



# PENENTUAN SANTRI PUTRI TERBAIK MENGGUNAKAN METODE K-MEANS CLUSTERING

Hihin Nurhayati<sup>1)</sup>, Sudin Saepudin<sup>2)</sup>

<sup>1, 2)</sup>Program Studi Sistem Informasi, Universitas Nusa Putra

Jl. Raya Cibolang No. 21 Cibolang Kaler, Cisaat, Sukabumi, Jawa Barat 43152  
e-mail: [hihin.nurhayati\\_si18@nusaputra.ac.id](mailto:hihin.nurhayati_si18@nusaputra.ac.id)<sup>1)</sup>, [Sudinsaepudin@nusaputra.ac.id](mailto:Sudinsaepudin@nusaputra.ac.id)<sup>2)</sup>

\* Korespondensi: e-mail: [Sudinsaepudin@nusaputra.ac.id](mailto:Sudinsaepudin@nusaputra.ac.id)

## ABSTRAK

*Pemilihan santri terbaik di suatu pondok pesantren sangat mempengaruhi banyak aspek penentu keberhasilan dari pondok pesantren tersebut. Salah satu hal terpenting dalam suatu pondok pesantren adalah pemilihan santri secara periodik sehingga untuk memacu semangat belajar para santri dalam menuntut ilmu. Namun pondok pesantren Al-jamiliyyah Sagaranten mendapat kendala memutuskan santri yang akan di prioritaskan. Hal ini disebabkan oleh belum tersedianya media yang dapat memproses penilaian santri dan memberikan rekomendasi dalam pemilihan santri terbaik. Selama ini pemilihan santri terbaik dilakukan dengan cara manual yaitu dengan cara menilai dari nilai kehadiran dan kedisiplinan, sehingga diperlukan penilaian dengan cara yang lebih akurat. Kriteria penelitian yang akan dilakukan adalah penentuan pemilihan santri putri terbaik di pondok pesantren Al-Jamiliyyah Sagaranten sukabumi dengan menggunakan metode K-means Clustering dengan 2 kriteria sebagai berikut: absensi dan nilai ulangan Pada penelitian ini peneliti mencoba melakukan pemilihan menggunakan metode K-means clustering. Hasil dari penelitian ini menunjukkan bahwa dengan adanya pemilihan santri terbaik menggunakan metode K-means lebih akurat dibandingkan dengan cara manual.*

**Kata Kunci:** *K-means, Clustering, pemilihan santri*

## ABSTRACT

*The selection of the best santri in an Islamic boarding school greatly influences many aspects that determine the success of the Islamic boarding school. One of the most important things in an Islamic boarding school is the periodic selection of students so as to spur the enthusiasm of students to study. However, the Al-Jamiliyyah Sagaranten Islamic Boarding School had problems deciding which students to prioritize. This is due to the unavailability of media that can process student assessments and provide recommendations in selecting the best students. So far, the selection of the best students is done manually, namely by assessing the value of attendance and discipline, so that an assessment in a more accurate way is needed. The research criteria to be carried out is determining the selection of the best female students at the Al-Jamiliyyah Sagaranten Islamic boarding school, Sukabumi by using the K-means Clustering method with 2 criteria as follows: attendance and test scores. The results of this study indicate that the selection of the best students using the K-means method is more accurate than the manual method.*

**Keywords:** *K-means, Clustering, Student Selection*

## I. PENDAHULUAN

Pentingnya pendidikan karakter ditemukan oleh para pakar, salah satunya yaitu Doni Koesoema A karakter sama dengan kepribadian dianggap sebagai ciri, atau karakteristik, atau gaya, atau sifat khas dari diri seseorang yang bersumber dari bentukan-bentukan yang diterima dari lingkungan.

Untuk pengembangan dan pembentukan karakter tersebut pondok pesantren sebagai salah satu sistem pendidikan yang tumbuh dan berkembang di dalam masyarakat dijadikan tumpuhan dan harapan untuk dijadikan suatu model pendidikan sebagai variasi lain dan bahkan dapat menjadi alternatif lain dalam pengembangan masyarakat guna menjawab tantangan masalah urbanisasi dan pembangunan dewasa ini. Pemilihan santri terbaik di suatu pondok pesantren sangat mempengaruhi banyak aspek penentu keberhasilan dari pondok pesantren tersebut. Salah satu hal terpenting dalam suatu pondok pesantren adalah pemilihan santri terbaik secara periodik sehingga untuk memacu semangat belajar para santri dalam menuntut ilmu.

