

Penerapan Data Mining Penyaluran Dana Bantuan Sosial Menggunakan Metode K-Means Clustering di Desa Rancaekek Kulon

Suci Rainaya¹, Sofia Dewi²

^{1,2}Sistem Informasi, Universitas Ma'soem, Indonesia
sucirainaya321@gmail.com

Received : Aug' 2025 Revised : Aug' 2025 Accepted : Aug' 2025 Published : Aug' 2025

ABSTRACT

Data Mining simplifies and accelerates the process of distributing social assistance funds, namely direct cash assistance from village funds and has a significant impact. One of the causes of the problem is due to inappropriate and inaccurate data, so that assistance is often not on target and also the lack of limited access to technology to manage population data updates which causes social assistance not to reach groups of people who really need it. The purpose of this study is to apply Data Mining for the distribution of social assistance funds to the community in the Rancaekek Kulon Village Area and to find out the accuracy of the results of the application of data mining using Rapidminer Tools and also to increase the efficiency and speed in the process of distributing social assistance funds in Rancaekek Kulon Village. Data analysis and implementation with the K-Means Clustering algorithm was carried out by manual calculation using Microsoft Excel and then tested using RapidMiner software. The data used in this study is 526 data. The application of data mining using the K-Means Clustering algorithm can be used to group people who are eligible and unworthy of receiving assistance by producing 2 clusters, namely C0 (Not Feasible) and C1 (Feasible). The results obtained from the RapidMiner test were unfeasible community groups, namely C0 (Not Feasible) as many as 206 data and C1 (Feasible) as many as 320 data. Where the Davies Bouldin Index (DBI) was obtained from the formation of the cluster as much as 0.701 which means that the cluster is quite good because it is close to 0. Based on the calculation of data using Microsoft Excel, namely the Unfit Community group (C0) as many as 232 data and the Decent Community group (C1) as many as 294 data with the accuracy results obtained from the validation that has been carried out through the Davies Bouldin Index (DBI) validation on Microsoft Excel, which is 0.378.

Keywords : *Data Mining; K-Means Clustering; Social Assistance Fund.*

ABSTRAK

*Data Mining mempermudah dan mempercepat proses dalam kegiatan penyaluran dana bantuan sosial yaitu bantuan langsung tunai dari dana Desa dan memberikan dampak yang signifikan. Salah satu penyebab permasalahan yang ada dikarenakan adanya data yang tidak sesuai dan tidak akurat, sehingga bantuan sering kali tidak tepat sasaran dan juga kurangnya keterbatasan akses ke teknologi untuk mengelola pembaruan data penduduk yang menyebabkan bantuan sosial tidak sampai kepada kelompok masyarakat yang benar-benar membutuhkan. Tujuan dari penelitian ini adalah untuk menerapkan *Data Mining* penyaluran dana bantuan sosial kepada Masyarakat di Wilayah Desa Rancaekek Kulon dan mengetahui hasil akurasi dari penerapan *data mining* menggunakan *Tools Rapidminer* dan juga meningkatkan efisiensi dan kecepatan dalam proses penyaluran dana bantuan sosial di Desa Rancaekek Kulon. Analisis dan implementasi data dengan algoritma *K-Means Clustering* dilakukan dengan perhitungan manual menggunakan *Microsoft Excel* kemudian diuji menggunakan *software RapidMiner*. Data yang digunakan pada penelitian ini yaitu 526 data. Penerapan *data mining* menggunakan algoritma *K-Means Clustering* dapat digunakan*

untuk mengelompokkan Masyarakat yang layak dan tidak layak menerima bantuan dengan menghasilkan 2 *cluster* yaitu C0 (Tidak Layak) dan C1 (Layak). Hasil yang didapatkan dari pengujian *RapidMiner* terdapat kelompok masyarakat tidak layak yaitu C0 (Tidak Layak) sebanyak 206 data dan C1 (Layak) sebanyak 320 data. Dimana diperoleh *Davice Bouldin Index* (DBI) dari pembentukan *cluster* tersebut sebanyak 0,701 yang berarti *cluster* tersebut cukup baik dikarenakan mendekati 0. Berdasarkan perhitungan data menggunakan *Microsoft Excel* yaitu kelompok Masyarakat tidak layak (C0) sebanyak 232 data dan kelompok Masyarakat yang layak (C1) sebanyak 294 data dengan hasil akurasi yang didapat dari validasi yang telah dilakukan melalui validasi *Davies Bouldin Index* (DBI) pada *Microsoft Excel* yaitu 0,378.

