

## Klasterisasi Pendidikan Masyarakat untuk mengetahui Daerah dengan Pendidikan Terendah menggunakan Algoritma K-Means

Nurahman<sup>1\*</sup>, Diana Dwi Aulia<sup>2</sup>

<sup>1,2</sup> Jurusan Sistem Informasi, Fakultas Ilmu Komputer,  
Universitas Darwan Ali

\*Email: nurrahman.ikhtiar@gmail.com

### Abstrak

Pendidikan adalah kebutuhan mendasar bagi tiap manusia yang berperan penting pada masa depan bangsa, karena bangsa yang dikatakan maju dapat dilihat dari sistem pembelajarannya yang baik. Pendidikan yang berhasil diukur dari rata-rata jumlah lulusan berbagai jenjang pendidikan pada berbagai wilayah. Namun tidak semua wilayah baik dalam kualitas pendidikannya. Salah satunya yaitu daerah yang ada di Indonesia seperti kabupaten Kapuas Kalimantan Tengah. Diketahui bahwa pada tahun-tahun sebelumnya daerah ini kurang dalam peningkatan pendidikannya, sehingga menyebabkan beberapa wilayah yang masyarakatnya tidak sekolah atau putus sekolah. Permasalahannya banyak disebabkan karena faktor ekonomi, rasa malas, kurangnya motivasi tentang pentingnya pendidikan, dan lain sebagainya. Pandemi covid-19 sebelumnya juga menjadi alasan peningkatan jumlah anak putus sekolah yang dikarenakan perekonomian keluarga menurun. Banyaknya wilayah yang ada di kab. Kapuas diperlukan pengelompokan jumlah desa yang ada. Pengelompokan bertujuan memudahkan pemerintah dalam memberikan perhatian khusus terhadap wilayah yang pendidikannya dinilai kurang dan tujuan lainnya untuk mengetahui desa mana saja yang tingkat pendidikannya rendah. Dalam pengelompokan, sistem yang diterapkan adalah data mining menggunakan metode Clustering algoritma K-Means yang diproses menggunakan software rapidminer. Adapun pengelompokan yang terbentuk pada data tingkat pendidikan sebanyak 229 record adalah 8 kluster yang dinyatakan desa pendidikannya terendah terdapat pada (C1) dengan jumlah 33 desa.

**Kata kunci:** Clusterisasi, Data Mining, K-Means, Pendidikan.

### Abstract

Education is a basic need for every human being who plays an important role in the future of the nation, because a nation that is said to be advanced can be seen from its good learning system. Successful education is measured by the average number of graduates at various levels of education in various regions. But not all regions are good in the quality of education. One of them is the area in Indonesia, such as the Kapuas district, Central Kalimantan. It is known that in previous years this area lacked improvement in education, causing several areas where people did not go to school or dropped out of school. Many of the problems are caused by economic factors, laziness, lack of motivation about the importance of education, and so on. The previous Covid-19 pandemic was also the reason for the increase in the number of children dropping out of school due to a declining family economy. The number of areas in Kapuas district requires grouping the number of existing villages. The grouping aims to make it easier for the government to pay special attention to areas where education is considered lacking and other purposes are to find out which villages have low levels of education. In grouping, the system applied is data mining using the K-Means Algorithm Clustering method which is processed using rapidminer software. The groupings formed on the education level data of 229 records are 8 clusters where the lowest education villages are stated in (C1) with a total of 33 villages.

**Keywords:** Clusterization, Data Mining, K-Means, Education

## PENDAHULUAN

Pertumbuhan Jumlah penduduk yang semakin meningkat di suatu daerah, terutama pada kabupaten Kapuas, tentu harus menjadi pusat perhatian tertinggi oleh pemerintah. Karena di wilayah ini memiliki jumlah penduduk paling tinggi diantara kabupaten lainnya yang ada di provinsi Kalimantan Tengah. Salah satunya yang harus menjadi perhatian khusus adalah masalah pendidikan. Pendidikan ini merupakan suatu kebutuhan yang mendasar bagi tiap manusia dalam kehidupan, dan yang terpenting pendidikan sangat berperan terhadap masa depan bangsa atau negara serta dalam sistem pendidikan yang dijalankan juga dapat mengukur kemajuan negara itu sendiri (Amirullah, 2019).

