



Penerapan K-Means Untuk Pengelompokkan Beasiswa Santri di Pondok Pesantren Miftahul Huda Bogor

Derman Janner Lubis^{1*}, M. Badru Tamam²

¹ STIKOM Binaniaga/Sistem Informasi
Email: derman@stikombinaniaga.ac.id

²STIKOM Binaniaga/Sistem Informasi
Email: mbadrutamam999@gmail.com

ABSTRACT

At this time the reflection of the quality of education can be seen from the high level of success and the low level of failure of santri in the world of education. One of the criteria for selecting prospective scholarship recipients is their high academic value. However, what has been happening so far is that high academic scores have come from outside of academic values, namely personality values. This is done so that santri can get scholarships according to their abilities. Therefore, the researcher tries to provide an alternative in the process of grouping scholarships by sorting out the variables of academic value data and personality value data. One way to select santri to get scholarships is to do segmentation by grouping data based on certain criteria. The K-Means Clustering Algorithm is one of the grouping methods that can group objects based on the similarity of their properties. The purpose of this study was to determine the application of the K-Means Clustering Algorithm method in selecting santri to get scholarships at Mifathul Huda Islamic Boarding School, Bogor. This research has conducted a feasibility test on the application that was built with a feasibility value of 87.59%, meaning that the application built is very feasible and has also been measured through MATLAB with a Silhouette Index with a value of 0.7030 (Strong Structure).

Keywords: *Clustering, Scholarship, K-Means Algorithm Method, Santri, Silhouette Index.*

ABSTRAK

Pada saat ini cerminan kualitas pendidikan terlihat dari tingginya tingkat keberhasilan dan rendahnya tingkat kegagalan santri di dunia pendidikan. Kriteria untuk menyeleksi santri calon penerima beasiswa salah satunya dilihat dari nilai akademik yang tinggi. Namun, yang selama ini terjadi tingginya nilai akademik yang tinggi, melainkan berasal dari luar nilai akademik yaitu nilai kepribadian. Hal ini dilakukan agar santri – santri dapat mendapatkan beasiswa sesuai dengan kemampuannya. Oleh karena itu, peneliti mencoba memberikan alternatif dalam proses pengelompokkan beasiswa dengan memilah variabel data nilai akademik dan data nilai kepribadian. Salah satu cara untuk menyeleksi santri untuk mendapatkan beasiswa adalah dengan melakukan segmentasi dengan mengelompokkan data berdasarkan kriteria tertentu. Algoritma K-Means Clustering adalah salah satu metode

pengelompokkan yang dapat mengelompokkan objek – objek berdasarkan kemiripan sifat yang dimilikinya. Tujuan dari penelitian ini adalah untuk mengetahui penerapan metode Algoritma K-Means Clustering dalam menyeleksi santri untuk mendapatkan beasiswa di Pondok Pesantren Mifathul Huda Bogor. Penelitian ini sudah melakukan uji kelayakan pada aplikasi yang dibangun dengan nilai kelayakan sebesar 87,59%, bermakna aplikasi yang dibangun sangat layak dan juga sudah dilakukan pengukuran melalui MATLAB dengan *Silhouette Index* dengan nilai 0,7030 (*Strong Structure*).

Kata Kunci: *Clustering, Beasiswa, Metode Algoritma K-Means, Santri, Silhouette Index*

A. PENDAHULUAN

1. Latar Belakang

Pondok Pesantren Miftahul Huda Bogor mempunyai cukup banyak santri atau bisa dibilang tidak sedikit santri disana yang mempunyai potensi untuk meraih prestasi dalam bidang keagamaan disana. Prestasi yang dimaksud adalah hasil yang diukur dengan menggunakan data nilai selama santri – santri menempuh pendidikan disana. Permasalahan dalam penelitian ini adalah pengelompokkan santri yang berhak atau tidaknya mendapatkan rekomendasi untuk mendapatkan beasiswa ketika sudah masuk ke dalam organisasi. Terdapat 5 kriteria untuk santri bisa masuk ke dalam organisasi yaitu nilai ibadah, nilai akhlak, nilai kesungguhan, nilai khidmah, nilai K3 (Kebersihan, Ketertiban, dan Kerapihan), dan nilai akademik. Beasiswa tersebut terbagi menjadi dua kelompok yaitu beasiswa belajar dengan beasiswa kebutuhan hidup selama menempuh studi di Pesantren. Beasiswa belajar adalah beasiswa pendidikan di Pesantren seperti biaya ujian dan SPP (Surat Pembinaan Pendidikan) sedangkan beasiswa kebutuhan hidup adalah beasiswa untuk keperluan asrama dan kebutuhan konsumsi selama di Pesantren. Kesulitan pihak manajemen yaitu pada saat proses penentuan santri yang berhak mendapatkan beasiswa. Proses prediksi untuk menyeleksi ini yang selama ini berjalan yaitu membutuhkan waktu yang lama karena pihak pada saat proses pengelompokkan harus diulang – ulang agar mendapatkan santri yang berprestasi dan lalu hasil dari seleksi tersebut dimusyawarahkan untuk mendapatkan hasil yang diharapkan. Pentingnya proses pengelompokkan santri yang berprestasi di pondok Pesantren, maka dibutuhkan suatu metode yang dapat mempermudah untuk proses prediksi dalam menyeleksi. Dari uraian diatas, K – Means Clustering diharapkan dapat menyelesaikan permasalahan dalam pengelompokkan santri yang berprestasi untuk mendapatkan beasiswa selama masa studi pendidikan di Pondok Pesantren Miftahul Huda Bogor.

