

ANALISIS DESA DI KECAMATAN KRANGKENG BERDASARKAN TINGKAT PENDIDIKAN MENGGUNAKAN ALGORITMA K-MEANS

Gustino, Tati Suprapti

Teknik Informatika, STMIK IKMI Cirebon
Jl. Perjuangan No. 10B, Karya Mulya Kec. Kesambi, Kota. Cirebon
gustinosaputra08@gmail.com

ABSTRAK

Tingkat pendidikan dapat menjadi indikator kemajuan suatu negara karena dari pendidikan dapat menciptakan inovasi-inovasi dari hasil ilmu pengetahuan. Pemetaan tingkat pendidikan suatu daerah memiliki peran yang sangat penting dalam mengidentifikasi tantangan dan peluang pendidikan yang perlu diatasi dan dimanfaatkan. Pemetaan tingkat pendidikan dapat memberikan gambaran tentang kondisi dan distribusi masyarakat yang memerlukan penanganan pendidikan yang lebih. Pemetaan ini juga membantu dalam perencanaan jangka panjang. Sumber data dari penelitian ini bersumber dari data penduduk di Kecamatan Krangkeng Tahun 2023. Sumber data yang didapat berupa sumber data set. Sumber data set didapat melalui kantor Kecamatan Krangkeng Kabupaten Indramayu. Yang sudah mendapat izin oleh Kepala Kantor Kecamatan Krangkeng. Metode penelitian yang digunakan adalah metode deskriptif dengan pendekatan kuantitatif. Penelitian kuantitatif ini merupakan penelitian yang di dasarkan pada dataset yang dianalisis menggunakan analisis deskriptif untuk menghitung jumlah numerik setiap kategori dan menggunakan algoritma k-means untuk *clustering*. *Clustering* K-Means mengelompokan data dari tingkat pendidikan di kecamatan krangkeng dan membaginya ke dalam tiga klaster desa. Hasil dari pengujian klastering K-Means Desa-desa di Kecamatan Krangkeng berdasarkan tingkat pendidikan menggunakan Rapid miner dihasilkan kategori 0 Kapringan, Srengseng, Krangkeng, Dukuhjati kategori 1 Kalianyar, Kedungwungu, Purwajaya, Singakerta, Tegal Mulya, Tanjakan Kategori 2 Luwung Gesik.

Kata Kunci: *K-Means*, Pemetaan, *Clustering*, Rapid Minner

1. PENDAHULUAN

Tingkat pendidikan dapat menjadi salah satu indikator kemajuan suatu negara karena pendidikan dapat menciptakan inovasi dari hasil ilmu pengetahuan. Inovasi-inovasi tersebut dapat membawa perubahan positif di berbagai sektor seperti ekonomi, teknologi, kesehatan dan lingkungan. Tingkat pendidikan yang merata dapat memperkuat fondasi pembangunan suatu negara. Selain itu, tingkat pendidikan yang sama juga menciptakan masyarakat yang lebih produktif dan kompetitif. Individu yang memiliki akses terhadap pendidikan berkualitas cenderung memiliki kesempatan kerja yang lebih baik dan dapat memberikan kontribusi yang lebih efektif terhadap pertumbuhan ekonomi.

Pemetaan tingkat pendidikan di suatu daerah mempunyai peranan yang sangat penting dalam mengidentifikasi tantangan dan peluang pendidikan yang perlu diatasi dan dimanfaatkan. Pemetaan tingkat pendidikan dapat memberikan wawasan mengenai kondisi dan sebaran masyarakat yang memerlukan perlakuan pendidikan lebih banyak. Pemetaan ini juga membantu perencanaan jangka panjang. Dengan memahami tren dan pola pendidikan di suatu wilayah, pemerintah dan lembaga pendidikan dapat merancang strategi untuk memenuhi kebutuhan masa depan. Hal ini termasuk mengembangkan program pelatihan bagi masyarakat. Secara keseluruhan, pemetaan pendidikan merupakan alat penting untuk mengelola kebijakan pendidikan dan alokasi sumber daya. Hal ini membantu menciptakan masyarakat yang terdidik,

terampil dan berkelanjutan yang akan membentuk masyarakat yang siap menghadapi tantangan. Pemetaan pendidikan memberikan pemahaman mengenai kondisi pendidikan di suatu daerah, kesenjangan pendidikan, dan kebutuhan masyarakat.

Rata-rata lama sekolah warga di Kabupaten Indramayu hanya mencapai 5,9 tahun. Ini artinya rata-rata pendidikan mereka tak sampai lulus SD. Rendahnya kesadaran warga akan pentingnya pendidikan menjadi salah satu penyebabnya. Hal itu terlihat dari data BPS Provinsi Jawa Barat, sebagaimana tercantum dalam website <https://jabar.bps.go.id>. Dalam website itu disebutkan rata-rata lama sekolah warga Kabupaten Indramayu pada 2018 hanya 5,98 tahun. Berdasarkan dari Data Tingkat Pendidikan Kepala Keluarga (KK) se Kecamatan Krangkeng Kabupaten Indramayu.[1], ing dibutuhkan dataset untuk dilakukannya pengujian. Dataset yang di dapat dari website data open jabar dari halaman (<https://opendata.jabarprov.go.id>). Dengan jumlah dataset 27 dari rata-rata setiap atribut dari tahun 2018-2020 dan 28 atribut di antaranya No. Kabupaten Kota, Anak Terlantar, Balita Terlantar, Anak Berhadapan dengan Hukum, Anak Jalanan, Anak dengan Kedisabilitas, Anak Korban Tindak Kekerasan, Anak yang memerlukan Perlindungan Khusus, Lanjut Usia Terlantar, Penyandang Disabilitas, Tuna Susila, Gelandangan, Pengemis, Kelompok Minoritas (Waria), Pemulung, Bekas Warga Binaan, Lembaga Pemasasyarakatan, Orang dengan HIV/AIDS (Odha), Korban Penyalahgunaan

