Jurnal Kecerdasan Buatan dan Teknologi Informasi
Vol. 3, No. 1, Januari 2024, hlm. 37-48
e-ISSN: 2964-2922, p-ISSN: 2963-6191

PENGELOMPOKAN SANTRI DI PONDOK AL-MA'RIFAH BERDASARKAN ASAL WILAYAH MENGGUNAKAN ALGORITMA K-MEAN CLUSTERING

Siti Nur Hasanah*[1], Nana Suarna[2], Denni Pratama[3]

[1]Manajemen Informatika, STMIK IKMI Cirebon, Indonesia
[2]Teknik Informatika, STMIK IKMI Cirebon, Indonesia
[3]Komputerisasi Akuntansi, STMIK IKMI Cirebon, Indonesia

Email: [1]nuy94496@gmail.com, [2]nana.ikmi@gmail.com, [3]denniikmi@gmail.com

(Naskah masuk : 29 Januari 2024, Revisi : 31 Januari 2024, Diterbitkan : 31 Januari 2024)

Abstrak

Pondok Al-Ma'rifah merupakan lembaga pendidikan Islam yang menampung santri dari berbagai wilayah. Penelitian ini bertujuan untuk mengelompokkan santri berdasarkan asal wilayah menggunakan algoritma K-Mean Clustering guna meningkatkan efisiensi manajemen pondok. mencerminkan kebutuhan akan pendekatan yang sistematis dalam mengorganisir santri untuk mendukung optimalisasi pelayanan dan pemenuhan kebutuhan mereka. Masalah utama adalah kompleksitas pengelolaan santri dari beragam wilayah yang dapat menghambat efisiensi dan efektivitas manajemen pondok. Tujuan penelitian ini adalah mengidentifikasi pola distribusi santri berdasarkan asal wilayah dan mengelompokkannya secara otomatis menggunakan algoritma K-Mean Clustering. Metode penelitian melibatkan pengumpulan data asal wilayah santri, pemrosesan data, dan implementasi algoritma K-Mean Clustering untuk membentuk kelompok homogen. Hasil penelitian ini memberikan gambaran yang jelas tentang pola distribusi geografis santri di Pondok Al-Ma'rifah dan menyajikan kelompok-kelompok yang mewakili kesamaan asal wilayah. Penelitian ini diharapkan dapat memberikan kontribusi signifikan dalam meningkatkan efisiensi manajemen pondok, memudahkan pemantauan dan pelayanan santri, serta membuka peluang pengembangan program pendidikan yang lebih terarah sesuai karakteristik kelompok wilayah tertentu.

Kata kunci: Analisis Algoritma, Klasterisasi, *K-Mean Clustering*

CLUSTERING STUDENTS IN AL-MA'RIFAH BOARDING SCHOOL BASED ON REGIONAL ORIGIN USING K-MEAN CLUSTERING ALGORITHM

Abstract

Pondok Al-Ma'rifah is an Islamic educational institution that accommodates students from various regions. This research aims to cluster students based on their regional origins using the K-Mean Clustering algorithm to enhance the efficiency of the boarding school's management. The introduction reflects the need for a systematic approach in organizing students to support the optimization of services and fulfillment of their needs. The main problem lies in the complexity of managing students from diverse regions, which can hinder the efficiency and effectiveness of boarding school management. The objective of this research is to identify patterns in the distribution of students based on their regional origins and automatically group them using the K-Mean Clustering algorithm. The research method involves collecting data on the regional origins of students, processing the data, and implementing the K-Mean Clustering algorithm to form homogeneous groups. The results of this research provide a clear picture of the geographical distribution patterns of students at Pondok Al-Ma'rifah and present groups that represent similarities in regional origins. This research is expected to make a significant contribution to improving the efficiency of boarding school management, facilitating the monitoring and service of students, and opening opportunities for the development of education programs that are more targeted to the characteristics of specific regional groups.

Keywords: *Algorithm Analysis, Clustering, K-Mean Clustering*

1. PENDAHULUAN

Pendidikan Islam yang mengasah pengetahuan agama, karakter, dan keterampilan peserta didik. Dalam konteks ini, pengelompokan santri berdasarkan asal wilayah menjadi penting untuk memahami keragaman

kultural, serta kebutuhan dan tantangan yang mungkin dihadapi oleh kelompok-kelompok tersebut. Algoritma k-means clustering adalah metode yang dapat digunakan untuk mengelompokkan data menjadi beberapa kelompok berdasarkan kesamaan fitur atau karakteristik tertentu.

Data asal wilayah santri di Pondok Al-Ma'rifah perlu dikumpulkan, mencakup informasi seperti nama santri, alamat asal, dan beberapa atribut lain yang relevan. Data yang dikumpulkan perlu diproses dan disiapkan untuk analisis. Ini melibatkan pembersihan data, penghapusan data yang tidak relevan, dan pengkodean ulang data. Sebelum menerapkan algoritma k-means, perlu memutuskan jumlah cluster yang diinginkan. Ini dilakukan dengan analisis data atau dengan menggunakan metode seperti elbow method untuk menentukan titik optimum jumlah cluster [1].

Penelitian terdahulu yang dilakukan oleh Dewi dkk dengan judul Implementasi Metode Clustering Algoritma K-Means untuk menentukan kelompok Tahfidz dan Tahsin di Pesantren Siswa Al-Ma'soem. Bahwa algoritma KMeans dapat diartikan sebagai algoritma pembelajaran yang sederhana untuk menentukan suatu permasalahan pengelompokkan yang bertujuan untuk meminimalkan kesalahan ganda [2]. Aplikasi yang digunakan untuk pengujian proses klasterisasi adalah RapidMiner. RapidMiner merupakan perangkat lunak yang bersifat terbuka (opensource). RapidMiner adalah sebuah solusi untuk melakukan analisis terhadap data mining, text mining dan analisis prediksi [3].