Pendidikan yang dikatakan berhasil dapat dilihat dari jumlah tingkat kelulusan pendidikan yang dimulai dari SD sampai dengan perguruan tinggi. Kualitas pendidikan di seluruh wilayah tidak dapat disamaratakan, maka dari itu perlu adanya perhatian di beberapa wilayah yang mungkin pendidikannya masih tergolong rendah. Banyaknya jumlah penduduk saat ini tentu menyulitkan pemerintah dalam mempertimbangkan wilayah mana saja yang menjadi fokus utama dalam pengembangan pendidikan. Dengan mengelompokkan suatu wilayah berdasarkan tingkat pendidikan akan mempermudah pemerintah untuk melihat daerah mana saja yang harus benar-benar diperhatikan dalam penanganan pendidikannya. Karena diketahui kondisi pendidikan beberapa tahun terakhir di kabupaten kapuas kurang dalam kemajuan dan peningkatannya.

Dari permasalahan yang ada, begitu sangat diperlukannya proses pengelompokan untuk mempertimbangkan daerah mana saja yang menjadi titik fokus agar dapat disalurkan bantuan atau biaya pendidikan. Pemerintah juga dapat mensosialisasikan ke wilayah yang tingkat pendidikannya rendah untuk memotivasi seluruh masyarakat dalam pentingnya menempuh dunia Pendidikan (Juariyah dan basrowi, 2010). Oleh karena itu, penulis mencoba menerapkan metode klasterisasi atau mengelompokkan jumlah data berdasarkan tingkat pendidikan dari tingkat belum sekolah sampai perguruan tinggi dengan menggunakan algoritmas K-Means. Alasan peneliti menggunakan k-means karena

algoritma ini cukup mudah diimplementasikan dan memiliki ketelitian yang cukup tinggi pada ukuran objeknya, sehingga algoritma ini relatif lebih terukur (Nurahman & Dwi Aulia, 2021). Dalam penelitian ini, penulis melakukan pengelompokan pada seluruh desa yang ada di Kabupaten Kapuas Kalimantan Tengah. Metode ini akan mempermudah dalam proses pengelompokan data pada tingkat pendidikan di setiap wilayah atau desa untuk menganalisis daerah mana saja yang menjadi pertimbangan untuk mendapatkan bantuan pendidikan.

Pengolahan data pada penelitian ini dikelompokkan berdasarkan tingkat pendidikan, yaitu : Tidak/belum sekolah, Belum tamat SD, Tamat SD, SLTP, SLTA, D1 dan D2, D3, S1, S2, dan S3 data yang diperoleh dari jumlah penduduk per desa di kabupaten kapuas pada 30 Juni tahun 2022. Melalui data ini peneliti akan mengelompokkan desa atau kelurahan yang jumlah tingkat pendidikannya tertinggi sampai paling rendah.

Penelitian oleh (Oktarian et al., 2020) Dengan judul Klasterisasi Penentuan Minat Siswa dalam Pemilihan Sekolah Menggunakan Metode Algoritma K-Means Clustering. Dalam penelitian ini, pengelompokan sekolah dilakukan menggunakan atribut jumlah Rombel, Jumlah dari Peserta Didik, serta Jumlah tenaga pendidik. Tujuan dari pengelompokannya adalah untuk melihat minat tertinggi siswa dalam menentukan sekolah. adapun hasil dari penelitian yang terbentuk adalah 3 klaster.

Penelitian yang dilakukan sebelumnya mengenai *k-means* oleh (Sari dkk., 2020) yang berjudul Penerapan algoritma *K-Means* untuk *Clustering* data kemiskinan provinsi Banten menggunakan *rapidminer*. Dalam penelitian ini atribut yang digunakan untuk dikelompokkan adalah Rata-rata lamanya sekolah, Rata-rata pengeluaran per-kapita, dan Rata-rata penduduk yang miskin dari kabupaten/kota di provinsi Banten. Tujuan dari penelitian yang dilakukan adalah untuk melihat tingkat kemiskinan terendah sampai tertinggi. Adapun hasilnya yaitu : tingkat terendah masuk pada *Cluster(2)*, tingkat sedang masuk *Cluster(0)*, dan tingkat kemiskinan tertinggi masuk pada *Cluster(1)*.

