



# Analisis Data Mining dengan Metode K-Means Clustering Dalam Pengelompokan Penggunaan Alat Kontrasepsi

Rahmad Rahmad\*, Sarjon Defit, Rini Sovia

Fakultas Ilmu Komputer, Magister Teknik Informatika, Universitas Putra Indonesia YPTK Padang, Padang, Indonesia

Email: <sup>1,\*</sup>iniemail.rahmad@gmail.com, <sup>2</sup>sarjon\_defit@upiptyk.ac.id, <sup>3</sup>rini\_sovia@upiptyk.ac.id

Email Penulis Korespondensi: iniemail.rahmad@gmail.com

**Abstrak**—Keluarga Berencana (KB) merupakan upaya strategis pemerintah dalam menekan laju pertumbuhan penduduk dan meningkatkan kualitas hidup masyarakat. Adanya beragam jenis alat kontrasepsi dapat menunda kehamilan yang tidak diinginkan, termasuk pada wanita yang menghadapi peningkatan risiko kehamilan. Berdasarkan hal tersebut maka, penelitian ini bertujuan untuk melakukan pengelompokan penggunaan alat kontrasepsi. Metode *K-Means Clustering* adalah salah satu algoritma *unsupervised learning* yang digunakan untuk mengelompokkan data ke dalam sejumlah kluster berdasarkan kemiripan karakteristik. Algoritma ini bekerja dengan cara meminimalkan jarak antara data dengan pusat kluster (*centroid*) dan kelebihan *K-Means* adalah kesederhanaan dan kecepatan prosesnya dalam mengolah data berukuran besar. Variabel penelitian ini menggunakan data dari hasil Pendataan Keluarga Tahun 2024 BKKBN Perwakilan Provinsi Sumatera Barat di Kabupaten Pasaman Barat. Berdasarkan penerapan metode *K-Means Clustering* pada data penggunaan alat kontrasepsi, diperoleh pengelompokan menjadi tiga kluster: kluster 0 (C0) dengan tingkat penggunaan alat kontrasepsi MKJP sedang yang mencakup 24 kelurahan/desa, kluster 1 (C1) dengan tingkat penggunaan MKJP tinggi yang terdiri atas 6 kelurahan/desa, serta kluster 2 (C2) dengan tingkat penggunaan MKJP rendah yang meliputi 60 kelurahan/desa dari total 90 kelurahan/desa di Kabupaten Pasaman Barat. Penelitian ini memberikan kontribusi dalam bentuk model analisis berbasis data mining yang mampu mengelompokkan pola penggunaan alat kontrasepsi secara lebih terstruktur dan objektif. Dengan menerapkan metode *K-Means Clustering*, penelitian ini menghasilkan informasi yang dapat digunakan untuk mengidentifikasi karakteristik tiap kelompok pengguna, sehingga instansi terkait dapat merancang strategi penyuluhan dan distribusi alat kontrasepsi yang lebih tepat sasaran.

**Kata Kunci:** Data Mining; K-Means Clustering; Alat Kontrasepsi; Pengelompokan; Keluarga Berencana

**Abstract**—Family Planning (KB) is a strategic government effort to suppress population growth and improve the quality of life. The availability of various types of contraceptives can delay unwanted pregnancies, including in women facing increased pregnancy risks. Based on this, this study aims to cluster contraceptive use. The K-Means Clustering method is an unsupervised learning algorithm used to group data into several clusters based on similar characteristics. This algorithm works by minimizing the distance between the data and the cluster center (centroid). The advantages of K-Means are its simplicity and speed in processing large data. This research variable uses data from the 2024 Family Data Collection of the BKKBN Representative Office of West Sumatra Province in West Pasaman Regency. Based on the application of the K-Means Clustering method to the contraceptive use data, the grouping is obtained into three clusters: low use of MKJP contraceptives, moderate use of MKJP contraceptives, and high use of MKJP contraceptives. This study contributes in the form of a data mining-based analysis model that is able to group contraceptive use patterns in a more structured and objective manner. By applying the K-Means Clustering method, this study produces information that can be used to identify the characteristics of each user group, so that relevant agencies can design more targeted contraceptive counseling and distribution strategies.

**Keywords:** Data Mining; K-Means Clustering; Contraceptives; Clustering; Family Planning

## 1. PENDAHULUAN

Keluarga Berencana (KB) merupakan upaya strategis pemerintah dalam menekan laju pertumbuhan penduduk dan meningkatkan kualitas hidup masyarakat. Tujuan dan manfaat dari KB adalah memperlambat pertumbuhan populasi, mengatur jarak atau menunda kehamilan dan mengurangi angka kelahiran [1]. Adanya beragam jenis alat kontrasepsi dapat menunda kehamilan yang tidak diinginkan, termasuk pada wanita yang menghadapi peningkatan risiko kehamilan [2]. Salah satu langkah intervensi yang menjadi fokus dalam program KB adalah peningkatan penggunaan Alat Kontrasepsi Jangka Panjang (MKJP), seperti Metode Operasi Wanita (MOW), Metode Operasi Pria (MOP), *Intrauterine Device* (IUD), dan Implan yang dinilai lebih efektif dan efisien dibandingkan dengan metode kontrasepsi jangka pendek [3].