Kata Kunci : *Data Mining*; Dana Bantuan Sosial; *K-Means Clustering*.

PENDAHULUAN

Kehidupan manusia berawal dari ketidakcukupan hingga saat ini semakin berkembangnya zaman yang maju dalam teknologi informasi dan komunikasi segala sesuatu bisa cepat dituntaskan dengan metode yang sederhana. Saat ini informasi tentang kemiskinan di Indonesia sedang marak dibicarakan oleh khalayak banyak orang, peranan pemerintah sangat berkaitan dalam upaya meningkatkan kesejahteraan rakyatnya dengan membuat berbagai program yang dapat mendukung kesejahteraan masyarakat dalam berbagai bidang seperti dalam bidang Pendidikan, Kesehatan dan Ekonomi yang bisa dijadikan sebagai tolak ukur kesejahteraan masyarakat. *Data mining* didefinisikan sebagai suatu dataset yang digunakan untuk mengeksplorasi data agar menentukan beberapa pola tersembunyi dari berbagai informasi yang bisa diprediksi kemungkinannya, maka dari itu peneliti menemukan peluang baru yaitu sebuah teknologi yang bisa menjawab beberapa kebutuhan.

Desa Rancaekek Kulon adalah Desa yang berada di Wilayah Kecamatan Rancaekek, Kabupaten Bandung, Provinsi Jawa Barat dengan Rukun Warga yang ada di Wilayah Desa Rancaekek Kulon sebanyak 13 Rw dan 63 Rt, total penduduk yang ada di Wilayah Desa Rancaekek Kulon ± 14.000 jiwa dengan 4.300 kepala keluarga dan rata-rata masyarakat di Wilayah Desa Rancaekek Kulon bekerja sebagai buruh.

Permasalahan utama khususnya di Desa Rancaekek Kulon ini dibuktikan dengan tingginya perbedaan pendapatan sehingga pihak desa mengadakan kegiatan bantuan sosial yang anggarannya langsung dari dana desa untuk para calon penerima bantuan sosial. Selain itu, sistem yang saat ini berjalan di Desa Rancaekek Kulon dalam kegiatan penyaluran dana bantuan sosial, data penduduk yang digunakan masih bergantung dari rekomendasi Rukun Tetangga dan Rukun Warga setempat kepada pihak desa, dengan menyesuaikan atribut para calon penerima apakah sudah memenuhi kriteria atau tidak, karena adanya data yang tidak sesuai dan tidak akurat, sehingga bantuan sering kali tidak tepat sasaran.

Saat ini masih ditemukan beberapa data penduduk yang kurang akurat sehingga menyebabkan ketidaksesuaian dalam penentuan penerima bantuan sosial dari dana Desa, kurangnya keterbatasan akses ke teknologi untuk mengelola pembaruan data penduduk yang menyebabkan bantuan sosial tidak sampai

kepada kelompok masyarakat yang benar-benar membutuhkan dan juga penyaluran bantuan sosial dari dana desa saat ini masih bergantung dari rekomendasi Rukun Tetangga dan Rukun Warga sehingga dapat memakan waktu dan kurang efisien.

Penelitian ini menggunakan metode *clustering* dengan algoritma *K-Means* dibantu dengan penggunaan *tools RapidMiner* yang diharapkan dapat membantu pihak desa dalam mengetahui kelompok Masyarakat yang layak dan tidak layak dalam penyaluran dana bantuan sosial dari dana desa. Selain itu juga diharapkan dapat meningkatkan efisiensi dan kecepatan dalam proses penyaluran dana bantuan sosial di Desa Rancaekek Kulon dengan menggunakan teknologi *data mining*.

METODE

Metode penelitian yang digunakan yaitu dengan menggunakan metode penelitian deskriptif kuantitatif. Digunakan untuk menggambarkan, memberikan penjelasan dengan hasil data dan fakta yang ada secara ringkas bagaimana mendeskripsikan suatu peristiwa atau fenomena tertentu pada objek penelitian. Adapun teknik pengumpulan data yang dilakukan penulis dalam penelitian ini meliputi:

1. Observasi; melakukan penelitian atau pengamatan secara langsung di Desa Rancaekek Kulon, mencatat kejadian yang berhubungan dengan skripsi, dan memperhatikan secara langsung mekanisme kerja khususnya yang berkaitan dengan kependudukan yang terjadi di bagian bidang pelayanan masyarakat.
2. Wawancara; dilakukan kepada bapak Taufik K Hermana sebagai sekretaris di kantor Desa Rancaekek Kulon untuk mengetahui terkait penyaluran dana bantuan sosial, proses pembagian penyaluran bantuan dana sosial, faktor-faktor penyebab kendala saat penyaluran dana bantuan sosial dan wawancara dilakukan juga kepada bapak Dadang Suherman bagian pelayanan untuk mendapatkan data *primer*.
3. Studi Pustaka; dilakukan studi pustaka untuk mengetahui data dan informasi yang berkaitan dengan penelitian melalui jurnal, skripsi, laporan dan buku. Seperti informasi dari Dinas Sosial mengenai penerimaan Bansos BLT, penelitian terdahulu dalam jurnal maupun skripsi yang berkaitan dengan data mining dan *algoritma K-Means Clustering*, buku-buku mengenai bantuan sosial, buku-buku mengenai teknik pengambilan data kuantitatif serta buku-buku terkait *data mining*.
4. Studi Dokumentasi; dilakukan untuk mendapatkan data secara langsung dari kantor Desa Rancaekek Kulon. Data yang didapatkan adalah data *primer* berupa data kependudukan pada tahun 2023. Data tersebut yang akan diolah dengan algoritma *K-means clustering* untuk menemukan pengelompokan penyaluran dana bantuan sosial.

Metode analisis data pada penelitian ini peneliti menggunakan algoritma *K-Means Clustering*. Adapun alur penerapannya sebagai berikut:

1. Identifikasi Masalah
-

Identifikasi masalah dilakukan untuk menemukan rumusan masalah yang akan diteliti oleh penulis. Permasalahan yang ada pada penelitian ini yaitu masih ditemukan beberapa data masyarakat yang masih kurang atau tidak sesuai untuk layak menerima bantuan sosial dari Desa.

2. Mempelajari Literatur

Dari beberapa permasalahan yang terjadi maka dilakukan pencarian dengan cara mempelajari jurnal-jurnal, skripsi, artikel yang berkaitan dengan pembahasan pada penelitian sebagai referensi untuk tujuan penulis menyelesaikan laporan penelitian ini.

3. Pengumpulan Data

Dilakukan dengan cara observasi yaitu mendatangi lokasi tempat penelitian dan meminta data penduduk kepada bagian pelayanan masyarakat. Data yang dikumpulkan merupakan data penduduk yang terdiri dari nama kepala keluarga, alamat, jenis kelamin, umur, agama, pendapatan, pekerjaan, status perkawinan dan tanggungan. Dari semua data yang telah terkumpul perlu diolah kembali sebelum diimplementasikan.

4. Pengolahan Data

Dikumpulkan dengan cara pengolahan data sesuai dengan tahapan *Knowledge Discovery in Database* (KDD). Kemudian data tersebut ditransformasikan dalam format yang sesuai dengan algoritma *K-Means Clustering*.

5. Seleksi Data, pada tahapan ini dilakukan untuk menghilangkan data yang tidak dibutuhkan pada saat pengolahan data, salah satunya seperti alamat, jenis kelamin dan agama, sehingga didapat data yang diambil yaitu nama kepala keluarga, usia, status perkawinan, pekerjaan, penghasilan dan tanggungan kepala keluarga.

6. Data *Cleaning* atau pembersihan data, pada tahapan ini digunakan untuk menghilangkan beberapa data yang duplikat atau sama. Dari data awal yang terkumpul sekitar 547 data penduduk. Kemudian berubah setelah dilakukan pembersihan data dari data yang duplikat yang akhirnya menjadi 526 data penduduk yang akan diolah ke tahapan berikutnya.

7. Data Transformasi ini merupakan sebuah proses untuk mengubah data dari struktur aslinya untuk di normalisasikan dari data berbentuk *numerik* ke dalam skala tertentu maka dibutuhkan tahapan transformasi data.

Analisis penduduk dilakukan berdasarkan data penduduk pada tahun dan 2023 untuk penerapan *data mining* dalam pembentukan *cluster* layak dan tidak layak dan dilakukan proses *pre-processing* data sehingga data akhir yang akan diolah berjumlah 526 data. Adapun atribut yang digunakan pada penelitian ini adalah sebagai berikut:

Tabel 1. Atribut Penelitian

No	Indikator	Bobot Penilaian
1	Usia Kepala Keluarga	1
2	Status Perkawinan	1
3	Pekerjaan	3
4	Penghasilan	3
5	Tanggungan	2

Berdasarkan tabel di atas, lima atribut yang digunakan untuk penerapan *data mining* dalam pembentukan *cluster* layak dan tidak layak akan dijelaskan sebagai berikut:

1. Usia Kepala Keluarga merupakan atribut sebagai acuan apakah kepala keluarga sudah lanjut usia atau masih muda.
2. Status Perkawinan merupakan atribut sebagai acuan apakah status dari kepala keluarga tersebut menikah atau sudah cerai.
3. Pekerjaan merupakan acuan untuk mengetahui tentang informasi terkait dengan pekerjaan dari kepala keluarga tersebut.