2. Permasalahan

Proses penentuan santri berprestasi merupakan bagian yang terpenting, sebab pada proses penentuan tersebut berpengaruh terhadap rekomendasi santri berprestasi untuk mendapatkan beasiswa belajar dan beasiswa bantuan santri selama menempuh masa pendidikan di Pesantren.

Penentuan santri berprestasi yang selama ini berjalan, ketika pihak manajemen akan merekomendasikan santri yang berhak atau tidaknya mendapatkan beasiswa dengan kategori. Proses seleksi tersebut diawali dengan wali kelas menilai santri dari segi akademik untuk memperoleh nilai akademik, lalu bagian pengasuhan menilai santri dari segi kepribadian untuk memperoleh nilai kepribadian, dan bagian manajemen santri yang menanangi bagian data diri untuk memperoleh data diri santri. Ketika semua pihak sudah menilai, maka semua pihak memberikan nilai – nilai tersebut kepada bagian UPZISWA (Unit Penghimpunan Zakat Infaq Sodaqoh dan Wakaf). Bagian UPZISWA kemudian mengadakan musyawarah untuk proses

seleksi, proses seleksi yang membutuhkan waktu yang cukup lama untuk membuat pemetaan siapa santi yang mendapatkan beasiswa untuk kategori prestasi belajar atau bantuan hidup. Dan ketika hasil dari dari pemetaan sudah selesai, maka pihak UPZISWA melakukan musyawarah kembali untuk menentukan siapa yang benar – benar berhak atau tidaknya mendapatkan beasiswa. Berdasarkan permasalahan tersebut maka diidentifikasi sebagai berikut:

- a. Belum akurat dalam pengelompokkan santri yang mendapat beasiswa;
- b. Belum efektif dalam proses pengelompokkan beasiswa santri yang mendapat beasiswa prestasi belajar dan beasiswa bantuan.

3. Tujuan

Adapun tujuan dari penelitian ini yaitu sebagai berikut :

1. Mendapatkan pengelompokkan santri yang mendapatkan beasiswa sesuai dengan kriteria yang dibutuhkan;
2. Mendapatkan proses pengelompokkan beasiswa santri yang lebih efektif;
3. Mengembangkan *prototype* aplikasi untuk rekomendasi santri berprestasi mendapat beasiswa dengan menggunakan metode K-Means;
4. Mengukur tingkat ketepatan dan efektifitas metode K-Means untuk mengelompokkan santri yang berprestasi untuk mendapatkan beasiswa.

4. Tinjauan Pustaka

a. Data Mining

Menurut Turban (2001), Data mining adalah proses yang mempekerjakan satu atau lebih teknik pembelajaran komputer (machine learning) untuk menganalisis dan mengekstrasi pengetahuan (knowledge) secara otomatis. Definisi lain diantaranya adalah pembelajaran berbasis induksi (induction – based learning) adalah proses pembentukan definisi-definisi konsep umum yang dilakukan dengan cara mengobservasi contoh-contoh spesifik dari konsep – konsep yang akan dipelajari. Knowledge Discovery in Databases (KDD) adalah penerapan metode saintifik pada data mining. Dalam konteks ini data mining merupakan satu langkah dari proses KDD. Menurut Tan (2006:10) Serangkaian proses tersebut memiliki tahap sebagai berikut:

1. Pembersihan data yaitu untuk membuang data yang tidak konsisten dan noise.
2. Integrasi data yaitu penggabungan data dari beberapa sumber
3. Transformasi data yaitu data diubah menjadi bentuk yang sesuai untuk di mining.
4. Aplikasi teknik data mining yaitu proses ekstraksi pola dari data yang ada.
5. Evaluasi pola yang ditemukan yaitu proses interpretasi pola menjadi pengetahuan yang dapat digunakan untuk mendukung pengambilan keputusan.
6. Presentasi pengetahuan yaitu dengan teknik visualisasi.

