

Algoritma K-Means Untuk Klasterisasi Kampung Di Desa Bojong Berdasarkan Tingkat Pendidikan

Yovi Yuliantin¹, Ahmad Faqih², Kaslani³

STMIK IKMI CIREBON¹²³

yoviyuliantin18@gmail.com¹, ahmadfaqih367@gmail.com², kaslani@ikmi.ac.id³

Abstrak - Desa Bojong menghadapi tantangan dalam memetakan tingkat pendidikan penduduknya secara terstruktur. Meskipun data tersedia, kurangnya pengorganisasian menyebabkan kesenjangan akses pendidikan, terutama di wilayah-wilayah tertentu. Penelitian ini bertujuan untuk mengelompokkan tingkat pendidikan warga Desa Bojong menggunakan algoritma K-Means, sehingga dapat memberikan wawasan yang lebih mendalam tentang kondisi pendidikan masyarakat. Data yang digunakan mencakup 6.027 penduduk dari 10 kampung, dengan atribut seperti usia, pendidikan terakhir, pekerjaan, dan status pernikahan. Proses analisis mengikuti tahapan *Knowledge Discovery in Database* (KDD) dan dilakukan menggunakan perangkat lunak RapidMiner. Hasil penelitian menunjukkan bahwa jumlah klaster yang ideal adalah 7, dengan nilai terbaik 0,467 untuk Davies-Bouldin Index (DBI). Setiap klaster menunjukkan tingkat pendidikan tertentu. Cluster 6 memiliki tingkat pendidikan yang sangat tinggi, dengan banyak penduduk yang sampai perguruan tinggi. Sementara itu, Cluster 0 terdiri dari orang-orang yang hanya tamat SD atau bahkan tidak sekolah. Studi ini menunjukkan distribusi pendidikan di Desa Bojong. Hasil ini dapat membantu pemerintah desa membuat program pendidikan yang lebih baik. Kampung dengan tingkat pendidikan rendah dapat berkonsentrasi pada program yang meningkatkan akses pendidikan dasar, seperti literasi dan subsidi pendidikan, sedangkan kampung dengan tingkat pendidikan tinggi dapat berkonsentrasi pada pengembangan program pendidikan lanjutan atau pelatihan vokasional. Metode ini diharapkan dapat membantu Desa Bojong mengatasi kesenjangan pendidikan dan meningkatkan kualitas hidup warganya dengan menyediakan program yang tepat sasaran. Selain itu, penelitian ini membantu implementasi algoritma K-Means dalam pengelompokan data pendidikan di daerah pedesaan.
Kata Kunci : K-Means Clustering, tingkat pendidikan, klasterisasi, pendidikan desa, RapidMiner

Abstract - Bojong village faces challenges in mapping the education level of its residents in a structured way. Although data is available, the lack of organization leads to disparities in access to education, especially in certain areas. This study aims to cluster the education level of Bojong Village residents using the K-Means algorithm, so as to provide a deeper insight into the educational condition of the community. The data used includes 6, residents from 10 villages, with attributes such as age, latest education, occupation, and marital status. The analysis process followed the Knowledge Discovery in Database (KDD) stage and was conducted using RapidMiner software. The results showed that the ideal number of clusters was 7, with the best value of 0.467 for the Davies-Bouldin Index (DBI). Each cluster shows a certain level of education. Cluster 6 has a very high level of education, with many residents attending university. Meanwhile, Cluster 0 consists of people who only finished elementary school or did not even go to school. This study shows the distribution of education in Bojong Village. These results can help village governments create better education programs. Villages with low education levels can concentrate on programs that improve access to basic education, such as literacy and education subsidies, while villages with high education levels can concentrate on developing further education programs or vocational training. This method is expected to help Bojong Village overcome the education gap and improve the quality of life of its citizens by providing targeted programs. In addition, this research helps implement the K-Means algorithm in clustering education data in rural areas.)

Keywords: K-Means Clustering, education level, clustering, village education, RapidMiner

I. PENDAHULUAN

Bidang informatika yang berkembang pesat telah mengubah banyak aspek kehidupan, seperti teknologi, bisnis, dan pendidikan. Metode seperti data mining dan klasterisasi dari analisis data berbasis informatika memungkinkan identifikasi kebutuhan pendidikan masyarakat yang lebih akurat dalam pendidikan. Metode ini memungkinkan pengambilan yang lebih tepat dan terarah dari berbagai kebijakan strategis pendidikan. Pada

akhirnya, ini akan menghasilkan peningkatan kualitas pendidikan di daerah tertentu. Algoritma *K-Means*, misalnya, dapat memberikan wawasan tentang distribusi pendidikan dan membantu dalam menentukan program mana yang harus diprioritaskan di daerah pedesaan yang selama ini kurang terjangkau oleh program pendidikan formal.