Metode hafalan Al Qur'an Tawazun adalah metode yang memaksimalkan penggunaan otak kanan dan otak kiri, memungkinkan seseorang dapat hafal, paham, dan mutqin. Tujuan dari penelitian ini untuk menentukan jumlah cluster optimum sekaligus anggota masing-masing cluster dengan pengukuran kinerja cluster menggunakan metode Davies Bouldin Index (DBI) dan implementasi algoritma K-Means [4]. Permasalahannya yaitu kompleksitas pengelolaan santri dari beragam wilayah di Pondok Al-Ma'rifah dapat menghambat efisiensi dan efektivitas manajemen. Tujuan penelitian ini adalah untuk mengelompokkan santri berdasarkan asal wilayah menggunakan algoritma K-Mean Clustering guna meningkatkan efisiensi manajemen pondok.

Penelitian ini menghasilkan identifikasi pola distribusi santri berdasarkan asal wilayah dan pembentukan kelompok homogen secara otomatis menggunakan algoritma K-Mean Clustering. Diharapkan kontribusi signifikan dalam meningkatkan efisiensi manajemen, memudahkan pemantauan dan pelayanan santri, serta membuka peluang pengembangan program pendidikan yang lebih terarah sesuai karakteristik kelompok wilayah tertentu [5].

2. METODE PENELITIAN

2.1. Sumber Data

Dalam pengumpulan sumber data, peneliti melakukan pengumpulan sumber data dalam wujud data primer dan data sekunder.

a) Data Primer

Data Primer ialah jenis dan sumber data penelitian yang di peroleh secara langsung dari sumber pertama (tidak melalui perantara),baik individu maupun kelompok. Jadi data yang di dapatkan secara langsung.Data primer secara khusus di lakukan untuk menjawab pertanyaan penelitian. Penulis mengumpulkan data primer dengan metode survey dan juga metode observasi. Metode survey ialah metode yang pengumpulan data primer yang menggunakan pertanyaan lisan dan tertulis. Penulis melakukan wawancara kepada kepala lembaga pondok pesantren Al-Ma'rifah untuk mendapatkan data atau informasi yang di butuhkan. Kemudian penulis juga melakukan pengumpulan data dengan metode observasi. Metode observasi ialah metode pengumpulan data primer dengan melakukan pengamatan terhadap aktivitas dan kejadian tertentu yang terjadi. Jadi penulis datang ke pondok pesantren kebon kelapa Al-Ma'rifah untuk mengamati aktivitas yang terjadi pada pesantren tersebut untuk mendapatkan data atau informasi yang sesuai dengan apa yang di lihat dan sesuai dengan kenyataan [6].

b) Data Sekunder

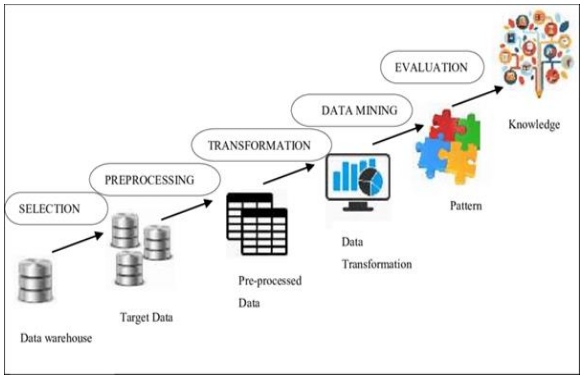
Data Sekunder merupakan sumber data suatu penelitian yang di peroleh peneliti secara tidak langsung melalui media perantara (di peroleh atau dicatat oleh pihak lain). Data sekunder itu berupa bukti,catatan atau laporan historis yang telah tersusun dalam arsip atau data dokumenter. Penulis mendapatkan data sekunder ini dengan cara melakukan permohonan ijin yang bertujuan untuk meminjam arsif data pada pondok pesantren Al-Ma'rifah dan buku yang di gunakan untuk pencatatan data setiap tahunnya [7].

c) Teknik Pengumpulan Data

Metode atau cara untuk melakukan pengumpulan data melalui tiga tahap yaitu tahap yang pertama dengan melakukan observasi pada pondok pesantren Al-Ma'rifah, kemudian memberikan pertanyaan-pertanyaan yang dapat menambah informasi bagi peneliti dengan melakukan wawancara,dan setelah itu mengumpulkan data-data atau dapat di sebut dengan istilah metode dokumentasi yang sangat berguna bagi yang membutuhkan.

2.2. Tahapan Perancangan

Langkah-langkah melakukan perancangan dari tahap awal hingga menyelesaikan laporan penelitian dengan menerapkan proses *Knowledge Discovery In Databases (KDD)* seperti pada gambar 1.



Gambar 1. Tahapan KDD 1

a) Data *Selection*

Data selection merupakan proses pengambilan data yang berhubungan dengan analisis dari basis data. Pada tahapan ini dilakukan teknik perolehan sebuah pengurangan representasi dari data dan meminimalkan hilangnya informasi data. Hal ini meliputi metode pengurangan atribut dan kompresi data. Pemilihan (seleksi) data dari sekumpulan data operasional harus dilakukan sebelum tahap KDD dimulai. Data hasil seleksi yang akan digunakan untuk proses data mining disimpan dalam suatu berkas, terpisah dari basis data operasional [8].

b) Data *prosessing*

Setelah melakukan tahapan data selection maka tahapan selanjutnya adalah melakukan data preprocessing. Tujuan utama dari pengolahan data adalah untuk mengubah data mentah menjadi informasi yang berguna atau yang mudah di pahami bagi pengguna.

