

ANALISIS *CLUSTER* MENGGUNAKAN ALGORITMA *K-MEANS* UNTUK *CLUSTERING* HASIL SELEKSI PENERIMAAN SANTRI BARU DI PONDOK PESANTREN SA'ADATUL IKHWAN NAHDLATUL WATHAN RENSING, KECAMATAN SAKRA BARAT, KABUPATEN LOMBOK TIMUR

^{1*}Lalu Ahmad Sya'roniy Zain, ²Lalu Muhammad Ridwan, ³Dara Puspita Anggraeni

^{1,2,3}Matematika, FMIPA, Universitas Nahdlatul Wathan Mataram, Mataram

Email: syaroniyahmad@gmail.com

Abstract

Indonesia is a country with the largest Muslim population in the world. Approximately 87% of Indonesia's population is Muslim and there are also many formal and non-formal educational institutions that prohibit Islamic Religious Education, namely Islamic boarding schools. Islamic boarding schools are religious institutions that provide education and teaching as well as develop and spread Islam. One of them is in the West Nusa Tenggara region, namely Rensing Village, West Sakra District, East Lombok Regency, namely the Sa'adatul Ikhwan Nahdlatul Wathan Rensing Islamic Boarding School. This Islamic boarding school organizes Islamic teaching and learning activities starting from the acceptance of new students, the selection process, class guidance, to the implementation of teaching and learning activities. In this process, a selection is made to read the Qur'an with several assessments to determine the learning class that suits the condition of the students, namely the value of fluency in reciting hijaiyah letters, the value of accuracy in practicing tajwid, the value of beauty in rhythm, and the value of adab. This class training can be carried out by applying one of the statistical methods, namely cluster analysis, which is a method of grouping data based on similarity in characteristics with the aim of summarizing, simplifying, and explaining data. In this study, the algorithm used is K-Means with the data used in the form of the results of the selection of 150 new students and the variables used are assessment indicators in the selection process. From the clustering carried out, 3 clusters were obtained with cluster 1 of 74 students, cluster 2 of 50 students, and cluster 3 of 26 students.

Keywords: Cluster Analysis, K-Means.

Abstrak

Indonesia merupakan negara dengan mayoritas penduduk muslim terbesar di dunia. Sekitar 87% penduduk Indonesia beragama Islam dan banyak juga lembaga pendidikan formal dan non-formal yang mengajarkan Pendidikan Agama Islam yakni pondok pesantren. Pondok pesantren adalah lembaga keagamaan yang memberikan pendidikan dan pengajaran serta mengembangkan dan menyebarkan agama Islam. Salah satunya di wilayah Nusa Tenggara Barat, yakni Desa Rensing, Kecamatan Sakra Barat, Kabupaten Lombok Timur, yaitu Pondok Pesantren Sa'adatul Ikhwan Nahdlatul Wathan Rensing. Pondok pesantren ini menyelenggarakan kegiatan belajar mengajar Islami yang dimulai dari penerimaan santri baru, proses seleksi, penentuan kelas belajar, hingga terlaksananya kegiatan belajar mengajar. Dalam proses tersebut, dilakukan seleksi membaca Al Qur'an dengan beberapa penilaian untuk menentukan kelas belajar yang sesuai dengan keadaan santri yaitu nilai kefasihan melafalkan huruf hijaiyah, nilai ketepatan mengamalkan tajwid, nilai keindahan dalam irama, dan nilai adab. Penentuan kelas tersebut dapat dilakukan dengan menerapkan salah satu metode statistika yaitu analisis *cluster* yang merupakan metode pengelompokan data berdasarkan kesamaan ciri yang dimilikinya dengan tujuan untuk meringkas, menyederhanakan, dan menjelaskan data. Dalam penelitian ini, algoritma yang digunakan adalah *K-Means* dengan data yang digunakan berupa nilai hasil seleksi 150 santri baru dan variabel yang digunakan adalah indikator penilaian dalam proses seleksi tersebut. Dari clustering yang dilakukan, diperoleh 3 cluster dengan *cluster* 1 sebanyak 74 santri, *cluster* 2 sebanyak 50 santri, dan *cluster* 3 sebanyak 26 santri

Kata kunci: Analisis Cluster, K-Means.