Penelitian sebelumnya mengenai algoritma K-means juga dilakukan oleh Ragil Kurniawan, M.Mukarobin, dan Mahradianur dengan penelitiannya yang berjudul Klasterisasi

tingkat pendidikan di DKI Jakarta pada tingkat kecamatan menggunakan algoritma *K-Means*. penelitian oleh (Kurniawan dkk., 2021) menggunakan atribut Kecamatan, SD, SMP, SMA, sampai Perguruan Tinggi. Dalam penelitian ini bertujuan menghasilkan data kelompok tertinggi dan terendah. Adapun kelompok yang paling rendah masuk pada *Cluster 2* dan pendidikan tertinggi masuk pada *Cluster 0*.

Beberapa penelitian terdahulu diatas terdapat perbedaan dengan penelitian yang telah dilakukan sekarang. Pada penelitian ini melakukan klasterisasi berdasarkan tingkat pendidikan dengan atribut yang digunakan antara lain desa/kelurahan, kemudian jenjang pendidikan mulai dari tidak/belum sekolah, SD, SMP, SMK dan perguruan tinggi lainnya. Pembaruan atau perbedaan lain dari penelitian ini yaitu melakukan pengujian akurasi menggunakan metode *Davies Bouldin Index*. dalam mengetahui tingkat performa terbaik dari hasil pengklasteran yang telah dibentuk agar menjadi klaster yang optimal.

## TINJAUAN PUSTAKA

### 2.1 Data Mining

Data mining adalah suatu alat bantu sebagai proses penggalian data pada basis data yang berukuran besar dalam spesifikasi tingkat kerumitannya. Pengertian data mining juga suatu proses dengan Teknik statistik, (kecerdasan buatan) / AI, *mathematics*, dan juga *Machine learning* untuk memproses maupun identifikasi informasi data yang digunakan untuk pengetahuan dari big data (Putra et al., 2022) .

### 2.2 Clustering

*Clustering* merupakan proses yang digunakan untuk mengelompokkan data ke beberapa kelompok atau klaster sehingga data yang ada dalam satu kelompok memiliki kemiripan maksimal dan data antar klaster memiliki kemiripan yang minimum. Analisis *clustering* juga merupakan cara untuk mengidentifikasi kelompok objek yang saling mirip yang akan membantu dalam menemukan pola penyebaran atau pola hubungan terhadap sekumpulan data yang terbilang besar (Wanto dkk., 2021).

### 2.3 Algoritma *K-Means*

Algoritma *K-Means* merupakan salah satu metode klaster non hirarki bertujuan membantu mengelompokkan variabel untuk dimasukkan dalam kelas-kelas yang letaknya dihasil akhir perhitungan. *K-means* juga salah satu metode klastering yang berusaha mempartisi data yang tersedia kedalam satu bentuk atau lebih kelompok, sehingga data yang karakteristik nya sama menjadi satu kelompok dalam satu cluster yang sama. *K-means* yaitu algoritma kalsifikasi dari algoritma clustering partisional yang paling banyak digunakan dan paling sederhana. Algoritma ini mulai dari memisahkan ruang data secara acak dan menunjukan sampel data yang ada ke dalam beberapa *cluster* berdasarkan kemiripannya sampel dan klaster, sampai sebuah data yang diuji menjadi data konvergen. Syarat dari criterion yang telah ditemukan yaitu saat tidak ada pemindahan sampel lagi dari satu klaster ke klaster lainnya yang menyebabkan berkurangnya jumlah *error* yang dikuadratkan. Algoritma *k-means* populer digunakan dikarenakan mudah mengimplementasikannya, dan memiliki kecepatan proses yang cukup baik (Wahyudi dkk., 2020).