Tahap ini merupakan bagian dari proses pencarian pengetahuan yang mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesa yang ada sebelumnya. Langkah terakhir KDD adalah mempresentasikan pengetahuan dalam bentuk yang mudah dipahami pengguna.

b. Clustering

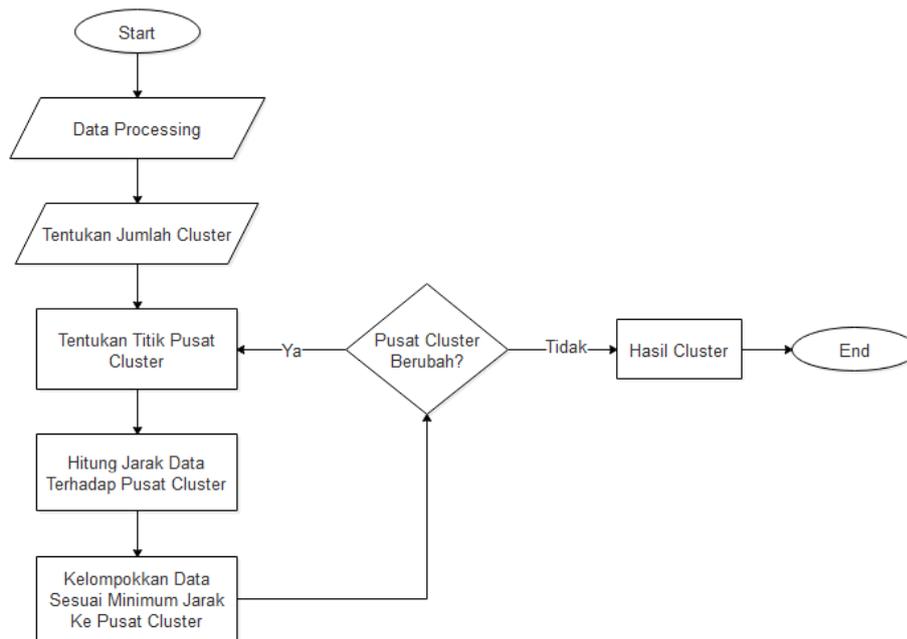
Menurut Tan (2006) Clustering adalah sebuah proses untuk mengelompokkan data ke dalam beberapa cluster atau kelompok sehingga data dalam satu cluster memiliki tingkat kemiripan yang maksimum dan data antar cluster memiliki kemiripan yang minimum. Adapun tujuan dari data clustering ini adalah untuk meminimalisasikan objektif function yang di-set dalam proses clustering, yang pada umumnya berusaha meminimalisasikan

variasi dalam suatu cluster. Dan meminimalisasikan variasi antar cluster. Secara garis besar, terdapat beberapa metode klasifikasi data. Pemilihan metode clustering tergantung pada tipe data dan tujuan *clustering* itu sendiri.

B. METODE

1. Algoritma K-Means

K-Means merupakan salah satu dari beberapa metode data clustering non hirarki dengan sistem kerja mempartisi data yang ada ke dalam bentuk satu atau lebih cluster/kelompok. Pembagian data ke dalam cluster/kelompok pada metode ini menggunakan data dengan karakteristik yang sama yang dikelompokkan ke dalam satu cluster yang sama (Purnamaningsih, Saptono, & Aziz, 2014). Berikut adalah tahapan dalam Algoritma K-Means:



Gambar 1. Tahapan Algoritma K-Means
 (Ahmad Chusyairi, Pelsri Ramadar Noor Saputra, 2019)

Keterangan :

1. Proses data untuk menjadi dataset yang akan digunakan untuk menghitung dengan menggunakan Algoritma K-Means;
2. Tentukan jumlah cluster yang akan dipergunakan dalam pembagian data;
3. Tentukan titik centroid awal yang diperoleh secara acak serta jumlah centroid sebanyak cluster yang akan dibuat. Pengertian centroid sendiri adalah titik pusat cluster atau awal pusat cluster;
4. Lakukan perhitungan jarak pada setiap inputan data terhadap pusat cluster hingga ditemukan jarak paling dekat dari setiap data terhadap centroid.

Pada tahap ini menggunakan rumus :

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

Keterangan :

$d(x,y)$: Jarak antara data x dan data y
 x_i : Data testing ke-i
 y_i : Data training ke-i

5. Kelompokkan setiap data terhadap jarak pada titik pusat centroid terdekat.
6. Ubah nilai centroid yang diperoleh dari rata-rata cluster yang bersangkutan
7. Jika anggota tiap cluster tidak ada yang berubah, maka iterasi selesai dan nilai rata-rata pusat cluster akan digunakan sebagai parameter dalam penentuan pembagian data.
8. Jika tidak, maka iterasi akan dilakukan dengan menggunakan langkah 4 hingga langkah 5.

