

Penerapan Metode K-Means untuk *Clustering* Data Anak Berdasarkan Kepemilikan Akta Kelahiran dan KIA

Diah Ayu Rahmayanti*, Ratna Juita**, Christian Dwi Suhendra***

Program Studi Teknik Informatika, Universitas Papua

*diahayurahmayanti@gmail.com, **r.juita@unipa.ac.id, ***c.suhendra@unipa.ac.id

ABSTRACT

Having a birth certificate and KIA is important for every child to have as proof of identity. However, in reality, the number of birth certificates and KIA in Manokwari Regency is not proportional to the number of children. This study aims to classify children's data based on the number of birth certificates and ownership of KIA in Manokwari district by using the K-Means clustering method to produce sub-district groupings with high coverage of non-ownership of birth certificates and KIA, so that it can assist the Department of Population and Civil Registry of Manokwari Regency in provide targeted service programs. This study uses the Davies Bouldin Index (DBI) method to determine the optimum number of clusters used. The optimum number of clusters obtained is 4 with the lowest DBI value of 0.069 so that 4 clusters are formed, namely cluster 0 with a fairly low coverage category, cluster 1 with a fairly high coverage category, cluster 2 in a high coverage category, and cluster 3 in a very high coverage category. The K-Means clustering method succeeded in grouping sub-districts in Manokwari district, as well as informing the number of sub-districts with high coverage that do not have birth certificates and KIA.

Keyword: *Clustering; Data Mining; Davies Bouldin Index; K-Means*

1. Pendahuluan

Pemenuhan hak atas identitas anak adalah pemenuhan hak asasi terhadap anak. Hak atas identitas seorang anak tegas dinyatakan dalam Pasal 5 dan 27 UU No. 23 Tahun 2002 tentang Perlindungan Anak, setiap anak berhak atas suatu nama sebagai identitas diri dan status kewarganegaraan yang dituangkan dalam akta kelahiran yang diterbitkan oleh Dinas Kependudukan dan Catatan Sipil (Disdukcapil). Saat ini dengan perkembangan yang ada, bukti identitas anak tidak terbatas pada akta kelahiran, tetapi juga termasuk Kartu Identitas Anak (KIA) yang diperuntukkan untuk anak usia dibawah 17 tahun yang belum menikah. Dalam pelaksanaannya, KIA dan akta kelahiran sangat penting untuk dimiliki setiap anak, bukan hanya untuk tujuan pendataan tetapi KIA bersama dengan akta kelahiran diharapkan mampu memberikan kepastian hukum dan jaminan atas pengakuan hak berikut dengan perlindungannya[1].

Namun, meskipun kepemilikan akta kelahiran anak dan KIA ini sangat penting pada kenyataannya jumlah kepemilikan akta kelahiran dan KIA di kabupaten Manokwari masih belum sebanding dengan jumlah anak yang ada. Berdasarkan data yang diperoleh di Disdukcapil kabupaten Manokwari ditemukan bahwa kepemilikan akta kelahiran pada anak mencapai 39.740 anak atau (60,7%) dari jumlah anak usia di bawah 17 tahun, sedangkan kepemilikan KIA baru mencapai 13.191 anak atau (21,65%) dari jumlah anak yang wajib KIA pada tahun 2021[2]. Apabila merujuk pada Rancangan Perencanaan Jangka Menengah Nasional (RPJMN) 2020-2024 terkait pemenuhan hak sipil anak atas kepemilikan akta kelahiran dengan target mencapai 100 persen dan juga target pemerintah pusat untuk setiap daerah minimal sudah sekitar 30 persen yang membuat KIA, maka cakupan kepemilikan akta kelahiran dan KIA masih jauh dari target yang telah ditetapkan.

Disdukcapil kabupaten Manokwari sebagai penyedia pelayanan administrasi kependudukan telah melakukan pelayanan masyarakat seperti, memberikan pelayanan keliling ke tiap-tiap daerah (jemput bola) di kabupaten Manokwari guna meningkatkan jumlah kepemilikan akta kelahiran dan KIA, akan tetapi dalam program pelayanan tersebut belum mendapatkan hasil yang maksimal karena hasil dari program pelayanan jemput bola tersebut masih jauh dari target yang telah ditetapkan pemerintah. Hal ini dikarenakan belum adanya pengelompokan daerah-daerah di kabupaten Manokwari berdasarkan tingkat pencatatan akta kelahiran anak dan KIA untuk membantu program pelayanan jemput bola tersebut dilakukan pada daerah yang tepat sasaran.