Langkah – langkah dasar yang diambil oleh algoritma *K-means* yaitu :

1. Menentukan jumlah kluster yaitu membuat catatan yang akan dibuat menjadi beberapa kelompok pada dataset, sebutlah sebanyak (**K**) klaster.
2. Secara acak pilih (**K**) klaster dalam catatan (dari sekian catatan yang ada) sebagai pusat kelompok awal. Tahap iterasi digunakan rumus pada persamaan (1) dibawah:

$$V_k = \frac{\sum_{i=1}^{N_i} X_i}{N_k} \quad (1)$$

Keterangan :

$V_k$  = *Centroid cluster* ke  $k$

$X_i$  = Data ke  $i$

$N_k$  = Banyaknya jumlah data yang jadi anggota klaster ke  $k$ .

3. Menentukan pusat kelompok yang terdekat jaraknya dengan centroid. Euclidean Distance yaitu merupakan jarak centroid yang digunakan, dengan rumus (2) persamaan di bawah ini :

$$D_E = \sqrt{(x_i - s_i)^2 + (y_i - t_i)^2} \quad (2)$$

Keterangan :

(x,y) = Koordinat Objek

(s,t) = Koordinat Centroid

DE = Euclidean Distance

i = Banyak Objek

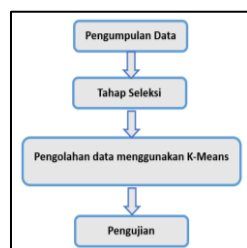
Lalu hitunglah rasio besaran *Between Cluster Variation* dengan *Within Cluster Variation*, kemudian bandingkan rasio yang sudah di hitung dengan rasio sebelumnya (jika sudah ada). Apabila rasio tersebut membesar, maka boleh dilanjutkan ke langkah keempat dan hentikan prosesnya jika tidak ada perubahan.

4. Pusat kelompok diperbaharui dengan mengelompokan objek berdasarkan jarak terdekatnya ke centroid. Kemudian Ulangi langkah 2 dan 3, dan lakukan iterasi hingga nilai centroid optimal.

## METODE PENELITIAN

### 3.1 Tahapan dalam penelitian

Dalam memudahkan penelitian tentu diperlukannya penentuan metodologi, karena metodologi dapat memecahkan sebuah masalah dan membantu mencapai tujuan yang diinginkan. Metodologi penelitian juga menggambarkan alur kerja dalam penelitian dari awal sampai selesai. Langkah penelitian yang dilakukan mulai dari mengumpulkan data, penyeleksian data, pengolahan atau memproses data, mengolah data dengan Algoritma, dan terakhir pengujian hasil (Nurdiawan dan Pratama, 2019). Agar permasalahan dapat diketahui dengan jelas, maka data harus dikumpulkan dan dianalisis terlebih dahulu hingga mencari referensi terkait. Adapun gambar kerangka kerja dalam penelitian yang dilakukan yaitu seperti pada Gambar 1.



Gambar 1. Kerangka kerja / Tahap penelitian

### 3.2 Isi Kerangka Kerja / Tahap Penelitian

#### 1. Pengumpulan Data

Data yang digunakan penulis dalam penelitian ini adalah data yang isinya jumlah tingkat pendidikan perkelurahan atau desa di wilayah kabupaten Kapuas pada 30 Juni 2022. Dataset diperoleh dari website “Visualisasi Data Kependudukan - GIS Dukcapil - Kemendagri” (<https://gis.dukcapil.kemendagri.go.id>) dengan jumlah data training sebanyak 229 data berdasarkan nama desa yang ada di kabupaten kapuas.

#### 2. Tahap Seleksi

Tahap penyeleksian data perlu dilakukan sebelum mengolah data yang diproses. Data yang telah di peroleh di seleksi menggunakan metode manual di ms. Excel, data yang diseleksi berupa data atribut yang dipilih berdasarkan kepentingan pada penelitian. Data yang pada awal nya memiliki atribut sebanyak 11 atribut yaitu Kecamatan, Tidak/belum sekolah, Belum tamat SD, Tamat SD, SLTP, SLTA, D1 dan D2, D3, S1, S2 dan S3 menjadi hanya 10 atribut dengan pengurangan pada atribut kecamatan.

#### 3. Proses Pengolahan Data dengan Clustering

Data yang telah diseleksi sesuai dengan keperluan penelitian kemudian diproses menggunakan metode cluster algoritma k-means. Adapun nilai K yang digunakan adalah sebanyak 8 klaster. Untuk fungsi Jarak menggunakan *Euclidean Distance*. Data diproses dengan menggunakan aplikasi data mining rapid miner.

