

# Pengelompokan Perguruan Tinggi berdasarkan Partisipasi pada Program Merdeka Belajar-Kampus Merdeka dengan Metode K-Means

**Titania Tasya Wananda\*, Jaka Nugraha**

Program Studi Statistika, Universitas Islam Indonesia, Jl. Kaliurang KM 14.5, DI Yogyakarta, Indonesia 55584

\*Corresponding author: [titania.wananda@alumni.uii.ac.id](mailto:titania.wananda@alumni.uii.ac.id).



E-ISSN: 2986-4178

## Riwayat Artikel

Dikirim:

3 Januari 2023

Direvisi:

29 April 2023

Diterima:

04 Juni 2023

## ABSTRAK

Merdeka Belajar-Kampus Mengajar (MBKM) merupakan program Kementerian Pendidikan dan Kebudayaan yang bertujuan agar menghasilkan kualitas Pendidikan yang lebih baik seperti mahasiswa yang memiliki kemampuan analisis yang tajam dan pemahaman komprehensif dalam mengembangkan diri. MBKM terdiri dari beberapa subprogram diantaranya yaitu magang, studi independen, kampus mengajar, dan lainnya. Jumlah mahasiswa peserta program magang, studi independen, dan kampus merdeka sebesar 0.45% dari total keseluruhan mahasiswa. Tujuan dari penelitian ini untuk mengetahui bagaimana tingkat partisipasi mahasiswa dari setiap perguruan tinggi dalam program MBKM dengan menggunakan analisis *cluster* K-Means. Pada penelitian ini terbentuk 2 *cluster* dengan jumlah anggota *cluster* 1 sebanyak 17 perguruan tinggi dan *cluster* 2 sebanyak 1359 perguruan tinggi. *Cluster* 1 memiliki karakteristik peserta program magang dan studi independen tergolong rendah, sedangkan peserta kampus mengajar tergolong tinggi. *Cluster* 2 memiliki karakteristik peserta program magang dan studi independen tergolong tinggi, sedangkan peserta kampus mengajar tergolong rendah.

**Kata Kunci:** Clustering, K-Means, MBKM, Perguruan Tinggi.

## ABSTRACT

*Merdeka Learning-Campus Teaching (MBKM) is a program of the Ministry of Education and Culture which aims to produce better quality education such as students who have sharp analytical skills and comprehensive understanding in developing themselves. MBKM consists of several subprograms including internships, independent studies, campus teaching, and others. The number of students participating in the internship program, independent study, and independent campus is 40,944 students, while the total number of students is 9,188,101 students. The purpose of this study was to find out how the level of student participation from each university in the MBKM program using K-Means cluster analysis. In this study, 2 clusters were formed with the number of members of cluster 1 as many as 17 universities and cluster 2 as many as 1359 universities. Cluster 1 has the characteristics of low class participants in the internship and independent study programs, while the campus teaching participants are classified as high. Cluster 2 has high characteristics of participants in the internship and independent study programs, while those on campus teaching are classified as low.*

**Keywords:** Clustering, K-Means, MBKM, College.

## 1. Pendahuluan

Merdeka Belajar merupakan salah satu program Menteri Pendidikan dan Kebudayaan agar menghasilkan kualitas pendidikan yang lebih baik seperti mahasiswa

yang memiliki kemampuan analisis yang baik dan dapat mengembangkan diri secara komprehensif [1]. Di era Pendidikan 4.0 ini Pendidikan Tinggi diharapkan mampu dalam menerapkan pembelajaran yang inovatif sehingga mahasiswa dapat mencapai hasil belajar secara optimal dan relevan. Kebijakan Merdeka Belajar-Kampus Merdeka diharapkan dapat mengatasi permasalahan yang ada di era Pendidikan 4.0. Program MBKM memberikan kesempatan kepada mahasiswa untuk mengasah kemampuan dengan terjun langsung ke dunia kerja. Kampus Merdeka sendiri terdiri dari beberapa program antara lain Magang, Studi Independen, Pertukaran Pelajar, dan Asisten Mengajar.