Namun pondok pesantren Al-jamiliyyah Sagaranten mendapat kendala dalam memutuskan santri yang akan di prioritaskan. Hal ini disebabkan oleh belum adanya media yang dapat memproses penilaian



santri dan memberikan rekomendasi dalam pemilihan santri terbaik. Selama ini pemilihan dengan cara manual yaitu dengan cara menilai dari nilai kehadiran dan kedisiplinan, sehingga diperlukan penilaian dengan cara yang lebih akurat. Untuk itu peneliti menyarankan penilaian dengan menggunakan k-means clustering untuk mendapatkan hasil yang lebih akurat karena menggunakan bobot penilaiannya dalam pemilihan santri terbaik. Kriteria penelitian yang akan dilakukan adalah penentuan pemilihan santri terbaik di pondok pesantren Al-jamiliyyah Sagaranten dengan menggunakan metode k-means clustering dengan 2 kriteria sebagai berikut: Kehadiran,serta Nilai Ulangan.

## II. TINJAUAN PUSTAKA

### A. Clustering

Clustering adalah metode penganalisaan data,yang sering dimasukan sebagai salah satu metode *Data Mining*, yang tujuannya untuk mengelompokan data dengan karakteristik yang sama pada suatu wilayah yang sama dan data dengan karakteristik yang berbeda ke wilayah yang lain [1].

### B. Algoritma K-Means

Algoritma K-Means merupakan suatu metode penganalisaan data atau data mining yang melakukan proses pemodelan tanpa supervisi dan merupakan salah satu metode yang melakukan pengelompokan data dengan system partisi. Metode k-means berusaha mengelompokan data yang ada ke dalam beberapa kelompok,dimana data dalam satu kelompok mempunyai karakteristik yang sama satu sama lainnya dan mempunyai karakteristik berbeda dengan data yang ada di dalam kelompok yang lain. Dengan kata lain metode ini berusaha untuk meminimalkan variasi antara data yang ada di dalam suatu cluster dan memaksimalkan variasi dengan data yang ada di cluster lainnya [2].

Algoritma K-means dimulai dengan pemilihan secara acak K, K disini merupakan banyaknya cluster yang ingin dibentuk. Kemudian tetapkan nilai-nilai K secara acak, untuk sementara nilai tersebut menjadi pusat cluster atau biasa disebut dengan centroid secara acak. Hitung jarak setiap data yang ada terhadap masing-masing centriroid menggunakan rumus Euclidean hingga dapat ditemukan jarak yang paling dekat dari setiap data dengan centroid. Terus lakukan langkah tersebut hingga nilai centriroid dapat ditemukan dan tidak berubah[3].

Rumusnya sebagai berikut

$$D(i,j) = \sqrt{(X1J - X1I)^2 + (X2J - X2I)^2 + \dots + (XKJ - XKI)^2} \dots (1)$$

$D(i,j)$  = Jarak data ke  $i$  ke pusat *cluster*  $j$

$X_{kj}$  = Data ke  $i$  pada atribut dat ke  $k$

$X_{Kj}$  = Data ke  $j$  pada atribut dat ke  $k$

## III. METODE PENELITIAN

A. Langkah-langkah dalam menentukan *clustering* dengan menggunakan metode *K-Means*.Terdiri dari beberapa tahapan antara lain sebagai berikut :

1. Tentukan jumlah cluster.
2. Menentukan titik awal cluster secara random.
3. Hitung jarak ke masing-masing centriroid.
4. Kelompokan objek berdasarkan jarak minimum
5. Kembali ke Step 2, apabila masih ada data yang berpindah cluster atau perubahan nilai centriroid,ada yang di atas nilai threshold yang telah ditentukan,atau apabila nilai pada objective function yang digunakan diatas nilai threshold yang ditentukan.

$$d = \sqrt{(x1 - x2)^2 + (y1 - y2)^2} \dots\dots\dots(2)$$

Keterangan :  $d$  = jarak

$x_1$  =Koordinat latitude 1

$x_2$  = Koordinat latitude 2

$y_1$  =Koordinat longitude 1

$y_2$  = Koordinat longitude 2

Koordinat latitude dan longitude dapat dijadikan variable guna melakukan perhitungan jarak antara dua



buah titik lokasi apabila dibentangkan garis lurus diantara keduanya (Yulianto,Ramadiani,& Kridalaksana,2018).