Meskipun demikian, tingkat penggunaan MKJP di berbagai daerah masih belum merata. Beberapa wilayah menunjukkan tingkat penggunaan MKJP yang tinggi, sementara di wilayah yang lain masih bergantung pada metode kontrasepsi jangka pendek. Ketimpangan ini dapat disebabkan oleh berbagai faktor, seperti tingkat pendidikan, sosial ekonomi, akses layanan kesehatan, serta budaya dan norma lokal [4].

Menurut penelitian yang dilakukan oleh Anwar et al [5], partisipasi wanita usia subur dalam program Keluarga Berencana menggunakan Metode Kontrasepsi Jangka Panjang (MKJP) dipengaruhi secara signifikan oleh tingkat pengetahuan dan dukungan suami. Penelitian yang sama juga mengidentifikasi bahwa faktor-faktor determinan penggunaan MKJP meliputi karakteristik demografis wanita (usia, tingkat pendidikan, dan jumlah paritas), pengetahuan serta sikap terhadap kontrasepsi, tingkat pendapatan keluarga, dukungan suami, peran tenaga kesehatan, serta tingkat kepuasan terhadap pelayanan yang diterima [6].

Implementasi pengelompokan penggunaan alat kontrasepsi pada penelitian ini dilakukan dengan menerapkan konsep *Machine Learning* (ML). *Machine Learning* merupakan salah satu cabang dari kecerdasan buatan (*Artificial Intelligence*) yang berfokus pada pengembangan algoritma dan model yang mampu mempelajari pola dari data serta



meningkatkan kinerjanya secara otomatis tanpa memerlukan pemrograman eksplisit untuk setiap tugas tertentu [7]. Dalam konteks penelitian ini, metode yang digunakan termasuk ke dalam kategori *unsupervised learning*.

*Unsupervised Learning* adalah pendekatan *machine learning* yang mempelajari pola dan struktur data tanpa menggunakan label atau variabel target [8]. Dalam konteks penelitian “Analisis Data Mining dengan Metode K-Means Clustering dalam Pengelompokan Penggunaan Alat Kontrasepsi”, metode ini digunakan untuk mengidentifikasi kelompok atau segmen pengguna berdasarkan kesamaan karakteristik.

Algoritma *K-Means Clustering*, sebagai salah satu metode *unsupervised learning* [9], bekerja dengan membagi data ke dalam sejumlah kluster berdasarkan kemiripan atribut jenis alat kontrasepsi yang digunakan. Prosesnya melibatkan perhitungan jarak setiap data terhadap pusat kluster (*centroid*) dan pembaruan posisi *centroid* secara iteratif hingga tercapai kesetimbangan [10].

Riwayat penelitian terdahulu dari metode *K-Means Clustering* dalam menyelesaikan masalah pengelompokan Pengelompokan Pendaftaran Mahasiswa Baru (Studi Kasus Universitas Abdurrab) [11]. Penelitian ini mengimplementasikan algoritma *K-Means Clustering* untuk mengelompokkan data pendaftaran mahasiswa baru di Universitas Abdurrab berdasarkan atribut demografis dan sosial ekonomi. Melalui evaluasi menggunakan *Davies-Bouldin Index* (DBI), diperoleh nilai optimal pada delapan *cluster*. Analisis hasil *clustering* menunjukkan bahwa algoritma ini efektif dalam mengidentifikasi pola karakteristik calon mahasiswa, yang diperkuat dengan visualisasi korelasi antar variabel melalui heatmap. Penelitian selanjutnya Perancangan Aplikasi Data Mining Untuk Menentukan Tingkat Kelarisan Produk Menggunakan Metode *Clustering* Dengan Algoritma K-Means [12]. Penelitian ini berhasil mengelompokkan produk susu bayi di Rima Market menjadi dua *cluster* menggunakan algoritma K-Means, yaitu produk laris dan tidak laris, berdasarkan data stok awal dan jumlah penjualan. Hasil ini diimplementasikan dalam bentuk aplikasi berbasis web, yang membantu pihak swalayan mengidentifikasi produk yang diminati konsumen serta mendukung pengambilan keputusan dalam pengelolaan stok dan strategi penjualan.

Penelitian selanjutnya Penerapan Metode *K-Means Clustering* untuk Pemetaan Pengelompokan Lahan Produksi Tandan Buah Segar [13]. Penelitian ini bertujuan untuk memetakan tingkat produktivitas lahan kelapa sawit di PT Surya Intisari Raya menggunakan algoritma *K-Means Clustering*. Dengan menggunakan delapan variabel utama (luas lahan, jumlah pohon, hasil panen, brondolan, curah hujan, pupuk, target produksi, dan persentase pencapaian), data 216 blok lahan dikelompokkan ke dalam tiga *cluster*: produktivitas tinggi, sedang, dan rendah. Penelitian selanjutnya Penerapan Metode *K-Means Clustering* Untuk Menentukan Status Gizi Baik Dan Gizi Buruk Pada Balita [14]. Penelitian ini mengimplementasikan algoritma *K-Means Clustering* untuk mengelompokkan status gizi balita di Kabupaten Rokan Hulu berdasarkan data tinggi dan berat badan. Melalui proses normalisasi dan iterasi hingga tahap konvergen, diperoleh dua *cluster* utama, yakni *cluster* gizi buruk (8 balita) dan *cluster* gizi baik (7 balita). Validasi menggunakan aplikasi RapidMiner menunjukkan hasil yang konsisten dengan perhitungan manual. Hasil ini menunjukkan bahwa metode K-Means efektif dalam mengidentifikasi pola status gizi balita, serta berpotensi menjadi alat bantu analisis dalam mendukung pengambilan keputusan di bidang kesehatan masyarakat, khususnya untuk intervensi dan pemantauan gizi.