Napza, Korban Trafficking, Korban Tindak Kekerasan, Pekerja Migrasi bermasalah Sosial, Perempuan Rawan Sosial Ekonomi, Fakir Miskin/Rumah Tangga Miskin, Keluarga Bermasalah Sosial Psikologi, Komunitas Adat Terpencil,[2] Tujuan penelitian ini adalah untuk memetakan tingkat pendidikan di Kecamatan Krangkeng Kabupaten Indramayu Provinsi Jawa Barat. Hasil yang ingin dicapai melalui penelitian ini adalah untuk menggolongkan penyebaran tingkat pendidikan pada setiap Desa Di Kecamatan Krangkeng. Desa-desa dengan tingkat pendidikan yang rendah dapat menjadi fokus program peningkatan pendidikan. Penelitian ini juga dapat menjadi acuan bagi pemerintah daerah untuk mengambil langkah-langkah yang tepat dalam meningkatkan tingkat pendidikan masyarakat di Kecamatan Krangkeng.

Proses pengelompokan diperlukan untuk membantu pemerintah menentukan daerah mana yang memerlukan dukungan berupa dukungan pendidikan.

Untuk itu penulis mengusulkan metode clustering dengan menggunakan algoritma K-Means terhadap jumlah lulusan dari berbagai jenjang pendidikan, mulai dari SD hingga SMP, dan dari SMA hingga Universitas. Metode clustering menyederhanakan proses pengelompokan dan pembagian data tingkat pendidikan tiap desa menjadi tiga cluster. Berdasarkan permasalahan di atas, maka perlu dilakukan penentuan sebaran tingkat pendidikan di desa-desa wilayah Kecamatan Klanken. Pemahaman yang lebih mendalam mengenai distribusi ini akan memberikan masukan bagi rancangan kebijakan pendidikan yang lebih efektif untuk mendistribusikan pendidikan secara adil di sub-kabupaten Klanken County. Penulis mengusulkan judul penelitian pemetaan desa di Kecamatan Klanken berdasarkan tingkat pendidikan dengan menggunakan algoritma K-means. Penelitian ini berfokus pada Pemetaan Desa. Maka judul penelitian yang dilakukan adalah "Pemetaan Desa Di Kecamatan Krangkeng Berdasarkan Tingkat Pendidikan Menggunakan Algoritma K-means". Dalam penelitian ini metode yang digunakan yaitu metode *clustering* dengan menggunakan algoritma K-Means dengan nilai k sebanyak 3 kluster yaitu tinggi, sedang dan rendah. Nilai k tersebut mewakili kualitas pendidikan di wilayah tersebut dan sekaligus akan menjadi bantuan kepada pemerintah dalam mempertimbangkan wilayah mana yang memerlukan bantuan. Setelah berhasil didapatkan jumlah anggota pada tiap *cluster*, selanjutnya akan divisualisasikan persebaran data sesuai *cluster* yang ditentukan dengan plot pada software Rapid Miner. Selanjutnya akan terlihat daftar keanggotaan dari masing-masing *cluster* yang ditentukan.

2. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Penelitian yang dilakukan oleh [3] Pada penelitian bertajuk "Pengelompokan Jenjang Pendidikan di DKI Jakarta Tingkat Kecamatan

Menggunakan Algoritma K-Means", hasil pengelompokan jumlah lulusan dari berbagai jenjang pendidikan mulai dari SD, SMP, dan SMA. dijelaskan. SMP/Sederajat, SMA/Sederajat, Diploma I/II, Diploma III, Diploma IV/Strata I, Strata II, Strata III. Penelitian yang dilakukan oleh

Penelitian yang dilakukan oleh [4] Judul penelitian "Analisis klasterisasi K-means pada pengelompokan provinsi/kota di Nusa Tenggara Timur berdasarkan indikator pendidikan" memaparkan hasil analisis tersebut. Rephrase Terdapat tiga variabel untuk mengelompokkan provinsi/kota di Nusa Tenggara Timur berdasarkan indikator pendidikan: angka partisipasi kasar (X_1), angka partisipasi murni (X_2), dan angka partisipasi murni (X_3). Sebelum menjalankan pengujian kluster K-means, data Anda harus memenuhi asumsi klasik pengelompokan. yaitu uji KMO untuk menentukan syarat kecukupan sampling, dan uji multikolinearitas untuk menguji apakah ada variabel independen yang mempunyai kemiripan antar variabel independen lainnya. Data tersebut memenuhi dua asumsi yang ada. Data dapat dilanjutkan dengan pengujian cluster K-Means.