c) Data *Transformation*

Setelah melakukan tahapan data *preprocessing* maka tahapan selanjutnya adalah melakukan data *transformation*. Pada tahapan data *transformation* penulis mengubah tipe atribut data yang semula *non-numeric* diubah menjadi tipe data *numeric atribute* yang di ubah adalah wilayah/kota. Tujuan menjadi angka menggunakan operator *Nominal to Numerical* agar sesuai dengan tipe data yang dibutuhkan Algoritma *K-Means* pada proses clustering .

d) Data Mining

Data mining pada penelitian ini di gunakan untuk mengidentifikasi dan mengelola data atau informasi yang berguna. Hal ini dapat melibatkan pengelompokan data menjadi kelompok-kelompok yang serupa *(clustering)* atau menemukan asosiasi antar variabel *(association rules).* Pilihan teknik metode,atau algoritma yang sesuai dan bergantung pada tujuan dan proses keseluruhan dalam *knowledge discovery in database* (KDD).Tahap ini merupakan inti dari proses dari keseluruhan proses KDD yang dilakukan untuk menganalisis data yang sudah diolah [9].

e) Intervetation atau evaluasion

Tahap ini digunakan untuk menilai hasil dari proses data mining dan memastikan apakah hasil tersebut dapat memenuhi tujuan yang telah ditetapkan.Dalam proses evaluasi ini ,mencakup pembuatan foto profil pada setiap cluster yang telah terbentuk ,dan dapat dilakukan menganalisis lebih lanjut untuk mengaitkan dengan astribut peminat .Yaitu informasi yang dihasilkan dari proses data mining perlu disajikan dalam format yang dapat dengan mudah dipahami oleh pihak pihak yang berkepentingan.

2.3. 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* atau kelompok. Pembagian data ke dalam *cluster* atau kelompok pada metode ini menggunakan data dengan karakteristik yang sama yang dikelompokkan ke dalam satu cluster yang sama [10].

Berikut adalah langkah-langkah dari algoritma *K-Means* :

1. Menentukan banyak *k-cluster* yang ingin dibentuk.

2. Membangkitkan nilai random untuk pusat *cluster* awal *(centroid)* sebanyak *k-cluster*.
3. Menghitung jarak setiap data input terhadap masing-masing centroid menggunakan rumus jarak *(Eucledian Distance*) hingga ditemukan jarak yang paling dekat dari setiap data dengan *centroid.*
   Berikut persamaan *Eucledian Distance*:

$$d(x_i, \mu_i) = \sqrt{(xi - \mu i)^2} \qquad (1)$$

4. Mengklasifikasikan setiap data berdasarkan kedekatannya dengan *centroid* (jarak terkecil).
5. Mengupdate nilai *centroid*. Nilai *centroid* baru diperoleh dari rata-rata *cluster* yang bersangkutan dengan menggunakan rumus:

$$C_k = \frac{1}{nk} \sum d_i \qquad (2)$$

   dimana:
   nk = jumlah data dalam cluster
   di = jumlah dari nilai jarak yang masuk dalam masing-masing cluster

6. Melakukan perulangan dari langkah 2 hingga 5 hingga anggota tiap *cluster* tidak ada yang berubah.
7. Jika langkah 6 telah terpenuhi, maka nilai rata-rata pusat cluster (μj) pada iterasi terakhir akan digunakan sebagaiparameter untuk menentukan klasifikasi data.

3. HASIL DAN PEMBAHASAN

3.1. Data *selection*

Tahapan pertama dalam proses KDD yaitu *data selection*. Pada tahapan ini penulis melakukan seleksi data yang akan digunakan dalam proses pengelompokkan. Data yang akan diproses yaitu datap pengelompokan jumlah santri per wilayah atau kota. Tahapan seleksi data bertujuan untuk memilih data yang akan digunakan dalam proses pengelompokkan melalui aplikasi *Rapidminer*, tahapan seleksi data ini dilakukan di *Microsoft Excel*.Data yang digunakan dalam penelitian ini yaitu data pengelompokan jumlah santri berdasarkan asal wilayah dengan data set jumlah 252 dan 4 atribut diantaranya Alamat atau kota, jumlah santri, kode pos atau kota, status keaktifan.

Tabel 1. Pengelompokan Jumlah santri 1

Alamat atau kota	Jumlah santri	Nilai keaktifan kelompok wilayah
Sumedang	60	80
Karawang	35	50
Indramayu	22	50
Subang	25	30
Cirebon	55	75
Bandung	5	20
Majalengka	30	75
Kuningan	10	20
Jakarta	1	10
Lampung	1	10
Bogor	3	10
Bekasi	2	20
Purwakarta	3	20
Brebes	1	10

Tabel 1 menunjukan ada 3 macam atribut yaitu Alamat atau kota, jumlah santri per wilayah, dan nilai keaktifan wilayah atau kelompok santri.

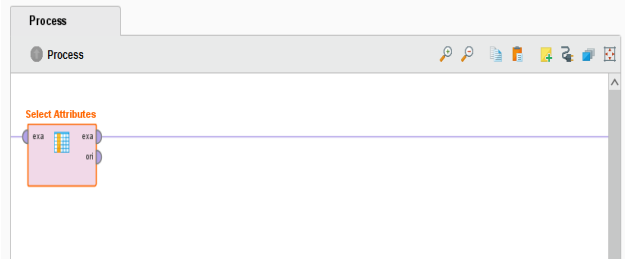
3.2. Data *Preprocessing*

Setelah melakukan tahapan data *selection* maka tahapan selanjutnya adalah melakukan data *preprocessing*. Tujuan dari data *preprocessing* yaitu untuk penghapusan atribut data yang tidak diperlukan, data yang null atau missing, menghilangkan data yang tidak konsisten, serta menambahkan id kedalam dataset yang akan digunakan.