#### IV. HASIL DAN PEMBAHASAN

##### A. Menentukan Jumlah Cluster

Data yang dianalisis yaitu data dari 30 nama santri putri ponpes Al-jamilyyah :

Tabel.1. Data Santri

NO	Nama	Parameter	
		ABSENSI	NILAI ULANGAN
1	SRI M	13	60
2	MAYANG	15	80
3	CUCU	14	70
4	AI RISMA	13	60
5	DIAS	14	70
6	NENENG	15	70
7	DEA VINA	14	70
8	LELA	14	80
9	LAILA	15	80
10	LUSIANI	15	70
11	LUSIANA	14	70
12	MIA	13	60
13	S. JENAB	13	60
14	S.AISAH	14	60
15	ROHIMAH	15	70
16	PUTRI	15	60
17	HELMI	13	70
18	HUMAIRA	14	70
19	JURIAH	15	80
20	NUR AZIZAH	14	80
21	NENG PUTRI	13	80
22	NAIVA	15	70
23	S.ALAWIYAH	13	70
24	HELENA	14	60
25	NASRIFAH	15	60
26	S.JAMILAH	15	70
27	S.MARYAM	14	80
28	ROHMAH	14	80
29	S.NURAIDAH	15	70
30	NURI HERNI	15	70

##### B. Menentukan titik awal cluster secara random

Berdasarkan tabel diatas maka kita akan bisa menentukan banyaknya cluster yang dibentuk tiga ( $k=3$ ).

Banyaknya cluster harus lebih kecil dari pada banyaknya data ( $k<n$ )

Tabel.2. Cluster

	Jumlah Cluster	3	
	Pusat Cluster	random	
C1	SRI M	13	60
C2	MAYANG	15	80
C3	CUCU	14	70



Inisialisasi centroid dataset pada tabel dataset diatas adalah :  $C1 = \{13,60\}$ ,  $C2 = \{15,80\}$ , dan  $C3 = \{14,70\}$

Untuk pengulangan berikutnya (pengulangan ke-1 sampai selesai), centroid baru berbeda dengan centroid sebelumnya, maka proses dilanjutkan ke langkah berikutnya. Namun jika centroid yang baru dihitung sama dengan centroid sebelumnya, maka proses clustering selesai.

### C. Hitung jarak ke masing-masing centroid dan mengelompokkan objek berdasarkan jarak minimum

Setelah didapat titik pusat awal cluster, kemudian dilakukan perhitungan jarak Euclidean, dan mengelompokkan berdasarkan jarak terkecil selanjutnya akan di dapat nilai centroid baru untuk acuan perhitungan berikutnya sampai nilai centroid sebelum dan sesudah bernilai sama. Perhitungan jarak Euclidean pada iterasi (1).

1. Berikut ini beberapa hasil perhitungan jarak Euclidean Distance, Pada titik pusat Cluster 1 (13,60)

$$[1] d(x_1, c_1) = \sqrt{(a_1 - c_1a)^2 + (b_1 - c_1b)^2} = \sqrt{(13 - 13)^2 + (60 - 60)^2} = 0$$

$$[2] d(x_2, c_1) = \sqrt{(a_2 - c_1a)^2 + (b_2 - c_1b)^2} = \sqrt{(15 - 13)^2 + (80 - 60)^2} = 20.09975124$$

$$[3] d(x_3, c_1) = \sqrt{(a_3 - c_1a)^2 + (b_3 - c_1b)^2} = \sqrt{(14 - 13)^2 + (70 - 60)^2} = 10.04$$