Standarisasi data dilakukan setelah data memenuhi dua prasyarat. Standarisasi data dilakukan apabila terdapat perbedaan satuan yang besar antar variabel yang diteliti. Terlihat bahwa variabel yang digunakan mempunyai satuan yang berbeda-beda. Oleh karena itu, penulis terlebih dahulu ingin melakukan standarisasi variabel yang digunakan sebelum mengolah datanya. Hasil pengelompokan data kemiskinan dengan menggunakan K-means adalah kluster wilayah yang tingkat kemiskinannya berada di bawah risiko atau C1 yaitu Kabupaten Xiamis, Kabupaten Sumedang, Kabupaten Purwakarta, Kabupaten Pangandaran, Kota Bogor, Kota Sukabumi, Kota Bandung; Diperoleh informasi tentang Cirebon Kota dan Depok. Kota, Kota Cimahi, Kota Tasikmalaya, Kota Banjar.[5] Clustering mengacu pada pengelompokan mencatat, mengamati atau memperhatikan dan menggambar kelas objek serupa. cluster merupakan kumpulan rekaman yang mempunyai kesamaan satu sama lain dan berbeda dari entri di cluster lain. Clustering berupaya mempartisi seluruh koleksi kepada kelompok yang mempunyai informasi yang relatif sedikit kesamaan dimana kesamaan catatan bersama kelompok mempunyai nilai tertinggi namun memiliki kesamaan dengan piring dari kelompok lain nilainya kecil Clustering dalam data mining berguna menemukan pola distribusi dalam kumpulan data berguna dalam proses analisis data[6]. Dari hasil 15 kali pengujian yang dilakukan dengan menggunakan 3 cluster data pengujian yang berbeda setiap pengujiannya, iterasi paling sedikit berhenti di iterasi ke-4 pada pengujian 1 dengan data uji 2,87, 100 dan pada pengujian 2 dengan data uji 1, 6, 9. Dengan iterasi yang paling banyak berhenti di iterasi ke9 pada pengujian ke-5 dengan data uji 57, 66, 84. Hasil analisis performansi k-means dari 15 pengujian dari

setiap uji coba yang dilakukan, diperoleh nilai rata-rata Presicion sebesar 76%, nilai Recall sebesar 76% dan Accuracy sebesar 81%. [7] Penelitian ini membahas metode K-means untuk mengelompokkan wilayah sebaran. Sumber data penelitian ini dikumpulkan berdasarkan dokumen informasi energi yang disiapkan oleh Direktorat Jenderal Bea dan Konsumen Data yang digunakan dalam penelitian ini adalah dari tahun 2011 hingga 2017 yang mencakup 34 negara bagian. Variabel yang digunakan sebanyak sebaran berdasarkan wilayah.

Metode yang digunakan untuk menyelesaikan kasus ini menggunakan teknik data mining k-means. Data diolah menggunakan software RapidMiner sehingga diperoleh dua cluster yaitu cluster dengan tingkat varian tinggi (C1) sebesar dan cluster dengan tingkat variance rendah (C2). [8] Dataset memiliki 495 data provinsi yang missing value. Data tersebut dihapus dari dataset, kemudian memilih atribut yang sesuai dengan pemodelan. Dataset juga dilakukan normalisasi. Data yang digunakan yaitu total kasus, total meninggal, total sembuh, total kasus aktif. Menambahkan peluang kematian (%) dan peluang kesembuhan (%). [9] Pada tahap ini dilakukan proses utama yaitu segmentasi pengelompokan data kependudukan. Berikut merupakan penerapan algoritma k-means dengan asumsi bahwa parameter input adalah jumlah dataset sebanyak n data dan jumlah inisialisasi centroid k adalah 2. Data yang diambil untuk penelitian berjumlah 823 data. [10] Ada berbagai jenis algoritma clustering, seperti k-means dan k-medoid. K-Means Algoritma K-means merupakan algoritma clustering yang mengelompokkan data berdasarkan pusat cluster (centroid) yang paling dekat dengan data. Ciri lainnya adalah mengelompokkan data dengan cara memaksimalkan kesamaan data dalam cluster dan meminimalkan kesamaan data antar cluster. Menurut penelitian yang dilakukan (Wu et al., 2008) dan (Macqueen, 1967), pengelompokan data menggunakan K-means cepat dan mudah.

Di sisi lain, k-Medoids menggunakan medoid (median), yang menentukan objek terjauh dari pusat cluster untuk menentukan perbedaan antara setiap objek, daripada menentukan nilai rata-rata objek dalam cluster sebagai titik referensi. Minimalkan jumlahnya. titik acuan Keuntungan clustering memungkinkan peneliti membuat perbandingan menggunakan dataset yang sama. Hal ini juga sangat penting ketika menetapkan kebijakan di perusahaan atau organisasi Anda. Karena dapat menentukan keberlangsungan suatu organisasi atau perusahaan. Studi ini membandingkan kumpulan data yang terdiri dari hasil survei terhadap 4.444 pelanggan minuman teh di Thailand. [11]

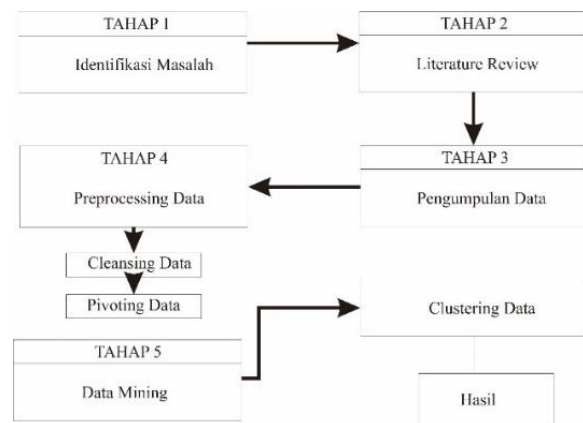
2.2. K-Means

K-Means merupakan suatu metode pengelompokan data menjadi dua kelompok atau lebih. Algoritma K-means merupakan teknik analisis kelompok yang membagi objek penelitian menjadi

kelompok. Setiap subjek yang diamati termasuk dalam kelompok data dengan nilai rata-rata yang berdekatan satu sama lain. Secara opsional, K digunakan untuk total konstanta pengelompokan dan mean berarti rata-rata kumpulan data. Oleh karena itu, K-means clustering merupakan teknik analisis data yang menggunakan sistem partisi untuk mengelompokkan data. K-means mengelompokkan data yang berupa satu atau lebih cluster, atau kelompok data yang mempunyai karakteristik yang sama, menjadi cluster, sehingga menghasilkan kelompok data non-hierarki dimana data yang sama dan karakteristik kelompok yang berbeda membentuk kelompok yang berbeda.

