Vol…. No…. Juli 2026
ISSN: 2252-4983

ANALISIS KARAKTERISTIK TUNGGAKAN PELANGGAN MENGGUNAKAN ALGORITMA K-MEANS UNTUK SEGMENTASI PENAGIHAN PADA PDAM TIRTA MUARE ULAKAN SAMBAS

Verla Aurelia[1], Theresia Widji Astuti[1],Widya Sartika[1], Fiqih Akbar[1],
Jurusan Manajemen Informatika, Prodi Manajemen Informatika
Politeknik Negeri Sambas
Email: theresiawidji@gmail.com

ABSTRAK

Penelitian ini bertujuan mengelompokkan pelanggan berdasarkan tingkat tunggakan menggunakan algoritma *K-Means*, menganalisis distribusi pelanggan berdasarkan ID_Golongan dan alamat pada setiap *cluster*, serta membangun aplikasi *clustering* berbasis *Python* sebagai pendukung segmentasi penagihan di PDAM Tirta Muare Ulakan Sambas. Penelitian menggunakan metode kuantitatif dengan tahapan *Knowledge Discovery in Databases* (KDD), yaitu data *selection*, data *preprocessing*, data *transformation*, data mining, dan *evaluation.* Data yang digunakan merupakan data tunggakan pelanggan PDAM Tirta Muare Ulakan Sambas. Proses clustering dilakukan menggunakan atribut jumlah bulan tunggakan (JML_REK) dan total tunggakan (TOTAL) yang telah dinormalisasi menggunakan MinMaxScaler. Penentuan jumlah *cluste*r dilakukan dengan Metode *Elbow* dan menghasilkan tiga *cluster*. Hasil penelitian menunjukkan bahwa algoritma, dan tinggi. Analisis berdasarkan ID_Golongan dan alamat memberikan *K-Means* mampu mengelompokkan pelanggan ke dalam kategori tunggakan rendah, sedang informasi mengenai karakteristik pelanggan serta wilayah dengan tingkat tunggakan yang lebih tinggi. Hasil *clustering* kemudian diimplementasikan dalam aplikasi berbasis *Python* untuk membantu PDAM Tirta Muare Ulakan Sambas menentukan prioritas penagihan secara lebih efektif.

Kata kunci: K-Means, Clustering, Segmentasi Penagihan, Tunggakan Pelanggan, PDAM..

ABSTRACT

This study aims to group customers based on the level of arrears using the K-Means algorithm, analyze customer distribution based on ID_Golongan and address in each cluster, and build a Python-based clustering application as a support for billing segmentation at PDAM Tirta Muare Ulakan Sambas. The research uses a quantitative method with the stages of Knowledge Discovery in Databases (KDD), namely data selection, data preprocessing, data transformation, data mining, and evaluation. The data used is the arrears of PDAM Tirta Muare Ulakan Sambas customers. The clustering process is carried out using the attributes of the number of months arrears (JML_REK) and total arrears (TOTAL) that have been normalized using MinMaxScaler. The determination of the number of clusters is carried out by the Elbow Method and produces three clusters. The results of the study show that the algorithm, and high. Analysis based on ID_Golongan and address gives K-Means the ability to group customers into categories of low arrears, while information about customer characteristics and regions with higher arrears rates. The clustering results are then implemented in a Python-based application to help PDAM Tirta Muare Ulakan Sambas determine billing priorities more effectively.

Keywords: *K-Means, Clustering, Billing Segmentation, Customer Arrears, PDAM*.

1. PENDAHULUAN

Air bersih merupakan kebutuhan dasar masyarakat yang penyediaannya dikelola oleh Perusahaan Daerah Air Minum (PDAM). Salah satu permasalahan yang sering dihadapi PDAM adalah tingginya tunggakan pembayaran pelanggan yang dapat memengaruhi pendapatan perusahaan serta menghambat kelancaran operasional. Di PDAM Tirta Muare Ulakan Sambas, data rekap tunggakan pelanggan periode 2024–2025 yang mencakup akumulasi tunggakan tahun 2006–2018 menunjukkan masih banyak pelanggan yang memiliki tunggakan. Namun, data tersebut selama ini lebih banyak dimanfaatkan sebagai laporan administratif dan belum digunakan sebagai dasar analisis dalam menentukan prioritas penagihan.