Peserta program magang, studi independen, dan kampus merdeka tahun 2021 berjumlah 40.944 mahasiswa. Sedangkan total mahasiswa secara keseluruhan sebesar 9.188.101 mahasiswa. Rasio antara peserta dan total mahasiswa kurang dari 0.05 persen. Sehingga tingkat partisipasi mahasiswa terhadap program MBKM masih sangat rendah. *Clustering* atau pengelompokan merupakan metode untuk mengetahui karakteristik mahasiswa dalam program MBKM dari masing-masing perguruan tinggi. Data yang memiliki karakteristik sama akan dikelompokkan kedalam satu *cluster* yang sama, sehingga Kemendikbud dapat mengetahui perguruan tinggi yang memiliki tingkat partisipasi rendah.

Penelitian ini membutuhkan beberapa riset sebelumnya terkait analisis *cluster non hierarki* dan metode *K-Means*. Penelitian pertama dilakukan untuk menentukan status gizi balita dengan menggunakan analisis *clustering* metode *K-Means* dan dikelompokkan dalam 5 *cluster* dengan label gizi buruk, gizi kurang, gizi baik, gizi lebih, dan obesitas [2]. Penelitian kedua menggunakan analisis *cluster* metode *K-Means* dan *Fuzzy C-Means* pada bidang ekonomi [3], Penelitian ketiga dilakukan oleh [4] dengan menggunakan metode *K-Means* pada bidang medis.

## **2. Metodologi Penelitian**

### **2.1. Data dan Sumber Data**

Penelitian ini menggunakan data sekunder yang diperoleh dari Direktorat Jendral Pendidikan Tinggi Kemendikbud. Data yang digunakan adalah data peserta MBKM Magang, Studi Independen, dan Kampus Merdeka tahun 2021. Variabel yang digunakan pada penelitian ini antara lain Perguruan Tinggi, Magang, Studi Independen, dan Kampus Mengajar.

### **2.2. Landasan Teori**

Penelitian ini menggunakan metode analisis deskriptif dan analisis *cluster non hierarki* dengan menggunakan metode *K-Means*. Analisis deskriptif digunakan untuk mengetahui gambaran data yang digunakan secara umum. Analisis *cluster K-means* digunakan untuk mengelompokkan data dengan karakteristik yang sama.

#### **Statistika Deskriptif**

Statistik secara garis besar dibagi menjadi statistik deskriptif dan statistik inferensia. Statistik deskriptif adalah analisis data dengan cara mendeskripsikan atau menggambarkan data tanpa membuat kesimpulan secara umum [5]. Nilai yang termasuk dalam statistika deskriptif adalah rata-rata, modus, median, varians, simbbangan baku, dan ukuran lainnya.

#### **Analisis Cluster**

Analisis cluster merupakan analisis yang digunakan untuk mengelompokkan elemen yang mirip menjadi satu kelompok (*cluster*). Analisis *cluster* termasuk dalam analisis statistik multivariat metode interdependen dengan mengelompokkan objek

berdasarkan karakteristik tertentu yang akan diteliti. Analisis cluster terbagi menjadi dua metode yaitu analisis cluster metode hierarki dan analisis cluster metode non hierarki [5].

### Analisis Cluster Non-Hierarki

Analisis *clustering* non hierarki terlebih dahulu menentukan jumlah cluster. Metode *clustering* non hierarki yang terkenal adalah metode *k-means* yang digunakan sebagai alternatif untuk ukuran data yang besar [6].

### Uji Sampel Mewakili Populasi

Uji sampel yang mewakili populasi merupakan sampel yang diuji apakah mewakili populasi yang ada dengan menggunakan sampel yang cukup besar. Pengujian sampel dilakukan dengan menggunakan uji Kaiser-Mayer-Olkin (KMO) yang mengukur kecukupan sampling pada setiap indikator secara menyeluruh. Nilai KMO *Measures of Sampling Adequacy* (MSA) berkisar 0.5 sampai 1 menunjukkan bahwa sampel representatif [3].

$$KMO = \frac{\sum_{i=1}^n \sum_j^n r_{ij}^2}{\sum_{i=1}^n \sum_j^n r_{ij}^2 + \sum_{i=1}^n \sum_j^n a_{ij}^2} \quad (1)$$

$$a_{ij} = \frac{-r_{ij}}{\sqrt{r_{ij}r_{ij}}}, \quad (2)$$

dengan  $n$  = banyaknya peubah,  $r_{ij}$ = koefisien korelasi antara peubah  $i$  dan peubah  $j$ ,  $a_{ij}$ = koefisien korelasi parsial antara peubah  $i$  dan peubah  $j$ .