3. METODE PENELITIAN

Metode penelitian yang digunakan adalah metode deskriptif dengan pendekatan kuantitatif. Penelitian kuantitatif ini merupakan penelitian yang didasarkan pada dataset yang dianalisis menggunakan analisis deskriptif untuk menghitung jumlah numerik setiap kategori dan menggunakan algoritma k-means untuk clustering. Dalam upaya mengumpulkan data dan menganalisis data penulis menggunakan tahapan penelitian sebagai berikut:



Gambar 1. Metode Penelitian

Dari Gambar 1 tentang metode penelitian dapat dijelaskan sebagai berikut:

- a. Identifikasi Masalah
Melakukan observasi secara langsung mengenai masalah yang terjadi Di Kecamatan Krangkeng terutama Tingkat Pendidikan.
- b. Literature Review
Mencari referensi dari artikel atau jurnal terdahulu dalam jangka waktu 5 tahun kebelakang.
- c. Pengumpulan Data
Mengumpulkan data yang dibutuhkan untuk penelitian berupa dataset.
- d. Data Preprocessing
Tugas dari prapemrosesan data adalah untuk meningkatkan kualitas data dan memastikan bahwa data tersebut sesuai untuk tugas spesifik yang ada.

e. Data Mining

- Cleansing Data (pembersihan) pada tahap ini dataset yang berisi berbagai data hanya diambil nama, tingkat pendidikan, dan desa
- Pivoting Data, tahap bertujuan untuk menghitung data tingkat pendidikan pada setiap desa
- Analisis Deskriptif pada bagian ini meringkas data hasil pivoting sehingga berkembang dan memenuhi kondisi data yang dibutuhkan untuk analisis data selanjutnya
- Clustering algoritma k-means, penerapan algoritma k-means untuk mengelompokkan data menjadi kelompok-kelompok berdasarkan kesamaan.

3.1. Sumber Data

Sumber data dari penelitian ini bersumber dari data penduduk di Kecamatan Krangkeng Tahun 2023. Sumber data yang didapat berupa sumber data set. Sumber data set didapat melalui kantor Kecamatan Krangkeng Kabupaten Indramayu. Yang sudah mendapat izin oleh Kepala Kantor Kecamatan Krangkeng.

3.2. Populasi dan Sampel

Populasi dalam penelitian ini adalah seluruh masyarakat Desa yang ada di Kecamatan Krangkeng yang teradata. Metode pengambilan sampel yang digunakan dalam penelitian ini ialah Quota Sampling, yaitu metode Populasi dibagi menjadi kelompok berdasarkan karakteristik tertentu, dan jumlah sampel diambil dari setiap kelompok berdasarkan kuota tertentu. Yang mana dalam penelitian ini terdapat 2 atribut, yaitu berdasarkan nama desa, tingkat pendidikan. (Tidak/Belum Sekolah, Tidak Tamat SD/Sederajat, Tamat SD/Sederajat, SLTP/Sederajat, SLTA/Sederajat, Diploma I/II, Diploma III, Diploma IV/Strata I, Strata II, Strata III

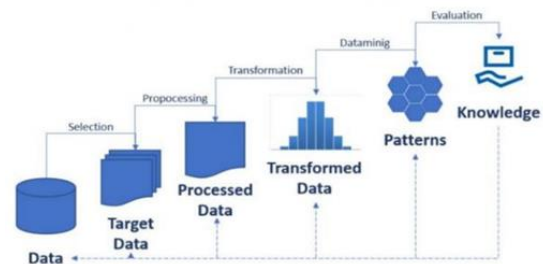
3.3. Teknik Pengumpulan Data

Data sekunder teknik pengumpulan data di dapatkan dari observasi yang sudah dilakukan meliputi:

- Mendatangi kantor Kecamatan Krangkeng untuk dimintai data penduduk masyarakat Kecamatan Krangkeng.
- Hasil dari pengumpulan data lalu disimpan dalam bentuk excel.

3.4. Teknik Analisa Data

Metode analisis data penelitian menggunakan proses KDD dan menggunakan algoritma K-means untuk clustering. Penggunaan proses KDD bekerja sedemikian rupa sehingga tahapan-tahapan yang dilakukan berlangsung secara sistematis.



Gambar 2. Proses KDD

Proses Knowledge Discovery Database (KDD) secara umum dapat dijelaskan sebagai berikut:

a. Data Selection

Sebelum tahap data mining, perlu dilakukan terlebih dahulu tahap pemilihan data dengan menyajikan data asli. Data yang telah dipilih disimpan dalam file terpisah dari database operasional. Pre-processing/Cleaning
Penyeleksian data yang bisa dipakai. Pembersihan data meliputi: menghapus duplikasi data, mengecek data inkonsisten dan membetulkan kesalahan data.

b. Transformastion

Proses transformasi informasi terpilih ke dalam wujud mining procedure, sehingga informasi tersebut cocok buat proses informasi mining ialah proses kreatif serta sangat bergantung pada tipe ataupun pola data yang hendak dicari dalam basis informasi.

c. Data Mining

Berbagai macam metode digunakan dalam tahap ini untuk mengekstrak pola-pola potensial menghasilkan data yang berguna. Data mining ialah proses mencari pola ataupun informasi menarik dalam data terpilih dengan menggunakan metode ataupun tata cara tertentu. Metode, tata cara, ataupun algoritma dalam data mining sangat bermacam-macam. Pemilihan metode ataupun algoritma dalam informasi mining sangat bermacam-macam. Pemilihan tata cara ataupun algoritma yang pas sangat tergantung pada tujuan serta proses KDD secara totalitas.

d. Evaluation

Dalam tahap ini, pola pola yang telah diidentifikasi dari proses data mining butuh ditampilkan dalam wujud yang mudah untuk dipahami oleh pihak yang berkepentingan.

e. Knowlegde

Tahap paling akhir dalam KDD adalah data-data yang sudah diproses divisualisasikan agar lebih mudah dipahami oleh pengguna dan diharapkan bisa diambil tindakan berdasarkan analisis. Presentasi data mining yang mudah dipahami oleh semua orang adalah tahapan yang diperlukan dalam data mining.