Pada tahapan *Preprocessing* ini penulis menggunakan dua jenis atribut yaitu *Select Attributes* dan *atribut Set Role*. Adapun langkah langkahnya sebagai berikut :

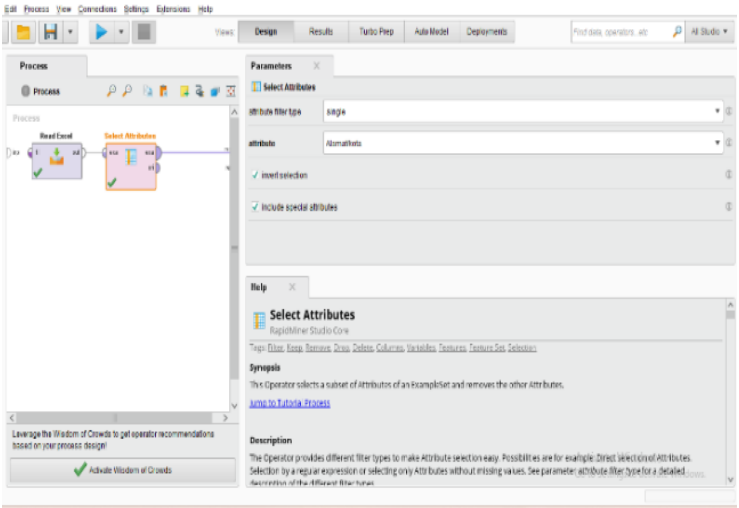
3.1. *Select Attributes*

Select Attributes adalah sebuah operator dalam Rapid Miner yang digunakan untuk memilih atribut atau fitur tertentu dari dataset. Tujuannya adalah untuk mengurangi kompleksitas *dataset* dengan mempertahankan hanya atribut-atribut yang dianggap penting atau relevan untuk analisis atau pemodelan yang sedang dilakukan. Tampilan operator *Select Attributes* dapat dilhat pada gambar 2 di bawah ini.



Gambar 2. *Select Attributes*

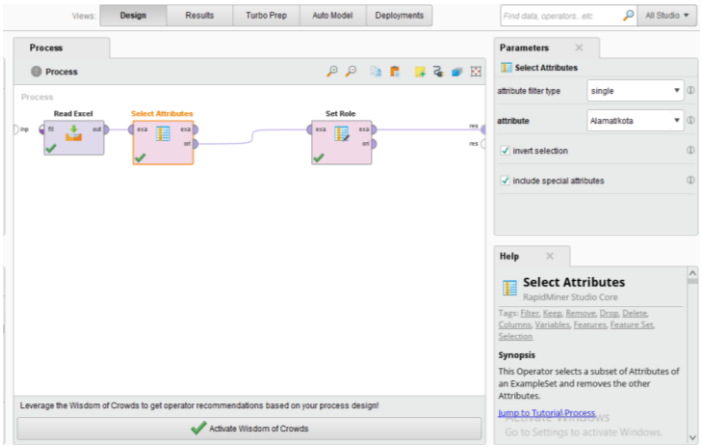
Pada operator *Select Attributes* ada parameter yang harus di sesuaikan terlebih dahulu. Tampilan proses menggunakan *Select Attributes* untuk menghilangkan parameter yang tidak di gunakan pada gambar 2.



Gambar 3. Parameter *Select Attributes*

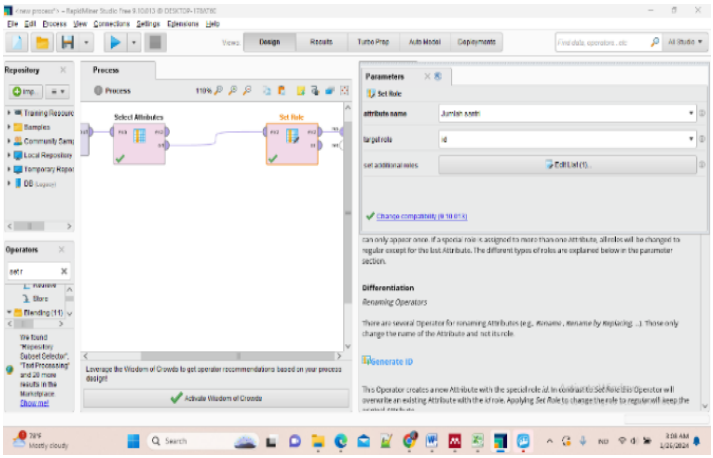
3.2. Operator *Set Role*

Langkah kedua dalam tahapan preprocessing ini yaitu menggunakan Opprator *Set Role*. Operator "*Set Role*" dalam *RapidMiner* digunakan untuk menentukan peran atau jenis data dari atribut dalam dataset. Peran atribut ini sangat penting karena dapat mempengaruhi bagaimana *RapidMiner* memperlakukan atribut tersebut selama proses analisis atau pemodelan data. Tampilan operator *Set Role* dapat dilihat pada gambar 4 di.