Dengan perkembangan teknologi infomasi yang semakin maju pada saat ini, banyak cabang ilmu komputer yang mampu memecahkan berbagai masalah, salah satunya adalah penerapan *data mining*

clustering dengan menggunakan metode K-Means. *Data mining* didefinisikan sebagai pembelajaran dari pengumpulan, pembersihan, analisis, dan mendapatkan pengetahuan yang sangat berguna dari sekumpulan data[3], secara sederhana *data mining* merupakan proses mengekstraksi atau menambang pengetahuan yang ada dari sekumpulan data[4] dan *clustering* adalah suatu teknik *data mining* dengan metode *unsupervised learning* yang tidak memerlukan contoh pelatihan berlabel kelas[5]. *Clustering* akan membagi data ke dalam *cluster* berdasarkan kesamaan[6]. Sedangkan, K-Means merupakan algoritma yang dapat mengklasterisasi data yang ada ke dalam beberapa *cluster*, dimana data dalam satu *cluster* memiliki karakteristik mirip satu sama lainnya dan memiliki karakteristik yang berbeda dengan data yang ada di dalam *cluster* lainnya[7]. Sehingga dalam hal ini dengan penerapan *data mining clustering* menggunakan algoritma K-Means dapat membantu Disdukcapil kabupaten Manokwari untuk mengelompokkan daerah-daerah di kabupaten Manokwari berdasarkan jumlah kepemilikan akta kelahiran anak dan KIA yang mana hasil *clustering* dapat digunakan sebagai bahan perencanaan dan evaluasi sasaran dalam pelayanan kepemilikan akta kelahiran dan KIA.

Penelitian-penelitian sebelumnya yang berkaitan dengan penelitian ini yaitu penelitian yang dilakukan oleh [8], dengan menjelaskan bagaimana penggunaan K-Means untuk mengelompokkan data kecamatan di Kabupaten Blora berdasarkan beberapa indikator kesehatan. Hasil pada penelitian ini berupa klasterisasi tahunan kecamatan kecamatan di Kabupaten Blora yang bisa dipakai untuk menggambarkan distribusi kecamatan berdasarkan profil penanganan dan pelayanan kesehatan masyarakat masing-masing. Selanjutnya penelitian yang dilakukan oleh [9], dengan tujuan dari penelitian ini adalah untuk memperoleh dan mengolah data COVID-19 yang terdapat pada Kaggle dengan menggunakan metode *data mining* yaitu K-Means *Clustering*. Dari hasil pengolahan data diperoleh dua klaster data, dan jumlah infeksi dan kematian pada klaster 2 lebih tinggi dibandingkan dengan klaster 1, sehingga wilayah klaster tersebut perlu diprioritaskan. Penelitian lain yang dilakukan oleh [10] dengan melakukan perbandingan metode-metode *clustering* dalam rangka mendapatkan metode terbaik sehingga bisa digunakan untuk pengelompokan Kecamatan di Kabupaten Bojonegoro, adapun hasil yang diperoleh metode *clustering* terbaik yaitu metode K-Means dengan *performance vector* sebesar -0,697 dalam membentuk 5 *cluster* yaitu *cluster* 1 dengan klasifikasi sangat aktif yang beranggotakan 4 kecamatan, *cluster* 3 dengan klasifikasi aktif beranggotakan 5 kecamatan, *cluster* 4 dengan klasifikasi cukup aktif beranggotakan 7 kecamatan, *cluster* 0 dengan klasifikasi kurang aktif beranggotakan 8 kecamatan, dan *cluster* 2 dengan klasifikasi tidak aktif yang beranggotakan 4 kecamatan.

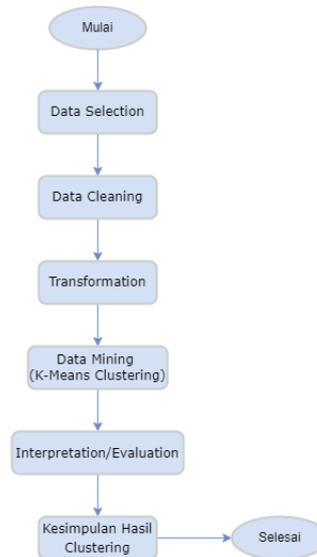
Berdasarkan penelitian-penelitian sebelumnya terdapat persamaan dengan penelitian ini yang bertujuan untuk melakukan pengelompokan daerah-daerah dengan penerapan *data mining clustering* menggunakan metode K-means. Namun terdapat beberapa perbedaan dimana pada penelitian sebelumnya daerah yang dikelompokkan berdasarkan data penanganan dan pelayanan kesehatan masyarakat, data kasus Covid-19 dan data kepemilikan dokumen kependudukan serta terdapat perbedaan pada penentuan jumlah *cluster* yang ditentukan. Sedangkan, pada penelitian ini dilakukan pengelompokan daerah berdasarkan jumlah kepemilikan akta kelahiran anak dan KIA serta penentuan jumlah *cluster* yang digunakan akan berdasarkan hasil evaluasi dengan menggunakan metode Davies Bouldin Index (DBI). DBI merupakan salah satu metode validitas internal dalam melakukan penilaian atau evaluasi terhadap suatu *cluster* dimana semakin rendah nilai DBI maka *cluster* yang dihasilkan akan semakin baik, yang berarti kemiripan antar data dalam suatu *cluster* akan semakin mirip[11].