4. HASIL DAN PEMBAHASAN

4.1. Hasil Penelitian

4.1.1. Data Selection

Pada tahap ini dilakukan data selection tahapan ini bertujuan untuk memilih data yang akan digunakan dalam proses Analisis melalui Microsoft Excel pada proses penelitian menggunakan data penduduk Desa Kexamatan Krangkeng Kabupaten Indramayu Tahun 2023 yang digunakan dalam proses data mining menganalisis penduduk berdasarkan pendidikan.

NAMA	PDDK AKHIR	NAMA KEL
A.D.ROHIM	Tamat SD/Sederajat	KRANGKENG
A.INATUL LAYYINAH	Tamat SD/Sederajat	KRANGKENG
A.JAMALUDIN AL MAKBUL	Tamat SD/Sederajat	KRANGKENG
A.KHOERUL MUBAROK	Tamat SD/Sederajat	KRANGKENG
A.MAULANA	Tamat SD/Sederajat	KRANGKENG
AAM AMINAH	Tamat SD/Sederajat	KRANGKENG
AAN ANISAH	Tamat SD/Sederajat	KRANGKENG
AAN KURNIAWAN	Tamat SD/Sederajat	KRANGKENG
AB. KODIR BIN WARMAN	Tamat SD/Sederajat	KRANGKENG
ABD.AJIZ	Tamat SD/Sederajat	KRANGKENG
ABD.RAHMAN	Tamat SD/Sederajat	KRANGKENG
ABDUL AZIZ	Tamat SD/Sederajat	KRANGKENG
ABDUL BARI BIN SARWAN	Tamat SD/Sederajat	KRANGKENG
ABDUL BASIT	Tamat SD/Sederajat	KRANGKENG
ABDUL BASIT	Tamat SD/Sederajat	KRANGKENG
ABDUL BASIT BIN H. SHOL	Tamat SD/Sederajat	KRANGKENG
ABDUL CARIM BIN WARIYAN	Tamat SD/Sederajat	KRANGKENG
ABDUL DULIMAN	Tamat SD/Sederajat	KRANGKENG

Gambar 3. Data Selection

4.1.2. Pivoting Data

Pivoting data adalah proses mengubah data dari format panjang ke format lebar dan sebaliknya. Proses ini umum terjadi dalam analisis data dan sering digunakan untuk mengatur ulang data guna pemahaman dan yang lebih dalam visualisasi. Metode spesifik untuk memutar data bergantung pada alat atau bahasa pemrograman yang digunakan.

	Tidak/Belum Sekolah	Tidak Tamat SD/Sederajat	Tamat SD/Sederajat	Strata III	Strata II	SLTP/Sederajat	SLTA/Sederajat	Diploma IV/Strata I	Diploma III	Diploma II	Akademik/Diploma III/S. Muda	Grand Total
DUKUHATI	1.777	544	4.833	0	5	1.394	488	744	13	23	9.821	
KALIANYAR	1.114	609	3.534	0	4	1.284	1.082	358	11	81	8.077	
KAPRINGAN	1.469	604	5.184	0	2	1.111	441	448	9	30	9.298	
KEDUNGWUNGU	1.227	471	3.235	2	7	1.492	714	504	20	24	7.696	
KRANGKENG	1.530	652	4.872	1	2	1.318	759	489	10	42	9.675	
LUWUNGSESIK	524	192	1.457	0	2	545	311	185	3	17	3.236	
PURWAJAYA	1.067	339	4.155	0	1	1.023	467	528	8	35	7.623	
SINGAKERTA	1.133	427	3.705	1	4	1.035	609	433	19	44	7.410	
SRENGENG	1.946	846	5.453	0	6	1.450	743	582	14	59	11.099	
TANJAKAN	882	364	2.964	1	3	820	410	283	9	25	5.761	
TEGALMULYA	1.257	502	3.352	0	0	1.383	519	596	8	24	7.641	
Grand Total	13.926	5.550	42.744	5	36	12.855	6.543	5.150	124	404	87.337	

Gambar 4. Pivoting Data

4.1.3. Data Transformation

pada tahap ini melibatkan proses transformasi data ke dalam format yang sesuai untuk diproses dalam data mining. Tujuannya untuk memudahkan penyesuaian algoritma dan alat yang digunakan dalam penelitian, khususnya data yang diproses oleh RapidMiner. Selama fase konversi, atribut number digunakan untuk menggantikan fungsionalitas atribut periode.