PENDAHULUAN

Indonesia merupakan negara dengan mayoritas penduduk muslim terbesar di dunia. Sekitar 87% penduduk Indonesia beragama Islam. Sehingga banyak lembaga pendidikan formal dan non-formal yang mengajarkan Pendidikan Agama Islam seperti pondok pesantren di Indonesia (Kemenag, 2012). Pondok pesantren adalah lembaga keagamaan yang memberikan pendidikan dan pengajaran serta mengembangkan dan menyebarkan agama Islam (Mujib, 2010). Salah satunya adalah Pondok Pesantren Sa'adatul Ikhwan Nahdlatul Wathan Rensing, Kecamatan Sakra Barat, Kabupaten Lombok Timur yang menyelenggarakan kegiatan seperti Tahfizh Al Qur'an, mempelajari kitab nahwu sharf, dan kitab-kitab hadits. Kegiatan ini diawali dari proses seleksi terhadap santri baru dengan tujuan agar bimbingan terhadap santri sesuai dengan kondisi dan kemampuan santri. Seleksi yang dilakukan berupa tes membaca Al Qur'an dengan penilaian meliputi kefasihan melafalkan huruf hijaiyah, ketepatan mengamalkan tajwid, keindahan dalam irama, dan adab.

Proses seleksi tersebut dapat dilakukan dengan salah satu metode pengelompokan dalam matematika, khususnya statistika yaitu analisis *cluster*. Analisis cluster merupakan metode pengelompokan data berdasarkan kesamaan ciri yang dimilikinya. Tujuan analisis *cluster* adalah untuk meringkas, menyederhanakan, dan menjelaskan data serta disajikan dalam bentuk data yang terdiri dari objek (baris) dan variabel (kolom) (Mattjik, 2011). Keunggulan dari analisis *cluster* adalah data yang digunakan dapat berupa data ordinal, interval, dan rasio dalam jumlah data dan variabel yang banyak. Data yang digunakan juga harus bersifat representatif, artinya dapat menjelaskan gambaran umum dari data agar hasil yang diperoleh dapat digeneralisasi (Abdillah, 2014).

Konsep kesamaan karakteristik pada analisis *cluster* didasarkan pada ukuran jarak antardata untuk mengetahui tingkat kesamaan atau perbedaan karakteristik antarobjek tersebut. Ukuran yang biasa digunakan adalah mengukur jarak antarseluruh objek (Supranto, 2004). Jarak yang besar menyatakan karakteristik antarobjek yang sedikit. Sedangkan jarak yang kecil menyatakan karakteristik antarobjek yang besar. Fungsi jarak antara objek a dan b dinotasikan dengan $d(a, b)$. Semakin kecil nilai d , maka kesamaan karakteristik antara kedua objek tersebut semakin besar. Sedangkan semakin besar nilai d , maka kesamaan karakteristik antara kedua objek tersebut semakin kecil (Sitepu, 2011).

Ada beberapa algoritma yang terdapat dalam analisis *cluster*. Salah satunya adalah *K-Means*. Penerapan algoritma K-Means dapat dilakukan pada data santri baru yang diseleksi di Pondok Pesantren Sa'adatul Ikhwan Nahdlatul Wathan Rensing, Kecamatan Sakra Barat, Kabupaten Lombok Timur. Variabel yang digunakan adalah indikator penilaian dalam proses seleksi, yaitu kefasihan melafalkan huruf hijaiyah, ketepatan mengamalkan tajwid, keindahan dalam irama, dan adab ketika membaca Al Qur'an.