#### 4. Pengujian Hasil dengan Performa DBI

Setelah melalui proses klasterisasi maka data diuji dengan melihat performanya menggunakan metode DBI (*Davies Bouldin Index*). Pengujian data ini bertujuan melihat tingkat akurasi terbaik dari pembagian klaster sebelumnya yang sudah dilakukan. Adapun *Davies Bouldin Index* ini merupakan fungsi dari jumlah rasio atau jarak antar klaster pada suatu pengelompokan dataset (Butsianto dan Saepudin, 2020). Untuk mendapatkan nilai DBI dapat dilakukan dengan menggunakan fungsi (3) :

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} R_{ij} \quad (3)$$

HASIL DAN PEMBAHASAN

4.1 Hasil Seleksi atau *Preprocessing* Data

Tahap *preprocessing* atau seleksi dilakukan sebelum dataset diproses menggunakan metode *clustering*. Adapun data valid yang diproses memiliki 229 data dengan 10 atribut. Data awal adalah 230 record dengan 11 atribut. Dimana pada atribut kecamatan dihilangkan dan penghapusan salah satu desa yang diketahui merupakan data missing. Tabel 1 merupakan dataset yang telah melalui tahap *Pre-processing*.

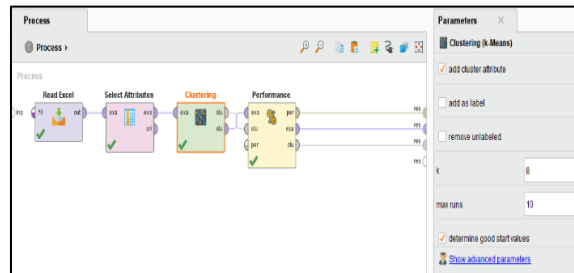
Tabel 1. Data *Preprocess*

No	Desa Kelurahan	Tidak belum sekolah	Belum tamat SD	Tamat SD	SLTP	SLTA	D1 dan D2	D3	S1	S2	S3
1	Basarang	824	577	890	416	292	13	5	39	0	0
2	Basarang Jaya	467	230	296	250	376	32	9	102	0	0
3	Basungkai	325	199	384	183	111	2	1	18	0	0
4	Batu Nindan	442	257	341	240	309	12	14	46	0	0
5	Batuah	423	211	527	173	126	1	1	15	0	0
6	Bungai Jaya	487	237	442	377	378	19	8	55	0	0
7	Lunuk Ramba	232	142	262	148	210	14	12	52	0	1
....	....	....	....	....	....	....	....	..	..	..	..
....	....	....	....	....	....	....	....	..	..	..	..
228	Timpah	986	654	1258	550	424	25	35	102	1	1
229	Tumbang Randang	218	103	356	128	76	2	1	7	0	0

K-Means 8 dengan nilai DBI terkecil. Selanjutnya Operator *performance* digunakan untuk mengetahui nilai pengujian atau performa dari analisis data menggunakan K-Means yang menghasilkan performa DBI (*Davies Bouldin Index*).

Pada proses pengelompokan data dengan menerapkan *Rapidminer* maka hasil data yang diperoleh sebagai berikut :

Gambar 2. dibawah merupakan desain pemodelan dataset dari jenjang tingkat pendidikan di kabupaten Kapuas dengan penerapan algoritma K-means. Kemudian pengujian dilakukan dengan memperhatikan nilai performa DBI (*Davies Bouldin Index*). Tujuannya untuk mendapatkan nilai DBI terbaik yang dilakukan dengan melihat hasil uji masing-masing nilai K dari K=3 hingga K=10.