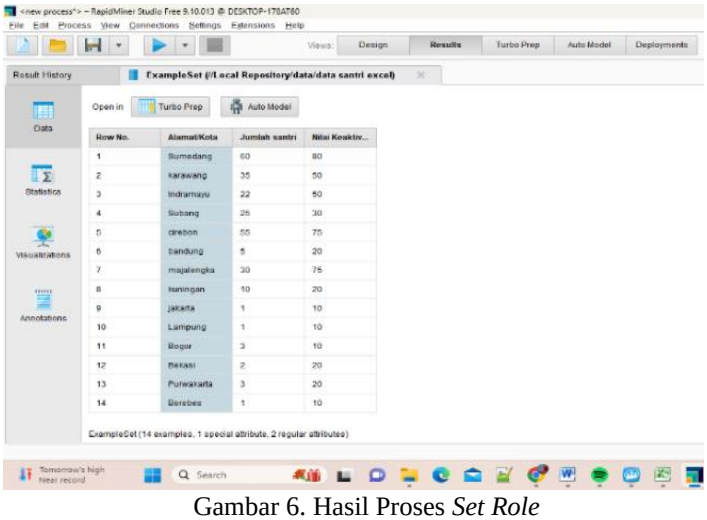
Gambar 4. Operator *set role*

Pada operator *Set Role* terdapat parameter yang harus disesuaikan terlebih dahulu. Tampilan proses menggunakan parameter *Set Role* untuk mengubah atribut menjadi id dapat dilihat pada gambar 5.



Gambar 5. Parameter Set Rol

Atribut data yang menjadi target role id yaitu atribut Jumlah santri. Hasil yang diperoleh dari proses operator ini dapat dilihat pada gambar 6.

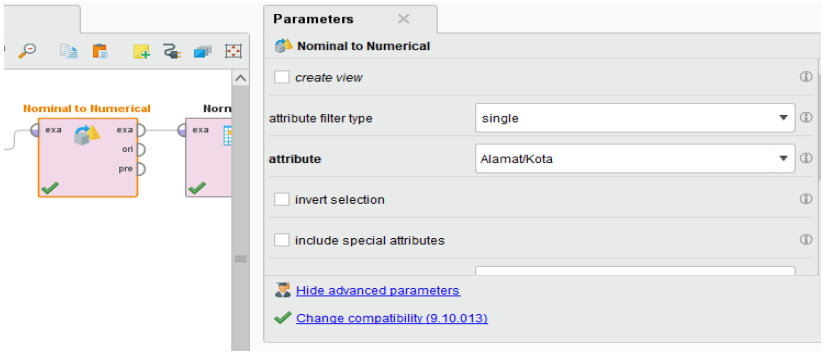


Gambar 6. Hasil Proses *Set Role*

3.3. Data *Transformation*

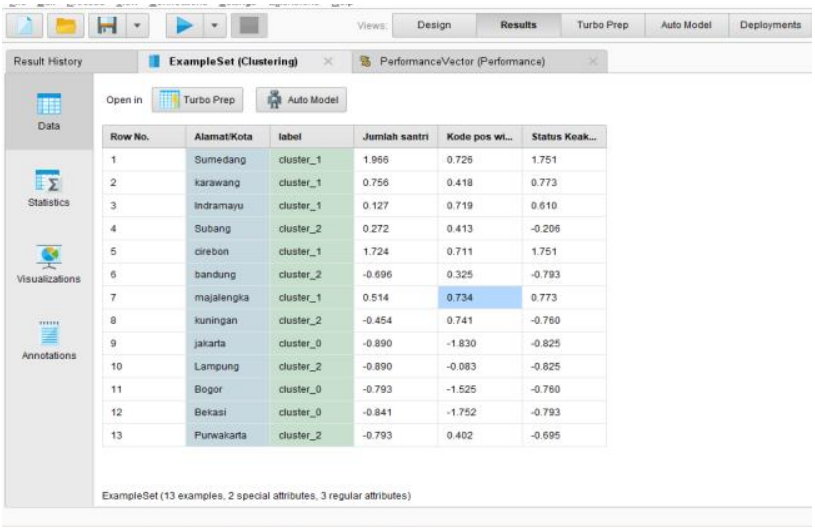
Setelah melakukan tahapan data preprocessing, tahap berikutnya adalah data *transformation*. Tahapan ini melibatkan manipulasi atau perubahan struktur data untuk mempersiapkan agar lebih sesuai dengan kebutuhan

analisis atau pembuatan model. Pada tahapan ini penulis mengubah atribut nominal menjadi atribut numerik. Atribut data yang di ubah dari nominal ke numerical adalah atribut Alamat/kota pada gambar 7.



Gambar 7. *Nominal to Numerical*

Pada gambar di atas menunjukan parameter *Nominal To Numerical* yang mengubah atribut Alamt/kota menjadi *numerical*, dengan *atribute type single* pada gambar 8.



Gambar 8. Hasil *Nominal to Numerical*

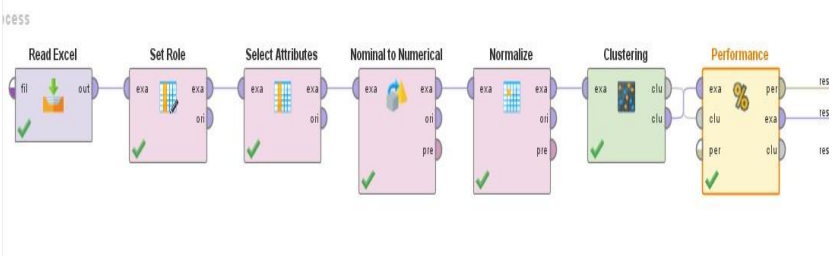
3.4. Data Mining

Setelah melakukan tahapan data transformation maka tahapan selanjutnya adalah melakukan penerapan pada data mining. Tahap berikutnya adalah melakukan proses mining data. Proses yang dilakukan bertujuan untuk memprediksi atau mencari nilai tertinggi dari suatu data. Untuk dapat mengumpulkan data, kita memerlukan cara atau proses tertentu. Hasil akhir yang akan di ambil adalah data jumlah santri Al-Ma'rifah sesuai kelompok wilayah.