Berdasarkan uraian sebelumnya, tujuan penelitian ini adalah untuk menganalisis data penggunaan alat kontrasepsi dengan menerapkan metode K-Means Clustering sebagai teknik data mining dalam proses pengelompokan. Melalui penerapan algoritma ini, penelitian diharapkan dapat menghasilkan pemetaan peserta KB aktif yang strategis sekaligus mengidentifikasi bentuk intervensi yang tepat di Kabupaten Pasaman Barat. Selain itu, implementasi metode K-Means pada proses klusterisasi kelurahan/desa diharapkan dapat mengoptimalkan perencanaan strategi edukasi serta distribusi Metode Kontrasepsi Jangka Panjang (MKJP) secara lebih efektif dan tepat sasaran. Keunikan penelitian ini terletak pada fokus penerapan K-Means Clustering dalam konteks pengelompokan penggunaan alat kontrasepsi, yang masih jarang dilakukan, sehingga mampu memberikan kontribusi baru dalam pemetaan pola pemakaian kontrasepsi dan mendukung perencanaan program keluarga berencana berbasis data.

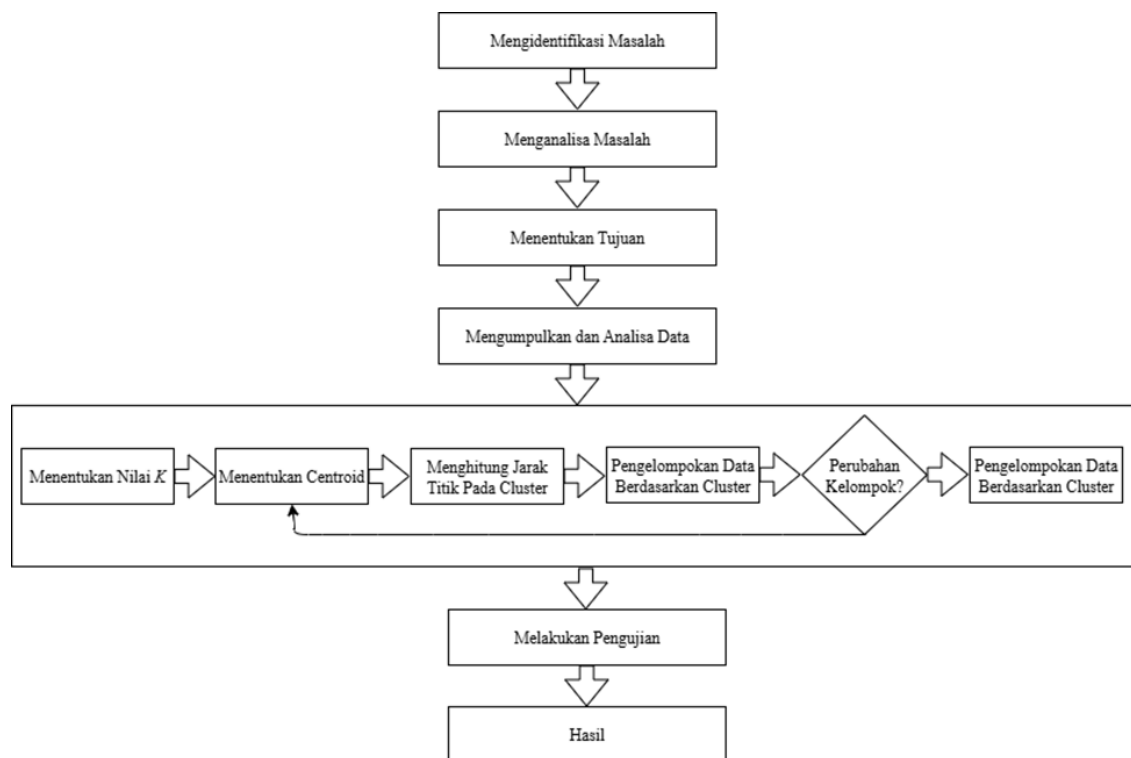
## 2. METODOLOGI PENELITIAN

Penelitian ini dilaksanakan dengan menggunakan pendekatan kuantitatif yang berorientasi pada pengolahan data secara numerik, sehingga memungkinkan dilakukannya analisis yang objektif dan terukur terhadap fenomena yang diteliti. Metode yang diterapkan adalah data mining dengan memanfaatkan algoritma K-Means Clustering, yakni salah satu teknik pengelompokan data yang efektif untuk menemukan pola atau karakteristik tertentu dari sekumpulan data yang besar dan kompleks. Penerapan algoritma ini ditujukan untuk mengelompokkan data penggunaan alat kontrasepsi Jangka Panjang ke dalam tiga kluster, sehingga dapat diperoleh gambaran yang lebih jelas mengenai kecenderungan dan variasi penggunaan alat kontrasepsi di lapangan. Adapun data yang dianalisis dalam penelitian ini bersumber dari Pendataan Keluarga Tahun 2024 yang dihimpun oleh Badan Kependudukan dan Keluarga Berencana Nasional (BKKBN) Perwakilan Provinsi Sumatera Barat, dengan fokus wilayah penelitian pada Kabupaten Pasaman Barat sebagai salah satu daerah yang menjadi lokus penting dalam perencanaan program kependudukan dan keluarga berencana.