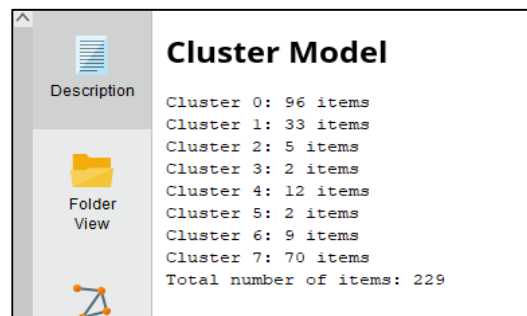
Gambar 2. Desain Processing Data

4.1 Hasil *Clustering* dan Pengujian data

Pengolahan dan pengujian data diproses menggunakan software *Rapidminer*. Data yang ada dikelompokkan dengan algoritma K-means. setelah data melewati proses klasterisasi maka dilanjutkan dengan pengujian menggunakan performa DBI (*Davies Bouldin Index*). Pada perhitungan awal algoritma K-means nilai K yang diberikan adalah 3 sehingga menghasilkan 3 Klaster pada proses pembagian partisinya. Pada nilai K=4 menghasilkan 4 klaster dalam pembagian partisinya. Demikian juga nilai K seterusnya pada nilai K=5, K=6, K=7, K=8, K=9, dan K=10 sehingga menemukan hasil DBI terkecil sampai nilai yang diperoleh (non-negatif  $\geq 0$ ).

Gambar 2. merupakan desain alur proses perhitungan Dataset tingkat pendidikan Kabupaten Kapuas. Operator select atribut digunakan untuk menyeleksi jenjang pendidikan yang tipe datanya integer, untuk Desa/ kelurahan dijadikan kolom ID pada dataset. Pada operator *clustering* berisikan nilai K=8, dengan melalui proses sebelumnya yaitu dari K=3 sampai dengan K=10. Maka diperoleh

Selanjutnya pada desain pemodelan yang dibentuk menjadi 8 *cluster*. Hasil yang didapat dari pengelompokan data 8 *cluster* adalah *cluster 0, cluster1, cluster2, cluster3, cluster4, cluster5, cluster6, dan cluster7*. Hasil masing-masing kelompok yang didapat dengan menggunakan *software Rapidminer* yaitu dilihat pada gambar 3.



Gambar 3. Model Klasterisasi

Gambar 4. merupakan tampilan dari seluruh data Tingkat pendidikan yang sudah dilakukan tahap proses klasterisasi menggunakan K-Means dengan Nilai K=8. Pada gambar dibawah juga dapat dilihat

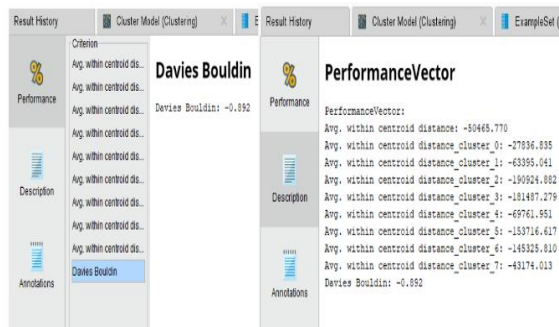
jumlah Tidak/Belum Sekolah dan Belum Tamat SD. Dimana *cluster\_3* pada kolom Tidak/Belum Sekolah dan Belum Tamat SD memiliki angka tertinggi dari *cluster* lainnya.

terletak di *cluster\_3*, sedangkan pendidikan terendah dengan tanda warna biru muda yang merupakan dari *cluster\_1*.

Row No.	Desa/Kelera...	cluster	Tidak/belum...	Belum tama...	Tamat SD	SLTP	SLTA	D1 dan D2	D3
15	Bamban Raya	cluster_1	682	489	624	101	44	0	1
16	Bangun Hajo	cluster_0	232	134	186	245	159	10	7
17	Budi Mufakat	cluster_1	730	352	828	130	89	3	0
18	Pulau Wamb...	cluster_2	1.051	653	1.202	717	554	16	8
19	Sei Jangot	cluster_2	1.419	825	1.655	385	166	1	1
20	Tamban Luar	cluster_1	1.013	451	852	548	356	14	12
21	Terusan Bag...	cluster_1	534	308	643	106	82	1	4
22	Terusan Karya	cluster_1	719	314	615	366	210	11	9
23	Terusan Mak...	cluster_7	573	342	377	267	146	12	9
24	Terusan Mulya	cluster_1	695	432	784	387	179	7	5
25	Terusan Raya	cluster_1	772	650	894	218	121	8	1
26	Terusan Ray...	cluster_7	471	242	536	122	88	3	2
27	Terusan Riy...	cluster_1	587	320	703	163	121	1	2
28	Pulau Kupang	cluster_2	1.738	1.053	2.827	661	520	8	8