Pendidikan merupakan faktor penting dalam pembangunan manusia, khususnya di pedesaan. Desa Bojong memiliki data

pendidikan yang tidak terorganisir dengan baik, yang menyebabkan kesenjangan pendidikan di beberapa wilayahnya. Penelitian ini bertujuan untuk menyediakan peta distribusi pendidikan yang lebih akurat dengan menggunakan algoritma K-Means, yang diharapkan dapat membantu pemerintah desa membuat program pendidikan yang lebih tepat sasaran

Penelitian sebelumnya seperti yang dilakukan oleh (Gustino & Suprati, 2024) di Kecamatan Krangkeng menunjukkan bahwa klasterisasi dengan K-Means memberikan hasil signifikan dalam pemetaan tingkat pendidikan masyarakat. Selain itu, (Setianingsih & Alli, 2023) menggunakan metode yang sama untuk klasterisasi pengeluaran desa dan menunjukkan bahwa K-Means efektif dalam mengelompokkan data tidak terstruktur. (Novitasari et al., 2023) menganalisis tingkat kemiskinan dengan menggunakan K-Means. Mereka menunjukkan bahwa teknik ini dapat menangani data yang sangat besar dan menghasilkan klaster yang representatif. Meskipun demikian, penelitian-penelitian tersebut belum fokus pada pemetaan pendidikan di wilayah pedesaan seperti Desa Bojong, sehingga penelitian ini bertujuan untuk melengkapi kesenjangan tersebut.

Penelitian ini bertujuan untuk mengidentifikasi dan mengelompokkan tingkat pendidikan penduduk di Desa Bojong menggunakan algoritma K-Means Clustering. Kontribusi utama dari penelitian ini adalah menyediakan peta distribusi pendidikan yang lebih akurat, yang dapat membantu pemerintah daerah dalam merancang program intervensi yang lebih efektif.

a. Data Mining

Proses pengolahan data menggunakan algoritma K-Means. Dalam menentukan jumlah cluster yang akan digunakan yaitu menggunakan Elbow Method. Kemudian akan dilakukan pengolahan data menggunakan hasil Elbow Method dan algoritma K-Means (Abdussalam Amrullah et al., 2022).

Beberapa penelitian telah menerapkan data mining dalam pendidikan. (Hasanah et al., 2021) menggunakan metode clustering untuk menentukan minat siswa terhadap mata pelajaran tertentu. Hasil penelitian mereka membantu sekolah dalam merancang program belajar yang lebih sesuai dengan kebutuhan siswa. Namun, penelitian ini masih terbatas pada pengelompokan minat dan tidak mempertimbangkan faktor sosial-ekonomi yang lebih luas.

Dalam konteks klasterisasi tingkat pendidikan, penelitian ini mengadopsi pendekatan serupa dengan mempertimbangkan faktor tambahan seperti

pekerjaan dan status pernikahan. Hal ini membedakannya dari penelitian sebelumnya yang hanya berfokus pada aspek pendidikan semata.

b. Clustering

Algoritma yang banyak digunakan dalam clustering adalah K-Means, yang bekerja dengan membagi data menjadi k kelompok berdasarkan jarak ke centroid yang ditentukan (Basalamah & Setyadi, 2023).

(Gustino & Suprati, 2024) mengelompokkan tingkat pendidikan masyarakat di Kecamatan Krangkeng menjadi tiga klaster. Penelitian ini menggunakan data dari kantor kecamatan dan menunjukkan bahwa klasterisasi dapat membantu perencanaan pendidikan daerah. Namun, penelitian ini tidak mempertimbangkan pekerjaan dan status pernikahan sebagai faktor yang dapat mempengaruhi pendidikan seseorang.

c. K-Means

K-means adalah salah satu algoritma clustering yang menggunakan metode partisi. K-means adalah algoritma clustering yang membagi masing-masing item data ke dalam satu cluster (R. Kurniawan et al., 2021).

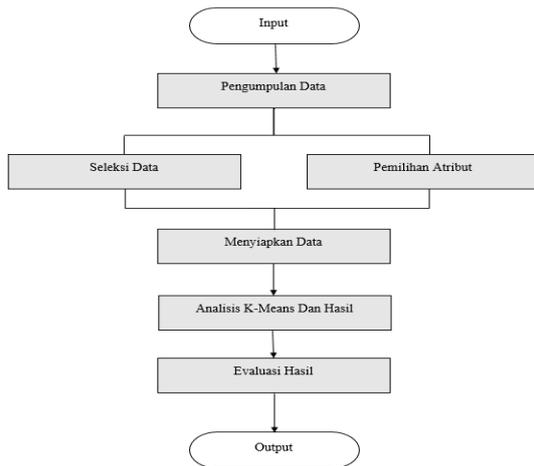
(Putri et al., 2024) Penelitian yang telah dilakukan, metode K-Means Clustering dapat mengelompokkan provinsi di Indonesia berdasarkan indikator pendidikan dengan cukup baik yang diukur dengan metode Silhouette Coefficient dengan struktur cluster standar dengan nilai 0.6308 dengan parameter yang sudah di seleksi sebelumnya. Selanjutnya, penentuan jumlah cluster terbaik dilakukan dengan melakukan percobaan menggunakan nilai K = 2 hingga K = 7.

d. RapidMiner

RapidMiner adalah perangkat lunak analisis data yang sering digunakan untuk data mining dan machine learning. Platform ini memungkinkan pengguna untuk melakukan analisis data secara visual tanpa perlu menulis kode (Saputro & Sucihermayanti, 2021).