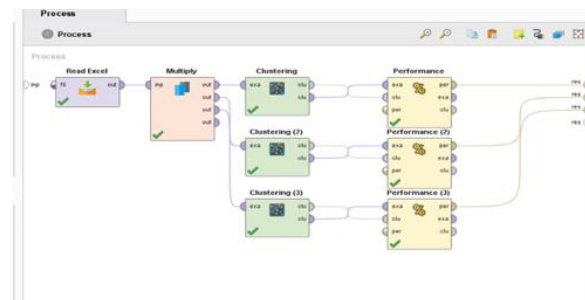
DESA	Tidak/Belum Sekolah	Tidak Tamat SD/Sederajat	Tamat SD/Sederajat	SLTP/Sederajat	SLTA/Sederajat	Diploma I/II	Akademik/Diploma III/S. Muda	Diploma IV/Strata I	Strata II	Strata III
DUKUHLATI	1777	544	4833	1394	488	13	23	744	5	0
KALIANYAR	1114	609	3534	1284	1082	11	81	358	4	0
KAPRINGAN	1469	604	5184	1111	441	9	30	448	2	0
KEDUNGWUNGU	1227	471	3235	1492	714	20	24	504	7	2
KRANGKENG	1530	652	4872	1318	759	10	42	489	2	1
LUWUNGSESIK	524	192	1457	545	311	3	17	185	2	0
PURWAJAYA	1067	339	4155	1023	467	8	35	528	1	0
SINGAKERTA	1133	427	3705	1035	609	19	44	433	4	1
SRENGENG	1946	846	5453	1450	743	14	59	582	6	0
TANJAKAN	882	364	2964	820	410	9	25	283	3	1
TEGALMULYA	1257	502	3352	1383	519	8	24	596	0	0

Gambar 5. Data Transformation

4.1.4. Data Mining

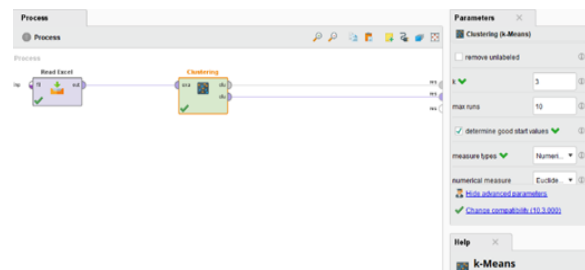
a. Operator Performance

Tahap ini merupakan proses pengolahan data yang menggunakan metode tertentu untuk memperoleh informasi dan wawasan. Metode yang digunakan adalah *clustering* dengan algoritma *k-means*. pengolahan data dilakukan dengan menggunakan rapid miner. Disini tahap pertama dari proses *clustering* K-means adalah menentukan jumlah *cluster* yang akan dibentuk. Penentuan jumlah *cluster* dilakukan dengan menggunakan operator performance yang menentukan jumlah *cluster* optimal yang akan dibentuk.



Gambar 6. Operator Performance

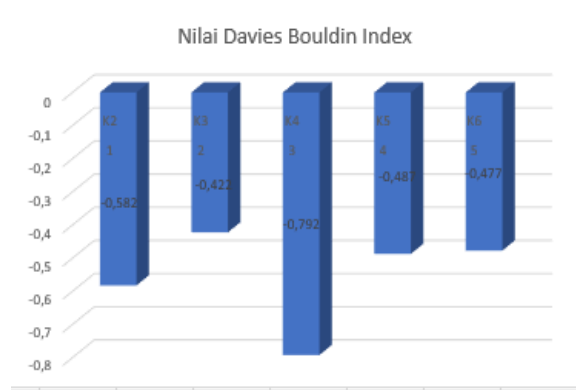
b. Proses Pengolahan Data Cluster yang dibentuk yaitu 3 cluster, diantaranya cluster rendah (C0) yang menunjukkan pendidikan rendah rendah, cluster sedang (C1) menunjukkan pendidikan sedang, dan cluster tinggi (C2).



Gambar 7. Proses pengolahan data

4.1.5. Cluster Model

Cluster yang dibentuk yaitu 3 cluster, diantaranya cluster rendah (C0) yang menunjukkan pendidikan rendah rendah, cluster sedang (C1) menunjukkan pendidikan sedang, dan cluster tinggi (C2)



Cluster Model

```
Cluster 0: 4 items
Cluster 1: 6 items
Cluster 2: 1 items
Total number of items: 11
```

Gambar 8. Cluster Model

4.4.6. Hasil Clustering K-Means

Pada tahap ini adalah hasil dari proses pengelompokan pada operator *Cluster* dengan menggunakan Algoritma K-means Aplikasi Data mining rapid minner yang mana tergolong ada 3 *cluster* rendah, sedang, tinggi.

DESA	cluster
DUKUJATI	cluster_0
KALIANYAR	cluster_1
KAPRINGAN	cluster_0
KEDUNGWU...	cluster_1
KRANGKENG	cluster_0
LUWUNGGE...	cluster_2
PURWAJAYA	cluster_1
SINGAKERTA	cluster_1
SRENGSENG	cluster_0
TANJAKAN	cluster_1
TEGALMULYA	cluster_1

Gambar 9. Hasil clustering

4.4.7. Evaluasi

Dalam tahap ini data di evaluasi untuk memastikan tingkat keakuratan dalam proses *clustering*. Evaluasi *clustering* dilakukan dengan tujuan untuk mengetahui seberapa baik kualitas dari hasil *clustering*. Pada penelitian ini, evaluasi hasil *clustering* yang digunakan adalah Davies Bouldin Index untuk mengetahui jumlah *cluster* yang paling optimal. Evaluasi dengan menggunakan Davies Bouldin Index ini memiliki skema evaluasi dari internal *cluster*, dimana baik atau tidaknya hasil *cluster* dilihat dari kuantitas dan kedekatan antar data hasil *cluster*. Evaluasi dilakukan di software Rapidminer dengan operator *Cluster Distance Performance* dengan percobaan pertama dengan 2 *cluster* dan percobaan kedua dengan 3 *cluster*.