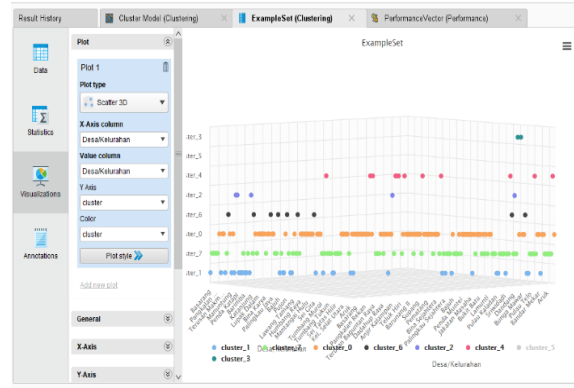
Gambar 4. Hasil klasterisasi data

Pengujian performa dengan metode *Davies Bouldin Index* menggunakan *software Rapidminer* pada nilai  $K=8$  menghasilkan nilai akurasi  $-0,892$ . Semakin kecil/rendah nilai DBI yang diperoleh (non-negatif  $\geq 0$ ), maka semakin baik klaster yang didapat dari menggunakan pengelompokan *K-Means*. Teori tersebut disebutkan oleh *Davies dan Bouldien* pada penelitiannya (*Atyanto et al., 2011*). Adapun pengujian nilai  $K=3$  sampai  $K=7$  dan juga  $K=10$  diperoleh nilai performa diatas  $-1$ , maka penulis hanya mengambil hasil perolehan dari pengujian  $K=8$  yang mendekati ketentuan performa DBI, seperti yang ditunjukkan pada Gambar 5.



Gambar 5. Performa (*Davies Bouldin Index*)

Pada gambar 6. menampilkan visualisasi dengan *plot type = Scatter 3D* yang bertujuan melihat daerah mana saja yang masuk ke 8 kelompok masing-masing. Dengan peringkat pendidikan terendah sampai tertinggi. Diketahui pada gambar 6 bahwa warna hijau tua merupakan tingkat pendidikan tertinggi yang



Gambar.6 visualisasi data

Berikutnya adalah tampilan nama desa/kelurahan yang terdiri 33 desa dengan tingkat pendidikan rendah. Dapat dilihat bahwa daerah tersebut memiliki jumlah anak terbanyak yang Tidak/Belum sekolah dan belum tamat SD. Hasil dapat dilihat pada Gambar 7

- cluster\_1
  - Basarang
  - Maluen
  - Pangkalan Rekan
  - Bamban Raya
  - Budi Mufakat
  - Tamban Luar
  - Terusan Baguntan Raya
  - Terusan Karya
  - Terusan Mulya
  - Terusan Raya
  - Terusan Raya Hulu
  - Saka Tamiang
  - Sei Kayu
  - Sei Pitung
  - Saka Batur
  - Sei Asam
  - Supang
  - Batanjung
  - Sei Bakut
  - Sei Teras
  - Mampai
- cluster\_2
  - Muara Dadahup
  - Marapit
  - Mantangai Hulu
  - Manusup
  - Pulau Kaladan
  - Tarantang
  - Bunga Mawar
  - Handiwung
  - Saka Lagun
  - Bandar Mekar
  - Bandar Raya
  - Warna Sari

Gambar 7. Cluster dengan Pendidikan Terendah

**SIMPULAN**

Berdasarkan dari hasil penelitian dan analisis perhitungan yang dilakukan pada pembahasan sebelumnya dengan menerapkan data mining menggunakan metode klasterisasi Algoritma K-Means terhadap tingkat pendidikan di Kabupaten Kapuas, maka dapat diambil kesimpulan bahwa dengan menggunakan metode klasterisasi dalam menerapkan algoritma K-Means, maka tingkat pendidikan diseluruh desa kabupaten Kapuas dikelompokkan menjadi 8 klaster. Dari 8 kelompok diketahui klaster yang memiliki pendidikan tertinggi dan rendah, yaitu pada *Cluster\_3* = 2 desa dengan Kategori Pendidikan tertinggi, sedangkan *Cluster\_1* = 33 desa merupakan kategori pendidikan terendah. Setelah melalui Proses pengelompokan, maka dapat dilihat daerah mana saja yang harus menjadi perhatian pemerintah. Daerah yang harus jadi perhatian lebih terdapat pada *cluster* yang ada dalam kategori pendidikan terendah yaitu desa yang terdapat di *cluster\_1*. Berdasarkan pengujian (DBI) *Davies Bouldin Index*, dengan jumlah data sebanyak 229, maka terbentuk 8 *cluster* diperoleh nilai pengujian sebesar -0,892.