Berikut merupakan data yang digunakan dalam proses penelitian ini yaitu:

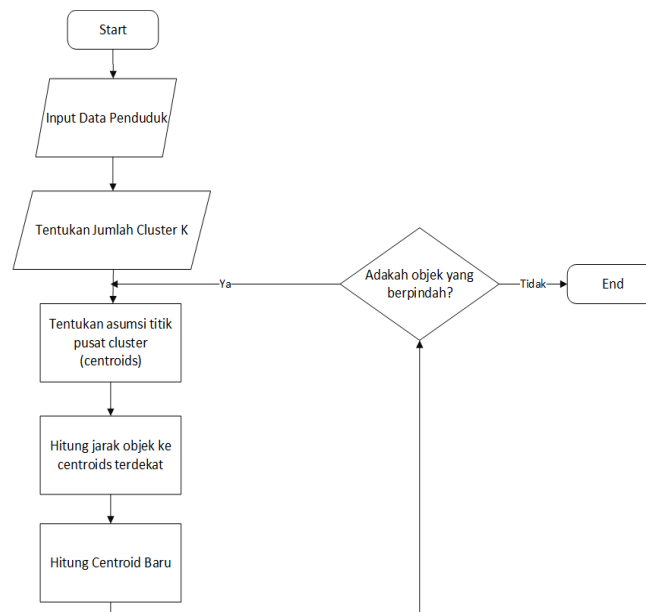
No	Nama Kepala Keluarga	Usia	Status Perkawinan	Pekerjaan	Penghasilan	Tanggungan
1	Entin Kartini	66	Cerai Mati	Mengurus Rumah Tangga	Rp500,000	2
2	Ujang Ahmad Solihin	40	Kawin Belum Tercatat	Belum/Tidak Bekerja	Rp -	2
3	Nani Rochani	75	Cerai Mati	Mengurus Rumah Tangga	Rp500,000	3
4	Jojo Warjo Permana	65	Kawin Belum Tercatat	Buruh Harian Lepas	Rp1,000,000	2
5	Ade Sutardi	58	Kawin Belum Tercatat	Buruh Harian Lepas	Rp1,000,000	4
s/d
545	Unduy Kosasih	72	Kawin Belum Tercatat	Buruh Harian Lepas	Rp1,000,000	3
546	Alan Jaelani	33	Kawin Belum Tercatat	Buruh Harian Lepas	Rp1,000,000	3
547	Marliah	81	Cerai Hidup	Karyawan Swasta	Rp3,500,000	3

Sumber: Hasil Peneliti

Gambar 1. Data Penduduk

K-means adalah algoritma mengelompokkan sebuah data yang digunakan untuk membagi kumpulan data menjadi beberapa kelompok yang unik. *K-means* dapat diartikan sebuah algoritma yang menentukan beberapa nilai dari *cluster* (k) secara acak yang dimana menjadi pokok utama yang disebut titik tengah suatu *cluster* atau *centroid* [1]. Algoritma *k-means* disusun dengan beberapa ide yang sederhana. Pertama *cluster* yang akan dibentuk harus ditentukan terlebih dahulu jumlah yang akan dibuat, untuk dijadikan sebagai *centroid* dan selanjutnya melakukan pengulangan langkah-langkah sampai terjadi kestabilan dan itu artinya berarti tidak ada objek yang dipindahkan. Sebuah objek data bisa terbilang ke dalam *cluster* jika memiliki jarak paling pendek atau paling dekat nilainya terhadap *centroid cluster*. *K-means* adalah algoritma pengelompokan yang mempunyai saran sederhana menggunakan cara *Sum of Squared Error* (SSE) antara beberapa objek data k *centroid* [2]. Langkah-langkah yang dilakukan dalam pengelompokan Masyarakat

yang layak dan tidak layak dalam penyaluran dana bantuan sosial dapat dideskripsikan pada gambar berikut:



Gambar 2. Alur Proses *K-Means Clustering*

HASIL DAN PEMBAHASAN

Penerapan algoritma *k-means clustering* dengan menggunakan *Microsoft Excel* yang kemudian diuji menggunakan *RapidMiner Studio*. *RapidMiner* merupakan sebuah *tools* yang dapat menganalisa sebuah setdata yang berukuran sangat besar[3]. *RapidMiner* menggunakan metode deskriptif dan prediktif untuk menyajikan informasi kepada pengguna sehingga mereka dapat membuat keputusan yang lebih baik. Pada penelitian ini juga digunakan untuk melakukan proses perhitungan *data mining* dengan metode *Clustering* dengan menggunakan algoritma *K-Means* terhadap data kependudukan di wilayah Desa Rancaekek Kulon. Adapun tahapan-tahapan *data mining* adalah *data selection*, *pre-processing/cleaning*, *data transformation*, *data mining*, *interpretation/evaluation*. Setelah melalui tahapan *data selection* sampai dengan *data transformation* data yang dihasilkan berjumlah 526 data.

Cluster yang akan dibentuk berjumlah 2 *cluster* yaitu *cluster 0* (Tidak Layak) dan *cluster 1* (Layak). Adapun data yang akan dihitung merupakan data dari kelima atribut yaitu usia kepala keluarga, status perkawinan, pekerjaan, penghasilan dan tanggungan.

NO	NKK	U	SP	PK	PH	T
1	ENTIN KARTINI	66	3	2	2	2
2	UJANG AHMAD SOLIHIN	40	1	1	1	2
3	NANI ROCHANI	75	3	2	2	3
4	JOJO WARJO PERMANA	65	1	3	3	2
5	ADE SUTARDI	58	1	3	3	4
6	MEMED JAENUDIN	71	1	3	3	3
7	APIP KUSUT	40	1	3	3	2
8	ENJANG RAHMAT	58	1	3	3	2
9	DIAN	38	1	3	3	3
10	DEDEN HERYANTO	45	1	9	9	3
11	YOYO SUPRIANA	78	1	3	3	5
12	YOYOH	67	3	2	2	3
13	AGUSTIAN	56	1	3	3	4
14	CARLAN	50	1	9	9	3
15	ASEP SENTANA PUTRA	41	1	9	9	3
16	AGUS KHAIDIR	56	1	9	9	3
17	DEDE MULYANA	50	1	1	1	3
18	ETI JUHAETI	66	2	2	2	3
19	IWAN SEHABUDIN	40	1	3	3	2
20	NONOK MARTINI	78	1	3	3	2

Gambar 3. Hasil Transformation Data Penduduk

Penerapan Algoritma K-Means

Adapun penerapan algoritma K-Means yaitu sebagai berikut :

1. Tentukan Jumlah Cluster

Langkah pertama yang dilakukan dalam perhitungan algoritma K-means clustering menggunakan Microsoft Excel yaitu menentukan berapa jumlah cluster yang akan digunakan. Pada penelitian ini peneliti menggunakan 2 cluster yaitu cluster 0 dan cluster 1.

2. Tentukan Centroid setiap Cluster

Pada saat penentuan centroid awal atau titik pusat cluster pada iterasi pertama dilakukan pemilihan secara acak untuk perhitungan iterasi pertama dan dengan menghitung nilai rata rata semua data yang ada pada masing-masing cluster untuk iterasi-iterasi selanjutnya. Berikut merupakan centroid awal yang dipilih:

Tabel 2. Centroid Awal

Centroid	Nama	U	SP	PK	PH	T
C0	HERI KUSAERI	52	1	10	10,00	3
C1	ROYASIH	74	2	2	2,00	2

3. Hitung Jarak Data dengan Centroid

Setelah menentukan centroid awal pada setiap cluster maka langkah berikutnya yaitu menghitung jarak data dengan centroid pada setiap cluster. Pada perhitungan ini peneliti menggunakan rumus Euclidean Distance, rumus ini digunakan untuk menghitung jarak terpendek dengan data yang telah ditentukan. Rumus tersebut adalah seagai berikut:

$$Cluster\ 0 = \sqrt{((data\ pertama\ U - centroid\ U\ C0)^2) + ((data\ pertama\ SP - centroid\ SP\ C0)^2) + ((data\ pertama\ PK - centroid\ PK\ C0)^2) + ((data\ pertama\ PH - centroid\ PH\ C0)^2) + ((data\ pertama\ T - centroid\ T\ C0)^2)}$$