### 2.1 Kerangka Kerja Penelitian

Kerangka kerja penelitian berfungsi sebagai panduan sistematis dalam pelaksanaan suatu studi, karena di dalamnya tercakup rangkaian langkah-langkah yang terencana dan saling berhubungan. Kerangka ini tidak hanya menggambarkan alur berpikir peneliti sejak tahap awal, yakni identifikasi masalah, tetapi juga mencakup proses perumusan tujuan,

penyusunan metode, pengumpulan dan pengolahan data, hingga tahap pengujian, analisis hasil, serta penarikan kesimpulan. Dengan adanya kerangka kerja yang jelas, penelitian dapat dilaksanakan secara lebih terarah, terstruktur, dan efektif, sehingga hasil yang diperoleh memiliki tingkat keandalan yang tinggi dan dapat dipertanggungjawabkan secara ilmiah. Adapun kerangka kerja penelitian yang dimaksud secara visual dapat dilihat pada Gambar 1



Gambar 1. Kerangka Kerja Penelitian

Penelitian ini diawali dengan mengidentifikasi masalah untuk mengetahui permasalahan yang terjadi pada penggunaan alat kontrasepsi, khususnya yang berkaitan dengan distribusi dan pemanfaatan MKJP. Selanjutnya dilakukan analisis masalah untuk memahami faktor-faktor yang memengaruhi pemakaian alat kontrasepsi dan potensi segmentasi data. Tahap berikutnya adalah menentukan tujuan penelitian yang menjadi arah dari keseluruhan proses analisis. Setelah itu dilakukan pengumpulan dan analisis data sebagai dasar penerapan algoritma *K-Means Clustering*.

Proses inti algoritma K-Means dimulai dengan menentukan jumlah kluster ( $K$ ) yang optimal, dilanjutkan dengan menentukan titik pusat awal (*centroid*) [15]. Kemudian, sistem menghitung jarak setiap data terhadap *centroid* menggunakan metode *Euclidean Distance*, untuk menentukan kedekatan masing-masing data dengan kluster yang ada. Data kemudian dikelompokkan berdasarkan kluster terdekat.

Setelah pengelompokan awal, dilakukan pengecekan apakah terjadi perubahan keanggotaan kluster. Jika ada perubahan, proses akan mengulang perhitungan *centroid* dan pengelompokan hingga komposisi kluster stabil. Tahap berikutnya adalah melakukan pengujian terhadap hasil klusterisasi untuk menilai kualitas dan keakuratannya. Proses ini menghasilkan output akhir berupa pemetaan penggunaan alat kontrasepsi yang dapat dijadikan dasar strategi distribusi dan edukasi yang lebih tepat sasaran.

## 2.2 Knowledge Discovery in Database (KDD)

*Knowledge Discovery in Databases* (KDD) adalah proses untuk menemukan pengetahuan baru yang bermanfaat dan dapat dipahami manusia dari kumpulan data yang besar [16]. Menurut Fayyad, KDD mencakup serangkaian tahapan mulai dari pemilihan data, pembersihan data, transformasi data, proses data mining, hingga evaluasi dan interpretasi hasil [17]. Proses ini sangat penting dalam era *big data* saat ini karena membantu pengguna menggali informasi tersembunyi yang tidak langsung terlihat dari data mentah.

## 2.3 Data Mining

Data Mining (penggalian data) adalah proses ekstraksi informasi yang berharga dan berguna dari data dalam jumlah besar dan kompleks [18]. Tujuannya adalah untuk menemukan pola-pola, hubungan, dan tren yang signifikan dalam data yang sulit dipahami atau tidak terdeteksi hanya dengan melihat secara kasat mata. Proses data mining melibatkan penggunaan teknik matematika, statistika, kecerdasan buatan (*artificial intelligence*), dan *machine learning* untuk mengidentifikasi pola-pola tersembunyi dalam data dan menggunakannya untuk membuat prediksi atau membuat keputusan penting [19].

## 2.4 Algoritma K-Means Clustering



Algoritma *K-Means Clustering* merupakan metode pengelompokan data non-hirarki yang membagi sekumpulan data ke dalam beberapa kluster [9]. Data yang memiliki karakteristik serupa akan dikelompokkan dalam kluster yang sama, sedangkan data dengan karakteristik berbeda ditempatkan pada kluster yang berbeda. Dengan demikian, variasi antar data dalam satu kluster menjadi relatif kecil. Metode ini bekerja dengan membagi data ke dalam kelompok-kelompok berbeda melalui algoritma partisi, sehingga jarak antara setiap data dengan pusat klasternya dapat diminimalkan [20]. Tahapan dalam algoritma *K-Means* adalah sebagai berikut:

- Tentukan berapa jumlah kluster yang ingin ditetapkan sebagai pusat *cluster*  $k$
- Menggunakan jarak *Euclidean Distance* dan kemudian menghitung setiap data ke pusat *cluster*.

$$d_{ik} = \sqrt{\sum_j^m (C_{ij} - C_{kj})^2} \quad (1)$$

Simbol  $d_{ik}$  merepresentasikan jarak antara data ke- $i$  dengan centroid pada kluster ke- $k$ . Nilai  $m$  menunjukkan jumlah variabel atau dimensi data yang digunakan dalam perhitungan. Sementara itu,  $C_{ij}$  adalah nilai data ke- $i$  pada variabel ke- $j$ , sedangkan  $C_{kj}$  menggambarkan nilai centroid kluster ke- $k$  pada variabel ke- $j$ .