**DAFTAR PUSTAKA**

Amirullah, I. (2019). Pemetaan Kelompok Kerja Siswa Dengan Metode Clustering K-Means Dan Algoritma Greedy. *Jurnal Informatika Dan Rekayasa Perangkat Lunak*, 1(2), 94–98.

Butsianto, S., & Saepudin, N. (2020). Penerapan Data Mining Terhadap Minat Siswa Dalam Mata Pelajaran Matematika Dengan Metode K-Means. *Jurnal Nasional Komputasi Dan Teknologi Informasi (JNKTI)*, 3(1). <https://doi.org/10.32672/jnkti.v3i1.2008>

Juariyah dan basrowi. (2010). Jurnal Ekonomi & Pendidikan, Volume 7 Nomor 1, April 2010. *Jurnal Ekonomi & Pendidikan*, 7(April), 60. <https://journal.uny.ac.id/index.php/jep/article/view/577/434>

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>

Nurahman, N., & Dwi Aulia, D. (2021). Algoritma K-Means Untuk Melihat Penularan Tertinggi Virus Covid-19 Diseluruh Provinsi Indonesia. *Jurnal Ilmiah Betrik*, 12(2), 162–168. <https://doi.org/10.36050/betrik.v12i2.331>

Nurdiawan, O., & Pratama, F. A. (2019). Implementasi Algoritma K-Means Dalam Penentuan Prioritas Rehabilitasi Daerah Aliran Sungai Cipunagara. *InfoTekJar (Jurnal Nasional Informatika Dan Teknologi Jaringan)*, 4(1). <https://doi.org/10.30743/infotekjar.v4i1.1633>

Oktarian, S., Defit, S., & Sumijan. (2020). Klasterisasi Penentuan Minat Siswa dalam Pemilihan Sekolah Menggunakan Metode Algoritma K-Means Clustering. *Jurnal Informasi Dan Teknologi*, 2.

Putra, J. W. P., Suganda, E. A., & Intan Purnamasari. (2022). Penerapan RapidMiner dengan Metode K-Means dalam Penentuan Kluster Gangguan Jaringan WIFI Provider PT . XYZ di Daerah Karawang. *Jurnal Informatika Dan Rekayasa Perangkat Lunak*, 4(1), 31–35.

Sari, Y. R., Sudewa, A., Lestari, D. A., & Jaya, T. I. (2020). Penerapan Algoritma K-Means Untuk Clustering Data Kemiskinan Provinsi Banten Menggunakan Rapidminer. *CESS (Journal of Computer Engineering, System and Science)*, 5(2), 192. <https://doi.org/10.24114/cess.v5i2.18519>

Wahyudi, M., Masitha, Saragih, R., & Solikhun. (2020). *Data Mining: Penerapan Algoritma K-Means Clustering dan K-Medoids Clustering Oleh Mochamad Wahyudi, Masi*.

Wanto, A., Hasan Siregar, M. N., & Dkk. (2021). *Data Mining: Algoritma dan Implementasi* - Google Books. In *Yayasan Kita Menulis* (pp. 37-42 (202)). [https://www.google.co.id/books/edition/Data\\_Mining\\_Algoritma\\_dan\\_Implementasi/gAnfDwAAQBAJ?hl=id&gbpv=1&dq=Data+Mining+:+Algoritma+dan+Implementasi&printsec=frontcover](https://www.google.co.id/books/edition/Data_Mining_Algoritma_dan_Implementasi/gAnfDwAAQBAJ?hl=id&gbpv=1&dq=Data+Mining+:+Algoritma+dan+Implementasi&printsec=frontcover)