Dalam penelitian ini, dilakukan penelitian dengan memanfaatkan data kependudukan pada Disdukpcapil kabupaten Manokwari yang diolah dengan penerapan *data mining clustering* menggunakan algoritma K-Means yang nantinya akan mengklasterisasi dan menggambarkan pengelompokan data anak berdasarkan jumlah kepemilikan akta kelahiran anak dan KIA pada tiap kecamatan di kabupaten Manokwari yang mana hasil *clustering* dapat digunakan sebagai bahan perencanaan dan evaluasi sasaran dalam pelayanan kepemilikan akta kelahiran dan KIA untuk mempercepat pencapaian target kepemilikan akta kelahiran dan KIA yang telah ditetapkan pemerintah.

2. Metode Penelitian

Penelitian ini dilakukan dengan cara mengumpulkan data, mengolah data, dan menyimpulkan hasil penelitian. Metode analisis data yang digunakan adalah *clustering* dengan menerapkan tahapan *Knowledge Discovery In Database* (KDD) untuk menganalisis data dalam penerapan *data mining* pada penelitian ini. KDD adalah proses mengekstrak pengetahuan atau informasi dari data tingkat bawah (basis data). Ekstraksi pengetahuan dari data mentah ini dilakukan dengan menerapkan metode *data mining*. KDD memiliki cakupan yang jauh lebih luas, di mana *data mining* merupakan salah satu tahapan terpenting KDD[12]. Secara keseluruhan terdapat lima tahapan dalam proses KDD yang dilakukan secara terurut yaitu *Data Selection* (Seleksi Data), *Data Cleaning* (Pembersihan Data), *Transformation* (Transformasi), *Data Mining* dan *Interpretation/Evaluation* (Interpresentasi/Evaluasi)[7]. Adapun tahapan analisis data yang dilakukan pada

penelitian ini disajikan pada Gambar 1.



Gambar 1. Tahapan Analisis Data

2.1. Studi Pustaka

Studi pustaka dilakukan dengan mempelajari literatur yang berkaitan dengan konsep data mining menggunakan metode *clustering* dan algoritma K-Means.

2.2. Pengumpulan Data

Untuk mengetahui informasi yang dibutuhkan, dilakukan pengumpulan data dengan teknik wawancara yang dilakukan secara langsung dengan pihak Disdukcapil kabupaten Manokwari untuk mendapatkan data yang dibutuhkan dalam penelitian.

2.3. Seleksi Data (*Data Selection*)

Seleksi data merupakan proses memilih kumpulan data, atau berfokus pada subset variabel atau sampel data, di mana penemuan akan dilakukan. Data hasil seleksi disimpan dalam suatu berkas, terpisah dari basis data operasional. Pada penelitian ini kriteria data yang digunakan adalah data anak pada tiap kecamatan di kabupaten Manokwari berdasarkan kepemilikan akta kelahiran dan KIA yang merupakan data angka.

2.4. Pembersihan Data (*Data Preprocessing/ Data Cleaning*)

Pembersihan data perlu dilakukan untuk menghilangkan kebisingan atau data yang tidak konsisten, mengumpulkan informasi yang diperlukan untuk memperhitungkan kebisingan, memutuskan strategi untuk menangani data yang hilang. Pada penelitian ini proses pembersihan data meliputi pembersihan data yang duplikat dan data yang kosong.

2.5. Transformasi (*Transformation*)

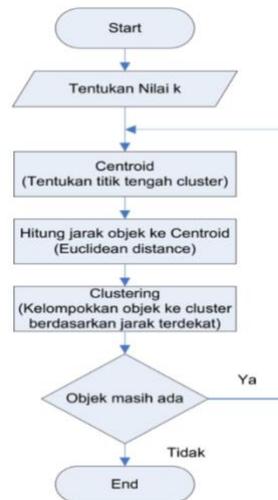
Pada proses transformasi, bentuk data akan diubah sehingga siap untuk dilakukan proses selanjutnya. Proses transformasi data pada penelitian ini akan mengubah data angka yang memiliki rentang nilai cukup jauh menjadi data yang memiliki rentang nilai antara 0 hingga 1 dengan menggunakan normalisasi data.

2.6. *Data Mining*

Data mining adalah proses mengekstraksi dan mengidentifikasi data yang terpilih dengan tujuan untuk mencari pola yang menarik dalam bentuk representasi tertentu atau sekumpulan representasi dengan menggunakan teknik atau metode tertentu seperti klasifikasi, prediksi, regresi maupun *clustering*. Teknik, metode, atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan. Tahap ini merupakan inti dari tahapan KDD yang dilakukan untuk menganalisis data.