METODE

Jenis penelitian dalam penelitian ini adalah penelitian deskriptif kuantitatif yaitu penelitian yang menjelaskan dan menganalisis data menggunakan metode statistika yaitu analisis *cluster* dengan algoritma *K-Means*. Data yang digunakan adalah data skunder, yaitu data yang diperoleh dari Pondok Pesantren Sa'adatul Ikhwan Nahdlatul Wathan Rensing, Kecamatan Sakra Barat, Kabupaten Lombok Timur berupa data 150 Hasil Seleksi Santri Baru Tahun 2022. Variabel yang digunakan adalah 4 indikator penilaian yang digunakan dalam proses seleksi membaca Al Qur'an yaitu nilai kefasihan dalam melafalkan huruf hijaiyah, nilai tajwid, nilai keindahan dalam irama, dan nilai adab. Setiap variabel diberikan nilai sesuai dengan hasil seleksi yang telah dilakukan. Nilai tersebut berupa angka 1, 2, 3, 4, dan 5. Di mana pemberian nilai 1 untuk hasil seleksi yang sangat kurang, nilai 2 untuk hasil seleksi yang kurang, nilai 3 untuk hasil seleksi yang cukup, nilai 4 untuk hasil seleksi yang baik, dan nilai 5 untuk hasil seleksi yang sangat baik.

Langkah penelitian diawali dengan pengumpulan data untuk dikaji berdasarkan teori yang ada, khususnya tentang analisis *cluster* menggunakan algoritma *K-Means*. Selanjutnya dilakukan analisis *cluster* menggunakan algoritma *K-Means* dengan langkah sebagai berikut (Gundono, 2011).

1. Menentukan K banyak *cluster* atau kelompok yang akan dibentuk.
2. Memilih sebanyak K data dari seluruh data yang ada untuk dijadikan sebagai pusat *cluster* (*centroid*) secara bebas.
3. Menghitung jarak antara objek dengan masing-masing *centroid* dengan menggunakan rumus jarak *Euclidean*.

$$d_{x_i, x_j} = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2}$$

di mana :

d_{x_i, x_j} : jarak objek ke- i dengan objek ke- j

p : jumlah variabel

x_{ik} : data objek ke- i pada variabel ke- k

x_{jk} : data objek ke- j pada variabel ke- k

4. Mengelompokkan objek berdasarkan jarak terdekat (minimum) dengan *centroid*.
5. Menentukan *centroid* baru dengan rumus rata-rata, yaitu:

$$C_{m(q)} = \frac{1}{n_m} \sum_{i=1}^{n_m} x_{i(q)}$$

di mana :

$C_{m(q)}$: rata-rata *centroid* baru

m : jumlah *centroid*

q : jumlah variabel

n_m : jumlah data pada *centroid* ke- m

$x_{i(q)}$: data ke- i variabel ke- q

6. Lakukan langkah 3 dan 4 hingga tidak ada lagi objek yang berpindah *cluster*.

Proses *clustering* dilakukan dengan beberapa iterasi (pengulangan) hingga diperoleh *cluster* untuk selanjutnya diinterpretasi dan ditarik kesimpulan.

HASIL DAN PEMBAHASAN

Tabel Data Jumlah Santri tiap Nilai pada Variabel

Nilai	Variabel			
	A	B	C	D
1	4	56	57	0
2	53	48	46	35
3	56	21	22	95
4	17	10	9	20
5	20	15	16	0
Total	150	150	150	150

Berdasarkan merupakan statistika deskriptig dari tabel tersebut di atas.

1. Pada variabel A yang merupakan nilai fasohah, terdapat 4 santri dinilai sangat kurang fasih, 53 santri dinilai kurang fasih, 56 santri dinilai cukup fasih, 17 santri dinilai fasih, dan 20 santri dinilai sangat fasih dalam melafalkan huruf hijaiyah.
2. Pada variabel B yang merupakan nilai tajwid, terdapat 56 santri dinilai sangat kurang tepat, 48 santri dinilai kurang tepat, 21 santri dinilai cukup tepat, 10 santri dinilai tepat, dan 15 santri dinilai sangat tepat dalam mengamalkan tajwid.
3. Pada variabel C yang merupakan nilai keindahan dalam irama, terdapat 57 santri dinilai sangat kurang indah, 46 santri dinilai kurang indah, 22 santri dinilai cukup indah, 9 santri dinilai indah, dan 16 santri dinilai sangat indah dalam irama.
4. Pada variabel D yang merupakan nilai adab, terdapat 35 santri dinilai kurang beradab, 95 santri dinilai cukup beradab, 20 santri dinilai beradab, serta tidak terdapat santri yang dinilai sangat kurang beradab dan sangat beradab.