II. METODOLOGI PENELITIAN

Melakukan klasterisasi data berdasarkan tingkat pendidikan, penelitian ini menggunakan pendekatan kuantitatif dan metode data mining, algoritma K-Means dipilih karena kemampuan untuk mengelompokkan data ke dalam klaster berdasarkan kesamaan fitur, seperti atribut pendidikan. Dalam penelitian data dan menganalisis data penulis menggunakan tahapan penelitian sebagai berikut:



Sumber : Olahan Penulis (2024)
Gambar 1. Metode penelitian

Dari Gambar 1 tentang metode penelitian dapat dijelaskan sebagai berikut :

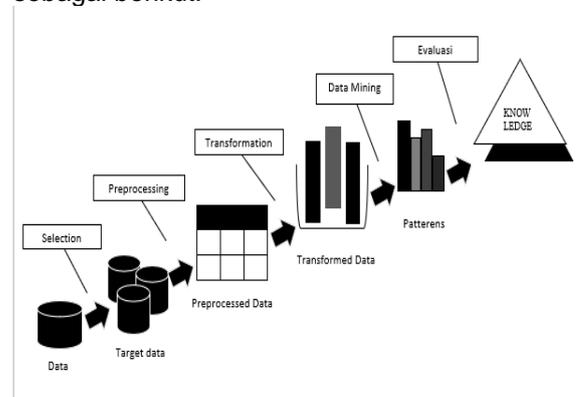
1. Pengumpulan Data
Mengumpulkan informasi pendidikan masyarakat kampung dari Kantor Desa Bojong.
2. Seleksi Data dan Pemilihan Atribut
Memastikan bahwa hanya data yang relevan dan berguna yang digunakan untuk analisis lebih lanjut
3. Mempersiapkan Data
Untuk menyediakan data untuk analisis, normalisasi atau standardisasi nilai diperlukan untuk data berbagai skala. Ini diperlukan agar algoritma K-Means dapat melakukan klusterisasi dengan baik.
4. Analisis K-Means dan Hasil
Algoritma K-Means digunakan untuk mengelompokkan data berdasarkan tingkat pendidikan. Penduduk Desa Bojong saat ini dikelompokkan ke dalam kluster berdasarkan kriteria tertentu, seperti tingkat pendidikan mereka. Sebelum ini, jumlah kluster k harus ditentukan
5. Evaluasi hasil
Keputusan dan tindakan proses data mining untuk nilai hasil Rapidminer dan DBI.

Data yang digunakan yaitu data yang diambil bersumber dari kantor Desa Bojong adalah data yang terkait dengan pengelompokan kampung di desa bojong berdasarkan tingkat Pendidikan. Dari 10 kampung dan 9 atribut diantaranya adalah nama, alamat, tempat tanggal lahir, usia, Pendidikan terakhir, pekerjaan dan status. Data ini langsung di ambil dari tangan pertama tanpa perantara .Peneliti mendapatkan data sekunder dengan izin untuk melihat data warga desa bojong. Penelitian ini mencakup semua warga Desa Bojong, dari anak-anak hingga orang dewasa,

dengan berbagai latar belakang pendidikan. Untuk menganalisis distribusi pendidikan dan menetapkan kebijakan yang diperlukan untuk meningkatkan kualitas pendidikan di desa. Dalam penelitian ini, metode purposive sampling digunakan karena peneliti ingin mengambil sampel berdasarkan kriteria tertentu, yaitu individu dengan latar belakang pendidikan yang jelas dan relevan untuk analisis kluster. Dengan menggunakan metode purposive sampling, dengan kriteria sebagai berikut : (1) Penduduk yang memiliki data pendidikan yang lengkap. (2) Minimal usia 6 tahun untuk memastikan data relevan dengan pendidikan formal dan (3) Jumlah data setiap kampung minimal 50 individu untuk memastikan klusterisasi memiliki distribusi yang representatif. Selain itu, sebelum dilakukan analisis, data mengalami proses normalisasi menggunakan operator multiply yang dilakukan uji dengan $k=3$ sampai $k=7$.

Pengumpulan data ini menggunakan data metode pengumpulan secara primer yang dimana peneliti memperoleh data langsung dari sumber yaitu dari kantor Desa Bojong. Tujuan utama observasi adalah untuk mengumpulkan data tingkat pendidikan dari setiap kampung di Desa Bojong dan kemudian disimpan dalam bentuk Excel.