- Seluruh komponen ini digunakan dalam proses perhitungan jarak guna menentukan kedekatan suatu data terhadap centroid kluster tertentu.

$$D(i,j) = \sqrt{(X_{1i} - X_{1j})^2 + (X_{2i} - X_{2j})^2 + \dots + (X_{ki} - X_{kj})^2} \quad (2)$$

Simbol  $D(i,j)$  menyatakan jarak antara data ke- $i$  dengan centroid kluster ke- $j$ . Notasi  $(X_{1i}, X_{2i}, \dots, X_{ki})$  merepresentasikan nilai atribut atau variabel dari data ke- $i$ , sedangkan  $(X_{1j}, X_{2j}, \dots, X_{kj})$  menggambarkan nilai atribut atau variabel dari centroid kluster ke- $j$ . Adapun  $k$  menunjukkan jumlah variabel atau dimensi data yang digunakan dalam proses perhitungan jarak tersebut.

- Menghitung pusat *cluster* dengan menggunakan persamaan

$$C_{kj} = \frac{\sum_{i=1}^p x_{ij}}{p} \quad (3)$$

Simbol  $C_{kj}$  menyatakan nilai centroid untuk kluster ke- $k$  pada variabel ke- $j$ . Nilai ini dihitung berdasarkan rata-rata dari seluruh data pada kluster tersebut. Adapun  $x_{ij}$  merepresentasikan nilai data ke- $i$  pada variabel ke- $j$ . Jumlah data yang termasuk ke dalam kluster ke- $k$  dituliskan sebagai  $p$ , sedangkan notasi  $\sum_{i=1}^p x_{ij}$  menunjukkan total nilai variabel ke- $j$  dari seluruh data yang berada dalam kluster ke- $k$ . Dengan demikian, centroid pada variabel ke- $j$  untuk kluster ke- $k$  diperoleh dengan membagi jumlah seluruh nilai variabel ke- $j$  dengan banyaknya data pada kluster tersebut.

- Menghitung kembali setiap objek yang menggunakan pusat kluster (*centroid*) yang baru. Dalam proses ini, apabila pusat kluster tidak berubah maka proses klasterisasi selesai, namun apabila mengalami perubahan maka perlu dilakukan pengulangan kembali ke langkah 3 sampai pusat kluster tidak berubah. Akan tetapi, klasterisasi juga dapat dihentikan dengan menggunakan batas maksimal terhadap suatu proses klasterisasi.

- Hitung *Davies-Bouldin Index* (DBI)

Langkah selanjutnya adalah menghitung nilai *Davies-Bouldin Index* (DBI) dengan menggunakan persamaan berikut:

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{j \neq i} \left( \frac{s_i + s_j}{d(c_i, c_j)} \right) \quad (4)$$

*Davies-Bouldin Index* (DBI) merupakan ukuran yang digunakan untuk menilai kualitas hasil klasterisasi. Simbol  $k$  menunjukkan jumlah kluster yang terbentuk. Nilai  $s_i$  menggambarkan rata-rata jarak antara setiap data pada kluster ke- $i$  dengan centroid-nya, yang disebut sebagai intra-cluster distance, sedangkan  $s_j$  adalah rata-rata jarak antara data pada kluster ke- $j$  dengan centroid-nya. Selanjutnya,  $c_i$  merepresentasikan centroid kluster ke- $i$  dan  $c_j$  merupakan centroid kluster ke- $j$ . Jarak antara centroid kluster ke- $i$  dengan centroid kluster ke- $j$  dinyatakan sebagai  $d(c_i, c_j)$  atau inter-cluster distance. Adapun notasi  $\max (j \neq i)$  berarti memilih nilai maksimum dari rasio antara jumlah intra-cluster distance dengan inter-cluster distance pada setiap pasangan kluster  $i$  dan  $j$ .

### 3. HASIL DAN PEMBAHASAN

Berdasarkan tahapan penelitian yang telah dilaksanakan, proses pengelompokan atau klasterisasi penggunaan alat kontrasepsi di Kabupaten Pasaman Barat dilakukan dengan memanfaatkan metode *K-Means Clustering*, yaitu salah satu algoritma dalam data mining yang berfungsi untuk membagi sekumpulan data ke dalam beberapa kelompok berdasarkan tingkat kemiripan karakteristiknya. Penerapan metode ini memungkinkan peneliti memperoleh pola yang lebih terstruktur mengenai distribusi dan kecenderungan penggunaan alat kontrasepsi di wilayah penelitian. Data yang digunakan dalam proses klasterisasi ini mencakup berbagai variabel penting, antara lain jumlah peserta KB aktif, jenis alat kontrasepsi yang digunakan oleh peserta, serta penyebarannya di tingkat kelurahan atau desa. Melalui data tersebut, dapat diidentifikasi perbedaan maupun kesamaan karakteristik antarwilayah, sehingga hasil analisis lebih mudah dipahami dan bermanfaat untuk perencanaan program Keluarga Berencana yang lebih tepat sasaran. Adapun data rinci mengenai jumlah peserta KB aktif berdasarkan kelurahan atau desa di Kabupaten Pasaman Barat secara lebih jelas dapat dilihat pada Tabel 1..