### Uji Multikolinearitas

Multikolinieritas merupakan keadaan dimana menggunakan lebih dari satu variabel predictor. Nilai Variance Inflation Factor (VIF) digunakan untuk mengetahui apakah terdapat multikolinearitas antar variabel [7].

$$a_{ij} = \frac{-r_{ij}}{\sqrt{r_{ij}r_{ij}}} \quad (3)$$

Multikolinearitas antar variabel dapat terjadi apabila nilai  $VIF \geq 10$  dan perlu dilakukan analisis lanjutan dengan menggunakan metode *Principal Component Analysis* (PCA).

### K-Means Cluster Analysis

*K-Means* merupakan salah satu metode *clustering non hierarki* yang mempartisi data menjadi dua kelompok atau lebih dengan mengelompokkan data yang memiliki karakteristik sama kedalam satu kelompok dan data yang berbeda karakteristik akan dikelompokkan pada kelompok lainnya [4]. Algoritma metode *K-Means Cluster Analysis* [2] :

1. Menentukan jumlah *cluster*
2. Menentukan nilai *centroid* dari data yang ada di setiap kelompok dengan menggunakan nilai rata-rata (*mean*).

$$v_{ij} = \frac{1}{N_i} \sum_{k=0}^{N_i} x_{kj}, \quad (4)$$

dengan  $v_{ij}$ = centroid atau rata-rata cluster ke- $i$  dan variabel ke- $j$ ;  $N_i$ = jumlah data anggota *cluster I*;  $i, k$  = indeks cluster; dan  $j$  = indeks variabel.

- Menghitung jarak setiap objek dengan titik centroid menggunakan *Euclidean Distance*.

$$d = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}, \quad (5)$$

dengan,  $d$ = Euclidean distance,  $x_i$  = objek x ke- $i$ , dan  $y_i$  = objek y ke- $i$ .

- Pengelompokan objek untuk menentukan anggota cluster dengan menghitung jarak minimum objek. Nilainya berkisar antara 0 sampai 1, dimana data dengan nilai 1 masukan kedalam cluster I, dan nilai 0 untuk cluster yang lain.
- Mengulangi iterasi pada persamaan (4) hingga nilai centroid hasil perulangan tetap dan anggota cluster tidak berpindah ke cluster lain.

### Metode *Elbow*

Metode *Elbow* merupakan metode untuk menentukan jumlah cluster atau nilai  $k$  secara optimal pada analisis cluster *K-Means*. *Total of within-cluster variation* atau *total of within-cluster sum of square* yang diminimalkan merupakan ide dasar dari *unsupervised* (salah satunya: *K-Means*) untuk menentukan banyaknya cluster. Nilai *Sum of Square* (SSE) digunakan untuk menentukan *cluster* optimal dengan rumus berikut [8].

$$SSE = \sum_{k=1}^k \sum_{x_i \in S_k} \|x_i - C_k\|^2, \quad (6)$$

dimana,  $k$  = Jumlah *cluster*,  $x_i$  : data ke- $i$ ,  $C_k$  = *Centroid* dari *cluster*. Semakin kecil nilai SSE maka jumlah cluster  $k$  akan semakin besar. Hasil dari perhitungan setiap cluster dengan metode *Elbow* dapat disajikan dengan menggunakan grafik.

## 3. Hasil dan Pembahasan

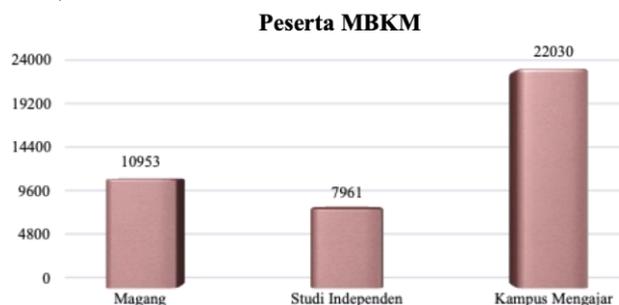
### 3.1. Analisis Deskriptif

Data peserta MBKM setiap perguruan tinggi di Indonesia dilakukan analisis deksriptif untuk mengetahui gambaran secara umum dari masing-masing variabel.