Lakukan perhitungan yang sama sampai data ke 30

2. Berikut ini beberapa hasil perhitungan jarak Euclidean Distance, Pada titik pusat Cluster 2 = (15, 80)

$$[1] d(x_1, c_1) = \sqrt{(a_1 - c_1a)^2 + (b_1 - c_1b)^2} = \sqrt{(13 - 15)^2 + (60 - 80)^2} = 20.09975124$$

$$[2] d(x_2, c_1) = \sqrt{(a_2 - c_1a)^2 + (b_2 - c_1b)^2} = \sqrt{(15 - 15)^2 + (80 - 80)^2} = 0$$

$$[3] d(x_3, c_1) = \sqrt{(a_3 - c_1a)^2 + (b_3 - c_1b)^2} = \sqrt{(14 - 15)^2 + (70 - 80)^2} = 10.04987562$$

Lakukan perhitungan yang sama sampai data ke 30

3. Berikut ini beberapa hasil perhitungan jarak Euclidean Distance, Pada titik pusat Cluster 3 = (14, 70)

$$[1] d(x_1, c_1) = \sqrt{(a_1 - c_1a)^2 + (b_1 - c_1b)^2} = \sqrt{(13 - 14)^2 + (60 - 70)^2} = 10.04987562$$

$$[2] d(x_2, c_1) = \sqrt{(a_2 - c_1a)^2 + (b_2 - c_1b)^2} = \sqrt{(15 - 14)^2 + (80 - 70)^2} = 10.04987562$$

$$[3] d(x_3, c_1) = \sqrt{(a_3 - c_1a)^2 + (b_3 - c_1b)^2} = \sqrt{(14 - 14)^2 + (70 - 70)^2} = 0$$

Lakukan perhitungan yang sama sampai data ke 30

Untuk seterusnya, hitung jarak pada setiap baris data, dan hasilnya seperti pada tabel dibawah.

Tabel.3. Data Parameter

NO	Nama	Parameter			C1	C2	C3
		ABSENS I	NILAI ULANGAN				
1	SRI M	13	60	0	20,09975124	10,04987562	
2	MAYANG	15	80	20,09975124	0	10,04987562	
3	CUCU	14	70	10,04987562	10,04987562	0	
4	AI RISMA	13	60	0	20,09975124	10,04987562	
5	DIAS	14	70	10,04987562	10,04987562	0	
6	NENENG	15	70	10,19803903	10	1	
7	DEA VINA	14	70	10,04987562	10,04987562	0	
8	LELA	14	80	20,02498439	1	10	
9	LAILA	15	80	20,09975124	0	10,04987562	
10	LUSIANI	15	70	10,19803903	10	1	
11	LUSIANA	14	70	10,04987562	10,04987562	0	
12	MIA	13	60	0	20,09975124	10,04987562	
13	S. JENAB	13	60	0	20,09975124	10,04987562	
14	S.AISAH	14	60	1	20,02498439	10	
15	ROHIMAH	15	70	10,19803903	10	1	
16	PUTRI	15	60	2	20	10,04987562	
17	HELMI	13	70	10	10,19803903	1	
18	HUMAIRA	14	70	10,04987562	10,04987562	0	
19	JURIAH	15	80	20,09975124	0	10,04987562	



20	NUR AZIZAH	14	80	20,02498439	1	10
21	NENG PUTRI	13	80	20	2	10,04987562
22	NAIVA	15	70	10,19803903	10	1
23	S.ALAWIYAH	13	70	10	10,19803903	1
24	HELENA	14	60	1	20,02498439	10
25	NASRIFAH	15	60	2	20	10,04987562
26	S.JAMILAH	15	70	10,19803903	10	1
27	S.MARYAM	14	80	20,02498439	1	10
28	ROHMAH	14	80	20,02498439	1	10
29	S.NURAIDAH	15	70	10,19803903	10	1
30	NURI HERNI	15	70	10,19803903	10	1

Pengelompokan data pada pengulangan ke 1

Tabel.4. Pusat Cluster baru

Pusat Cluster baru					
C1		C2		C3	
ABSENSI	NILAI ULANGAN	ABSENSI	NILAI ULANGAN	ABSENSI	NILAI ULANGAN
13	60				
		15	80		
				14	70
13	60				
				14	70
				15	70
				14	70
		14	80		
		15	80		
				15	70
				14	70
13	60				
13	60				
14	60				
				15	70
15	60				
				13	70
				14	70
		15	80		
		14	80		
		13	80		
				15	70
				13	70
14	60				
15	60				
				15	70
		14	80		
		14	80		
				15	70
				15	70

Kemudian kelompokan data sesuai dengan cluster-nya data yang mempunyai jarak paling pendek.

