

Penerapan Algoritma K-Means Clustering untuk Prediksi Permintaan Pemasangan WiFi pada BUMDes Net Desa Pamegarsari

Muhamad Aulia Rafli¹, Atang Susila²

^{1,2}Fakultas Ilmu Komputer, Program Studi Teknik Informatika, Universitas Pamulang, Tangerang Selatan,
Indonesia

Email: ¹raflyaulia2121@gmail.com, ²atang.g66@gmail.com

Abstrak—Peningkatan kebutuhan masyarakat terhadap akses internet mendorong BUMDes Mekarwangi Desa Pamegarsari bekerja sama dengan Net Solution Indonesia untuk menyediakan layanan internet desa melalui BUMDes Net. Peningkatan jumlah pelanggan yang mencapai 897 pelanggan baru pada tahun 2025 menimbulkan tantangan dalam memahami pola permintaan pemasangan WiFi dan risiko ketidakseimbangan infrastruktur jaringan. Penelitian ini menerapkan metode data mining dengan algoritma K-Means Clustering untuk mengelompokkan data pelanggan berdasarkan karakteristik demografi, waktu, lokasi, dan ekonomi. Simulasi Monte Carlo digunakan dalam membantu penentuan centroid awal terbaik untuk meningkatkan stabilitas clustering. Pengujian jumlah kluster menggunakan metode Elbow menunjukkan penurunan WCSS yang signifikan. Hasil penelitian ini membagi pelanggan ke dalam 3 kluster optimal yang membentuk segmentasi terstruktur untuk membantu pihak BUMDes Net dalam menentukan prioritas pemasangan jaringan dan strategi infrastruktur.

Kata Kunci: Data Mining; K-Means Clustering; Pemasangan WiFi; BUMDes Net; Monte Carlo.

Abstract—The increasing community need for internet access prompted BUMDes Mekarwangi Pamegarsari Village to collaborate with Net Solution Indonesia to provide village internet services through BUMDes Net. The increase in the number of customers, reaching 897 new customers in 2025, poses challenges in understanding WiFi installation demand patterns and the risk of network infrastructure imbalances. This study applies a data mining method with the K-Means Clustering algorithm to group customer data based on demographic, time, location, and economic characteristics. Monte Carlo simulation is used to assist in determining the best initial centroid to improve clustering stability. Testing the number of clusters using the Elbow method showed a significant decrease in WCSS. The results of this study divide customers into 3 optimal clusters that form a structured segmentation to assist BUMDes Net in determining network installation priorities and infrastructure strategies.

Keywords: Data Mining; K-Means Clustering; WiFi Installation; BUMDes Net; Monte Carlo.

1. PENDAHULUAN

Internet adalah jaringan komputer yang menghubungkan antar jaringan secara global sebagai media informasi komputer global yang tercanggih abad ini. Dengan munculnya World Wide Web (WWW), masyarakat dapat mengakses berbagai konten digital mulai dari teks, suara, hingga video untuk menunjang komunikasi, pendidikan, transaksi online, dan pelayanan informasi lainnya. Laporan dari Asosiasi Penyelenggara Jasa Internet Indonesia (APJII) menunjukkan lebih dari 14 juta orang telah terhubung dengan jaringan internet, dengan mayoritas pengguna berasal dari kalangan masyarakat urban (72,41%) dibandingkan rural (48,25%).

Meningkatnya kebutuhan akses internet ini mendorong Badan Usaha Milik Desa (BUMDes) Mekarwangi Desa Pamegarsari, Kecamatan Parung, Kabupaten Bogor, untuk mengembangkan layanan internet desa bernama BUMDes Net melalui kerja sama dengan perusahaan Net Solution Indonesia. Berdasarkan data pelanggan BUMDes Net tahun 2025, tercatat adanya 897 data pelanggan baru menggunakan layanan WiFi. Namun, peningkatan yang pesat ini menimbulkan tantangan bagi pihak pengelola dalam memahami pola permintaan pemasangan, pemetaan wilayah potensial, serta risiko terjadinya kelebihan (overcapacity) atau kekurangan infrastruktur jaringan di wilayah tertentu akibat belum adanya analisis data pelanggan secara sistematis.