2.6.1. K-Means

K-Means termasuk dalam salah satu metode *non-hierarchical clustering*. K-Means berusaha membagi data yang ada ke dalam kelompok-kelompok, dimana data dalam satu kelompok memiliki karakteristik yang mirip satu sama lain dan karakteristik yang berbeda dengan data pada kelompok lain[13]. Pada Gambar 2 disajikan diagram alir dari Algoritma K-Means.



Gambar 2. Diagram Alir Algoritma K-Means

Adapun langkah-langkah dari algoritma K-Means dijelaskan sebagai berikut[13]:

1. Menentukan berapa jumlah *cluster* (k) yang diinginkan pada dataset.
2. Tentukan nilai pusat (*centroid*).
Penentuan *centroid* pada tahap awal dilakukan secara acak, sedangkan pada tahap iterasi digunakan persamaan berikut ini:

$$V_{ij} = \frac{1}{N_i} = \sum_{k=0}^{N_i} X_{kj} \quad (1)$$

Keterangan:

- V_{ij} = *Centroid* rata-rata pada *cluster* ke – i untuk variabel ke – j
 N_i = Jumlah anggota *cluster* ke – i
 i, k = Indeks dari *cluster*
 j = Indeks variabel
 X_{kj} = Nilai data ke – k variabel ke – j untuk *cluster* tersebut

3. Menghitung jarak terpendek tiap data dengan *Centroid*. Gunakan rumus *Euclidean distance* (d) untuk menghitung jarak terpendek dengan *Centroid* tersebut. Adapun rumus tersebut adalah:

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

Keterangan:

- de = *Euclidean distance*
 i = Banyaknya objek
 (x, y) = Koordinat objek
 (s, t) = Koordinat *centroid*

4. Kelompokkan objek berdasarkan jarak ke *centroid* terpendek.
5. Ulangi langkah ke-3 hingga ke-4, lakukan iterasi hingga *centroid* bernilai optimal.

Dalam perhitungan K-Means jika ada data dengan perbedaan besaran angka yang cukup jauh maka data perlu dinormalisasi. Adapun rumus yang digunakan adalah:

$$X = \frac{X_i - X_{\min}}{X_{\max} - X_{\min}} \quad (3)$$

Keterangan:

X_i = Bobot atribut ke i

X_{\min} = Bobot minimum atribut i

X_{\max} = Bobot maksimum atribut i

Normalisasi data ini bertujuan agar nilai dari masing-masing data berada dalam rentang yang sama yakni rentang 0 sampai 1[14].

2.6.2. Davies Bouldin Index

Davies Bouldin Index (DBI) adalah salah satu metode validitas internal dalam melakukan penilaian atau evaluasi terhadap suatu *cluster*. Validitas Internal yang dilakukan DBI adalah seberapa baik *cluster* dilakukan dengan menghitung kuantitas dan fitur turunan dari data set. Tahapan dari perhitungan Davies Bouldin Index adalah sebagai berikut[11]:

Sum of square within cluster (SSW) sebagai metrik kohesi dalam sebuah *cluster* ke- i diformulasikan pada persamaan sebagai berikut:

$$SSW_i = \frac{1}{m_i} \sum_{j=1}^{m_i} d(x_j, c_i) \quad (4)$$

Keterangan:

SSW = *Sum of square within cluster*

m_i = Jumlah data dalam *cluster* ke - i

$d(x, c)$ = Jarak data x ke *centroid* c

Sementara metrik untuk separasi antara dua *cluster* digunakan formula *Sum of square between cluster* (SSB) dengan mengukur jarak antar *centroid* c_i dan c_j seperti pada persamaan berikut ini:

$$SSB_{i,j} = d(c_i, c_j) \quad (5)$$

Keterangan:

SSB = *Sum of square between cluster*

$d(c_i, c_j)$ = Jarak *centroid* c_i dengan *centroid* c_j

Langkah selanjutnya adalah menghitung nilai R_{ij} . R_{ij} adalah ukuran rasio seberapa baik nilai perbandingan antara *cluster* ke- i dan *cluster* ke- j . Nilainya didapatkan dari komponen kohesi dan separasi. *Cluster* yang baik adalah yang mempunyai kohesi yang sekecil mungkin dan separasi yang sebesar mungkin. R_{ij} di formulasikan dalam persamaan sebagai berikut:

$$R_{ij} = \frac{SSW_i + SSW_j}{SSB_{i,j}} \quad (6)$$

Keterangan:

R_{ij} = Nilai Perbandingan *cluster* i dengan *cluster* j

Setelah kita mendapatkan nilai SSW, SSB dan R_{ij} . Nilai Davies Bouldin Index (DBI) dapat dihitung menggunakan formula dalam persamaan berikut ini:

$$DBI = \frac{1}{K} \sum_{j=1}^K \max(R_{i,j}) \quad (7)$$

Keterangan:

k = Jumlah *cluster* yang digunakan

Seperti yang terlihat dari persamaan di atas, semakin rendah nilai DBI maka *cluster* yang dihasilkan akan semakin baik, yang berarti kemiripan antar data dalam suatu *cluster* akan semakin mirip.