Berdasarkan statistika deskriptif tersebut di atas, selanjutnya dilakukan analisis *cluster* dengan algoritma *K-Means* dengan dengan langkah sebagai berikut.

- Iterasi 1

1. Menentukan K banyak *cluster* (kelompok) yang akan dibentuk

Pada tahap ini, jumlah *cluster* (K) yang akan dibentuk disesuaikan dengan jumlah kelas di Pondok Pesantren Sa'adatul Ikhwan Nahdlatul Wathan Rensing, Kecamatan Sakra Barat, Kabupaten Lombok Timur yaitu sebanyak 3 kelas. Oleh karena itu, banyak *cluster* yang akan dibentuk adalah 3.

2. Memilih K data dari seluruh data secara bebas sebagai pusat *cluster* (*centroid*)

3.

Pada tahap ini, jumlah *centroid* disesuaikan dengan banyak K yaitu 3. Karena dipilih secara bebas, maka *centroid* yang dipilih tidak mempengaruhi hasil. Sehingga, untuk *centroid* 1 (C_1) dipilih data ke-50, *centroid* 2 (C_2) dipilih data ke-75, dan *centroid* 3 (C_3) dipilih data ke-100. Berikut merupakan tabel data variabel untuk tiap *centroid*.

Tabel Data *Centroid* Iterasi 1

Santri	Centroid	Variabel			
		A	B	C	D
S050	1	2	1	1	2
S075	2	2	1	2	3
S100	3	3	2	1	3

4. Menghitung jarak semua data dengan masing-masing *centroid* menggunakan rumus jarak *Euclidean*

a. Untuk C_1 , maka dihitung jarak semua data dengan *centroid* 1

Berikut merupakan perhitungan jarak data ke-1 dengan *centroid* 1.

$$\begin{aligned}
 d_{x_1, x_{50}} &= \sqrt{(x_{1,1} - x_{50,1})^2 + (x_{1,2} - x_{50,2})^2 + (x_{1,3} - x_{50,3})^2 + (x_{1,4} - x_{50,4})^2} \\
 &= \sqrt{(4 - 2)^2 + (3 - 1)^2 + (2 - 1)^2 + (3 - 2)^2} \\
 &= 3.16227766
 \end{aligned}$$

Sehingga jarak data ke-1 dengan *centroid* 1 adalah 3.16227766. Selanjutnya dihitung jarak data ke-2 dengan *centroid* 1 sebagai berikut.

$$\begin{aligned}
 d_{x_2, x_{50}} &= \sqrt{(x_{2,1} - x_{50,1})^2 + (x_{2,2} - x_{50,2})^2 + (x_{2,3} - x_{50,3})^2 + (x_{2,4} - x_{50,4})^2} \\
 &= \sqrt{(2 - 2)^2 + (1 - 1)^2 + (1 - 1)^2 + (3 - 2)^2} \\
 &= 1
 \end{aligned}$$

Sehingga jarak data ke-2 dengan *centroid* 1 adalah 1. Selanjutnya dihitung jarak data ke-3 dengan *centroid* 1, data ke-4 dengan *centroid* 1, hingga data ke-150 dengan *centroid* 1 dan dilakukan dengan cara yang sama.

b. Untuk C_2 , maka dihitung jarak semua data dengan *centroid* 2

Berikut merupakan perhitungan jarak data ke-1 dengan *centroid* 2.

$$\begin{aligned}
 d_{x_1, x_{75}} &= \sqrt{(x_{1,1} - x_{75,1})^2 + (x_{1,2} - x_{75,2})^2 + (x_{1,3} - x_{75,3})^2 + (x_{1,4} - x_{75,4})^2} \\
 &= \sqrt{(4 - 2)^2 + (3 - 1)^2 + (2 - 2)^2 + (3 - 3)^2} \\
 &= 2.828427125
 \end{aligned}$$

Sehingga jarak data ke-1 dengan *centroid* 2 adalah 2.828427125. Selanjutnya dihitung jarak data ke-2 dengan *centroid* 2 sebagai berikut.