PerformanceVector

```
PerformanceVector:
Avg. within centroid distance: -611438.744
Avg. within centroid distance_cluster_0: -163614.188
Avg. within centroid distance_cluster_1: -867338.490
Davies Bouldin: -0.582
```

Gambar 10. Performance (K)=2

PerformanceVector

```
PerformanceVector:
Avg. within centroid distance: -212435.371
Avg. within centroid distance_cluster_0: -280388.722
Avg. within centroid distance_cluster_1: -163614.188
Avg. within centroid distance_cluster_2: -0.000
Davies Bouldin: -0.422
```

Gambar 11. Performance (K)=3

4.2. Pembahasan

- Penerapan algoritma K-Means pengelompokan anggota

Penerapan algoritma K-Means pada pengelompokan anggota menggunakan Metode K-Means dan aplikasi rapid miner langkah pertama yaitu upload dataset yang berjumlah 11 Desa 87, 303 penduduk ke dalam rapid miner, kemudian langkah kedua tentukan jumlah klaster, Pada langkah kedua saya memilih 3 Cluster Rendah, Sedang, Tinggi. langkah ketiga hasil dari Penerapan algoritma K-Means pada pengelompokan anggota dengan hasil penelitian Dukuhjati, Kapringan, Krangkeng, Srengseng tergolong pada cluster 0 selanjutnya Kalianyar, Kedungwungu, Purwajaya, Singakerta, Tanjakan, tegal Mulya tergolong kelompok 1 Lalu Luwung gesik masuk pada cluster 2. Tahap selanjutnya adalah setelah semua proses data mining sudah selesai dilakukan, akan dibuatkan sebuah peta dalam bentuk visualisasi pemetaan menggunakan QGIS. Hasil dari pengklasteran pada proses modeling dilakukan visualiasi berupa pemetaan menggunakan QGIS agar mudah dipahami dalam Peta Desa di Kecamatan Krangkeng Berdasarkan Tingka Pendidikan.



Gambar 12. Peta Kec. Krangkeng

b. Evaluasi data Clustering

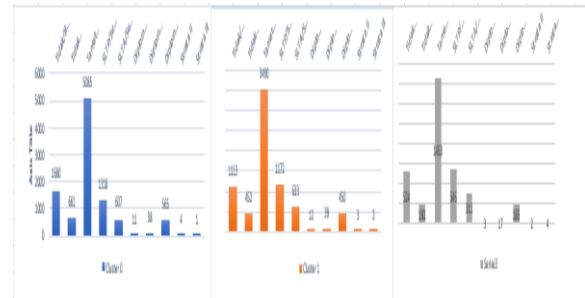
Pada tahap evaluasi peneliti membahas hasil DBI. Pada tahap evaluasi, peneliti membahas hasil DBI (Davies-Bouldin Index) untuk mengukur sejauh mana kualitas klustering yang telah dilakukan. Hasil DBI memberikan gambaran tentang seberapa baik kluster-klusternya terbentuk dan seberapa jelas batas antar-kluster. Selain itu, peneliti juga dapat mengevaluasi apakah jumlah kluster yang dipilih telah optimal atau masih memerlukan penyesuaian. Hasil DBI melibatkan interpretasi terhadap nilai yang diperoleh. Jika nilai DBI mendekati nol, itu menunjukkan bahwa kluster-klusternya sangat baik terbentuk dengan batas antar-kluster yang jelas. Sebaliknya, nilai DBI yang tinggi dapat mengindikasikan adanya tumpang tindih antar-kluster atau ketidakjelasan batas cluster. Setelah dilakukan proses evaluasi didapatkan tingkat performance pada $K(2) = -0.582$, $K(3) = -0.422$, $K(4) = -0.792$, $K(5) = -0.487$, $K(6) = -0.477$ sehingga dipilih K yang akan digunakan yaitu $K=3$ karena nilai yang paling mendekati nol.

c. Hasil analisis cluster

Dari hasil analisis kluster, kluster 0 yang merupakan tingkat sekolah dasar memiliki nilai rata-rata yang lebih tinggi dibandingkan dengan kluster lainnya, dan kluster 1 yang merupakan tingkat sekolah menengah atas memiliki nilai rata-rata yang lebih tinggi dibandingkan dengan kluster lainnya. Selain itu, Kluster 2 berada pada tingkat Strata III. Ini memiliki nilai rata-rata yang lebih tinggi dibandingkan cluster lainnya. Berdasarkan penelitian yang telah dilakukan, proses clustering data tingkat pendidikan Kecamatan Krangkeng tahun 2023 berjumlah 11 desa dan 3 cluster, sehingga cluster 0 berjumlah 4 desa, maka dapat disimpulkan cluster 1 mencakup total empat desa. Dari keenam desa tersebut, termasuk kluster 2 yang berjumlah satu desa. Berdasarkan hasil analisis kluster yang dilakukan diketahui bahwa kluster 2 merupakan kelompok desa dengan tingkat pendidikan tinggi, sedangkan kluster 0 merupakan kelompok desa dengan tingkat pendidikan rendah. Setelah dilakukan implementasi dan diskusi, disimpulkan bahwa kelompok Cluster 0 memerlukan pertimbangan khusus berupa dukungan pendidikan dari pemerintah agar tidak tertinggal dengan daerah lain.

Berdasarkan nilai rata-rata perbandingan karakteristik gambar 13 bahwa cluster 0 mempunyai karakteristik Belum sekolah, Tidak Tamat SD/ sederajat, Tamat SD/ Sederajat, Tamat SLTP/ Sederajat, Diploma IV/ Strata 1, Strata 2 mempunyai nilai rata-ratanya lebih tinggi dari pada cluster lainnya, Sedangkan Cluster 1 nilai rata-rata Tingkat SLTA & Diploma I/II lebih

tinggi dari pada cluster lainnya, selanjutnya cluster 2 pada tingkat Strata III nilai rata-rata nya lebih tinggi dari pada cluster-cluster lainnya.