Pada tahap ini menggunakan rumus :

$$\mu_k = \frac{1}{N_k} \sum_{i=1}^{N_k} x_i$$

Keterangan :

μ_k : Titik centroid dari cluster ke-K
 N_k : Banyaknya data pada cluster ke-K
 x_i : Data ke-i pada cluster ke-K

2. Teknik Analisa Data

Menurut Eko Prasetyo (2014, p, 283) *Silhouette Index* dapat digunakan untuk memvalidasi baik sebuah data, cluster tunggal (satu cluster dari sejumlah cluster) atau bahkan seluruh cluster, metode ini paling banyak digunakan untuk memvalidasi cluster. Untuk menghitung nilai *silhouette coefficient* diperlukan jarak antar dokumen dengan menggunakan rumus *Euclidean Distance*. Setelah itu tahapan untuk menghitung nilai *silhouette coefficient* adalah sebagai berikut :

1. Untuk setiap objek i , hitung rata-rata jarak dari objek i dengan seluruh objek yang berada dalam satu cluster. Akan didapatkan nilai rata-rata yang disebut a .
2. Untuk setiap objek i , hitung rata-rata jarak dari objek i dengan objek yang berada di cluster lainnya. Dari semua jarak rata-rata tersebut ambil nilai yang paling kecil. Nilai ini disebut b .
3. Setelah itu maka untuk objek i memiliki nilai *silhouette coefficient*.

$$S = (b-a) / \text{Max}(a,b)$$

Keterangan:

S = Silhouette Index
 a = rata-rata jarak dari objek i dengan seluruh objek yang berada dalam satu cluster
 b = jarak rata-rata yang paling kecil

Hasil perhitungan nilai *silhouette coefficient* dapat bervariasi antara -1 hingga 1. Hasil clustering dikatakan baik jika nilai *silhouette coefficient* bernilai positif ($a_i < b_i$) dan a_i mendekati 0, sehingga akan menghasilkan nilai *silhouette coefficient* yang maksimum yaitu 1 saat $a_i = 0$. Maka dapat dikatakan, jika $s_i = 0$ maka objek i berada di antara dua cluster sehingga objek tersebut tidak jelas harus dimasukkan ke dalam cluster A atau cluster B. Akan tetapi, jika $s_i = -1$ artinya struktur cluster yang dihasilkan overlapping., sehingga objek i lebih tepat dimasukkan ke dalam cluster yang lain. Nilai rata-rata *silhouette coefficient* dari tiap objek dalam suatu cluster adalah suatu ukuran yang menunjukkan seberapa ketat data

dikelompokkan dalam cluster tersebut. Nilai silhouette berdasarkan tabel Kaufman dan Rousseeuw.

Tabel 1. Kaufman dan Rousseeuw

Skala	Keterangan
$0.7 < SC \leq 1$	<i>Strong Structure</i>
$0.5 < SC \leq 0.7$	<i>Medium Structure</i>
$0,25 < SC \leq 0.5$	<i>Weak Structure</i>
$SC \leq 0.25$	<i>No Structure</i>

C. HASIL DAN PEMBAHASAN

1. HASIL

Berikut ini adalah langkah-langkah dalam menghitung Algoritma K-Means Clustering:

a. Siapkan dataset

Dataset yang digunakan pada perhitungan ini menggunakan variabel data rata – rata dan variabel angka mutu santri kelas 11 semester 1. Dataset tersebut terdiri dari 55 baris dan 4 kolom.

Tabel 2. Dataset Perhitungan K-Means Beasiswa Santri

No	Nama Santri	Rata - Rata Nilai Akademik	Angka Mutu
1	Ahmed Michail Ghibran	6.20	3.00
2	Akmal Muhammadan	5.85	3.00
...
55	Zahra	5.10	2.00

b. Tentukan jumlah cluster (K)

Dari dataset perhitungan K-means beasiswa santri akan dibagi menjadi 2 kelompok/*cluster* sehingga nilai K pada perhitungan ini adalah 2. Dimana cluster 1 untuk beasiswa belajar dan cluster 2 untuk beasiswa bantuan.

c. Pilih titik centroid secara acak

Pada langkah 2 di atas, dataset perhitungan K-means beasiswa santri akan dibagi menjadi 2 (dua) kelompok/*cluster* sehingga titik centroid yang dipilih juga sejumlah 2 (dua). Titik centroid yang digunakan dapat dilihat pada tabel 3 titik centroid awal.

Tabel 3. Titik Centroid Awal

Data ke-i	Rata – Rata Nilai Akademik	Angka Mutu
9	5.05	3.00
18	4.35	4.00

d. Kelompok data sehingga terbentuk K buah cluster dengan titik centroid dari setiap cluster

Pada tahap ini menggunakan rumus :

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

Keterangan:

$d(x, y)$: Jarak antara data x dan data y

x_i : Data testing ke-i

y_i : Data training ke-i

Data ke – 1

$$C1 = \sqrt{(6,20 - 5,05)^2 + (3,00 - 3,00)^2} = 1,15$$

$$C2 = \sqrt{(6,20 - 4,35)^2 + (3,00 - 4,00)^2} = 2,10$$

Tabel 4. Hasil Pengelompokkan Data Iterasi Pertama

No	Nama Santri	Rata - Rata Nilai Akademik	Angka Mutu	Cluster		Jarak Terdekat	Hasil
				C1	C2		
1	Ahmed Michail Ghibran	6.20	3.00	1.15	2.10	1.15	Belajar
2	Akmal Muhammadan	5.85	3.00	0.80	1.80	0.80	Belajar
...
55	Zahra	5.10	2.00	1.00	2.14	1.00	Belajar