Penelitian ini menggunakan algoritma K-Means untuk clustering. Tujuannya adalah untuk menyatukan data penduduk berdasarkan tingkat pendidikan di berbagai kampung di Desa Bojong. Dengan menggunakan teknik ini, diharapkan dapat ditemukan kelompok yang memiliki karakteristik pendidikan tertentu (Oktarian et al., 2020). Karakteristik-karakteristik ini dapat digunakan sebagai dasar untuk perencanaan dan intervensi pendidikan yang lebih baik. Mengubah data mentah menjadi informasi bermanfaat adalah tujuan dari proses KDD, yang terdiri dari sejumlah tahapan yang sistematis. Tahapan KDD dapat sebagai berikut:



Sumber: Olahan Penulis (2024)
Gambar 2. Proses KDD

Proses *Knowledge Discovery Database (KDD)* Sebagai Berikut:

- a. *Pre-processing*
Pre-processing/Cleaning Penyeleksian data yang bisa dipakai. Pembersihan data meliputi: menghapus duplikasi data, mengecek data inkonsisten dan membetulkan kesalahan data (Gustino & Suprpti, 2024).
- b. *Transformation*
Transformation ini yang dilakukan adalah mentransformasikan bentuk data yang belum memiliki entitas yang jelas kedalam bentuk data yang valid atau siap untuk dilakukan proses Data Mining (Setianingsih & Alli, 2023).
- c. *Data Mining*
Melakukan pengujian dengan menerapkan algoritma K-means dan tools Rapidminer untuk membentuk model cluster (Awaliah et al., 2024).
- d. *Evaluation*
Pola informasi yang dihasilkan dari proses data mining perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dari proses KDD yang disebut dengan interpretation. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesa yang ada sebelumnya (M. D. Kurniawan et al., 2023).
- e. *Knowledge*
Di mana gambaran teknik visualisasi dan pengetahuan digunakan untuk memberikan pengetahuan yang telah diberikan kepada user (Romli & Dewi Puspita, 2021).

III. HASIL DAN PEMBAHASAN

1. Hasil Penelitian

- a. *Data Selection*
Dengan menggunakan Algoritma K-Means, kami akan mengolah data pengelompokan kampung dari Desa Bojong, yang terdiri dari 6.027 data warga yang memiliki 9 atribut, yaitu nama, alamat, tempat tanggal lahir, usia, pendidikan terakhir, pekerjaan, dan status. Gamabaran lengkap dataset sebelum rekap ditunjukkan di bawah ini:

Tabel 1. Select data

No.	Nama	Alamat	Rt/Rw	Jenis Kelamin	Tempat Lahir	Tanggal Lahir	Agama	Pendidikan	Pekerjaan	Usia	Status	Status
1	Cahaya	Kp.Pasir H	3/1	Laki-Laki	Bandung	08/08/1967	Islam	SLT A/S ede raja y a	Karyawan	57	Suami	Kawin

2	Alhasanah	Kp.Pasir H u u t	3/1	Perempuan	Bandung	04/01/1968	Islam	Diploma IV/S trat a l			PNS	57	Istri	Kawin
3	Rifka Diani Ekas	Kp.Pasir H u u t	3/1	Perempuan	Bandung Bandung Bandung	27/12/1999	Islam	SLT A/S ede raja t			Mahasiswa	25	Anak	Belum Kawin
4	Anita Rahman	Kp.Pasir H u u t	4/1	Perempuan	Bandung	08/02/1989	Islam	SLT P/S ede raja t			Mengurus Rumah Tangga	36	Istri	Kawin
5	Detra N Hermansyah	Kp.Pasir H u u t	3/1	Laki-Laki	Bandung	12/02/2012	Islam	Tidak Sekolah			Belum Berkarya	13	Anak	Belum Kawin
--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
6	Asep Kusn	Kp.Cibisoro	3/10	Laki-Laki	Bandung	04/12/1972	Islam	Tamat Sd/ Sed erajat			Buruh Har	52	Suami	Kawin

6	F	Kp	3	Lak	Ba	07/	I	Belu	B	1	A	B
0	a	.Ci	/	i-	nd	07/	s	m	e	6	n	e
2	d	bi	1	Lak	un	20	l	Tam	l		a	l
7	l	so	0	i	g	08	a	at	i		k	u
.	i	ro					m	Sd/				m
	a	ro						Sed				
	l							eraj				
	A							at				
	k											
	b											
	a											
	r											

Sumber: Olahan Penulis (2024)

Tabel di atas menunjukkan gambar data lengkap yang diberikan oleh kantor desa Bojong dalam format Excel. Data ini belum direkap untuk membentuk kelompok tingkat pendidikan di desa Bojong. Dari data tersebut hanya perlu rekapan data dari pendidikan terakhir warga kampung di desa bojong saja saja dan hasilnya menjadi 11 atribut dan dapat dilihat pada gambar 4. di bawah.

Tabel 2. Hasil Rekapan Data

No.	Kampung	Belum sekolah	Tidak tamat SD	Tamat SD	SLTP	SLTA	Strata III	Strarata II	Diploma IV/Strata I	Diploma I/II	Akademi/Diploma III/S. Muda	Grand Total
1	kp.Pasir Huut (Rw 01)	61	41	47	72	94	3	2	23	7	1	451
2	kp.Durung Tengah (rw 02)	88	87	28	18	13	0	0	12	4	8	679
3	Kp.Pamepeuk (Rw 03)	127	47	69	89	78	0	0	87	7	1	626
4	kp. Durung Kidul 1 (Rw 04)	139	96	35	18	94	0	0	10	3	3	828
5	kp. Durung Kidul 2 (Rw 05)	81	69	36	08	37	0	0	30	0	0	582
6	kp. Cibisoro 1 (Rw 06)	63	77	04	46	54	0	0	02	1	1	547
7	kp. Cibisoro 2 (Rw 07)	97	86	36	22	28	0	0	01	3	3	793