2.7. Interpretasi/Evaluasi (*Interpretation/Evaluation*)

Dalam tahap ini, dilakukan penerjemahan pola-pola yang merupakan hasil dari tahap data mining. Pola informasi hasil dari data mining tersebut perlu ditampilkan atau divisualisasikan ke dalam bentuk pengetahuan yang mudah untuk dimengerti.

3. Hasil dan Analisis

Sumber data yang digunakan untuk penelitian ini diperoleh dari Buku Agregat Kependudukan Semester 2 Tahun 2021 yang berasal dari Disdukcapil Kabupaten Manokwari.

3.1. Seleksi Data

Data yang ada pada Buku Agregat Kependudukan Semester 2 Tahun 2021 ini tidak semuanya dipakai, oleh karena itu hanya data yang sesuai untuk penelitian. Pada tahap seleksi data ini, atribut yang digunakan untuk proses *data mining* adalah kecamatan, jumlah anak yang belum memiliki akta kelahiran, dan jumlah anak yang belum memiliki KIA. Adapun hasil seleksi data dalam penelitian ini dapat dilihat pada Tabel 1.

Tabel 1. Data Hasil Seleksi Data

No.	Kecamatan	Belum Akta Kelahiran	Belum KIA
1	Warmare	1066	1995
2	Prafi	1115	2438
3	Masni	1098	2866
4	Manokwari Barat	13866	25242
5	Manokwari Timur	2480	4012
6	Manokwari Utara	1162	1509
7	Manokwari Selatan	3684	7646
8	Tanah Rubuh	888	1321
9	Sidey	374	715
	Jumlah	25733	47744

3.2. Pembersihan Data

Data yang diperoleh Disdukcapil Kabupaten Manokwari tidak terdapat duplikasi data ataupun data yang kosong, sehingga data tersebut tidak perlu dilakukan pembersihan data.

3.3. Transformasi

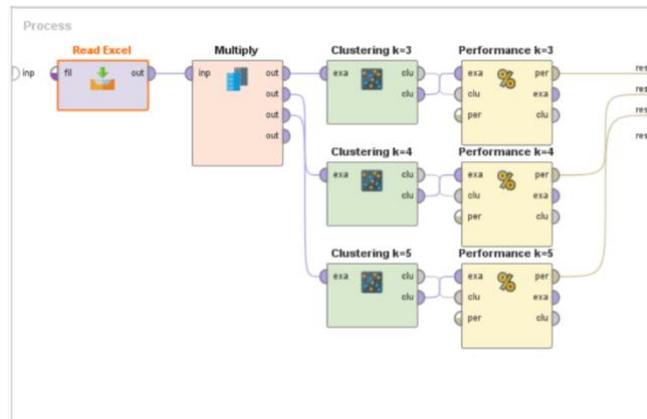
Pada tahapan transformasi dilakukan dengan normalisasi data agar nilai dari masing-masing data berada dalam rentang yang sama yaitu 1 hingga 0. Adapun hasil dari transformasi data pada penelitian ini dapat dilihat pada Tabel 2.

Tabel 2. Data Hasil Transformasi

No.	Kecamatan	Belum Akta Kelahiran	Belum KIA
1	Warmare	0.051289653	0.05218739
2	Prafi	0.054921435	0.07024911
3	Masni	0.053661429	0.08769927
4	Manokwari Barat	1	1
5	Manokwari Timur	0.156092499	0.13442329
6	Manokwari Utra	0.058404981	0.03237249
7	Manokwari Selatan	0.245330566	0.28258654
8	Tanah Rubuh	0.03809665	0.02470747
9	Sidey	0	0

3.4. Data Mining

Setelah data siap, dilanjutkan ke proses berikutnya yaitu tahapan *data mining*. Data akan diolah menggunakan algoritma K-Means. Pada penelitian ini *clustering* dilakukan dengan melakukan beberapa percobaan jumlah *cluster* yang berbeda-beda, yaitu dengan jumlah *cluster* 3, 4 dan 5. Nantinya jumlah *cluster* yang digunakan pada penelitian ini akan ditentukan berdasarkan hasil evaluasi dengan menggunakan metode Davies Bouldin Index (DBI). Penerapan *data mining* pada penelitian ini dilakukan dengan menggunakan RapidMiner sebagai perangkat lunak untuk melakukan proses *clustering*. Adapun penerapan metode *clustering* dengan algoritma K-Means pada Rapidminer di sajikan pada Gambar 3.