$$Cluster\ 1 = \sqrt{((data\ pertama\ U - centroid\ U\ C1)^2) + ((data\ pertama\ SP - centroid\ SP\ C1)^2) + ((data\ pertama\ PK - centroid\ PK\ C1)^2) + ((data\ pertama\ PH - centroid\ PH\ C1)^2) + ((data\ pertama\ T - centroid\ T\ C1)^2)}$$

$$Cluster\ 0 = \sqrt{(40 - 52)^2 + (1 - 1)^2 + (1 - 10)^2 + (1,00 - 10,00)^2 + (2 - 3)^2} = 17,521$$

$$Cluster\ 1 = \sqrt{(58 - 74)^2 + (1 - 2)^2 + (3 - 2)^2 + (3,00 - 2,00)^2 + (4 - 2)^2} = 11,618$$

Gambar 9. Perhitungan Jarak Data dengan Centroid

4. Tentukan Centroid Baru

Setelah melakukan perhitungan jarak maka data akan terbagi ke dalam cluster 0 dan cluster 1 berdasarkan jarak terdekat data pada setiap cluster. Langkah selanjutnya yaitu menentukan centroid baru dengan menghitung nilai rata-rata semua data yang ada pada masing-masing cluster.

a. Iterasi 1 menghasilkan 393 data untuk cluster 0.

Tabel 3. Centroid Baru Cluster 0

Centroid Baru	U	SP	PK	PH	T
C0	0,1169	0,0020	0,0060	0,0060	0,0080

b. Iterasi 1 menghasilkan 133 data untuk cluster 1.

Tabel 4. Centroid Baru Cluster 1

Centroid Baru	U	SP	PK	PH	T
C1	85,676	2,0333	3,0333	3,0333	2,3333

Setelah centroid baru ditentukan, kemudian hitung kembali jarak data dengan centroid baru. Perhitungan ini menggunakan rumus yang sama dengan rumus pada langkah sebelumnya. Lakukan perhitungan ini sampai jumlah anggota cluster tidak mengalami perubahan atau hanya mengalami sedikit perubahan.

Validasi Davies Bouldin Index

Davies Bouldin Index (DBI) merupakan metode untuk mengevaluasi kinerja pengelompokan suatu data. Pada tahapan ini data yang akan diuji yaitu 526 data dimana centroid pada iterasi terakhir akan disesuaikan dengan cluster data. Adapun langkah-langkah untuk menghitung nilai DBI adalah sebagai berikut:

1. Menghitung Sum of Square Within-cluster (SSW)

Data yang akan diuji yaitu 526 data dimana centroid pada iterasi terakhir akan disesuaikan dengan cluster data.

Tabel 5. Centroid Iterasi Terakhir

Centroid	U	SP	PK	PH	T
C0	41	1,034	6,120	6,120	2,887
C1	63	1,363	4,704	4,704	2,663

Perhitungan jarak data *cluster* tersebut dilakukan dengan rumus sebagai berikut:

$$\text{Jarak Cluster} = \sqrt{(40 - 41)^2 + (1 - 1,034)^2 + (1 - 6,120)^2 + (1,00 - 6,120)^2 + (2 - 2,887)^2} = 7,365 \quad (1)$$

Dari 526 data yang terbagi menjadi 2 *cluster* kemudian akan dihitung rata-rata pada setiap *clusternya* seperti di bawah ini:

6,521	10,170
SSW0	SSW1

Gambar 4. Hasil Perhitungan SSW

2. Menghitung Sum of Square Between-cluster (SSB)

Sama halnya dengan menentukan SSW, pada perhitungan SSB menggunakan *centroid* terakhir untuk data uji validasi dikarenakan mempunyai kualitas *centroid* yang baik dan lebih stabil. Untuk itu perhitungan ini *centroid* pada iterasi terakhir digunakan. Perhitungan SSB dilakukan dengan rumus sebagai berikut:

$$SSB_{(0,1)} = \sqrt{(41 - 63)^2 + (1,034 - 1,363)^2 + (6,120 - 4,704)^2 + (6,120 - 4,704)^2 + (2,887 - 2,663)^2} = 22 \quad (2)$$

3. Menghitung Nilai Rasio

Tahapan berikutnya adalah menghitung rasio untuk mengetahui nilai perbandingan antara *cluster* ke-*i* dan *cluster* ke-*j*. Perhitungan ini menggunakan persamaan dengan rumus sebagai berikut:

$$R_{(i,j)} = \frac{SSW_i + SSW_j}{SSB_{(i,j)}} \\ R_{(0,1)} = \frac{6,521 + 10,170}{22} = 0,755 \quad (3)$$