**Tabel 1.** Data Kelurahan/Desa dengan Jumlah Peserta KB Aktif

No	Kelurahan/ Desa	Implan	IUD	Vasektomi	Tubektomi
1	Aia Bangih	2897	382	0	312
2	Ujung Gading	568	519	24	450
3	Brastagi Ujung Gading	50	83	0	0
4	Tampus Damai Ujung Gading	189	204	14	89
5	Taluak Ambun Ujung Gading	65	103	6	21
⋮	⋮	⋮	⋮	⋮	⋮
85	Jambak Selatan	169	130	12	103
86	Pujorahayu	1868	157	0	202
87	Sasak	444	29	17	180
88	Padang Harapan	342	26	12	151
89	Ranah Pasisie	268	12	17	321
90	Maligi	231	0	0	39

Setelah data dikumpulkan dan dianalisa maka proses selanjutnya:

- Menentukan jumlah kluster C di mana peneliti akan membagi jumlah pengelompokkan kedalam 3 kluster yaitu: tinggi, sedang dan rendah.
- Menentukan nilai titik tengah (*centroid*) secara random sebanyak jumlah kluster C. Berikut Tabel 2 penentuan *centroid* awal.

**Tabel 2.** Centroid Awal

DATA	Centroid Awal	Implan	Iud	Vasektomi	Tubektomi
Data ke 3	C0	50	83	0	0
Data ke 34	C1	3024	720	36	242
Data ke 22	C2	1078	47	0	354

Setelah nilai titik tengah (*centroid*) secara random, langkah selanjutnya dilakukan penghitungan menggunakan *centroid* awal pada Tabel 2.

- Menghitung jarak dari *centroid* awal

Tahap selanjutnya adalah menghitung jarak antara setiap titik data dengan titik pusat kluster (*centroid*). Setiap data dibandingkan dengan seluruh *centroid* yang telah ditentukan pada tahap sebelumnya untuk mengetahui jarak terdekat, dengan menggunakan rumus *Euclidean Distance*. Adapun proses perhitungannya adalah sebagai berikut:

$$D_{(1,0)} = \sqrt{(2897 - 50)^2 + (382 - 83)^2 + (0 - 0)^2 + (312 - 0)^2}$$

$$= 2.879,61$$

$$D_{(2,0)} = \sqrt{(568 - 50)^2 + (519 - 83)^2 + (24 - 0)^2 + (450 - 0)^2}$$

$$= 813,32$$

$$D_{(3,0)} = \sqrt{(50 - 50)^2 + (83 - 83)^2 + (0 - 0)^2 + (0 - 0)^2}$$

$$= 0$$

$$D_{(1,1)} = \sqrt{(2897 - 3024)^2 + (382 - 720)^2 + (0 - 36)^2 + (312 - 242)^2}$$

$$= 369,55$$

$$D_{(2,1)} = \sqrt{(568 - 3024)^2 + (519 - 720)^2 + (24 - 36)^2 + (450 - 242)^2}$$

$$= 2.473,00$$

$$D_{(3,1)} = \sqrt{(50 - 3024)^2 + (83 - 720)^2 + (0 - 36)^2 + (0 - 242)^2}$$

$$= 3.051,28$$

$$D_{(1,2)} = \sqrt{(2897 - 1078)^2 + (382 - 47)^2 + (0 - 0)^2 + (312 - 354)^2}$$

$$= 1.850,07$$

$$D_{(2,2)} = \sqrt{(568 - 1078)^2 + (519 - 47)^2 + (24 - 0)^2 + (450 - 354)^2}$$

$$= 701,91$$

$$D_{(3,2)} = \sqrt{(50 - 1078)^2 + (83 - 47)^2 + (0 - 0)^2 + (0 - 354)^2}$$

$$= 1.087,84$$

- Pengelompokkan jarak setiap data ke *centroid*



Pada tahap ini dilakukan pengelompokan data berdasarkan jarak terdekat antara setiap objek dengan titik pusat kluster (*centroid*). Diperoleh jarak antar data dengan pusat kluster (C0), pusat kluster (C1) dan pusat kluster (C2) sehingga hasil perhitungan dari keseluruhan data dengan masing-masing kluster dapat dilihat pada Tabel 3.

**Tabel 3.** Hasil Nilai Kluster Iterasi 1

No	Nama Kelurahan/Desa	C0	C1	C2	Jarak Terdekat
1	Aia Bangih	2.879,61	369,55	1.850,07	369,55
2	Ujung Gading	813,32	2.473,00	701,91	701,91
3	Brastagi Ujung Gading	-	3.051,28	1.087,84	-
4	Tampus Damai Ujung Gading	205,13	2.885,72	940,95	205,13
5	Taluak Ambun Ujung Gading	33,20	3.030,86	1.067,82	33,20
⋮	⋮	⋮	⋮	⋮	⋮
85	Jambak Selatan	164,69	2.918,74	946,74	164,69
86	Pujorahayu	1.830,68	1.286,93	811,98	811,98
87	Sasak	436,85	2.671,72	657,91	436,85
88	Padang Harapan	333,85	2.771,93	763,87	333,85
89	Ranah Pasisie	394,84	2.846,65	811,61	394,84
90	Maligi	202,91	2.891,67	904,90	202,91

- e. Menghitung kembali setiap objek yang menggunakan pusat kluster (*centroid*) yang baru

Pada tahap ini diperoleh hasil akhir perhitungan nilai kluster pada iterasi ke-5. Iterasi tersebut menunjukkan tercapainya kondisi konvergensi, ditandai dengan tidak adanya lagi perpindahan data antar kluster. Hal ini mengindikasikan bahwa proses K-Means telah stabil dan menghasilkan pembagian kluster yang optimal. Untuk hasil kluster iterasi ke-5 dapat dilihat pada Tabel 4.