8	kp.Cibisoro 3 (Rw 08)	63	44	26	16	36	8	0	0	5	1	1	560
9	kp. Cijauh (Rw 09)	48	48	18	18	02	81	0	0	2	0	2	461
10	kp. Cibisoro (Rw 10)	49	71	21	43	54	1	0	0	0	0	0	500
11	Grand Total	816	666	240	430	822	3	3	63	25	20	27	6027

Sumber: Olahan Penulis (2024)

b. *Preprocessing*

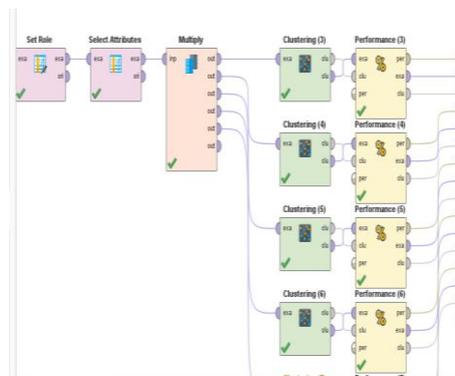
Dalam tahapan *Preprocessing* yaitu fungsinya adalah untuk memilih atribut mana saja yang digunakan dalam proses clustering data mining.

c. *Transformation*

Pada *Transformation* yaitu data diubah atau dimanipulasi untuk menyiapkannya agar lebih optimal dalam proses data mining.

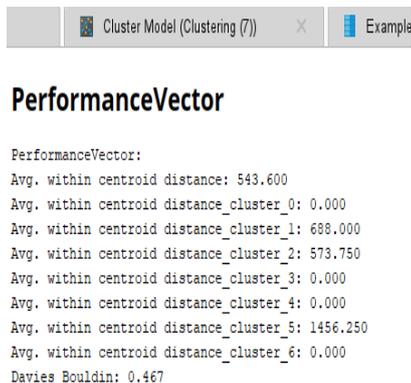
d. *Data Mining*

- 1) Operator *Performance*, data mining yaitu menggunakan algoritma k-means untuk mengelompokkan kampung didesa bojong berdasarkan tingkat pendidikan menggunakan Rapidminer.Operator *Performance* melibatkan proses pencarian nilai jarak dan nilai DBI untuk mengukur kinerja dari suatu sistem atau perangkat komunikasi secara efektif .
- 2) Proses pengolahan Data cluster yang dibentuk yaitu 7 cluster diantaranya Sangat Rendah (C0) Menengah Rendah (C1) Rendah (C2) Sedang (C3) Tinggi (C4) Menengah Tinggi (C5) dan Sangat Tinggi (C6).



Sumber: Olahan Penulis (2024)
Gambar 5. Operator performance

3) Setelah melalui serangkaian tahapan performa yang cermat dan teliti seperti analisis data, pengukuran kinerja dan evaluasi berbagai parameter yang relevan akhirnya ditemukan hasil rata-rata (Avg) yang mencerminkan tingkat efisiensi suatu sistem atau proses



Sumber: Olahan Penulis (2024)
Gambar 6. Hasil Avg. within centroid distance

4) *Evaluasi*
Selama proses pembuatan model, kumpulan data dikumpulkan untuk melakukan evaluasi model. Proses ini dilakukan untuk mendapatkan cluster terbaik dengan melihat hasil evaluasi DBI, yang merupakan indeks Davies Bouldien hasil evaluasi DBI adalah hasil nilai evaluasi paling kecil. Nilai kinerja dapat dilihat dan dievaluasi menggunakan model k-means, yang mencakup masing-masing algoritma k-means dari k=3 hingga k=7.

k	Jumlah Anggota cluster	Hasil Nilai DBI	Nilai Avg. Within Centroid Distance
3	Cluster 0: 5 items Cluster 1: 2 items Cluster 2: 3 items Total number of items: 10	0.805	3210.570
4	Cluster 0: 3 items Cluster 1: 2 items Cluster 2: 2 items Cluster 3: 3 items Total number of items: 10	0.841	2217.433
5	Cluster 0: 2 items Cluster 1: 2 items Cluster 2: 2 items Cluster 3: 2 items Cluster 4: 2 items Total number of items: 10	0.699	1416.350
6	Cluster 0: 2 items Cluster 1: 1 items Cluster 2: 2 items Cluster 3: 2 items Cluster 4: 1 items Cluster 5: 2 items Total number of items: 10	0.556	911.650
7	Cluster 0: 1 items Cluster 1: 2 items Cluster 2: 2 items Cluster 3: 1 items Cluster 4: 1 items Cluster 5: 2 items Cluster 6: 1 items Total number of items: 10	0.467	543.600

Sumber: Olahan Penulis (2024)
Gambar 7. Hasil Evaluasi DBI

Setelah semua tahapan selesai dan telah ditemukan cluster terbaik yang terbentuk yaitu pada k=7. Maka tergolong ada 7 cluster yaitu Sangat Rendah, Menengah Rendah, Rendah, Sedang, Tinggi, Menengah dan Sangat Tinggi. Dapat dilihat pada gambar berikut:

Kampung	cluster ↑
kp. Cibusoro ...	cluster_0
kp.Pasir Huut..	cluster_1
kp. Cijauh (R...	cluster_1
kp. Durung Ki...	cluster_2
kp. Cibusoro ...	cluster_2
Kp.Pamemp...	cluster_3
kp. Cibusoro (...)	cluster_4
kp.Durung Te...	cluster_5
kp.Cibusoro 3...	cluster_5
kp. Durung Ki...	cluster_6

Sumber: Olahan Penulis (2024)
Gambar 8. Hasil Clustering

2. Pembahasan
a. Mengimplementasikan algoritma K-Means pada masyarakat desa Bojong

Penerapan Algoritma K-Means dalam pengelompokan kampung desa Bojong berdasarkan tingkat pendidikan menggunakan Rapidminer pada data pendidikan kampung di Desa Bojong, telah terjadi kemajuan besar dan berhasil di implementasikan dalam menemukan pengelompokan pendidikan masyarakat yang dibantu dengan tahapan KDD dengan periapan data menggunakan data 6.237 penduduk dari 10 kampung Data diproses melalui set role untuk menentukan ID yang akan digunakan (Nama Kampung) dan select atribut tingkat pendidikan seperti Tidak sekolah, Tamat Sd, Tamat SLTP, Tamat SLTA, Diploma (I, II, III), Strata I, Strata II dan Strata III. kemudian diolah menggunakan RapidMiner dengan nilai k bervariasi untuk menentukan jumlah cluster optimal. Penelitian ini berhasil menentukan bahwa 7 kelompok adalah jumlah yang ideal untuk mengelola kampung berdasarkan tingkat pendidikan dengan menggunakan DBI untuk mengevaluasi kualitas kelompok. Hasilnya dapat digunakan untuk membuat program pendidikan yang lebih sesuai dengan kebutuhan masing-masing kelompok. Dengan evaluasi Nilai DBI semakin kecil nilai DBI maka semakin optimal cluster yang terbentuk. Pengelompokkan yang

optimal yaitu K=7 dengan niai DBI 4,67 terbentuk menjadi 7 cluster yaitu cluster 0 1 kampung (kp. Cibisoro 02 Rw.07) ,cluster 1 2 kampung (kp.Pasir Huut Rw.01 dan kp. Cijauh Rw.09) ,cluster 2 2 kampung (kp Durung Kidul 2 Rw.05 dan Kampung Cibisoro 1 Rw.06) , cluster 3 1 kampung (kp. Pamempeuk Rw.03) ,cluster 4 1 kampung (kp. Cibisoro Rw.10) , cluster 5 2 kampung kp.Durung Tengah Rw.02 dan Kp.Cibisoro 3 Rw.08) dan cluster 6 1 kampung (kp. Durung Kidul 1 Rw.04) dengan semua total menjadi 10 kampung. Dari Cluster 7 paling tinggi dan cluster 0 paling rendah.

Tahapan ini mengikuti model penelitian sebelumnya oleh (Gustino & Suprapti, 2024) yang menerapkan k-means untuk menganalisis tingkat pendidikan di kecamatan krangkeng. Studi ini menunjukkan bahwa klasterisasi memberikan hasil yang signitif dalam memahami pola dsitribusi pendidikan untuk mendukung perencanaan intervensi.

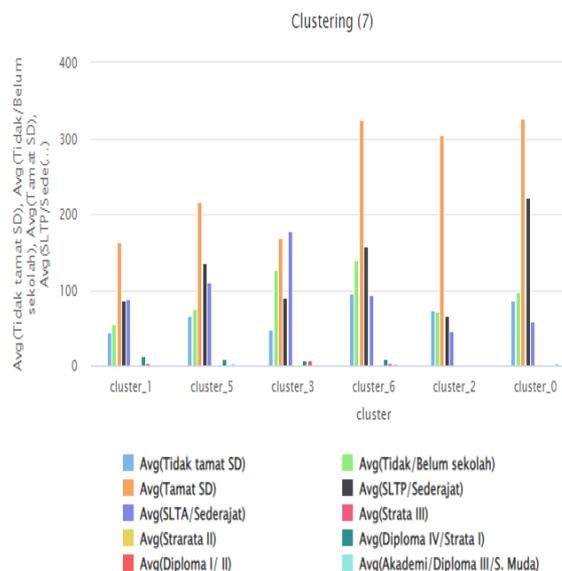
b. Menentukan klaster tingkat pendidikan yang terbentuk dari hasil analisis

Berdasarkan hasil uji dengan nilai K=3 sampai K=7 bahwa nilai Davies-Bouldin Index (DBI) terkecil yaitu diperoleh pada K=7 dengan nilai DBI 0.467 menunjukkan bahwa cluster yang terbentuk lebih optimal, semakin kecil nilai hasil DBI semakin optimal DBI yang terbentuk. Oleh karena itu, K=7 dipilih sebagai nilai optimal untuk pengelompokkan kampung berdasarkan tingkat pendidikan.