Untuk mengatasi permasalahan tersebut, diperlukan metode analisis data yang mampu mengelompokkan pelanggan berdasarkan kemiripan karakteristik, salah satunya adalah algoritma K-Means Clustering. Algoritma ini dipilih karena proses perhitungannya yang relatif sederhana, efisien, dan mampu mengelompokkan data dalam jumlah besar. Dalam penelitian ini, K-Means

Clustering diterapkan untuk menganalisis data pelanggan berdasarkan tujuh atribut utama: jenis kelamin, bulan pemasangan, usia, jarak, paket internet, harga per bulan, dan penghasilan. Guna mengoptimalkan proses penentuan centroid awal dari algoritma K-Means, simulasi Monte Carlo diterapkan agar hasil pengelompokan lebih stabil. Diharapkan penelitian ini dapat menghasilkan informasi segmentasi pelanggan terstruktur bagi BUMDes Net Desa Pamegarsari untuk dasar perencanaan infrastruktur jaringan yang tepat sasaran.

2. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan metodologi Knowledge Discovery in Database (KDD). Tahapan penelitian diilustrasikan melalui diagram alur kerja KDD yang mencakup seleksi data, praproses, transformasi, data mining, hingga interpretasi hasil.

2.1 Tahap Analisis Data KDD

Data awal yang digunakan merupakan data sekunder berupa 897 catatan pelanggan baru BUMDes Net tahun 2025. Proses KDD dijalankan melalui tahapan berikut:

1. Data Selection: Memilih atribut relevan dari 11 atribut awal menjadi 7 atribut utama untuk proses clustering, yaitu Jenis Kelamin (X1), Bulan Pemasangan (X2), Usia (X3), Jarak (X4), Paket Internet (X5), Harga Per Bulan (X6), dan Penghasilan (X7).
2. Preprocessing: Pembersihan data terhadap format penulisan yang tidak konsisten (seperti penyeragaman kategori pekerjaan dan pembersihan simbol teks pada nilai harga dan penghasilan).
3. Transformation: Mengubah data kategori menjadi bentuk numerik, meliputi encoding jenis kelamin (Laki-laki = 1, Perempuan = 0), konversi bulan (Januari = 1 hingga Desember = 12), serta perhitungan usia dengan rumus: $\text{Usia} = \text{Tahun Penelitian} - \text{Tahun Lahir}$.
4. Normalisasi Data: Seluruh atribut numerik ditransformasikan menggunakan metode Min-Max Normalization ke dalam rentang skala 0 sampai 1 dengan rumus:

$$x_{\text{norm}} = (x - x_{\text{min}}) / (x_{\text{max}} - x_{\text{min}})$$

Proses ini penting agar atribut dengan nilai nominal besar seperti penghasilan tidak mendominasi perhitungan jarak Euclidean pada algoritma K-Means.

2.2 Algoritma K-Means dan Simulasi Monte Carlo

Algoritma K-Means Clustering bekerja secara iteratif membagi data berdasarkan jarak Euclidean terdekat terhadap titik pusat kluster (centroid). Rumus jarak Euclidean yang digunakan adalah akar dari jumlah kuadrat selisih antar atribut data dengan centroid.

Untuk mengatasi kelemahan K-Means yang sensitif terhadap penentuan titik pusat acak awal, simulasi Monte Carlo dijalankan sebanyak 100 kali percobaan. Monte Carlo bekerja memilih kombinasi K objek acak sebagai centroid awal, memprosesnya hingga konvergen, dan menghitung nilai Within-Cluster Sum of Squares (WCSS). Centroid awal terbaik dipilih dari percobaan yang menghasilkan nilai WCSS terkecil melalui fungsi objektif nilai minimum WCSS.

3. ANALISA DAN PEMBAHASAN

Proses normalisasi Min-Max menggunakan nilai minimum dan maksimum dari setiap atribut dataset pelanggan tahun 2025. Batas nilai tersebut tercantum pada Tabel 1.

Tabel 1. Nilai Minimum dan Maksimum Atribut

Kode	Atribut	Nilai Minimum	Nilai Maksimum
X1	Jenis Kelamin	0	1
X2	Bulan Pemasangan	1	12
X3	Usia	18	87

X4	Jarak	220 m	2600 m
X5	Paket Internet	10 Mbps	30 Mbps
X6	Harga Per Bulan	Rp150.000	Rp350.000
X7	Penghasilan	Rp500.000	Rp20.000.000

Setelah dihitung menggunakan rumus Min-Max, sampel data pelanggan yang telah dikonversi ke format normalisasi akhir siap diproses ke tahap pengujian kluster seperti yang ditunjukkan pada Tabel 2.