Contoh :

$d_{c1} < d_{c2} < d_{c3}$  maka  $x_1$  masuk ke dalam cluster  $c_1$ . Pada tabel diatas data  $N=1$  masuk kedalam cluster 1, karena  $d_{c1} < d_{c2} < d_{c3}$ , sedangkan  $N = 2,8,9,10$  masuk ke dalam cluster  $c_2$ .

Setelah mendapatkan label cluster untuk masing-masing data  $N= 1,2,3,4-30$  maka dicari nilai rata-



ratanya dengan menjumlahkan seluruh anggota masing-masing cluster dan dibagi jumlah anggotanya.

Tabel.5. Hasil Pusat Cluster baru

Pusat Cluster	ABSENSI	NILAI ULANGAN
C1	13,75	60
C2	14,25	80
C3	14,3571428 6	70

Dari tabel diatas hasil iterasi 1 dapat dilihat centroid baru yang dihasilkan dari iterasi 1 dengan titik pusat awal tidak sama, maka perlu dilakukannya perhitungan ulang, dengan perhitungan jarak euclidean pada iterasi 2.

1. Berikut ini beberapa hasil perhitungan jarak Euclidean Distance, Pada titik pusat Cluster 1 (13.75,60)

$$[1] d(x1,c1) = \sqrt{(a1 - c1a)^2 + (b1 - c1b)^2} = \sqrt{(13 - 13,75)^2 + (60 - 60)^2} = 0$$

$$[2] d(x2,c1) = \sqrt{(a2 - c1a)^2 + (b2 - c1b)^2} = \sqrt{(15 - 13,75)^2 + (80 - 60)^2} = 20.09975124$$

$$[3] d(x3,c1) = \sqrt{(a3 - c1a)^2 + (b2 - c1b)^2} = \sqrt{(14 - 13,75)^2 + (70 - 60)^2} = 10.04$$

Lakukan perhitungan yang sama sampai data ke 30

2. Berikut ini beberapa hasil perhitungan jarak Euclidean Distance, Pada titik pusat Cluster 2 = (14.25, 80)

$$[1] d(x1,c1) = \sqrt{(a1 - c1a)^2 + (b1 - c1b)^2} = \sqrt{(13 - 14,25)^2 + (60 - 80)^2} = 20.09975124$$

$$[2] d(x2,c1) = \sqrt{(a2 - c1a)^2 + (b2 - c1b)^2} = \sqrt{(15 - 14,25)^2 + (80 - 80)^2} = 0$$

$$[3] d(x3,c1) = \sqrt{(a3 - c1a)^2 + (b2 - c1b)^2} = \sqrt{(14 - 14,25)^2 + (70 - 80)^2} = 10.04987562$$

Lakukan perhitungan yang sama sampai data ke 30

3. Berikut ini beberapa hasil perhitungan jarak Euclidean Distance, Pada titik pusat Cluster 3 = (14.35, 70)

$$[1] d(x1,c1) = \sqrt{(a1 - c1a)^2 + (b1 - c1b)^2} = \sqrt{(13 - 14,35)^2 + (60 - 70)^2} = 10.04987562$$

$$[2] d(x2,c1) = \sqrt{(a2 - c1a)^2 + (b2 - c1b)^2} = \sqrt{(15 - 14,35)^2 + (80 - 70)^2} = 10.04987562$$

$$[3] d(x3,c1) = \sqrt{(a3 - c1a)^2 + (b2 - c1b)^2} = \sqrt{(14 - 14,35)^2 + (70 - 70)^2} = 0$$