Gambar 3. Clustering K-Means menggunakan Rapidminer

3.5. Evaluasi/Interpresentasi

Evaluasi *cluster* dilakukan untuk mengetahui berapa jumlah *cluster* yang menghasilkan nilai DBI terendah, artinya jumlah *cluster* dengan tingkat kemiripan data yang paling mirip. Dari hasil percobaan dengan menggunakan 3, 4 dan 5 *cluster* pada RapidMiner, rekapitulasi nilai DBI yang diperoleh ditunjukkan dalam Tabel 3.

Tabel 3 Hasil Evaluasi Percobaan Cluster

Percobaan	Jumlah Cluster	DBI
1	3	0.185
2	4	0.069
3	5	0.161

Berdasarkan hasil rekapitulasi Davies Bouldin Index (DBI) yang ditunjukkan pada Tabel, dengan hasil dari nilai k=3 yaitu 0,185, k=4 yaitu 0,069 dan k=5 yaitu 0,161, maka nilai k yang akan dipilih dan menjadi jumlah *cluster* adalah pada percobaan kedua dengan nilai k=4 karena memiliki nilai DBI paling kecil yaitu 0,069, karena semakin kecil nilai DBI maka semakin optimum *cluster* yang dihasilkan, sehingga pada penelitian ini *cluster* yang akan digunakan sebanyak 4 *cluster*. Adapun hasil *clustering* K-Means yang diperoleh dengan menggunakan 4 *cluster* sebagai berikut:



Gambar 4 Hasil Davies Bouldin Index dengan 4 cluster

Pada Gambar 4 menunjukkan nilai DBI terendah yaitu dari *cluster* 4. Adapun hasil *cluster model* yang diperoleh ditunjukkan pada Gambar 5.

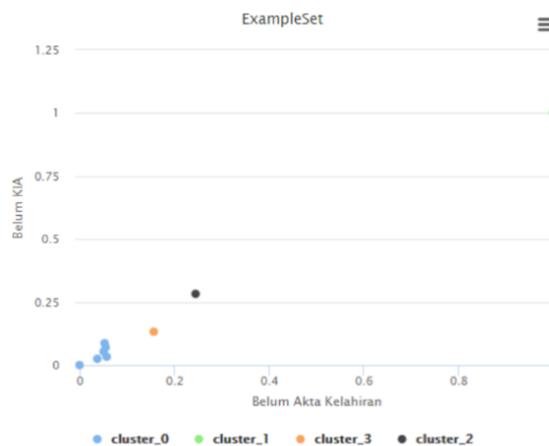


Gambar 5 Hasil *Clustering* dengan 4 *cluster*

Attribute	cluster_0	cluster_1	cluster_2	cluster_3
Belum Akta Kelahiran	0.043	1	0.245	0.156
Belum KIA	0.045	1	0.283	0.134

Gambar 6 *Centroid Table* Algoritma K-Means

Pada Gambar 6 merupakan nilai *centroid* yang diperoleh, *centroid* adalah titik atau tolak ukur dalam menentukan anggota *cluster* atau kelompok, karena proses *clustering* ini merupakan suatu proses pembentukan *cluster* atau kelompok dengan menentukan jarak terdekat dari masing-masing dataset. Berdasarkan hasil nilai *centroid* yang diperoleh dari proses *clustering* pada perangkat lunak RapidMiner menunjukkan *cluster* terendah adalah *cluster* 0, sedangkan *cluster* tertinggi terdapat pada *cluster* 1.



Gambar 7 Visualisasi *Scatter* Hasil *Clustering*

Pada Gambar 7 merupakan visualisasi hasil *clustering* menggunakan chart dengan jenis *scatter* yang menunjukkan nilai dengan bentuk titik-titik berwarna. Warna biru menyatakan *cluster* 0, warna hijau menyatakan *cluster* 1, warna oranye menyatakan *cluster* 3 dan warna hitam menyatakan *cluster* 2.

3.6. Pembahasan Hasil Analisis *Cluster*

Dari nilai DBI yang diperoleh didapatkan bahwa jumlah *cluster* optimum adalah 4, dimana berdasarkan hasil *clustering* K-Means yang dilakukan pada RapidMiner hasil *clustering* K-Means dapat dilihat pada Tabel 4.