$$\begin{aligned} d_{x_2, x_{75}} &= \sqrt{(x_{2,1} - x_{75,1})^2 + (x_{2,2} - x_{75,2})^2 + (x_{2,3} - x_{75,3})^2 + (x_{2,4} - x_{75,4})^2} \\ &= \sqrt{(2 - 2)^2 + (1 - 1)^2 + (1 - 2)^2 + (3 - 3)^2} \\ &= 1 \end{aligned}$$

Sehingga jarak data ke-2 dengan *centroid* 2 adalah 1. Selanjutnya dihitung jarak data ke-3 dengan *centroid* 2, data ke-4 dengan *centroid* 2, hingga data ke-150 dengan *centroid* 2 dan dilakukan dengan cara yang sama.

c. Untuk C_3 , maka dihitung jarak semua data dengan *centroid* 3

Berikut merupakan perhitungan jarak data ke-1 dengan *centroid* 3.

$$\begin{aligned} d_{x_1, x_{100}} &= \sqrt{(x_{1,1} - x_{100,1})^2 + (x_{1,2} - x_{100,2})^2 + (x_{1,3} - x_{100,3})^2 + (x_{1,4} - x_{100,4})^2} \\ &= \sqrt{(4 - 3)^2 + (3 - 2)^2 + (2 - 1)^2 + (3 - 3)^2} \\ &= 1.732050808 \end{aligned}$$

Sehingga jarak data ke-1 dengan *centroid* 3 adalah 1.732050808. Selanjutnya dihitung jarak data ke-2 dengan *centroid* 3 sebagai berikut.

$$\begin{aligned} d_{x_2, x_{100}} &= \sqrt{(x_{2,1} - x_{100,1})^2 + (x_{2,2} - x_{100,2})^2 + (x_{2,3} - x_{100,3})^2 + (x_{2,4} - x_{100,4})^2} \\ &= \sqrt{(2 - 3)^2 + (1 - 2)^2 + (1 - 1)^2 + (3 - 3)^2} \\ &= 1.414213562 \end{aligned}$$

Sehingga jarak data ke-2 dengan *centroid* 3 adalah 1.414213562. Selanjutnya dihitung jarak data ke-3 dengan *centroid* 3, data ke-4 dengan *centroid* 3, hingga data ke-150 dengan *centroid* 3 dan dilakukan dengan cara yang sama.

5. Mengelompokkan data berdasarkan jarak terdekat (minimum) dengan *centroid*

Pada tahap ini, dibandingkan jarak antara masing-masing data dengan masing-masing *centroid*. Untuk data ke-1 dengan *centroid* 1, diperoleh jarak sebesar 3.16227766. Sedangkan data ke-1 dengan *centroid* 2 diperoleh jarak sebesar 2.828427125. Kemudian data ke-1 dengan *centroid* 3 diperoleh jarak sebesar 1.732050808. Karena jarak data ke-1 dengan *centroid* 3 merupakan jarak terdekat (minimum) di antara *centroid* yang lain, maka data ke-1 dikelompokkan ke dalam *cluster*

3.

Selanjutnya dilakukan perbandingan untuk data ke-2 terhadap masing-masing *centroid*. Untuk data ke-2 dengan *centroid* 1, diperoleh jarak sebesar 1. Sedangkan data ke-2 dengan *centroid* 2 diperoleh jarak sebesar 1. Kemudian data ke-2 dengan *centroid* 3 diperoleh jarak sebesar 1.414213562. Karena jarak data ke-2 dengan *centroid* 1 merupakan jarak terdekat (minimum) di antara *centroid* yang lain, maka data ke-2 dikelompokkan ke dalam *cluster* 1.

Untuk data ke-3, data ke-4, hingga data ke-150 dilakukan dengan cara yang sama dan hasil yang diperoleh dilanjutkan ke iterasi 2.

- Iterasi 2

1. Menentukan *centroid* baru

Pada tahap ini, *centroid* baru dibentuk dari rata-rata masing-masing variabel anggota masing-masing *cluster* iterasi sebelumnya, yaitu iterasi 1 dengan hasil sebagai berikut.

