mining with the K-Means Clustering Method and Davies Bouldin Index for Grouping IMDB Movies. *Journal of Applied Informatics and Computing (JAIC)* [4]
- Dinah Adillah, Nuriadi Manurung, Ari Dermawan (2022). Implementation Of K-Means Clustering Analysis To Determine Barriers To Online Learning Case Study: Swasta Yapendak Tinjowan Junior High School. *Jurnal Teknik Informatika (JUTIF)* [5]
- Ari Sulistiyawati, Eko Supriyanto (2021). Implementasi Algoritma K-means Clustering dalam Penentuan Siswa Kelas Unggulan. *Jurnal TEKNO KOMPAK* [6]
- Desy Noor Permata Saria, YL. Sukestiyarno (2021). Analisis Cluster dengan Metode K-Means pada Persebaran Kasus COVID-19 Berdasarkan Provinsi di Indonesia. *PRISMA, Prosiding Seminar Nasional Matematika* [7]
- Lili Kartikawati (2022). Analisis Kualitas Pengelompokan Algoritma K-Means di Knime dan Excel untuk PTMT Pasca Vaksinasi Covid-19. *Ideguru : Jurnal Karya Ilmiah Guru* [8]
- Heri Susanto, Sudiyatno (2014). Data Mining Untuk Memprediksi Prestasi Siswa Berdasarkan Sosial Ekonomi, Motivasi, Kedisiplinan Dan Prestasi Masa Lalu. *Jurnal Pendidikan Vokasi* [9]
- Gustientiedina, M.Hasmil Adiya, Yenny Desnelita (2019). Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan Pada RSUD Pekanbaru. *Jurnal Nasional Teknologi dan Sistem Informasi* [10]
- Permendikbud RI (2020). Peraturan Menteri Pendidikan Dan Kebudayaan Republik Indonesia Nomor 26 Tahun 2020 Tentang Organisasi Dan Tata Kerja Unit Pelaksana Teknis Kementerian Pendidikan Dan Kebudayaan.[11]