Setelah didapatkan nilai dari masing – masing *clustering* maka selanjutnya adalah melakukan pengecekan data antara *clustering* C1 dan C2 untuk mendapatkan jarak terdekat dengan pusat data *clustering* dapat dilihat pada tabel 1.4. Dari tabel 4 terlihat bahwa santri pada cluster 1 yaitu 53 orang, sedangkan santri pada cluster 2 yaitu 2 orang.

- e. **Jika anggota tiap cluster tidak ada yang berubah, maka iterasi selesai dan nilai rata-rata pusat cluster akan digunakan sebagai parameter dalam penentuan pembagian data.**

Pada tahap ini menggunakan rumus :

$$\mu_k = \frac{1}{N_k} \sum_{i=1}^{N_k} x_i$$

Keterangan :

μ_k : Titik centroid dari cluster ke-K

N_k : Banyaknya data pada cluster ke-K

x_i : Data ke-i pada cluster ke-K

Pada tabel 1.4 menunjukkan titik centroid lama dan tabel 4 menunjukkan titik centroid baru untuk iterasi berikutnya. Perhitungan manual titik centroid lama menjadi titik centroid baru adalah sebagai berikut :

- a. Titik Centroid 1 baru iterasi 2

$$X = \frac{6.20 + 5.85 + 7.15 + \dots + 5.10}{53} = 6.67$$

$$Y = \frac{3.00 + 3.00 + 3.00 + \dots + 2.00}{53} = 2.92$$

b. Titik Centroid 2 baru iterasi 2

$$X = \frac{4.80 + 4.35}{2} = 4.58$$

$$Y = \frac{4.00 + 4.00}{2} = 4.00$$

Tabel 5. Titik Centroid Baru Untuk Iterasi 2

Centroid Baru	X	Y
C1	6,67	2,92
C2	4,58	4,00

f. **Ulangi langkah c dan e sampai nilai dari titik centroid tidak lagi berubah**

Untuk iterasi kedua dapat dilihat pada perhitungan sebagai berikut :

Data ke – 1

$$C1 = \sqrt{(6,20 - 6,47)^2 + (3,00 - 2,92)^2} = 0,48$$

$$C2 = \sqrt{(6,20 - 4,58)^2 + (3,00 - 4,00)^2} = 1,91$$

Tabel 6. Hasil Pengelompokkan Data Iterasi Kedua

No	Nama Santri	Rata - Rata Nilai Akademik	Angka Mutu	Cluster		Jarak Terdekat	Hasil
				C1	C2		
1	Ahmed Michail Ghibran	6.20	3.00	0.48	1.91	0.48	Belajar
2	Akmal Muhammadan	5.85	3.00	0.82	1.62	0.82	Belajar
...
55	Zahra	5.10	2.00	1.82	2.07	1.82	Belajar

Setelah didapatkan nilai dari masing – masing clustering maka selanjutnya adalah melakukan pengecekan data antara clustering C1 dan C2 untuk mendapatkan jarak terdekat dengan pusat data clustering dapat dilihat pada tabel 6. Dari tabel 6 terlihat bahwa santri pada cluster 1 yaitu 44 orang, sedangkan santri pada cluster 2 yaitu 11 orang. Dikarenakan hasil iterasi pertama dan kedua berubah, maka selanjutnya diharuskan untuk memperbarui titik centroid di iterasi ketiga yang dapat dilihat pada perhitungan sebagai berikut:

a. Titik Centroid 1 baru iterasi 2

$$X = \frac{6.20 + 5.85 + 7.15 + \dots + 5.10}{44} = 6.98$$

$$Y = \frac{3.00 + 3.00 + 3.00 + \dots + 2.00}{2} = 2.91$$

b. Titik Centroid 2 baru iterasi 2

$$X = \frac{4.80 + 5.35 + 5.05 + \dots + 4.80}{11} = 5.12$$

$$Y = \frac{4.00 + 3.00 + 3.00 + \dots + 4.00}{11} = 3.18$$

Tabel 7 Titik Centroid Baru Untuk Iterasi 3

Centroid Baru	X	Y
C1	6.98	2,91
C2	5.12	3.18

Untuk iterasi ketiga dapat dilihat pada perhitungan sebagai berikut :