Perkembangan teknologi informasi memungkinkan data tunggakan diolah menjadi informasi yang lebih bernilai melalui teknik data mining. Salah satu algoritma yang banyak digunakan untuk proses pengelompokan data adalah *K-Means* karena mampu mengelompokkan objek berdasarkan tingkat kemiripan karakteristik. Beberapa penelitian terdahulu menunjukkan bahwa algoritma *K-Means* efektif diterapkan untuk segmentasi pelanggan dan pengelompokan data pada berbagai bidang. Akan tetapi, sebagian besar penelitian hanya menghasilkan kelompok data tanpa menganalisis karakteristik pelanggan berdasarkan golongan dan wilayah, serta belum mengimplementasikan hasil pengelompokan ke dalam aplikasi yang dapat dimanfaatkan secara langsung.

Berdasarkan permasalahan tersebut, penelitian ini menerapkan algoritma *K-Means* untuk mengelompokkan pelanggan PDAM Tirta Muare Ulakan Sambas berdasarkan jumlah bulan tunggakan (JML_REK) dan total tunggakan (TOTAL). Selain menghasilkan kelompok pelanggan dengan tingkat tunggakan rendah, sedang, dan tinggi, penelitian ini juga menganalisis distribusi pelanggan berdasarkan ID_Golongan dan alamat pada setiap *cluster*. Hasil analisis kemudian diimplementasikan dalam aplikasi berbasis *Python* menggunakan Streamlit untuk membantu PDAM menentukan prioritas penagihan secara lebih efektif, efisien, dan tepat sasaran.

2. METODOLOGI PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan teknik data mining. Proses analisis data dilakukan menggunakan kerangka kerja *Knowledge Discovery in Databases* (KDD) dan algoritma *K-Means* untuk mengelompokkan pelanggan berdasarkan tingkat tunggakan. Data yang digunakan berupa data tunggakan pelanggan PDAM Tirta Muare Ulakan Sambas periode 2024–2025 yang mencakup akumulasi tunggakan pelanggan dari tahun 2006–2018 sebanyak 12.560 data.

2.1 Tahapan Knowledge Discovery in Databases (KDD)

1) Data *Selection*, yaitu memilih data dan atribut yang relevan untuk proses analisis.
2) Data *Preprocessing*, yaitu membersihkan data dengan menghapus data duplikat, menangani nilai kosong, serta mengubah tipe data numerik agar siap diproses.
3) Data *Transformation*, yaitu membentuk dataset berdasarkan ID_Pelanggan dan melakukan normalisasi atribut JML_REK dan TOTAL menggunakan metode *MinMaxScaler* sehingga memiliki rentang nilai yang sama.
4) Data Mining, yaitu menentukan jumlah *cluster* menggunakan Metode *Elbow*, kemudian menerapkan algoritma *K-Means* untuk mengelompokkan pelanggan berdasarkan tingkat tunggakan.
5) Evaluation, yaitu menganalisis hasil *clustering* berdasarkan jumlah anggota setiap *cluster*, nilai *centroid*, visualisasi *scatter plot*, serta distribusi pelanggan berdasarkan ID_Golongan dan alamat..

Hasil *clustering* selanjutnya diimplementasikan ke dalam aplikasi berbasis *Python* menggunakan *Streamlit* untuk membantu PDAM Tirta Muare Ulakan Sambas dalam menentukan prioritas penagihan pelanggan secara lebih efektif dan efisien.