Tahapan ini adalah inti dari tahapan KDD *(Knowledge Discovery In Database).* Dalam tahapan ini penulis menggunakan perangkat lunak Rapidminer dengan metode Algoritma K-Means. Pada tahapan data mining penulis menggunakan dua proses, proses yang pertama menggunakan Algoritma K-Means Clustering.Dalam proses ini oprator yang di gunakan yaitu Clustering (K-Means) lalu tahapan kedua penulis melakukan pengujian dan mengevaluasi hasil dari clustering menggunakan operator Cluster Disance Performance dengan metode yang digunakan evaluasi David Bouldin Index (DBI) [11].

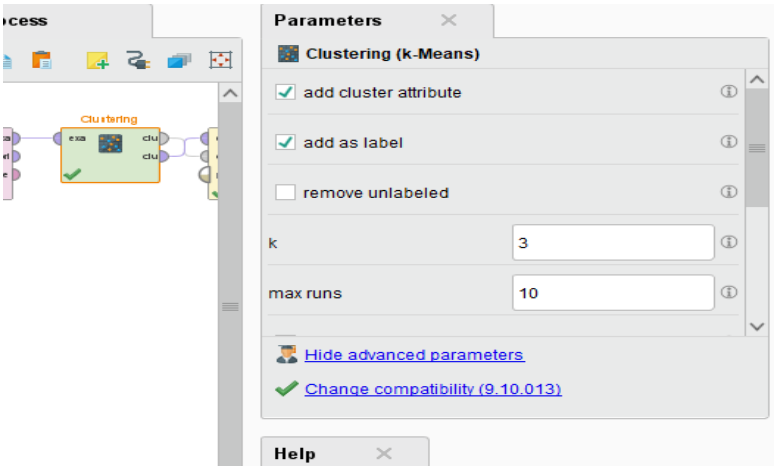
a. Tahapan *Clustering K-Means*

Tahapan *Clustering K-Means* adalah proses langkah-langkah yang dilakukan dalam algoritma *K-Means* untuk mengelompokkan data menjadi beberapa kelompok atau klaster. Pemodelan data mining pada pengelompokkan Jumlah data santri Al-Ma;rifah dapat dilihat pada gambar 9.



Gambar 9. Model Penerapan Algoritma *K-Means*

Pada proses *clustering* terdapat parameter yang harus disesuaikan terlebih dahulu. Tampilan parameter clustering dapat dilhat pada gambar 10.



Gambar 10. Tampilan parameter *K-Means clustering*

K adalah jumlah cluster yang digunakan. Jumlah percobaan cluster yang digunakan oleh penulis yaitu dari (3-10 *cluster*).*Max* runs adalah Jumlah maksimum run *K-Means (default)*. Jumlah yang digunakan penulis yaitu 10.

b. *Evaluasi David Bouldin Index (DBI).*

Tahapan yang kedua yaitu melakukan pengujian hasil *clustering* dengan menerapkan nilai DBI pada parameter *Cluster Distance Performance*. Pada pengujian DBI, *cluster* yang memiliki nilai DBI terkecil atau mendekati 0 dijadikan sebagai cluster yang terbaik. Untuk mencari *cluster* terbaik penulis melakukan percobaan dari cluster 1 sampai dengan 10 rekapitulasi. Jumlah k *cluster* yang dihasilkan dari nilai *Davies Bouldin Index* dapat dilihat dari tabel berikut:

Tabel 2. Hasil *Cluster* dari k2 - k10

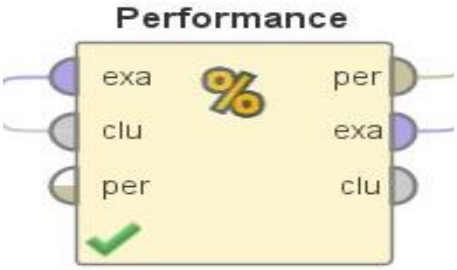
Clustering	Jumlah Anggota *Cluster*	Hasil Nilai DBI
2	*Cluster 0: 3 items*	0,571
	Cluster 1: 8 items	
3	*Cluster 0: 3 items*	0,143
	Cluster 1: 8 items	
	Cluster 2: 2 items	
4	*Cluster 0: 3 items*	0.786
	Cluster 1: 8 items	
	Cluster 2: 2 items	
	Cluster 3: 1 items	
5	*Cluster 0: 3 items*	0.425
	Cluster 1: 8 items	
	Cluster 2: 2 items	
	Cluster 3: 1 items	

	Cluster 4: 5 items	
6	*Cluster 0: 3 items*	0.495
	Cluster 1: 8 items	
	Cluster 2: 2 items	
	Cluster 3: 49 items	
	Cluster 4: 22 items	
	Cluster 5: 126 items	
7	*Cluster 0: 288 items*	0.462
	Cluster 1: 40 items	
	Cluster 2: 6 items	
	Cluster 3: 1 items	
	Cluster 4: 5 items	
	Cluster 5: 2 items	
	Cluster 6: 4 items	
8	*Cluster 0: 3 items*	0.485
	Cluster 1: 8 items	
	Cluster 2: 2 items	
	Cluster 3: 1 items	
	Cluster 4: 5 items	
	Cluster 5: 12 items	
	Cluster 6: 38 items	
	Cluster 7: 126 items	
9	*Cluster 0: 135 items*	0.44
	Cluster 1: 284 items	
	Cluster 2: 6 items	
	Cluster 3: 4 items	
	Cluster 4: 40 items	
	Cluster 5: 10 items	
	Cluster 6: 126 items	
	Cluster 7: 74 items	
	Cluster 8: 21 items	
10	*Cluster 0: 284 items*	0.586
	Cluster 1: 23 items	
	Cluster 2: 135 items	
	Cluster 3: 6 items	
	Cluster 4: 10 items	
	Cluster 5: 74 items	
	Cluster 6: 126 items	
	Cluster 7: 17 items	
	Cluster 8: 4 items	
	Cluster 9: 21 items	