K	DBI
3	0.805
4	0.841
5	0.699
6	0.556
7	0.467

Sumber: Olahan Penulis (2024)
 Gambar 9. Hasil DBI

Dari tabel hasil niai DBI dapat dilihat cluster yang optimal terdapat pada K=7 dengan nilai DBI terkecil yaitu 0,467 semkain rendah nilai DBI maka semakin baik cluster yang terbentuk. Dengan K=7 data terbagi menjadi 7.



Sumber: Olahan Penulis (2024)
 Gambar 10. Grafik Hasil Cluster

Dari gambar Hasil Visualisasi diatas dapat disimpulkan bahwa grafik yang diperoleh data bahwa ada beberapa karakteristik dalam data penelitian ini.

- 1) Cluster 0 memiliki tingkat pendidikan yang sangat rendah, didominasi Tamat SD dan memiliki sedikit di tingkat SLTA ke atas.
 - 2) Cluster 1 memiliki tingkat pendidikan menengah rendah, masih didominasi Tamat SD tetapi sedikit lebih tinggi dari pada Cluster 0.
 - 3) Cluster 2 memiliki tingkat pendidikan menengah rendah, didominasi Tamat SD dan memiliki distribusi yang lebih merata dibandingkan Cluster 0 dan 1.
 - 4) Cluster 3 memiliki tingkat pendidikan menengah, dengan variasi yang lebih dengan peningkatan pada SLTA dan Strata I dan peningkatan pada tidak tamat SD dan tamat SD.
 - 5) Cluster 4 terdiri dari pendidikan tinggi, didominasi Tamat SD dan SLTA, dengan sedikit representasi pendidikan tinggi.
 - 6) Cluster 5 terdiri dari pendidikan menengah tinggi, didominasi Tamat SD, tetapi ada yang mencapai SLTA dan Strata I.
 - 7) Cluster 6 memiliki tingkat pendidikan sangat tinggi dengan rata-rata tertinggi dan distribusi merata di semua tingkatan, menunjukkan populasi yang lebih beragam dan cenderung lebih berpendidikan dibandingkan dengan klaster lain.
- Klasterisasi ini sejalan dengan penelitian (Setianingsih & Alli, 2023)

oleh , yang menggunakan K-means untuk klasterisasi pengeluaran desa. Hasilnya menunjukkan bahwa k-means efektif dalam menangkap pola dan karakteristik data multidimensi.

c. Menganalisis Hasil klasterisasi untuk memberikan gambaran distribusi tingkat pendidikan yang dapat digunakan sebagai dasar pengambilan kebijakan pendidikan di Desa Bojong

Hasil analisis klasterisasi menunjukkan bahwa tingkat pendidikan masyarakat Desa Bojong bervariasi, yang dapat digunakan sebagai dasar untuk membuat kebijakan pendidikan. Kelompok 0 dan 1 menunjukkan tingkat pendidikan sangat rendah, didominasi oleh orang yang hanya tamat SD atau bahkan tidak sekolah. Cluster 2 hingga 4 menunjukkan tingkat pendidikan menengah, dengan sebagian besar penduduk menyelesaikan SLTA, Cluster 5 dan 6 menunjukkan distribusi pendidikan menengah yang sangat tinggi, dengan beberapa orang mencapai Strata I dan II. Mereka juga dapat menawarkan akses teknologi pendidikan dan program pelatihan vokasional di desa yang lebih maju. Hasil penelitian ini memberi gambaran dasar kampung mana yang diperhatikan secara priotas untuk meningkatkan pemerataan pendidikan di Desa Bojong.

Berdasarkan pada model yang disampaikan oleh (Novitasari et al., 2023) yang menggunakan klasterisasi K-means untuk mengidentifikasi tingkat kemiskinan di Jawa Barat. Pendekatan berbasis data ini menunjukkan efektivita intervensi yang spesifik dan terarah.

IV. KESIMPULAN

Dengan menggunakan tahapan KDD dan alat bantu RapidMiner, algoritma K-Means berhasil digunakan untuk mengelompokkan sepuluh kampung di Desa Bojong berdasarkan tingkat pendidikan. Hasil pemeriksaan menunjukkan bahwa nilai ideal untuk jumlah cluster adalah $K = 7$ $K=7$, yang menghasilkan nilai terkecil dari Davies-Bouldin Index (DBI), yaitu 0,467. Pengelompokan ini membantu mengidentifikasi distribusi tingkat pendidikan secara lebih rinci, karena setiap kelompok mewakili pola tertentu pada tingkat pendidikan masyarakat.

Hasil analisis menunjukkan bahwa cluster tersebut memiliki kualitas terbaik, dengan nilai DBI terendah ditemukan pada $K = 7$ $K=7$. Setiap kelompok terdiri dari tujuh cluster, dengan cluster mulai dari sangat rendah

(Cluster 0) hingga sangat tinggi (Cluster 6). Setiap kelompok menunjukkan perbedaan besar dalam distribusi pendidikan. Ini dapat menjadi informasi penting untuk membuat rencana peningkatan pendidikan desa.