3. HASIL DAN PEMBAHASAN

3.1. Data Peneltian

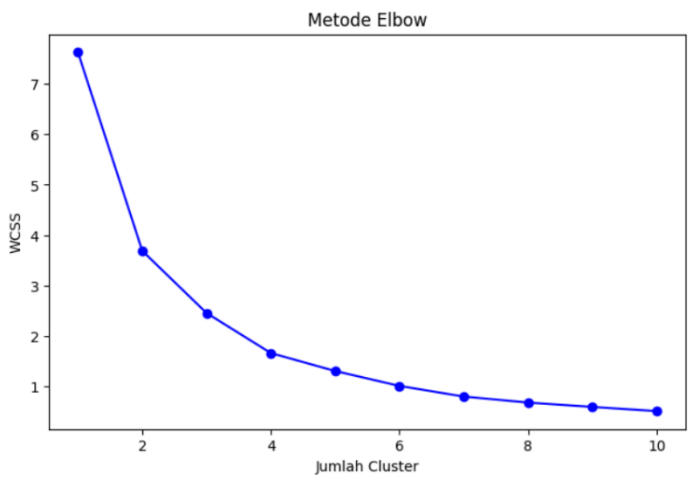
Data yang digunakan dalam penelitian ini merupakan data tunggakan pelanggan PDAM Tirta Muare Ulakan Sambas periode 2024–2025 yang mencakup akumulasi tunggakan pelanggan dari tahun 2006–2018 sebanyak 12.560 data. Sebelum dilakukan proses *clustering*, data terlebih dahulu melalui tahap *preprocessing* yang meliputi penghapusan data duplikat, penanganan *missing value*, dan penyesuaian format data. Hasil *preprocessing* menghasilkan 6.280 data yang layak digunakan dalam proses analisis. Selanjutnya, data ditransformasikan berdasarkan ID_Pelanggan sehingga diperoleh 102 pelanggan yang menjadi objek pada proses *clustering*.

Tabel 1 Hasil *Preprocessing* Data

	ID_PELANGGAN	JML_REK	TOTAL	ID_GOLONGAN	ALAMAT
0	3010006	47	6191850.0	B3	JL.G. HAMZAH
1	3010042	16	1612185.0	B4	JL.G. HAMZAH
2	3010200	17	762905.0	B3	JL.G. HAMZAH
3	3010271	24	2418475.0	C3	JL.G. HAMZAH
4	3011140	3	219527.0	B3	JL.G.HAMZAH

3.2. Penentuan Jumlah Cluster

Penentuan jumlah *cluster* dilakukan menggunakan Metode Elbow dengan menghitung nilai *Within Cluster Sum of Squares* (WCSS) pada beberapa nilai **k**. Berdasarkan hasil pengujian, diperoleh titik siku (*elbow*) pada **k = 3**, sehingga jumlah *cluster* yang digunakan pada penelitian ini adalah tiga. Jumlah *cluster* tersebut dianggap mampu memberikan hasil pengelompokan yang optimal terhadap karakteristik tunggakan pelanggan.

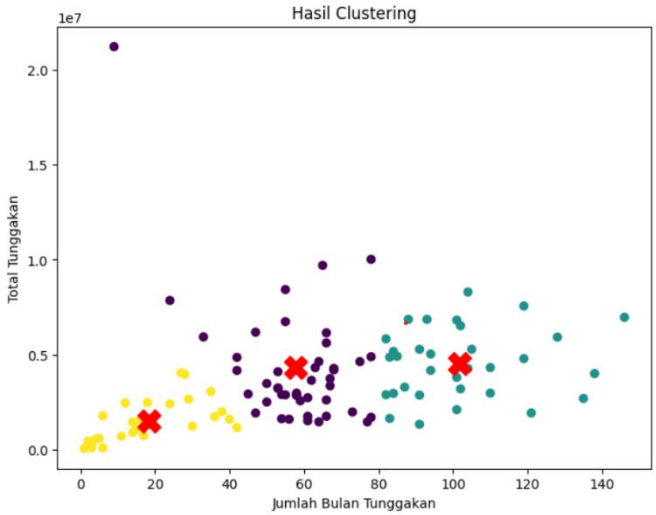


Gambar 1 *Metode Elbow*

3.3. Hasil Clustering Menggunakan Algoritma K-Means

Proses *clustering* dilakukan menggunakan atribut JML_REK (jumlah bulan tunggakan) dan TOTAL (total nominal tunggakan) yang telah dinormalisasi menggunakan metode *MinMaxScaler*. Hasil pengelompokan menunjukkan bahwa seluruh pelanggan berhasil dikelompokkan ke dalam tiga *cluster*, yaitu pelanggan dengan tingkat tunggakan rendah, sedang, dan tinggi. Pengelompokan tersebut didasarkan

pada tingkat kemiripan karakteristik setiap pelanggan sehingga pelanggan dengan pola tunggakan yang serupa berada pada *cluster* yang sama.