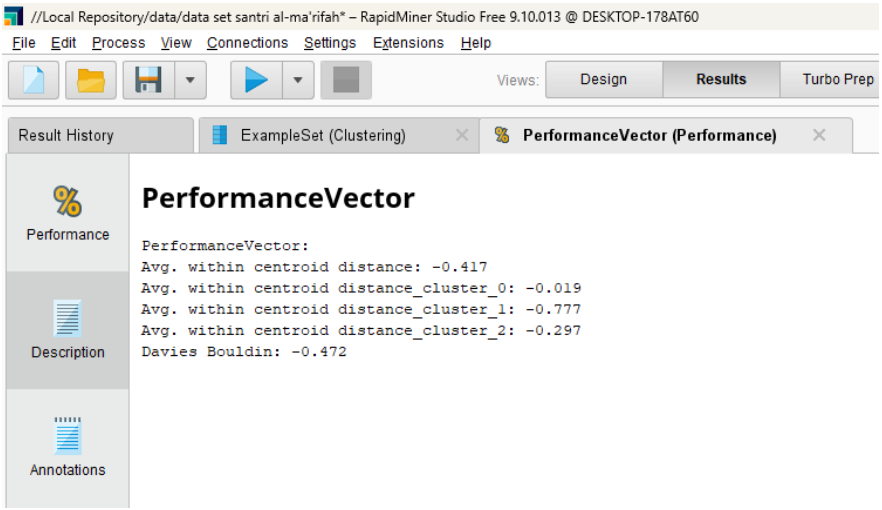
Berdasarkan tabel 2 Menunjukan bahwa *cluster* data pengelompokan jumlah santri menggunakan perhitungan *Davies Bouldin Index* nilai yang paling mendekati angka 0 dengan percobaan cluster 2 sampai *cluster* 10 menghasilkan nilai k terbaik pada *cluster* 3 yaitu 0.143 dengan jumlah anggota *Cluster* 0: 3 *items*, *Cluster* 1: 8 items, *Cluster* 2: 2 *items*.

3.5. *Intervetation atau evaluasion*

Dalam penelitian ini, peneliti menggunakan operator *Cluster Distance Performance*. Operator *K-means* merupakan operator *performance* untuk mengevaluasi kerja metode *clustering* berbasis *centroid* berperan dalam mengamati nilai *Davies Bouldin Index (DBI)* dan rata rata jarak dalam *cluster centroid* berdarkan jarak terdekat yang telah dikelompokkan oleh algoritma *K-Means*.berikut ini gambar *performance* dikelompokkan oleh algoritma *K-Means*.



Gambar 11. Operator *Cluster Distance Performance*



Gambar 12. Hasil *Performance Vector*

Pada hasil gambar 12 Avg.i di atas menunjukan :

1. *within centroid distance*: -0.078 Ini adalah rata-rata jarak antara titik-titik dalam semua kluster dengan centroidnya masing-masing. Nilai negatif mungkin menunjukkan bahwa titik-titik dalam kluster cenderung berada lebih dekat satu sama lain daripada dengan centroid.
2. *Avg. within centroid distance_cluster*_0: -0.201 Ini adalah rata-rata jarak antara titik-titik dalam kluster 0 dengan centroidnya. Nilai negatif menunjukkan bahwa titik-titik dalam kluster 0 cenderung berada lebih dekat satu sama lain daripada dengan centroid kluster 0.
3. *Avg. within centroid distance_cluster*_1: -0.055 Ini adalah rata-rata jarak antara titik-titik dalam kluster 1 dengan centroidnya. Nilai negatif menunjukkan bahwa titik-titik dalam kluster 1 cenderung berada lebih dekat satu sama lain daripada dengan centroid kluster 1.
4. *Avg. within centroid distance_cluster*_2: -0.024 Ini adalah rata-rata jarak antara titik-titik dalam kluster 2 dengan centroidnya. Nilai negatif menunjukkan bahwa titik-titik dalam kluster 2 cenderung berada lebih dekat satu sama lain daripada dengan centroid kluster 2.
5. *Avg. within centroid distance_cluster*_3: -0.000 Ini adalah rata-rata jarak antara titik-titik dalam kluster 3 dengan centroidnya. Nilai yang sangat mendekati nol mungkin menunjukkan bahwa titik-titik dalam kluster 3 sangat dekat dengan centroidnya.
6. Davies Bouldin: -0.379 *Davies-Bouldin Index* adalah metrik evaluasi klustering yang mengukur seberapa baik sebuah klustering telah dilakukan. Nilai negatif mungkin menunjukkan bahwa klustering yang dilakukan dianggap baik sesuai dengan metrik ini.

3.6. Hasil Analisis

Proses pengelompokkan data Pengelompokan santri berdasarkan wilayah menggunakan algoritma *K-Means*, selanjutnya yaitu mengetahui nilai k terbaik dari masing-masing *cluster* dari data pengelompokan tersebut, terakhir penulis menganalisa hasil dari data pengelompokkan tersebut dengan menggunakan algoritma *K-Means*. Pada hasil *clustering* diperoleh nilai k terbaik yaitu pada *cluster* 3 dan mendapatkan 3 kelompok terbaik pada

pengelompokkan data santri berdasarkan wilayah menggunakan algoritma *K-Means*. Berikut tabel hasil pengclusteran berdasarkan jumlah santri per wilayah yaitu:

Tabel 3. Anggota *Cluster* 0

No	*Cluster* 0
1	karawang = 0,828
2	indramayu = 0,192
3	Subang = 0,339
4	Majalengka = 0,585

Tabel 4. Anggota *Cluster* 1

No	*Cluster* 1
1	Bandung = -0,639
2	kuningan = -0,395
3	Jakarta = -0,835
4	Lampung = -0,835
5	Bogor = -0,737
6	Bekasi = -0,786
7	Purwakarta = -0,737
8	Berebes = -0,835