Tabel 4. Hasil *Clustering* K-Means

No.	Kecamatan	Belum Akta Kelahiran	Belum KIA	Cluster
1	Warmare	1066	1995	cluster_0
2	Prafi	1115	2438	cluster_0
3	Masni	1098	2866	cluster_0
4	Manokwari Barat	13866	25242	cluster_1
5	Manokwari Timur	2480	4012	cluster_3
6	Manokwari Utara	1162	1509	cluster_0
7	Manokwari Selatan	3684	7646	cluster_2
8	Tanah Rubuh	888	1321	cluster_0
9	Sidey	374	715	cluster_0

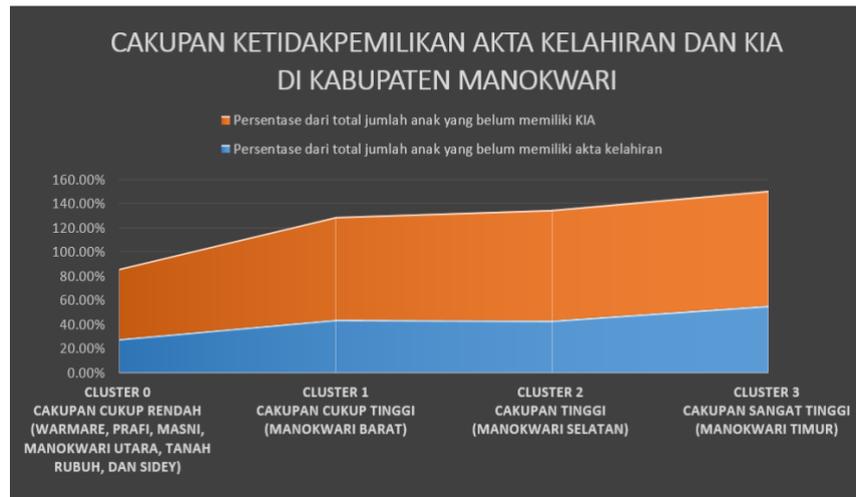
Dari hasil *clustering K-Means* dari setiap *cluster* yang telah terbentuk maka dapat dianalisa rincian total anak yang belum memiliki akta kelahiran dan KIA pada setiap *cluster* adalah sebagai berikut:

1. *Cluster* 0 terdiri dari kecamatan Warmare, Prafi, Masni, Manokwari Utara, Tanah Rubuh, dan Sidey dengan jumlah anak yang belum memiliki akta kelahiran sebanyak 5.703 anak dan jumlah anak yang belum memiliki KIA sebanyak 10.844 anak.
2. *Cluster* 1 terdiri dari kecamatan Manokwari Barat dengan dengan jumlah anak yang belum memiliki akta kelahiran sebanyak 13.866 anak dan jumlah anak yang belum memiliki KIA sebanyak 25.242 anak.
3. *Cluster* 2 terdiri dari kecamatan Manokwari Selatan dengan jumlah anak yang belum memiliki akta kelahiran sebanyak 3.684 anak dan jumlah anak yang belum memiliki KIA sebanyak 7.646 anak.
4. *Cluster* 3 terdiri dari kecamatan Manokwari Timur dengan jumlah anak yang belum memiliki akta kelahiran sebanyak 2.480 anak dan jumlah anak yang belum memiliki KIA sebanyak 4.012 anak.

Adapun hasil *clustering K-Means* yang terbentuk menghasilkan kelompok-kelompok kecamatan pada setiap *cluster* yang dimasukkan ke dalam kategori berdasarkan cakupan ketidakpemilikan akta kelahiran dan KIA. Dimana didapatkan pada *cluster* 0 dengan persentase jumlah anak yang belum memiliki akta kelahiran sebanyak 27.64% dan anak yang belum memiliki KIA sebanyak 58.40% masuk kedalam kategori cakupan ketidakpemilikan akta kelahiran anak dan KIA yang cukup rendah. *Cluster* 1 dengan persentase jumlah anak yang belum memiliki akta kelahiran sebanyak 44.01% dan anak yang belum memiliki KIA sebanyak 84.75% masuk kedalam kategori cakupan ketidakpemilikan akta kelahiran anak dan KIA yang cukup tinggi. *Cluster* 2 dengan persentase jumlah anak yang belum memiliki akta kelahiran sebanyak 43.47% dan anak yang belum memiliki KIA sebanyak 91.49% masuk kedalam kategori cakupan ketidakpemilikan akta kelahiran anak dan KIA yang tinggi. *Cluster* 3 dengan persentase jumlah anak yang belum memiliki akta kelahiran sebanyak 55.83% dan anak yang belum memiliki KIA sebanyak 94.93% masuk kedalam kategori cakupan ketidakpemilikan akta kelahiran anak dan KIA yang sangat tinggi. Adapun hasil kategori pada setiap *cluster* disajikan pada Tabel 5. serta visualisasi hasil kategori pada setiap *cluster* disajikan pada Gambar 8.