**Tabel 1** Deskriptif Statistik Data Peserta MBKM

Variabel	N		Minimum	Maximum	Mean
	Valid	Missing			
Perguruan Tinggi	1376	0	NA	NA	NA
Magang	1376	0	0	493	8
Studi Independen	1376	0	0	268	6
Kampus Mengajar	1376	0	0	1022	16

Berdasarkan Tabel 1 diperoleh hasil statistik deskriptif dari data peserta MBKM dengan variabel Perguruan Tinggi, Magang, Studi Independen, dan Kampus Mengajar terdapat 1376 data valid dan tidak terdapat data yang hilang (*missing value*). Variabel Perguruan Tinggi merupakan data string sehingga tidak dapat dihitung untuk nilai minimum, maximum, dan rata-rata.



**Gambar 1.** Jumlah Peserta MBKM

Berdasarkan grafik pada Gambar 1. menunjukkan jumlah peserta MBKM pada setiap program. Program MBKM yang memiliki paling banyak peserta adalah Kampus Mengajar dengan total peserta 22030 mahasiswa.

### 3.2. Uji Sampel Mewakili Populasi

Uji sampel mewakili populasi digunakan untuk mengetahui apakah sampel representatif atau tidak dengan menggunakan statistik uji KMO. Berdasarkan Output SPSS didapatkan nilai KMO MSA sebesar 0.533 dan nilai signifikansi sebesar 0.000. Diketahui bahwa nilai KMO lebih dari 0.500 dan nilai signifikansi kurang dari 0.05. Sehingga, data yang ada menunjukkan bahwa sampel mewakili populasi atau sampel representatif.

### 3.3. Uji Sampel Mewakili Populasi

Uji Multikolinearitas digunakan untuk mengetahui ada atau tidaknya multikolinearitas antar variabel dengan menggunakan nilai VIF.

Tabel 2 Uji Multikolinearitas

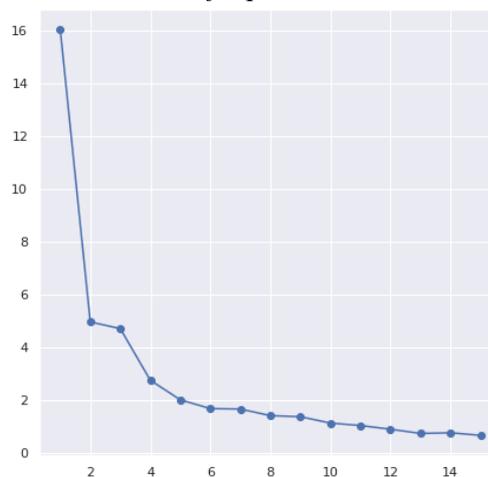
Variabel	VIF
Magang	1.222
Studi Independen	1.222
Kampus Mengajar	1.117

Berdasarkan Output SPSS pada Tabel 2. didapatkan nilai VIF pada ketiga variabel kurang dari 10 sehingga dapat disimpulkan bahwa tidak terdapat multikolinearitas antar variabel dan asumsi uji multikolinearitas terpenuhi.

### 3.4. K-Means Cluster Analysis

#### 3.4.1. Jumlah Cluster Optimal

*Clustering* dengan menggunakan metode K-Means terlebih dahulu dengan menentukan jumlah *k cluster* yang paling optimal dengan menggunakan metode *Elbow*. Metode *Elbow* menggunakan nilai *Sum of Square Error* untuk menentukan *k* optimal.



Gambar 2. Grafik Elbow

Berdasarkan Gambar 2. dapat dilihat bahwa garis yang mengalami patahan membentuk *elbow* atau siku pada saat  $k=2$ . Sedangkan garis pada nilai  $k$  lainnya cenderung

landai dan tidak membentuk siku sehingga, pada penelitian kali ini menggunakan  $k=2$  untuk melakukan analisis cluster.