Setelah nilai rasio didapatkan, kemudian hitung *Davies Bouldin Index* (DBI) yang didapatkan dari rumus berikut:

$$DBI = \frac{1}{2} \times 0,755 = 0,37 \quad (4)$$

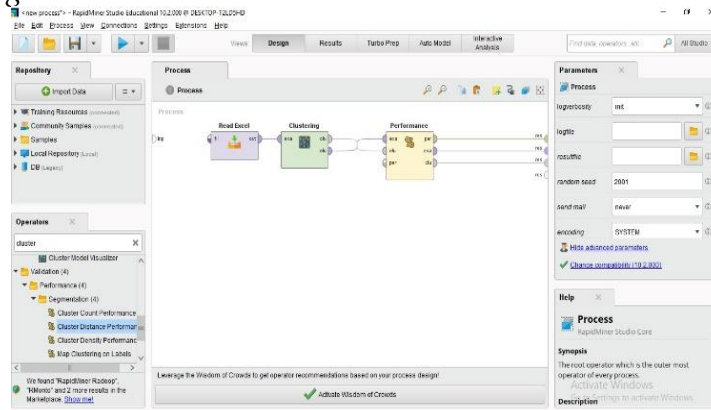
Ini berarti *cluster* yang dibentuk pada perhitungan ini dapat dikatakan baik atau optimal karena nilai DBI mendekati 0 yaitu 0,378.

Hasil Implementasi

Setelah melakukan perhitungan menggunakan *Microsoft Excel* selanjutnya melakukan pengujian pada *RapidMiner* untuk mengetahui akurasi dari *cluster* tingkat pemahaman yang telah terbentuk sebelumnya. Adapun langkah-langkahnya adalah sebagai berikut:

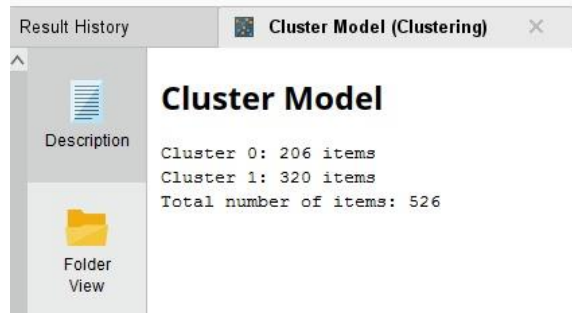
1. Buka *software RapidMiner*, kemudian pilih *Blank Process* untuk memulai proses pengujian.
2. Tambahkan operator *Read Excel* untuk membaca data *excel* yang akan digunakan dengan cara buka *Operators* - cari *Read Excel* - klik 2x pada *Read Excel*.

3. *Import* data yang akan digunakan. Klik 2x *Read Excel* – pilih data – klik *Next* – *select* kolom dan baris yang akan diproses – klik *next*. Kemudian sesuaikan tipe data yang akan diproses.
4. Kemudian tambahkan operator *K-Means* untuk membentuk *cluster* dan operator *Performance/Cluster Distance Performance* untuk menampilkan hasil dari pembentukan *cluster* tersebut. Atur jumlah *cluster* sesuai dengan *cluster* yang telah ditentukan sebelumnya yaitu 2 pada operator *clustering*. Setelah itu klik tombol *running*.



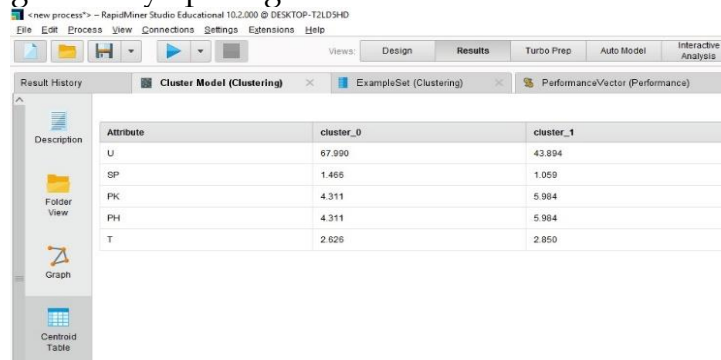
Gambar 5. Desain Clustering

5. Setelah proses sukses secara otomatis akan terlihat hasil *cluster* yang terbentuk dan informasi ringkas pada menu *results*.



Gambar 6. Hasil Cluster Model

Kemudian dapat dilihat nilai *centroid* yang secara otomatis dipilih untuk masing-masing *clusternya* pada gambar di bawah ini:



Gambar 7. Centroid Table

Pada halaman *ExampleSet* dapat terlihat data anggota masing-masing *cluster* yang terbentuk.