Data ke – 1

$$C1 = \sqrt{(6,20 - 6,96)^2 + (3,00 - 2,91)^2} = 0.77$$

$$C2 = \sqrt{(6,20 - 5,02)^2 + (3,00 - 3,40)^2} = 1.10$$

Tabel 8 Hasil Pengelompokkan Data Iterasi Ketiga

No	Nama Santri	Rata - Rata Nilai Akademik	Angka Mutu	Cluster		Jarak Terdekat	Hasil
				C1	C2		
1	Ahmed Michail Ghibran	6.20	3.00	0.77	1.10	0.77	Belajar
2	Akmal Muhammadan	5.85	3.00	1.12	0.75	0.75	Bantuan
...
55	Zahra	5.10	2.00	2.07	1.18	1.18	Bantuan

Setelah didapatkan nilai dari masing – masing clustering maka selanjutnya adalah melakukan pengecekan data antara clustering C1 dan C2 untuk mendapatkan jarak terdekat dengan pusat data clustering dapat dilihat pada tabel 8. Dari tabel 8 terlihat bahwa santri pada cluster 1 yaitu 36 orang, sedangkan santri pada cluster 2 yaitu 19 orang. Dikarenakan hasil iterasi kedua dan ketiga berubah, maka selanjutnya diharuskan untuk memperbarui titik centroid di iterasi keempat yang dapat dilihat pada perhitungan sebagai berikut:

a. Titik Centroid 1 baru iterasi 4

$$X = \frac{6.20 + 7.15 + 8.30 + \dots + 7,40}{36} = 7,29$$

$$Y = \frac{3.00 + 3.00 + 3.00 + \dots + 3.00}{36} = 2.97$$

b. Titik Centroid 2 baru iterasi 4

$$X = \frac{5.85 + 4.80 + 5.40 + \dots + 5.10}{19} = 5.28$$

$$Y = \frac{3.00 + 4.00 + 2.00 + \dots + 2.00}{19} = 2.95$$

Tabel 9. Titik Centroid Baru Untuk Iterasi 4

Centroid Baru	X	Y
C1	7.29	2.97
C2	5.28	2,95

Untuk iterasi keempat dapat dilihat pada perhitungan sebagai berikut :
 Data ke – 1

$$C1 = \sqrt{(6,20 - 7,29)^2 + (3,00 - 2,97)^2} = 1.09$$

$$C2 = \sqrt{(6,20 - 5,28)^2 + (3,00 - 2,95)^2} = 0,92$$

Tabel 10. Hasil Pengelompokkan Data Iterasi Keempat

No	Nama Santri	Rata - Rata Nilai Akademik	Angka Mutu	Cluster		Jarak Terdekat	Hasil
				C1	C2		
1	Ahmed Michail Ghibran	6.20	3.00	1.09	0.92	0.92	Bantuan
2	Akmal Muhammada n	5.85	3.00	1.44	0.57	0.57	Bantuan
...
55	Zahra	5.10	2.00	2.39	0.97	0.97	Bantuan

Setelah didapatkan nilai dari masing – masing clustering maka selanjutnya adalah melakukan pengecekan data antara clustering C1 dan C2 untuk mendapatkan jarak terdekat dengan pusat data clustering dapat dilihat pada tabel 1.10. Dari tabel 1.10 terlihat bahwa santri pada cluster 1 yaitu 31 orang, sedangkan santri pada cluster 2 yaitu 24 orang. Dikarenakan hasil iterasi kedua, ketiga, dan keempat berubah, maka selanjutnya diharuskan untuk memperbarui titik centroid di iterasi kelima yang dapat dilihat pada perhitungan sebagai berikut:

a. Titik Centroid 1 baru iterasi 5

$$X = \frac{7.15 + 8.30 + 7.55 + \dots + 7.40}{31} = 7.47$$

$$Y = \frac{3.00 + 3.00 + 3.00 + \dots + 3.00}{31} = 3.00$$

b. Titik Centroid 2 baru iterasi 5

$$X = \frac{6.20 + 5.85 + 4.80 + \dots + 5.10}{24} = 5.46$$

$$Y = \frac{3.00 + 3.00 + 4.00 + \dots + 2.00}{24} = 2.92$$

Tabel 11. Titik Centroid Baru Untuk Iterasi 5

Centroid Baru	X	Y
C1	7.47	3.00
C2	5.46	2.92

Untuk iterasi kelima dapat dilihat pada perhitungan sebagai berikut :

Data ke – 1

$$C1 = \sqrt{(6,20 - 7,47)^2 + (3,00 - 3,00)^2} = 1.27$$

$$C2 = \sqrt{(6,20 - 5,46)^2 + (3,00 - 2,92)^2} = 0.75$$

Tabel 12. Hasil Pengelompokkan Data Iterasi Kelima

No	Nama Santri	Rata - Rata Nilai Akademik	Angka Mutu	Cluster		Jarak Terdekat	Hasil
				C1	C2		
1	Ahmed Michail Ghibran	6.20	3.00	1.27	0.75	0.75	Bantuan
2	Akmal Muhammadan	5.85	3.00	1.62	0.40	0.40	Bantuan
...
55	Zahra	5.10	2.00	2.58	0.98	0.98	Bantuan