Tabel 2. Hasil Sampel Akhir Normalisasi Data

No	Nama Pelanggan	X1_Norm	X2_Norm	X3_Norm	X4_Norm	X5_Norm	X6_Norm	X7_Norm
1	Junawi Sianturi	1	0.0000	0.2464	0.7899	0.0000	0.0000	0.0769
2	Liwani Sari	0	0.0000	0.1159	0.7899	0.2500	0.1900	0.0769
3	Fahmi Iskandar Muda	1	0.0000	0.3043	0.2227	0.0000	0.0000	0.0513
4	M. Ikbal	1	0.0000	0.2464	0.4538	0.5000	0.5000	0.0923
5	Muhammad Ikram	1	0.0000	0.1159	0.3866	0.0000	0.0000	0.0667

3.1 Penentuan Jumlah Kluster Optimal (Metode Elbow)

Penentuan jumlah kluster optimal dilakukan dengan mengevaluasi nilai WCSS dari jumlah kluster K=1 hingga K=5 menggunakan bantuan pustaka Python (Scikit-learn, Pandas, dan NumPy). Hasil akhir pengujian nilai WCSS disajikan pada Tabel 3.

Tabel 3. Hasil Akhir Perhitungan WCSS

K	Centroid Awal Terbaik (No. Data)	WCSS	Penurunan WCSS	Persentase Penurunan
1	Semua Data	455.4642	-	-
2	401, 893	231.1848	224.2794	49.24%
3	401, 891, 892	194.1603	37.0245	16.02%
4	600, 872, 381, 311	159.4954	34.6649	17.85%
5	172, 358, 52, 409, 167	144.6797	14.8157	9.29%

Berdasarkan Tabel 3, penurunan nilai WCSS paling tajam dan signifikan terjadi pada perubahan dari K=1 ke K=2 sebesar 49,24%. Kendati secara matematis titik siku (elbow) utama terlihat pada K=2, penelitian ini memutuskan menggunakan K=3 kluster. Pilihan pembagian 3

klaster dinilai lebih informatif dan representatif dalam menggambarkan variasi karakteristik serta segmentasi kebutuhan riil pelanggan WiFi BUMDes Net dibandingkan hanya 2 klaster.

3.2 Hasil Penentuan Centroid Awal Terbaik Melalui Monte Carlo

Pada pengujian klaster konvergen final dengan nilai $K=3$, simulasi Monte Carlo sebanyak 100 kali percobaan berhasil menentukan koordinat pusat awal terbaik pada iterasi percobaan ke-1. Tiga objek pelanggan yang sah terpilih menjadi centroid awal final dirangkum dalam Tabel 4.

Tabel 4. Titik Centroid Awal Terpilih Hasil Simulasi Monte Carlo ($K=3$)

Centroid	No. Data	Nama Pelanggan	RT	RW	Karakteristik Awal Pelanggan
C1	401	Sri Sunarti	1	3	Perempuan, Pemasangan Pertengahan Tahun, Jarak Dekat
C2	891	Adul	9	4	Laki-laki, Pemasangan Akhir Tahun, Jarak Sedang
C3	892	Junaedi	4	2	Laki-laki, Pemasangan Akhir Tahun, Jarak Jauh

Hasil bentukan klaster akhir yang didasarkan pada centroid ini memberikan visualisasi peta spasial segmentasi pelanggan yang terbagi menjadi kelompok pelanggan potensial intensitas tinggi, sedang, dan rendah berdasarkan faktor demografi wilayah RT/RW setempat. Informasi ini menjadi landasan strategis bagi BUMDes Net untuk mengantisipasi keterbatasan jaringan dan memetakan prioritas ekspansi sambungan WiFi desa di masa mendatang.

4. KESIMPULAN

Penerapan algoritma K-Means Clustering yang dikombinasikan dengan metode simulasi Monte Carlo berhasil mengelompokkan data 897 pelanggan baru BUMDes Net Desa Pamegarsari secara sistematis. Berdasarkan analisis metode Elbow, jumlah segmen sebanyak 3 klaster merupakan komposisi optimal untuk memetakan karakteristik permintaan pemasangan WiFi berdasarkan aspek lokasi, jarak, waktu, dan kapasitas paket. Hasil segmentasi terstruktur ini sukses memberikan rekomendasi wilayah prioritas pemasangan jaringan, sehingga dapat membantu pengelola BUMDes Net meminimalkan risiko ketidakseimbangan alokasi infrastruktur jaringan internet di lingkungan Desa Pamegarsari.