Tabel 5. Kategori Hasil *Clustering* K-Means

<i>Cluster</i>	Persentase dari total jumlah anak yang belum memiliki akta kelahiran	Persentase dari total jumlah anak yang belum memiliki KIA	Kategori
<i>Cluster</i> 0	27.64%	58.40%	Cakupan cukup rendah
<i>Cluster</i> 1	44.01%	84.75%	Cakupan cukup tinggi
<i>Cluster</i> 2	43.37%	91.49%	Cakupan tinggi
<i>Cluster</i> 3	55.83%	94.93%	Cakupan sangat tinggi



Gambar 8. Visualisasi Kategori Hasil *Clustering* K-Means

4. Kesimpulan

Penerapan data mining dalam *clustering* data anak menggunakan algoritma K-Means didapatkan jumlah cluster yang digunakan yaitu 4 *cluster* berdasarkan hasil evaluasi DBI. Diperoleh jumlah *cluster* 4 memiliki nilai DBI yang terendah yaitu 0,069. Dari 4 *cluster* yang terbentuk diperoleh *cluster* 0 termasuk kategori cakupan cukup rendah yang terdiri dari 6 kecamatan. *Cluster* 1 termasuk kategori cakupan cukup tinggi yang terdiri dari 1 kecamatan. *Cluster* 2 termasuk cakupan tinggi yang terdiri dari 1 kecamatan. Dan *cluster* 3 termasuk cakupan sangat tinggi yang terdiri dari 1 kecamatan. Adapun untuk mendapatkan hasil yang lebih baik, penelitian ini dapat dikembangkan dengan menggunakan data per-kelurahan agar penyebaran data dapat dilihat lebih detail dan dikembangkan dengan melakukan perbandingan dengan algoritma yang lain.

References

- [1] F. Alihar, "No Titleעלון הגמטע מצב, תמונת מצב: ענף הקיור", vol. 66, pp. 37–39, 2018, [Online]. Available: https://www.fairportlibrary.org/images/files/RenovationProject/Concept_cost_estimate_accepted_031914.pdf.
- [2] Dinas Kependudukan dan Pencatatan Sipil Kota Serang, "Data Agregat Kependudukan," 2020.
- [3] C. C. Aggarwal, *Data Mining: The Textbook*. Springer International Publishing, 2015.
- [4] A. Nur Khormarudin, "Teknik Data Mining: Algoritma K-Means Clustering," *J. Ilmu Komput.*, pp. 1–12, 2016, [Online]. Available: <https://ilmukomputer.org/category/datamining/>.
- [5] A. Jamal, A. Handayani, A. A. Septiandri, E. Ripmiatin, and Y. Effendi, "Dimensionality Reduction using PCA and K-Means Clustering for Breast Cancer Prediction," *Lontar Komput. J. Ilm. Teknol. Inf.*, no. December, p. 192, 2018, doi: 10.24843/lkjiti.2018.v09.i03.p08.
- [6] W. Purba, S. Tamba, and J. Saragih, "The effect of mining data k-means clustering toward students profile model drop out potential," *J. Phys. Conf. Ser.*, vol. 1007, no. 1, pp. 0–6, 2018, doi: 10.1088/1742-6596/1007/1/012049.
- [7] S. S. M. K. Indah Werdiningsih, S. S. M. K. Barry Nuqoba, and S. S. M. S. Muhammadun, *Data Mining Menggunakan Android, Weka, dan SPSS*. Airlangga University Press, 2020.
- [8] N. Cahyana and A. Aribowo, "Metode Data Mining K-Means Untuk Klasterisasi Data Penanganan Dan Pelayanan Kesehatan Masyarakat," *Semin. Nas. Inform. Medis*, no. 5, pp. 24–31, 2018.
- [9] R. A. Indraputra and R. Fitriana, "K-Means Clustering Data COVID-19," *J. Tek. Ind.*, vol. 10, no. 3, p. 3, 2020.
- [10] N. A. Sholikhah, "Studi Perbandingan Clustering Kecamatan di Kabupaten Bojonegoro Berdasarkan Keaktifan Penduduk Dalam Kepemilikan Dokumen Kependudukan dikeluarkan oleh population reference mempunyai data kependudukan yang Pemerintah melalui Dinas Kependudukan dan Catatan," vol. 1, no. 1, pp. 42–53, 2022.
- [11] S. Retno, "Peningkatan Akurasi Algoritma K-Means Dengan Clustering Purity Sebagai Titik Pusat Cluster Awal (Centroid)," 2019, [Online]. Available: <https://repositori.usu.ac.id/bitstream/handle/123456789/16782/177038001.pdf?sequence=1&isAllowed=y>.
- [12] C. Kenneth and O. Chinecherem, "Knowledge Discovery in Databases (KDD): An Overview," vol. 15, no. 12, pp. 13–16, 2017.
- [13] A. Wanto, *Data Mining : Algoritma dan Implementasi*. Yayasan Kita Menulis, 2020.
- [14] R. Prasojo, Y. R. W. Utami, and R. T. Vulandari, "Implementasi K-Means Clustering Pada Pengelompokan Potensi Kerjasama Pelanggan," *J. Teknol. Inf. dan Komun.*, vol. 7, no. 2, 2019, doi: 10.30646/tikomsin.v7i2.435.