### 3.4.2. Clustering

Proses clustering dilakukan dengan membagi cluster sebanyak 2 kelompok sesuai hasil *within cluster sum of square* dengan menggunakan *software* SPSS. Hasil iterasi didapatkan hasil seperti pada Tabel 3.

**Tabel 3 Iterasi**

Iterasi	Change in cluster centers	
	1	2
1	324.873	508.452
2	51.217	0.576
3	24.015	0.248
4	24.607	0.257
5	21.042	0.232
6	17.668	0.220
7	16.482	0.212
8	0.000	0.000

Proses iterasi dilakukan sebanyak delapan kali dan dapat dilihat bahwa pada iterasi pertama sampai ketujuh nilai *centroid* belum signifikan karena nilai perubahan dalam pusat cluster lebih dari 0.05. Pada iterasi ke-delapan didapatkan nilai perubahan pada setiap cluster sebesar 0.000 dimana kurang dari 0.05 sehingga *centroid* sudah signifikan dan proses iterasi dapat dihentikan.

**Tabel 4 Jumlah Anggota Cluster**

Cluster	Anggota
1	17
2	1359

Berdasarkan pada Tabel 4 didapatkan jumlah anggota pada setiap cluster yang dikelompokkan berdasarkan kemiripan karakteristik. Diketahui anggota perguruan tinggi pada cluster 1 seperti pada Tabel 5.

**Tabel 5 Anggota Cluster**

Perguruan Tinggi	Cluster	Magang	Studi Independen	Kampus Mengajar
1. Universitas Ahmad Dahlan	1	39	29	477
2. Universitas HKBP Nommensen	1	12	13	349
3. Universitas Jambi	1	33	16	492
4. Universitas Jember	1	122	123	401
5. Universitas Lampung	1	115	106	289
6. Universitas Muhammadiyah Surakarta	1	59	56	491
7. Universitas Negeri Makassar	1	60	52	1022
8. Universitas Negeri Malang	1	104	102	898
9. Universitas Negeri Medan	1	53	39	802
10. Universitas Negeri Padang	1	32	27	382
11. Universitas Negeri Semarang	1	110	99	742
13. Universitas Negeri Yogyakarta	1	66	53	299
14. Universitas Nusa Cendana	1	4	4	562
15. Universitas Pendidikan Indonesia	1	248	180	994

Perguruan Tinggi	Cluster	Magang	Studi Independen	Kampus Mengajar
16. Universitas Sultan Ageng Tirtayasa	1	69	70	283
17. Universitas Sumatera Utara	1	166	127	297

Tabel 5 berisi perguruan tinggi yang merupakan anggota cluster 1, sedangkan 1359 Perguruan Tinggi lainnya masuk kedalam cluster 2. Terdapat rata-rata total setiap cluster yang ditampilkan pada tabel 6.

**Tabel 6** Final Cluster Centers

Karakteristik	Cluster	
	1	2
Peserta Magang	85	7
Peserta Studi Independen	72	5
Peserta Kampus Mengajar	546	9

Cluster 1 memiliki rata-rata jumlah mahasiswa pada ketiga program lebih tinggi dibandingkan dengan cluster 2. Jarak titik pusat antara cluster 1 dan 2 sebesar 546.42 yang menunjukkan cluster cukup baik karena memiliki nilai *between cluster* yang tinggi. *Between cluster* merupakan nilai perbedaan antar *cluster*. Oleh karena itu, semakin besar nilai *between cluster* maka cluster satu dengan yang lain akan semakin renggang dan terlihat jelas perbedaannya.

### 3.4.3. Profiling

Proses profiling cluster digunakan untuk menjelaskan karakteristik masing-masing cluster berdasarkan pendaftar setiap program MBKM yaitu Magang, Studi Independen, dan Kampus Mengajar yang disajikan dalam Tabel 7 seperti berikut.

**Tabel 7** Profiling Cluster

Karakteristik	Cluster	
	1	2
Peserta Magang	Tinggi	Rendah
Peserta Studi Independen	Tinggi	Rendah
Peserta Kampus Mengajar	Tinggi	Rendah

*Profiling* dari masing-masing cluster mengenai peserta program MBKM dijelaskan seperti berikut.