**Tabel 4.** Hasil Nilai Kluster Iterasi ke-5

No	Nama Kelurahan/Desa	C0	C1	C2	Jarak Terdekat
1	Aia Bangih	2.630,49	84,82	1.735,61	84,82
2	Ujung Gading	614,68	2.300,34	781,00	614,68
3	Brastagi Ujung Gading	267,94	2.838,46	1.157,55	267,94
4	Tampus Damai Ujung Gading	160,71	2.683,20	1.011,00	160,71
5	Taluak Ambun Ujung Gading	246,28	2.819,51	1.138,79	246,28
⋮	⋮	⋮	⋮	⋮	⋮
85	Jambak Selatan	131,40	2.705,89	1.024,45	131,40
86	Pujorahayu	1.581,20	1.012,79	683,81	683,81
87	Sasak	174,44	2.436,21	747,26	174,44
88	Padang Harapan	83,94	2.539,70	850,74	83,94
89	Ranah Pasisie	217,24	2.607,26	928,86	217,24
90	Maligi	129,07	2.663,36	977,71	129,07

Setelah diperoleh perhitungan jarak dengan menggunakan *centroid* baru pada tabel 4 di atas maka selanjutnya akan dilakukan pengelompokan data sesuai dengan pembentukan ketiga kluster. Hasil yang diperoleh dari proses pengolahan data menggunakan metode K-Means Clustering menunjukkan bahwa dari total 90 kelurahan atau desa yang dianalisis, data berhasil dikelompokkan ke dalam tiga kluster utama berdasarkan tingkat penggunaan alat kontrasepsi jangka panjang di masing-masing wilayah. Metode ini secara efektif membagi data ke dalam kelompok-kelompok yang memiliki karakteristik serupa dalam hal variabel yang dianalisis.

Adapun hasil klusterisasi tersebut menghasilkan tiga kluster, yaitu:

1. Kluster 0 terdiri atas 24 kelurahan/desa dengan tingkat penggunaan MKJP pada kategori sedang. Wilayah dalam kluster ini menunjukkan adanya partisipasi yang cukup baik, namun masih memerlukan penguatan program untuk meningkatkan capaian penggunaan kontrasepsi jangka panjang.
2. Kluster 1 mencakup 6 kelurahan atau desa dengan tingkat penggunaan MKJP tinggi. Kelompok ini mencerminkan keberhasilan dalam implementasi program KB, sehingga dapat dijadikan contoh praktik baik sekaligus model pembelajaran bagi wilayah lain.
3. Kluster 2 mencakup 60 kelurahan/desa dengan tingkat penggunaan MKJP rendah. Wilayah dalam kluster ini membutuhkan intervensi yang lebih intensif, baik melalui edukasi, penyuluhan, maupun peningkatan akses layanan, agar pemanfaatan kontrasepsi jangka panjang dapat ditingkatkan secara signifikan.

### 3.1 Implementasi Rapidminer

Implementasi dilakukan untuk mengevaluasi kinerja algoritma *K-Means Clustering* dalam mengelompokkan data penggunaan alat kontrasepsi. Pengujian ini menggunakan aplikasi *RapidMiner* versi 7.1. Adapun hasil pengujian *rapidminer* dapat disajikan pada Gambar 4.



## Cluster Model

Cluster 0: 24 items  
 Cluster 1: 6 items  
 Cluster 2: 60 items  
 Total number of items: 90

**Gambar 2.** Hasil *Clustering* Data

Berdasarkan penelitian yang telah dilakukan, analisis data menggunakan perangkat lunak *RapidMiner* menunjukkan hasil yang konsisten dengan perhitungan *K-Means Clustering* secara manual. Temuan ini mengindikasikan bahwa implementasi metode *K-Means* melalui *RapidMiner* dapat mereplikasi proses klasterisasi dengan akurasi yang tinggi, sekaligus mempercepat proses pengolahan data dibandingkan dengan perhitungan manual. Hal ini juga memperkuat validitas penggunaan alat bantu berbasis teknologi dalam mendukung analisis data untuk penelitian yang bersifat kuantitatif.