Sebagai hasil dari analisis klasterisasi, tingkat pendidikan di Desa Bojong bervariasi. Kampung dengan tingkat pendidikan sangat rendah (Cluster 0 dan 1) harus memprioritaskan peningkatan akses ke pendidikan dasar. Sebaliknya, kampung dengan tingkat pendidikan menengah hingga tinggi (Cluster 5 dan 6) dapat menjadi model untuk pengembangan fasilitas pendidikan lanjutan, termasuk teknologi pendidikan dan pelatihan vokasional. Hasil penelitian ini digunakan sebagai dasar strategis untuk membuat program pendidikan di Desa Bojong lebih seragam dan sesuai dengan kebutuhan masyarakat.

V. REFERENSI

- Abdussalam Amrullah, Intam Purnamasari, Betha Nurina Sari, Garno, & Apriade Voutama. (2022). Analisis Cluster Faktor Penunjang Pendidikan Menggunakan Algoritma K-Means (Studi Kasus: Kabupaten Karawang). *Jurnal Informatika Dan Rekayasa Elektronik*, 5(2), 244–252. <https://doi.org/10.36595/jire.v5i2.701>
- Awaliah, L., Rahaningsih, N., & Dana, R. D. (2024). Implementasi Algoritma K-Means Dalam Analisis Cluster Korban Kekerasan Di Provinsi Jawa Barat. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(1), 188–195. <https://doi.org/10.36040/jati.v8i1.8332>
- Basalamah, A. T., & Setyadi, R. (2023). Penerapan Algoritma K-Means Clustering Pada Tingkat Penyelesaian Pendidikan Di Provinsi Indonesia. *04(02)*, 114–121.
- Febriansyah, F., & Muntari, S. (2023). Penerapan Algoritma K-Means untuk Klasterisasi Penduduk Miskin pada Kota Pagar Alam. *JISKA (Jurnal Informatika Sunan Kalijaga)*, 8(1), 66–77.
- Gustino, & Suprapti, T. (2024). Analisis Desa Di Kecamatan Krangkeng Berdasarkan Tingkat Pendidikan Menggunakan Algoritma K-Means. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(2), 1861–1868. <https://doi.org/10.36040/jati.v8i2.9220>
- Hasanah, M., Defit, S., & Nurcahyo, G. W. (2021). Implementasi Algoritma K-Means untuk Klasterisasi Peserta Olimpiade Sains Nasional Tingkat SMA. *Jurnal Sistim Informasi Dan Teknologi*, 3(1), 30–35. <https://doi.org/10.35134/jsisfotek.v1i3.6>
- Kurniawan, M. D., Priyatna, B., & Nurapriana,

- F. (2023). Implementasi Algoritma K-Means Untuk Klasterisasi Data Obat Puskesmas Kotabaru. *Jurnal Sains Komputer & Informatika (J-SAKTI)*, 7(2), 882–890.
<https://tunasbangsa.ac.id/ejurnal/index.php/jsakti> Implementasi
- Kurniawan, R., M. Mukarrobil, M. M., & Mahradianur, M. (2021). Klasterisasi Tingkat Pendidikan Di Dki Jakarta Pada Tingkat Kecamatan Menggunakan Algoritma K-Means. *Technologia: Jurnal Ilmiah*, 12(4), 234.
<https://doi.org/10.31602/tji.v12i4.5633>
- Novitasari, N., Nuris, N. D., & Herdiana, R. (2023). Penerapan Algoritma K-Means untuk Clustering Data Jumlah Penduduk Miskin Berdasarkan Kota/Kabupaten di Jawa barat menggunakan Rapidminer. *Jurnal Informatika Terpadu*, 9(1), 68–73.
<https://doi.org/10.54914/jit.v9i1.660>
- Oktarian, S., Defit, S., & Sumijan. (2020). Clustering Students' Interest Determination in School Selection Using the K-Means Clustering Algorithm Method. *Jurnal Informasi Dan Teknologi*, 2(3), 68–75.
<https://doi.org/10.37034/jidt.v2i3.65>
- Putri, M. R., Nugraha, G. S., & Dwiyanaputra, R. (2024). *Pengelompokan Provinsi di Indonesia Berdasarkan Indikator Pendidikan Menggunakan Metode K-Means Clustering*. 7(1), 1–8.
- Romli, I., & Dewi Puspita, F. R. (2021). Penerapan Data Mining Menggunakan Algoritma K-Means Untuk Klasifikasi Penyakit Ispa. *Indonesian Journal of Business Intelligence (IJUBI)*, 4(1), 10.
<https://doi.org/10.21927/ijubi.v4i1.1727>
- Saputro, D. T., & Sucihermayanti, W. P. (2021). Penerapan Klasterisasi Menggunakan K-Means untuk Menentukan Tingkat Kesehatan Bayi dan Balita di Kabupaten Bengkulu Utara Daniel. *Jurnal ABDIMAS-KU: Jurnal Pengabdian Masyarakat Kedokteran*, 12(2), 146–155.
<https://doi.org/10.30659/abdimasku.1.3.126-133>
- Setianingsih, I., & Alli, I. (2023). *Klasterisasi Pengeluaran Kas Di Desa Pamengkang Menggunakan Metode K-Means*. 7(1), 784–787.