Gambar 13. Perbandingan karakteristik

5. KESIMPULAN DAN SARAN

Berdasarkan penelitian yang dilakukan, proses clustering data tingkat pendidikan Kecamatan Klanken tahun 2023 berjumlah 11 desa dan 3 cluster, sehingga anggota cluster 0 berjumlah 4 desa, cluster. Dapat kita simpulkan bahwa 1 memiliki total 4 anggota. Dua dari enam desa dan kluster memimpin total satu desa. Berdasarkan hasil analisis kluster yang dilakukan diketahui bahwa kluster 2 merupakan kelompok desa dengan tingkat pendidikan tinggi, sedangkan kluster 0 merupakan kelompok desa dengan tingkat pendidikan rendah. Setelah dilakukan implementasi dan diskusi, disimpulkan bahwa kelompok Cluster 0 memerlukan pertimbangan khusus berupa dukungan pendidikan dari pemerintah agar tidak tertinggal dengan daerah lain.

Berdasarkan hasil penelitian ini menyadari adanya kekurangan pada penelitian ini, karena keterbatasan baik dalam perihal waktu ataupun pengetahuan. Dalam rangka memperbaiki kekurangan serta guna penyempurnaan penelitian ini maka memberikan beberapa saran sebagai berikut : Penelitian ini dapat dikembangkan lagi bukan hanya megacu pada tingkat pendidikan tapi bisa juga dikembangkan lagi seperti prediksi bantuan sosial pada penduduk kecamatan krangkeng, Pada penelitian selanjutnya bisa menggunakan dataset yang jauh lebih banyak lagi.

DAFTAR PUSTAKA

- [1] E. Asidah, "Pernyataan Mengenai Skripsi Dan Sumber Informasi," *sidayu.indramayukab.go.id*, vol. 1, no. 2, pp. 105–112, 2019, [Online]. Available: https://sidayu.indramayukab.go.id/assets/resume/ANWAR_ASAI_118090120_SKRIPSI_eka.pdf
- [2] L. Rohmaniah, A. Faqih, and T. Suprpti, "Vol . 15 No . 1 September 2022 ISSN : 1979-8415 JAWA BARAT MENGGUNAKAN K-MEANS DAN FUZZY C-MEANS ISSN : 1979-8415," vol. 15, no. 1, pp. 1–7, 2022.
- [3] R. Kurniawan, M. M. M. Mukarrob, and M. Mahradianur, "KLAUSTERISASI TINGKAT

- PENDIDIKAN DI DKI JAKARTA PADA TINGKAT KECAMATAN MENGGUNAKAN ALGORITMA K-MEANS,” *Technol. J. Ilm.*, vol. 12, no. 4, p. 234, 2021, doi: 10.31602/tji.v12i4.5633.
- [4] C. Kamila, “Systematic Literature Review: Penggunaan Algoritma K-Means Untuk Clustering di Indonesia dalam Bidang Pendidikan,” *Intech*, vol. 2, no. 1, pp. 19–24, 2021, doi: 10.54895/intech.v2i1.866.
- [5] N. Bloom and J. Van Reenen, “Clustering Pemetaan Tingkat Kemiskinan Di Provinsi Jawa Barat Menggunakan Algoritma K-Means,” *NBER Work. Pap.*, vol. 9, no. 2, p. 89, 2023, [Online]. Available: <http://www.nber.org/papers/w16019>
- [6] Z. Nabila, A. Rahman Isnain, and Z. Abidin, “Analisis Data Mining Untuk Clustering Kasus Covid-19 Di Provinsi Lampung Dengan Algoritma K-Means,” *J. Teknol. dan Sist. Inf.*, vol. 2, no. 2, p. 100, 2021, [Online]. Available: <http://jim.teknokrat.ac.id/index.php/JTSI>
- [7] R. K. Dinata, S. Safwandi, N. Hasdyna, and N. Azizah, “Analisis K-Means Clustering pada Data Sepeda Motor,” *INFORMAL Informatics J.*, vol. 5, no. 1, p. 10, 2020, doi: 10.19184/isj.v5i1.17071.
- [8] C. Astria, A. P. Windarto, A. Wanto, and E. Irawan, “Metode K-Means pada Pengelompokan Wilayah Pendistribusian Listrik,” *Semin. Nas. Sains Teknol. Inf.*, pp. 306–312, 2019, [Online]. Available: <http://prosiding.seminar-id.com/index.php/sensasi/issue/archivePage%7C306>
- [9] T. Hardiani, “Analisis Clustering Kasus Covid 19 di Indonesia Menggunakan Algoritma K-Means,” *J. Nas. Pendidik. Tek. Inform.*, vol. 11, no. 2, pp. 156–165, 2022, doi: 10.23887/janapati.v11i2.45376.
- [10] D. Bahtiar *et al.*, “PEMETAAN PENDUDUK PENERIMA BANTUAN SOSIAL DESA WARU JAYA MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING,” 2023. [Online]. Available: <http://pijarpemikiran.com/index.php/Scientia>
- [11] A. Amalia, D. R., Narasatu, R., Faqih, “Perbandingan Hasil Klasifikasi Rasa Minuman Thai Tea yang Paling Digemari Menggunakan K-means dan K-medoids,” *Pros. Semin. Nas. Unimus*, vol. 2, pp. 401–407, 2019.