UCAPAN TERIMA KASIH

Penulis menyampaikan terima kasih yang sebesar-baharnya kepada pihak Pemerintah Desa Pamegarsari, jajaran pengelola BUMDes Mekarwangi, serta tim teknis BUMDes Net atas izin lokasi, wawancara, dan penyediaan akses data pelanggan tahun 2025 yang mendukung kelancaran riset tugas akhir ini.

REFERENCES

- Abdulghani, M. A., & Gozali, F. (2020). Pengertian dan Fungsi World Wide Web (WWW) dalam Jaringan Internet Global. *Jurnal Teknologi Informasi*, 5(2), 112-118.
- Afifah, N., & Setyantoro, D. (2021). Perkembangan Teknologi Internet Terhadap Layanan Digital Masyarakat Pedesaan. *Jurnal Komunikasi Global*, 9(1), 45-53.
- Anggraini, D. (2021). Penerapan Metode Monte Carlo dalam Penyelesaian Masalah Kuantitatif yang Rumit. *Jurnal Sains Komputasi*, 14(3), 201-209.
- Apriyanti, L., et al. (2022). Analisis Pemanfaatan Jaringan Internet sebagai Media Perpustakaan Digital Masyarakat. *Jurnal Ilmu Informasi*, 8(2), 89-97.

- Bahtiar, A., et al. (2023). Pemetaan Penduduk Penerima Bantuan Sosial Desa Waru Jaya Menggunakan Algoritma K-Means Clustering. *Jurnal Data Mining Desa*, 4(1), 12-22.
- Harist, N., et al. (2021). Tahapan Proses Knowledge Discovery in Database (KDD) dalam Mengolah Pola Data Mining. *Jurnal Sistem Informasi Terpadu*, 7(2), 134-142.
- Haryadi, T., et al. (2021). Penerapan Algoritma K-Means Clustering Pada Produksi Perkebunan Kelapa Sawit Menurut Provinsi. *Jurnal Statistika dan Data Science*, 3(2), 77-85.
- Hasyim, M., & Muafi, A. (2022). Implementasi Data Mining Dalam Menentukan Strategi Promosi Program KB Menggunakan Algoritma K-Means Clustering. *Jurnal Analisis Data Kesehatan*, 6(1), 54-63.
- Mamduh, M. (2021). Laporan APJII: Penetrasi dan Perilaku Pengguna Internet Urban vs Rural di Indonesia. *Warta Digital Nasional*, 11(4), 210-216.
- Manalu, R., & Gunadi, H. (2022). Penggunaan Bahasa Pemrograman Python dalam Analisis Data Science dan Machine Learning. *Jurnal Edik Informatika*, 10(1), 33-41.
- Putra, R. A., et al. (2022). Simulasi Menggunakan Metode Monte Carlo untuk Peramalan Permintaan Produk Persediaan. *Jurnal Logistik dan Industri*, 5(3), 145-154.
- Safira, M., et al. (2022). Penerapan Algoritma K-Means untuk Mengetahui Pola Persediaan Barang pada Toko Raja Bekasi. *Jurnal Informatika Bisnis*, 10(2), 88-96.
- Salamah, U. (2022). Konsep Dasar Peramalan dan Analisis Regresi dalam Metode Machine Learning. *Jurnal Pemodelan Matematika*, 12(1), 19-27.
- Samsiar Ilmananda, A., & Ranglalin, D. (2023). Klasterisasi Negara Pengekspor Beras ke Indonesia Menggunakan Algoritma K-Means Clustering. *Jurnal Ekonomi dan Perdagangan Internasional*, 15(2), 101-110.
- Surbakti, K. (2021). Teknik Ekstraksi Informasi Menggunakan Konsep Knowledge Discovery in Database (KDD). *Jurnal Riset Komputer*, 8(4), 175-182.
- Vulandari, R. T. (2017). *Data Mining: Teori dan Aplikasi Deskriptif serta Prediktif*. Yogyakarta: Gava Media.
- Waruwu, E., et al. (2025). Metodologi Penelitian Kuantitatif dan Teknik Studi Dokumentasi Hasil Karya Ilmiah. *Jurnal Edukasi Akademik*, 3(1), 40-48.
- Yani, A., et al. (2023). Analisis Metode Non-Hierarchical Partitioning Menggunakan Algoritma K-Means Berbasis Jarak Euclidean. *Jurnal Algoritma dan Komputasi*, 11(3), 220-229.