Row No.	id	NKK	cluster	U	SP	PK	PH	T
1	1	ENTIN KARTL...	cluster_0	96	3	2	2	2
2	2	UJIANG AHMA	cluster_1	40	1	1	1	2
3	3	NANI ROCHA	cluster_0	75	3	2	2	3
4	4	JOJO WARJO	cluster_0	95	1	3	3	2
5	5	ADE SUTARDI	cluster_0	58	1	3	3	4
6	6	MEMED JAE...	cluster_0	71	1	3	3	3
7	7	APIP KUSUT	cluster_1	40	1	3	3	2
8	8	ENJIANG RA...	cluster_0	58	1	3	3	2
9	9	DIAN	cluster_1	38	1	3	3	3
10	10	DEDEN HER...	cluster_1	45	1	9	9	3
11	11	YOYO SUPRI...	cluster_0	78	1	3	3	5
12	12	YOYOH	cluster_0	67	3	2	2	3
13	13	AGUSTIAN	cluster_0	56	1	3	3	4
14	14	CARLAN	cluster_1	50	1	9	9	3

Gambar 8. Data Anggota Cluster

Validasi Performance

Penerapan algoritma *K-Means* pada software *RapidMiner* akan menghasilkan validasi *performance* berupa *Davies Bouldin Index* (DBI) yang dideskripsikan pada halaman *PerformanceVector*.

Metric	Value
PerformanceVector	
Avg. within centroid distance	-87.020
Avg. within centroid distance_cluster_0	-112.387
Avg. within centroid distance_cluster_1	-70.690
Davies Bouldin	-0.701

Gambar 9. Validasi Performance

Pada gambar diatas menunjukkan nilai *Davies Bouldin Index* (DBI) dari proses *clustering* untuk kelompok Masyarakat penerimaan bantuan sosial dari dana desa sebesar 0,701 itu berarti *cluster* yang terbentuk cukup optimal dikarenakan mendekati 0.

PENUTUP

Berdasarkan penelitian yang telah peneliti lakukan mengenai penerapan *data mining* menggunakan metode *k-means clustering* untuk menentukan penerimaan

dana bantuan sosial dari dana desa dengan tepat sasaran kepada masyarakat yang layak menerima bantuan dapat disimpulkan bahwa:

1. Berdasarkan hasil yang didapat dalam perhitungan *Microsoft Excel* dan pengujian *RapidMiner*, yang dimana untuk *cluster 0* merupakan kelompok masyarakat yang tidak layak dan untuk *cluster 1* merupakan kelompok masyarakat yang layak, maka didapati bahwa kelompok masyarakat tidak layak yang dinyatakan dengan *cluster 0* berjumlah 232 data dan untuk kelompok masyarakat layak menerima bantuan dari dana desa yang dinyatakan dengan *cluster 1* berjumlah 294 data menurut perhitungan pada *Microsoft Excel*. Sedangkan untuk hasil *RapidMiner* yaitu *cluster 0* berjumlah 206 data dan *cluster 1* berjumlah 320 data.
2. Hasil akurasi yang didapat dari validasi yang telah dilakukan melalui validasi *Davies Bouldin Index (DBI)* pada *Microsoft Excel* yaitu 0,378 dan hasil validasi *Davies Bouldin Index* pada *RapidMiner* yaitu 0,701 dimana semakin kecil nilai yang didapat maka semakin baik dan optimal *cluster* yang terbentuk.
3. Melalui penerapan *data mining* ini memberikan gambaran kepada pihak Desa Rancaekek Kulon dalam kegiatan penerimaan bantuan kepada Masyarakat berdasarkan kategori yang ada sehingga bantuan sosial yang diberikan bisa tepat sasaran kepada penerimanya.

DAFTAR PUSTAKA

- [1] Vulandari, R. T. (2017). *Data Mining Teori dan Aplikasi Rapidminer*. Yogyakarta : Gavamedia.
- [2] Suyanto. (2017). *Data Mining Untuk Klasifikasi dan Klasterisasi Data*. Bandung : Informatika.
- [3] Wulandari, F., Jusia, P. A., & Jasmir, J. (2020). Klasifikasi Data Mining Untuk Mendiagnosa Penyakit ISPA Menggunakan Metode Naïve Bayes Pada Puskesmas Jambi Selatan. *Jurnal Ilmiah Mahasiswa Sistem Informasi*, 2(3), 214-227.