## 4. KESIMPULAN

Berdasarkan hasil analisis data penggunaan alat kontrasepsi jangka panjang (MKJP) di Kabupaten Pasaman Barat dengan menerapkan metode *K-Means Clustering*, dapat disimpulkan bahwa algoritma ini berhasil mengelompokkan sebanyak 90 kelurahan/desa ke dalam tiga klaster utama, yaitu Klaster 0 (C0) dengan tingkat penggunaan kontrasepsi MKJP sedang sebanyak 24 kelurahan/desa, Klaster 1 (C1) dengan tingkat penggunaan kontrasepsi MKJP tinggi sebanyak 6 kelurahan/desa, dan Klaster 2 (C2) dengan tingkat penggunaan kontrasepsi MKJP rendah sebanyak 60 kelurahan/desa. Hasil klasterisasi yang diperoleh menunjukkan adanya variasi atau heterogenitas pola penggunaan kontrasepsi jangka panjang antarwilayah, yang mencerminkan kondisi riil distribusi serta kecenderungan pemanfaatan kontrasepsi di tingkat lokal. Temuan ini penting karena dapat memberikan gambaran yang lebih terarah mengenai wilayah mana saja yang memerlukan perhatian khusus, wilayah yang cukup stabil dalam pemanfaatan kontrasepsi, serta wilayah yang dapat dijadikan model atau contoh praktik baik dalam penerapan program keluarga berencana. Lebih lanjut, informasi yang dihasilkan dari proses klasterisasi ini memiliki nilai strategis sebagai dasar dalam penyusunan kebijakan dan perencanaan program oleh Perwakilan BKKBN Provinsi Sumatera Barat maupun pemerintah daerah Kabupaten Pasaman Barat. Hasil penelitian dapat mendukung perumusan strategi intervensi yang lebih tepat sasaran, baik dalam bentuk edukasi, penyuluhan, maupun peningkatan distribusi alat kontrasepsi, sehingga efektivitas program KB dapat terus ditingkatkan. Dari sisi metodologis, penelitian ini juga menegaskan bahwa algoritma *K-Means* terbukti efektif sebagai pendekatan data mining untuk analisis data kependudukan, khususnya dalam pemetaan penggunaan alat kontrasepsi. Dengan demikian, penelitian ini tidak hanya memberikan kontribusi empiris terhadap pemetaan kondisi lapangan, tetapi juga kontribusi metodologis dalam pengembangan model analisis berbasis teknologi guna mendukung perencanaan program pembangunan kependudukan dan keluarga berencana di masa mendatang.

## REFERENCES

- [1] D. & S. Yanti, E. M., Wirastri, "Edukasi Pentingnya Keluarga Berencana (KB) dalam meningkatkan Pengetahuan dan Pemilihan Alat Kontrasepsi pada Wanita Usia Subur (WUS) di Dusun Anjani Timur Desa Anjani Kecamatan Suralaga Kabupaten Lombok Timur," *Indones. J. Community Dedication*, vol. 5, no. 1, pp. 7–12, 2023.
- [2] S. Sumarsih, "Hubungan Karakteristik Ibu Nifas Terhadap Pemilihan Metode Kontrasepsi Pascasalin Di Puskesmas Selompang Kabupaten Temanggung," *Sinar J. Kebidanan*, vol. 5, no. 1, pp. 1–14, 2023, doi: 10.30651/sinar.v5i1.17321.
- [3] BKKBN, "Peraturan Badan Kependudukan Dan Keluarga Berencana Nasional Republik Indonesia Nomor 1 Tahun 2023 Tentang Pemenuhan Kebutuhan Alat dan Obat Kontrasepsi Bagi Pasangan Usia Subur Dalam Pelayanan Keluarga Berencana," *Peratur. Menteri Kesehat. Republik Indones. Nomor 1 Tahun 2023*, vol. 151, no. 2, pp. 10–17, 2023.
- [4] A. Fatchiya, A. Sulistyawati, B. Setiawan, and R. Damanik, "Peran Penyuluhan Keluarga Berencana dalam Meningkatkan Pengetahuan KB pada Pasangan Usia Subur (PUS) Kelompok Masyarakat Miskin," *J. Penyul.*, vol. 17, no. 1, pp. 60–71, 2021, doi: 10.25015/17202134151.
- [5] Y. Anwar, E. Amin, and A. H. Tiyas, "Hubungan Pengetahuan Ibu dan Dukungan Suami terhadap Pemakaian Metode Kontrasepsi Jangka Panjang (MKJP) di Kecamatan Mamuju," *J. Akad. Kebidanan Kesehat. Baru*, vol. 2024, 2024, [Online]. Available: <http://jurnal.stikeskb.ac.id/index.php/akbid/article/view/29%0Ahttp://jurnal.stikeskb.ac.id/index.php/akbid/article/download/29/27>
- [6] C. Setyorini, A. Dewi Lieskusumastuti, and L. Hanifah, "Faktor-Faktor Yang Mempengaruhi Penggunaan Metode Kontrasepsi Jangka Panjang (Mkjp): Scoping Review," *Avicenna J. Heal. Res.*, vol. 5, no. 1, pp. 132–146, 2022, doi: 10.36419/avicenna.v5i1.600.
- [7] B. Hakim, F. J. Kaunang, C. Susanto, J. Salim, and R. Indradjaja, "Implementasi Machine Learning Dalam Pengelompokan Musik Menggunakan Algoritma K-Means Clustering," *IDEALIS Indones. J. Inf. Syst.*, vol. 8, no. 1, pp. 74–83, 2025, doi: 10.36080/idealis.v8i1.3357.
- [8] Y. B. Pratama and A. Setiawan, "Implementasi Machine Learning Menggunakan Algoritma K-Means Untuk Klasifikasi Sekolah Dasar," *Resolusi Rekayasa Tek. Inform. dan Inf.*, vol. 4, no. 3, pp. 249–257, 2024, doi: 10.30865/resolusi.v4i3.1591.
- [9] M. Norshahlan, H. Jaya, and R. Kustini, "Penerapan Metode Clustering Dengan Algoritma K-means Pada Pengelompokan Data