Tabel Data *Centroid* Baru Iterasi 2

<i>Centroid</i>	Variabel			
	A	B	C	D
1	1,911111111	1,022222222	1,044444444	2,511111111
2	2,64	1,56	2,52	2,96
3	3,675	3,0625	2,7625	3,1

2. Menghitung jarak semua data dengan masing-masing *centroid* menggunakan rumus jarak *Euclidean*

a. Untuk C_1 , maka dihitung jarak semua data dengan *centroid* 1 pada *centroid* baru iterasi 2

Berikut merupakan perhitungan jarak data ke-1 dengan *centroid* 1

$$\begin{aligned}
 d_{x_1, C_1} &= \sqrt{(x_{1,1} - C_{1,1})^2 + (x_{1,2} - C_{1,2})^2 + (x_{1,3} - C_{1,3})^2 + (x_{1,4} - C_{1,4})^2} \\
 &= \sqrt{(4 - 1.911111111)^2 + (3 - 1.022222222)^2 + (2 - 1.044444444)^2 + (3 - 2.511111111)^2} \\
 &= 3.070368137
 \end{aligned}$$

Sehingga jarak data ke-1 dengan *centroid* 1 adalah 3.070368137. Selanjutnya dihitung jarak data ke-2 dengan *centroid* 1 dengan perhitungan sebagai berikut.

$$\begin{aligned}
 d_{x_2, C_1} &= \sqrt{(x_{2,1} - C_{1,1})^2 + (x_{2,2} - C_{1,2})^2 + (x_{2,3} - C_{1,3})^2 + (x_{2,4} - C_{1,4})^2} \\
 &= \sqrt{(2 - 1.911111111)^2 + (1 - 1.022222222)^2 + (1 - 1.044444444)^2 + (3 - 2.511111111)^2} \\
 &= 0.499382335
 \end{aligned}$$

Sehingga jarak data ke-2 dengan *centroid* 1 adalah 0.499382335. Selanjutnya dihitung jarak data ke-3, data ke-4, hingga data ke-150 masing-masing dengan *centroid* 1 dan dilakukan dengan cara yang sama.

- b. Untuk C_2 , maka dihitung jarak semua data dengan *centroid* 2 pada *centroid* baru iterasi 2
- c.

Berikut merupakan perhitungan jarak data ke-1 dengan *centroid* 2

$$\begin{aligned} d_{x_1, C_2} &= \sqrt{(x_{1,1} - C_{2,1})^2 + (x_{1,2} - C_{2,2})^2 + (x_{1,3} - C_{2,3})^2 + (x_{1,4} - C_{2,4})^2} \\ &= \sqrt{(4 - 2.64)^2 + (3 - 1.56)^2 + (2 - 2.52)^2 + (3 - 2.96)^2} \\ &= 2.048218738 \end{aligned}$$

Sehingga jarak data ke-1 dengan *centroid* 2 adalah 2.048218738. Selanjutnya dihitung jarak data ke-2 dengan *centroid* 2 dengan perhitungan sebagai berikut.

$$\begin{aligned} d_{x_2, C_2} &= \sqrt{(x_{2,1} - C_{2,1})^2 + (x_{2,2} - C_{2,2})^2 + (x_{2,3} - C_{2,3})^2 + (x_{2,4} - C_{2,4})^2} \\ &= \sqrt{(2 - 2.64)^2 + (1 - 1.56)^2 + (1 - 2.52)^2 + (3 - 2.96)^2} \\ &= 1.742182539 \end{aligned}$$

Sehingga jarak data ke-2 dengan *centroid* 2 adalah 1.742182539. Selanjutnya dihitung jarak data ke-3, data ke-4, hingga data ke-150 masing-masing dengan *centroid* 3 dan dilakukan dengan cara yang sama.