Gambar 2. Hasil *Clustering* Menggunakan Algoritma *K-Means*

3.4. Jumlah Anggota Setiap Cluster

Setelah proses *clustering* menggunakan algoritma *K-Means* dengan tiga *cluster* selesai dilakukan, diperoleh jumlah anggota pada setiap *cluster* seperti yang ditunjukkan pada Tabel 2. Hasil tersebut menunjukkan bahwa seluruh pelanggan berhasil dikelompokkan berdasarkan kemiripan karakteristik jumlah bulan tunggakan (JML_REK) dan total tunggakan (TOTAL).

Tabel 1 Hasil Pelanggan Per *Cluster*

Cluster	count
0	44
1	32
2	26

Berdasarkan Tabel 2, *Cluster* 0 memiliki jumlah pelanggan terbanyak, sedangkan *Cluster* 2 memiliki jumlah pelanggan paling sedikit. Perbedaan jumlah anggota pada setiap *cluster* menunjukkan bahwa karakteristik tingkat tunggakan pelanggan tidak tersebar secara merata. Hasil ini dapat dimanfaatkan sebagai dasar dalam menentukan prioritas penagihan sesuai tingkat tunggakan masing-masing pelanggan.

3.5. Analisis Hasil Clustering

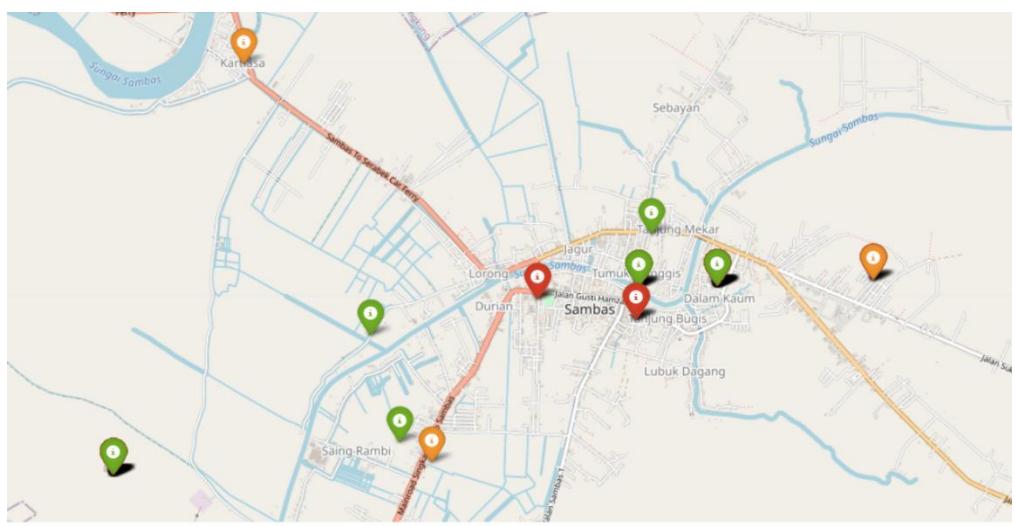
Hasil *clustering* menunjukkan bahwa setiap *cluster* memiliki karakteristik yang berbeda. *Cluster* dengan tingkat tunggakan rendah terdiri atas pelanggan yang memiliki jumlah bulan tunggakan dan total tunggakan relatif kecil. *Cluster* sedang menunjukkan pelanggan dengan

tingkat tunggakan menengah, sedangkan *cluster* tinggi terdiri atas pelanggan yang memiliki jumlah bulan tunggakan dan total tunggakan terbesar sehingga menjadi prioritas utama dalam proses penagihan.

Analisis selanjutnya dilakukan berdasarkan ID_Golongan dan Alamat pelanggan. Hasil analisis menunjukkan bahwa distribusi pelanggan pada setiap *cluster* berbeda-beda sehingga dapat memberikan informasi mengenai golongan pelanggan maupun wilayah yang memiliki tingkat tunggakan lebih tinggi. Informasi tersebut dapat dimanfaatkan oleh PDAM Tirta Muare Ulakan Sambas dalam menentukan strategi penagihan yang lebih efektif dan tepat sasaran.