Berdasarkan perhitungan nilai titik centroid baru iterasi kelima, didapatkan hasil sama dengan nilai titik centroid iterasi keempat. Alhasil, pengulangan (iterasi) akan dihentikan dan dari hasil pengelompokkan data iterasi kelima telah didapatkan data masing – masing kelompok yang dapat dilihat pada tabel 12. Dapat disimpulkan bahwa hasil clustering dari dataset perhitungan K-Means bahwa beasiswa santri setelah dihitung dengan nilai rata – rata nilai akademik dan angka mutu yang tinggi termasuk ke dalam cluster 1 yaitu beasiswa belajar dengan jumlah 31 santri dan rata – rata nilai akademik dan angka mutu yang cukup masuk ke dalam cluster 2 yaitu beasiswa belajar dengan jumlah 24 santri yang dapat dilihat pada tabel 13.

Tabel 13. Hasil Pengelompokkan Data Iterasi Kelima

No	Nama Santri	Hasil Cluster
1	Ahmed Michail Ghibran	Beasiswa Bantuan
2	Akmal Muhammadan	Beasiswa Bantuan
...
55	Zahra	Beasiswa Bantuan

2. PEMBAHASAN

Analisa metode *Silhouette* ini dilakukan dengan melihat besar nilai s dari hasil perhitungan dengan menggunakan bantuan *software* MATLAB. Hasil perhitungan *Silhouette* berada pada nilai 0.7030 (Strong Structure). Nilai s dikonversikan ke 1 jika nilai *Silhouette*-nya lebih besar dari 0 maka artinya hasil cluster yang dihasilkan merupakan cluster yang terbaik karena semakin sedikit nilai s yang nilainya di bawah 0 berarti objek i sudah berada pada cluster yang tepat di setiap clusternya. Mengacu pada tabel 3.8 *Kaufman* dan *Rousseeuw* nilai rata-rata validasi dari setiap cluster berada pada poin 0.7030 (Strong Structure), nilai rata-rata ini berada pada nilai $s = 1$, ini menunjukkan ketatnya pengelompokkan dalam cluster tersebut. Berikut ini adalah langkah-langkah dalam menerapkan Algoritma K-Means Clustering dengan menggunakan aplikasi MATLAB untuk mengelompokkan beasiswa santri dan *silhouette score* di Pondok Pesantren Miftahul Huda Bogor yaitu :

1. Buatlah folder untuk menyimpan data;
2. Siapkan dataset.

Tabel 14. Dataset Nilai Santri

akademik	Pribadi
6.2	3
...	...
5.1	2

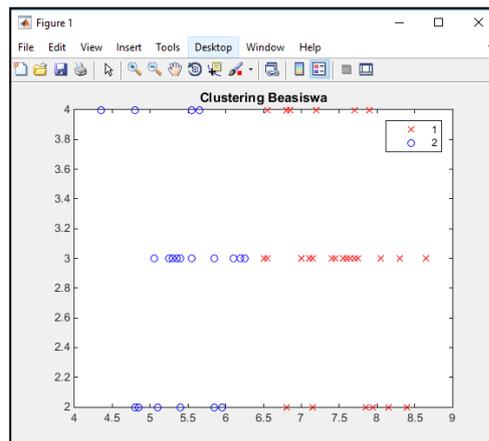
3. Ubah dataset tersebut ke dalam bentuk excel
4. Buka aplikasi matlab, lalu isi source code seperti di bawah ini

```

clc;clear all;close all;
data=readtable('data2.xlsx')
X=[data.akademik,data.pribadi];
jumlah_cluster = 2;
rand('seed',0);
opts = statset('Display','Final');
[label,centroid]= kmeans(X,jumlah_cluster,'Options',opts);
figure;
gscatter(X(:,1),X(:,2),label,'rbg','xod'),title('Clustering
Beasiswa')
    
```