- a. *Cluster 1*  
Berdasarkan hasil *clustering* peserta program MBKM dengan 17 Perguruan Tinggi memiliki karakteristik yang sama yaitu jumlah peserta magang, studi independen, dan kampus mengajar tergolong tinggi.
- b. *Cluster 2*  
Berdasarkan hasil *clustering* peserta program MBKM dengan 1359 Perguruan Tinggi memiliki karakteristik yang sama yaitu peserta magang, studi independen, dan kampus mengajar tergolong rendah.

## 4. Kesimpulan

Kesimpulan yang didapatkan dari penelitian terkait partisipasi mahasiswa setiap perguruan tinggi pada program MBKM dengan menggunakan *analisis cluster non hierarki* metode *K-Means* adalah sebagai berikut :

1. Pengelompokan perguruan tinggi terhadap Program Merdeka Belajar Kampus Merdeka (MBKM) berdasarkan peserta program Magang, Studi Independen, dan Kampus Merdeka didapatkan 2 *cluster* dengan cluster 1 sebanyak 17 perguruan tinggi dan *cluster 2* sebanyak 1359 perguruan tinggi.

2. Karakteristik dari setiap *cluster* pada perguruan tinggi terhadap Program Merdeka Belajar Kampus Merdeka (MBKM) seperti berikut.
  - a. *Cluster* 1 merupakan hasil *clustering* peserta program MBKM dengan 17 perguruan tinggi memiliki karakteristik yang sama yaitu peserta magang, studi independen, dan kampus mengajar tergolong tinggi.
  - b. *Cluster* 2 merupakan hasil *clustering* peserta program MBKM dengan 1359 perguruan tinggi memiliki karakteristik yang sama yaitu peserta magang, studi independenkampus mengajar tergolong rendah.

Sehingga Ditjen Dikti dapat mengetahui perguruan tinggi yang memiliki tingkat partisipasi rendah pada program MBKM dan dapat mendorong mahasiswa dari perguruan tinggi tersebut untuk berpartisipasi dalam MBKM.

## 5. Daftar Pustaka

- [1] M. Saleh, "Merdeka Belajar di Tengah Pandemi Coivd-19," pp. 51-56, 2020.
- [2] W. M. P. Duhita, "Clustering Menggunakan Metode K-Means untuk Menentukan Status Gizi Balita," *Jurnal Informatika*, pp. 160-174, 2015.
- [3] D. R. Ningrat, D. A. I. Maruddani and T. Wuryandari , "Analisis Cluster dengan Algoritma K-Means dan Fuzzy C-Means Clustering untuk Pengelompokan Data Obligasi Korporasi," *Jurnal Gaussian*, pp. 641-650, 2016.
- [4] G. M. H. Adiya and Y. Desnelita, "Penerapan Algoritma K-Means untuk Clustering Data Obat-Obatan pada RSUD Pekanbaru," *Jurnal Nasional Teknologi dan Sistem Informasi*, pp. 17-24, 2019.
- [5] S. Metode Penelitian Pendidikan (Pendekatan Kuantitatif, Kualitatif, dan R&D), Bandung: Alfabeta, 2015.
- [6] R. Sitepu, I. and B. Gultom, "Analisis Cluster terhadap Tingkat Pencemaran Udara pada Sektor Industri di Sumatera Selatan," *Jurnal Penelitian Sains*, pp. 11-17, 2011.
- [7] R. Windasari, "Analisis Cluster Hirarki Merode Average Linkage Berdasarkan Jumlah Kriminalitas di Indonesia Tahun 2019," 2019.
- [8] F. M. Izzadin, "Optimasi Jumlah Cluster K-Means dengan Metode Elbow dan Silhouette pada Produktivitas Tanaman Pangan di Provinsi Jawa Tengah Tahun 2018," *dspace*, 2020.
- [9] X. Li, "Comparison and Analysis between Holt Exponential Smoothing and Brown Exponential Smoothing Used for Freight Turnover Forecast," in *Third International Conference on Intelligent System Design and Engineering Applications*, 2013.