- d. Untuk C_3 , maka dihitung jarak semua data dengan *centroid* 3 pada *centroid* baru iterasi 2

Berikut merupakan perhitungan jarak data ke-1 dengan *centroid* 3

$$\begin{aligned} d_{x_1, C_3} &= \sqrt{(x_{1,1} - C_{3,1})^2 + (x_{1,2} - C_{3,2})^2 + (x_{1,3} - C_{3,3})^2 + (x_{1,4} - C_{3,4})^2} \\ &= \sqrt{(4 - 3.675)^2 + (3 - 3.0625)^2 + (2 - 2.7625)^2 + (3 - 3.1)^2} \\ &= 0.837220102 \end{aligned}$$

Sehingga jarak data ke-1 dengan *centroid* 3 adalah 0.837220102. Selanjutnya dihitung jarak data ke-2 dengan *centroid* 3 dengan perhitungan sebagai berikut.

$$\begin{aligned} d_{x_2, C_3} &= \sqrt{(x_{2,1} - C_{3,1})^2 + (x_{2,2} - C_{3,2})^2 + (x_{2,3} - C_{3,3})^2 + (x_{2,4} - C_{3,4})^2} \\ &= \sqrt{(2 - 3.675)^2 + (1 - 3.0625)^2 + (1 - 2.7625)^2 + (3 - 3.1)^2} \\ &= 3.18997453 \end{aligned}$$

Sehingga jarak data ke-2 dengan *centroid* 3 adalah 3.18997453. Selanjutnya dihitung jarak data ke-3, data ke-4, hingga data ke-150 masing-masing dengan *centroid* 3 dan dilakukan dengan cara yang sama.

3. Mengelompokkan data berdasarkan jarak terdekat (minimum) dengan *centroid*

4.

Pada tahap ini, dibandingkan jarak antara masing-masing data dengan masing-masing *centroid* iterasi 2. Untuk data ke-1 dengan *centroid* 1, diperoleh jarak sebesar 3.070368137. Sedangkan data ke-1 dengan *centroid* 2 diperoleh jarak sebesar 2.048218738. Kemudian data ke-1 dengan *centroid* 3 diperoleh jarak sebesar 0.837220102. Karena jarak data ke-1 dengan *centroid* 3 merupakan jarak terdekat (minimum) di antara *centroid* yang lain, maka data ke-1 dikelompokkan ke dalam *cluster* 3.

Selanjutnya dilakukan perbandingan untuk data ke-2 terhadap masing-masing *centroid* iterasi 2. Untuk data ke-2 dengan *centroid* 1, diperoleh jarak sebesar 0.499382335. Sedangkan data ke-2 dengan *centroid* 2 diperoleh jarak sebesar 1.742182539. Kemudian data ke-2 dengan *centroid* 3 diperoleh jarak sebesar 3.18997453. Karena jarak data ke-2 dengan *centroid* 1 merupakan jarak terdekat (minimum) di antara *centroid* yang lain, maka data ke-2 dikelompokkan ke dalam *cluster* 1.

Untuk data ke-3, data ke-4, hingga data ke-150 dilakukan dengan cara yang sama dan berdasarkan hasil dilanjutkan ke iterasi 3 karena hasil *clustering* pada iterasi 2 tidak sama dengan iterasi 1. Proses berlanjut hingga iterasi 4, iterasi 5, dan iterasi 6. Pada iterasi 7, proses *clustering* berhenti dan selesai karena hasil *clustering* pada iterasi 7 sama dengan iterasi 6. Maka selanjutnya dapat diinterpretasi hasil yang diperoleh.

Setelah dilakukan analisis *cluster* menggunakan algoritma *K-Means*, maka diperoleh anggota *cluster* dari 3 *cluster* yang telah ditentukan sebelumnya. Berikut merupakan interpretasi hasil yang diperoleh dengan anggota *cluster* terlampir.

1. Pada *cluster* 1, diperoleh 74 santri dengan karakteristik yang sama. Kesamaannya diamati berdasarkan kemiripan interval nilai hasil seleksi yang diperoleh santri. Sehingga pada *cluster* 1 ini dapat menjadi kelompok untuk santri yang memperoleh nilai hasil seleksi kurang.