Tabel.6. Perhitungan Pusat Cluster baru

Pusat Cluster baru					
C1		C2		C3	
ABSENSI	NILAI ULANGAN	ABSENSI	NILAI ULANGAN	ABSENSI	NILAI ULANGAN
13	60				
		15	80		
				14	70
13	60			14	70
				15	70
				14	70
		14	80		
		15	80		
				15	70
				14	70
13	60				
13	60				



14	60				
				15	70
15	60				
				13	70
				14	70
		15	80		
		14	80		
		13	80		
				15	70
				13	70
14	60				
15	60				
				15	70
		14	80		
		14	80		
				15	70
				15	70

Setelah mendapatkan label cluster untuk masing-masing data N=1,2,3,4,5 sampai dengan data 30 maka dicari nilai rata-ratanya dengan menjumlahkan masing-masing cluster dan dibagi jum. Anggotanya.

Tabel.7. Hasil Pusat Cluster Baru

Pusat Cluster	absensi	Nilai ulangan
C1	13,75	60
C2	14,25	80
C3	14,35714286	70

Nilai rata-rata centroid pada pengulangan ke 2

Karena Centroid tidak mengalami perubahan, maka proses clustering selesai  
**Hasil:**

Tabel.8. Hasil Pusat Cluster Baru

NO	Nama	Kelompok cluster	Nilai rata rata dari absensi dan ulangan
1	SRI MULYANI	C1	73
2	MAYANG	C2	95
3	CUCU	C3	84
4	AI RISMA	C1	73
5	DIAS	C3	84
6	NENENG	C3	85
7	DEA VINA	C3	84
8	LELA	C2	94
9	LAILA	C2	95
10	LUSIANI	C3	85
11	LUSIANA	C3	84
12	MIA	C1	73
13	S. JENAB	C1	73
14	S.AISAH	C1	74
15	ROHIMAH	C3	85
16	PUTRI	C1	75
17	HELMI	C3	83
18	HUMAIRA	C3	84



19	JURIAH	C2	95
20	NUR AZIZAH	C2	94
21	NENG PUTRI	C2	93
22	NAIVA	C3	85
23	S.ALAWIYAH	C3	83
24	HELENA	C1	74
25	NASRIFAH	C1	75
26	S.JAMILAH	C2	85
27	S.MARYAM	C2	94
28	ROHMAH	C3	94
29	S.NURAIDAH	C3	85
30	NURI HERNI	C3	85

4. Kriteria prestasi :

Tabel.9. Kriteria prestasi

Cukup	<	80
Baik	<	90
Baik sekali	>	90

## V. KESIMPULAN

Berdasarkan penelitian ini dapat disimpulkan sebagai berikut:

1. Telah berhasil dilakukan pengelompokan santri terbaik menggunakan algoritma k-means clustering menjadi 3 kelompok yaitu untuk kelompok 1 berjumlah 8 siswi, Kel. 2 berjumlah 8 siswi, Kelas 3 berjumlah 14 siswi.
2. Penentuan titik pusat awal (centroid) sangat berpengaruh terhadap iterasi yang akan dihitung
3. Berdasarkan penelitian ini dapat dijadikan acuan sebagai pemilihan santri terbaik di ponpes Al-jamiliyyah sagaranten.

## DAFTAR PUSTAKA

- [1] Agusta, Y. (2017). *Clustering*. Wordpress.Com. <https://yudiagusta.wordpress.com/clustering/>
- [2] Agusta, Y. (2017). *Clustering*. Wordpress.Com. <https://yudiagusta.wordpress.com/clustering/>
- [3] [Witten, Ian H. dan Frank, Eibe. 2005. Data Mining Practical Machine Learning Tools and Techniques, Second Edition. Morgan Kaufmann, San Francisco](#)
- [4] Dede Sobari, Sumanto Sumanto, Lia Mazia, L. S. M. (n.d.). *PEMILIHAN SANTRI TERBAIK MENGGUNAKAN METODE SIMPLE ADDITIVE WEIGHTING*. 2017. Retrieved June 26, 2021, from <https://seminar.bsi.ac.id/knist/index.php/UnivBSI/article/view/200>