Gambar 1. Script MATLAB Terkait Algoritma K-Means Clustering

5. Klik “Run”
6. Jika *source code* pada langkah ke-4 dijalankan, maka akan muncul keluaran seperti pada gambar 1



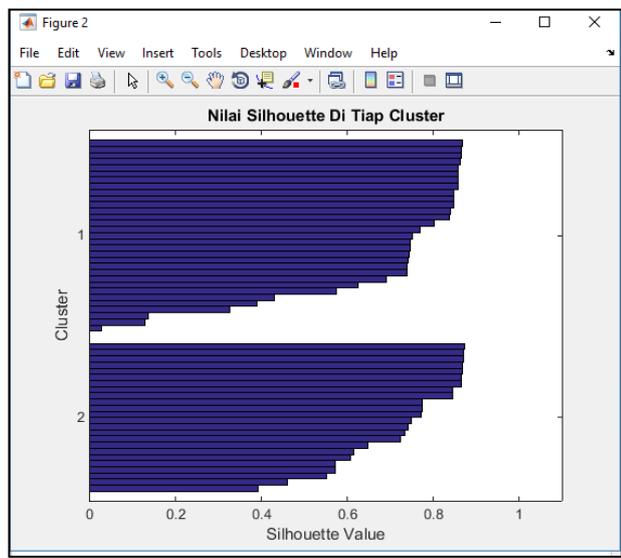
Gambar 2 Plot Clustering

7. Untuk memvalidasi sebuah data, cluster tunggal (satu cluster dari cluster) atau bahkan seluruh cluster maka dapat digunakan *Silhouette Index*, karena metode ini paling banyak digunakan untuk memvalidasi cluster. Berikut ini merupakan *source code* untuk memvalidasi cluster dengan Algoritma K-Means dengan metode *Silhouette Index* setiap cluster bisa melalui script berikut :

```
figure
[S,H] = silhouette (X, label);
title('Nilai Silhouette Di Tiap Cluster')
info = array2table ([S,label], 'variablenames',
{'nilai','label'});
disp (info)
score = mean(S)
```

Gambar 3. Script Terkait *Silhouette Index*

8. Jika source code pada langkah ke-7 dijalankan, maka akan muncul keluaran seperti pada gambar 4



Gambar 4. Plot *Silhouette Index*

9. Hasil dari silhouette index setiap cluster pada tabel 15

Tabel 15 Hasil *Silhouette Index* Dengan Perhitungan MATLAB

Nilai	Label
0.45904	2
...	...
0.8681	2

Score = 0.7030

Keterangan :

Nilai : Nilai Silhouette Score 1 : Beasiswa Belajar
 Label : Cluster 2 : Beasiswa Bantuan

D. KESIMPULAN

Berdasarkan hasil penelitian yang dilakukan, kesimpulan yang bisa diuraikan antara lain :

1. Dapat memberikan hasil yang dibutuhkan untuk mengelompokkan beasiswa dengan menggunakan metode Algoritma K-Means Clustering, dari hasil pengelompokkan diperoleh 31 santri untuk beasiswa belajar dan 24 santri untuk beasiswa bantuan dengan menggunakan variabel nilai rata – rata akademik dan angka mutu nilai kepribadian;
2. Dalam pembuatan aplikasi untuk menentukan pengelompokkan santri untuk mendapat beasiswa dengan menerapkan metode Algoritma K-Means Clustering, telah dilakukan uji pengguna dengan menggunakan kuesioner PSSUQ, sesuai dengan kategori PSSUQ diantaranya yaitu nilai keputus secara keseluruhan (Overall) sebesar 87,59%, kegunaan sistem (Sysuse) sebesar 85,71%, kualitas informasi (Infoqual) sebesar 87,76%, dan kualitas antar muka (Interqual) sebesar 90,48% yang artinya aplikasi ini layak digunakan;
3. Dalam menerapkan metode Algoritma K-Means Clustering pada sistem pengelompokkan santri untuk mendapat beasiswa telah dilakukan uji hasil dengan evaluasi cluster dengan menggunakan metode *Silhouette Index* melalui aplikasi MATLAB dan diperoleh nilai 0.7030 (Strong Structure), nilai rata-rata ini berada pada nilai $s = 1$, ini menunjukkan ketatnya pengelompokkan dalam cluster tersebut.

E. DAFTAR PUSTAKA

- [1] Ahmad Chusyairi, P. R. (2019). Pengelompokan Data Puskesmas Banyuwangi Dalam Pemberian Imunisasi Menggunakan Metode K-Means Clustering. *Telematika*, 12 (2), Hal. 139-148.
- [2] Chandra Purnamaningsih, S. R. (2014). Pemanfaatan Metode K-Means Clustering dalam Penentuan Penjurusan Siswa SMA. *Jurnal ITSmart*, 3(1), Hal. 27 – 33.
- [3] Ghaniy, R., & Indriyaningsih, F. (2020). Penerapan Metode Fuzzy C-Means dalam Pemilihan Program Studi Mahasiswa Baru di Perguruan Tinggi. *Teknois : Jurnal Ilmiah Teknologi Informasi dan Sains*, 10(2), 19-30. doi:<https://doi.org/10.36350/jbs.v10i2.84>
- [4] Prasetyo, Eko. (2014). *Data Mining - Mengolah Data Menjadi Informasi Menggunakan Matlab*. Yogyakarta: Andi Offset.
- [5] Supriyatna, A., & Suryanto, A. (2019). Penerapan Metode Weighted Product Untuk Penentuan Penerima Beasiswa Peserta Didik. *Teknois : Jurnal Ilmiah Teknologi Informasi dan Sains*, 9(1), 73-82. doi:<https://doi.org/10.36350/jbs.v9i1.8>
- [6] Tan, P. S. (2006). *Introduction to Data Mining*. Boston: Pearson Education.
- [7] Turban. (2001). *Decision Support System and Intelligent System (Sistem Pendukung Keputusan dan Sistem Cerdas)*. Yogyakarta: Andi.