Tabel Hasil Cluster 1

Kode Santri				
S002	S037	S068	S099	S125
S006	S038	S069	S100	S128
S007	S039	S072	S106	S129
S011	S040	S075	S108	S130
S015	S042	S076	S112	S132
S018	S043	S078	S113	S135
S020	S044	S079	S114	S137
S022	S047	S080	S115	S138
S025	S048	S083	S116	S140
S027	S050	S087	S117	S141

S029	S055	S088	S119	S144
S031	S061	S093	S120	S147
S032	S062	S095	S121	S148
S035	S065	S096	S122	S149
S036	S067	S097	S123	

2. Pada *cluster* 2, diperoleh 50 santri dengan karakteristik yang sama. Kesamaannya diamati berdasarkan kemiripan interval nilai hasil seleksi yang diperoleh santri. Sehingga pada *cluster* 2 ini dapat menjadi kelompok untuk santri yang memperoleh nilai hasil seleksi cukup.

Tabel Hasil *Cluster* 2

Kode Santri				
S001	S019	S056	S084	S118
S003	S021	S057	S091	S124
S004	S023	S058	S092	S127
S005	S024	S059	S094	S131
S008	S028	S060	S101	S133
S009	S033	S064	S103	S134
S010	S034	S071	S105	S139
S012	S046	S073	S107	S143
S013	S051	S074	S110	S145
S014	S053	S077	S111	S150

3. Pada *cluster* 3, diperoleh 26 santri dengan karakteristik yang sama. Kesamaannya diamati berdasarkan kemiripan interval nilai hasil seleksi yang diperoleh santri. Sehingga pada *cluster* 3 ini dapat menjadi kelompok untuk santri yang memperoleh nilai hasil seleksi baik.

Tabel Hasil *Cluster* 3

Kode Santri				
S016	S049	S070	S089	S109
S017	S052	S081	S090	S126
S026	S054	S082	S098	S136
S030	S063	S085	S102	S142
S041	S066	S086	S104	S146
S045				

SIMPULAN

Berdasarkan hasil dan pembahasan yang telah dilakukan, dapat ditarik kesimpulan antara lain sebagai berikut.

1. Analisis cluster menggunakan algoritma K-Means pada data hasil seleksi penerimaan santri baru di Pondok Pesantren Sa'adatul Ikhwan Nahdlatul Wathan Rensing, Kecamatan Sakra Barat, Kabupaten Lombok Timur dapat dilakukan sebanyak 7 iterasi.
2. Anggota cluster yang terbentuk pada cluster 1 berjumlah 74 santri, pada cluster 2 berjumlah 50 santri, dan pada cluster 3 berjumlah 26 santri.
3. Berdasarkan jumlah anggota cluster yang diperoleh, dapat diketahui bahwa masih banyak santri yang memerlukan bimbingan khusus di Pondok Pesantren Sa'adatul Ikhwan Nahdlatul Wathan Rensing, Kecamatan Sakra Barat, Kabupaten Lombok Timur. Baik dari segi fasahah, tajwid, irama, dan adab.

SARAN

Untuk penelitian selanjutnya, diharapkan dapat melakukan pengelompokan dengan analisis cluster menggunakan algoritma yang lain sehingga dapat menambah referensi dan menjadi literatur untuk penelitian selanjutnya.

REFERENSI

- Abdillah, A, 2014, Analisis Cluster pada Grup Sunspot Klasifikasi Mc. Intosh yang Berpotensi Membangkitkan Flare, Malang: UIN Maulana Malik Ibrahim Malang.
- Gundono, 2011, Analisis Data Multivariat, Yogyakarta: BPFE.
- Kementerian Agama Republik Indonesia, 2012, Jakarta.
- Mattjik, A. A. dan Sumertajaya, I. M., 2011, Sidik Peubah Ganda dengan Menggunakan SAS, Bogor: IPB Press.
- Mujib, A, 2010, Ilmu Pendidikan Islam, Jakarta: PT. Kencana.
- Sitepu, R, 2011, Analisis Cluster Terhadap Tingkat Pencemaran Udara pada Sektor Industri di Sumara Selatan, Palembang: Jurnal Penelitian Sains.
- Supranto, 2004, Analisis Multivariat: Arti dan Interpretasi, Jakarta: PT. Rineka Cipta.