Tabel 5. Anggota *Cluster* 2

No	*Cluster* 2
1	Sumedang = 2.050
2	Cirebon = 1.806

Tabel 6. Hasil Perhitungan *Cluster* Menggunakan DBI

Hasil DBI	
Cluster	Nilai
2	0,571
3	0.143
4	0.786
5	0.425
6	0.495
7	0.462
8	0.485
9	0.441
10	0.586

Dari hasil nilai DBI di atas dapat disimpulkan bahwa nilai k yang mendekati optimum adalah k=3 dengan hasil nilai DBI 0.143. Pada k3 diperoleh pengelompokan sebanyak 3 kelompok dimana data pengelompokan wilayah tersebut berdasarkan. Jumlah santri *Cluster* 0 = 3 item

4. KESIMPULAN

Dalam kesimpulannya, pengelompokan santri berdasarkan wilayah dapat memengaruhi integrasi sosial dengan cara-cara yang beragam. Sementara identitas lokal dapat memperkuat solidaritas, perlu ada usaha untuk memastikan bahwa interaksi antar-kelompok tetap terbuka dan inklusif, sehingga mempromosikan pemahaman, toleransi, dan kerjasama di antara santri. Pembentukan kelompok tertutup dapat memperkuat identitas kelompok, yang dapat menciptakan rasa solidaritas dan kebersamaan di antara anggota kelompok. Penyediaan layanan konseling dan dukungan psikologis untuk santri yang menghadapi tantangan dalam beradaptasi atau merasa terisolasi adalah langkah yang sangat penting.

DAFTAR PUSTAKA

[1] P. Alkhairi and A. P. Windarto, "Penerapan K-Means Cluster pada Daerah Potensi Pertanian Karet Produktif di Sumatera Utara," *Semin. Nas. Teknol. Komput. Sains*, pp. 762–767, 2019.

[2] M. Benri, H. Metisen, and S. Latipa, "Analisis Clustering Menggunakan Metode K-Means Dalam Pengelompokkan Penjualan Produk Pada Swalayan Fadhila," *J. Media Infotama*, vol. 11, no. 2, pp. 110–118, 2015, [Online]. Available: https://core.ac.uk/download/pdf/287160954.pdf

[3] U. S, "Penerapan Data Mining Dengan Mengimplementasikan Algoritma K-Means Dalam Proses Clustering Untuk Pengelompokan Mahasiswa Calon Penerima Beasiswa KIP," *Build. Informatics, Technol. Sci.*, vol. 5, no. 1, 2023, doi: 10.47065/bits.v5i1.3411.

[4] D. Marlina, N. Lina, A. Fernando, and A. Ramadhan, "Implementasi Algoritma K-Medoids dan K-Means untuk Pengelompokkan Wilayah Sebaran Cacat pada Anak," *J. CoreIT J. Has. Penelit. Ilmu Komput. dan Teknol. Inf.*, vol. 4, no. 2, p. 64, 2018, doi: 10.24014/coreit.v4i2.4498.

[5] … Preddy, P. Marpaung, I. Pebrian, and W. Putri, "Penerapan Data Mining Untuk Pengelompokan Kepadatan Penduduk Kabupaten Deli Serdang Menggunakan Algoritma K-Means," *J. Ilmu Komput. dan Sist. Inf.*, vol. 6, no. 2, pp. 64–70, 2023.

[6] I. Nasution, A. P. Windarto, and M. Fauzan, "Penerapan Algoritma K-Means Dalam Pengelompokan Data Penduduk Miskin Menurut Provinsi," *Build. Informatics, Technol. Sci.*, vol. 2, no. 2, pp. 76–83, 2020, doi: 10.47065/bits.v2i2.492.

[7] F. S. Napitupulu, I. S. Damanik, I. S. Saragih, and A. Wanto, "Algoritma K-Means untuk Pengelompokkan Dokumen Akta Kelahiran pada Tiap Kecamatan di Kabupaten Simalungun," *Build. Informatics, Technol. Sci.*, vol. 2, no. 1, pp. 55–63, 2020, doi: 10.47065/bits.v2i1.323.

[8] H. Tusyakdiah *et al.*, "IMPLEMENTASI METODE K-MEANS DAN K- MEDOIDS PADA PENGELOMPOKAN PROVINSI INDONESIA BERDASARKAN ASPEK PENDIDIKAN PEMUDA," *Community Serv. Soc. Work Bull.*, vol. 3, no. 1, pp. 1–10, 2023, doi: 10.31000/cswb.v3i1.10153.

[9] R. Riadi and Mesran, "Penerapan Data Mining Menggunakan Algoritma K-Means Untuk Analisa Penjualan Parfume," *J. Informatics, Electr. Electron. Eng.*, vol. 2, no. 4, pp. 138–145, 2023, doi: 10.47065/jieee.v2i4.1181.

[10] S. Sindi, W. R. O. Ningse, I. A. Sihombing, F. I. R.H.Zer, and D. Hartama, "Analisis Algoritma K-Medoids Clustering Dalam Pengelompokan Penyebaran Covid-19 Di Indonesia," *J. Teknol. Inf.*, vol. 4, no. 1, pp. 166–173, 2020, doi: 10.36294/jurti.v4i1.1296.

[11] B. Imran, H. Hambali, A. Subki, Z. Zaeniah, A. Yani, and M. R. Alfian, "Data Mining Using Random Forest, Naïve Bayes, and Adaboost Models for Prediction and Classification of Benign and Malignant Breast Cancer," *J. Pilar Nusa Mandiri*, vol. 18, no. 1, pp. 37–46, 2022, doi: 10.33480/pilar.v18i1.2912.