Tabel 3 Tabel Berdasarkan Golongan

ID_GOLONGAN	B3	B4	C1	C2	C3
Kategori					
Rendah	20	1	1	1	3
Sedang	42	1	1	0	0
Tinggi	31	1	0	0	0



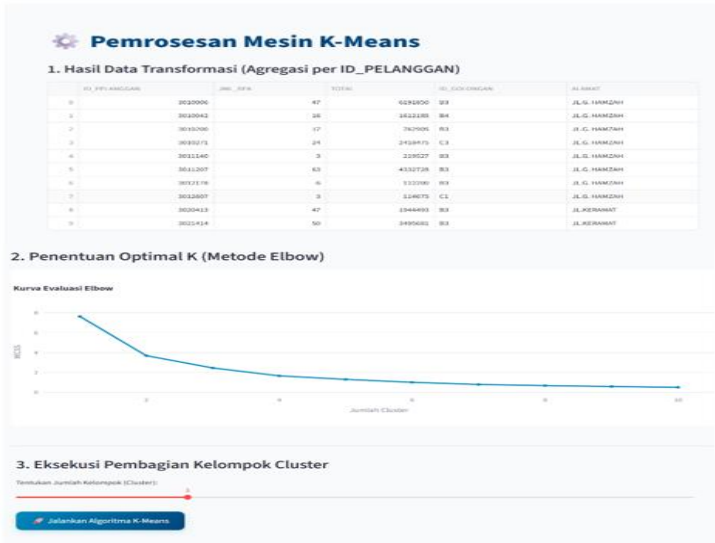
Gambar 3 Peta Persebaran Pelanggan Berdasarkan Hasil Clustering

3.6. Implementasi Aplikasi

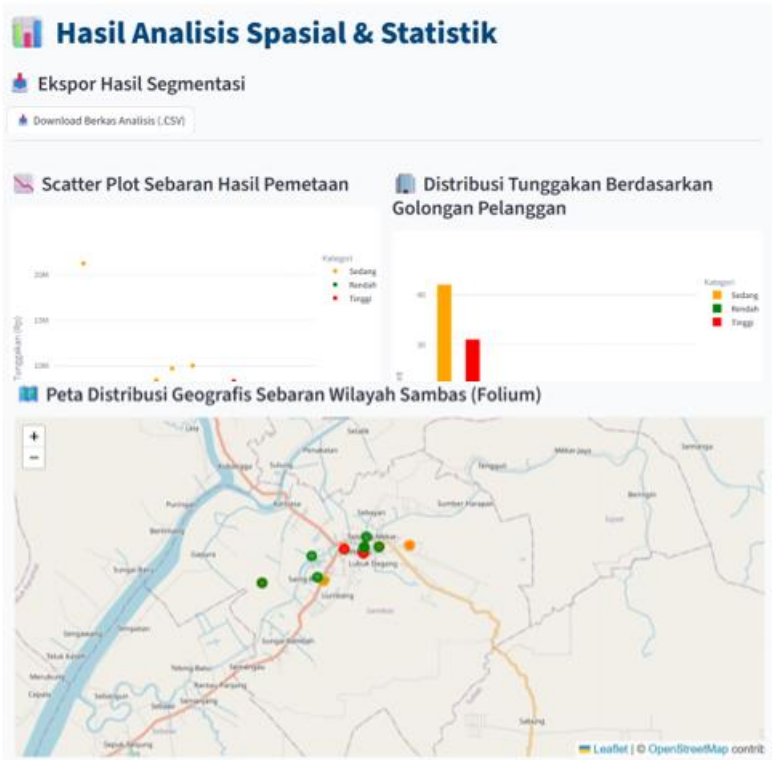
Hasil *clustering* diimplementasikan ke dalam aplikasi berbasis *Python* menggunakan *Streamlit*. Aplikasi menyediakan fitur unggah data, *preprocessing*, transformasi data, proses *clustering*, visualisasi hasil, serta pengunduhan hasil analisis. Melalui aplikasi ini, pengguna dapat melakukan proses pengelompokan pelanggan secara otomatis sehingga membantu PDAM Tirta Muare Ulakan Sambas dalam menentukan prioritas penagihan berdasarkan tingkat tunggakan pelanggan.



Gambar 4 *Dashboard* Aplikasi *Cluster*



Gambar 5 Tampilan Analisis Data



Gambar 6 Tampilan Hasil Analisis

4. KESIMPULAN

Penelitian ini berhasil menerapkan algoritma *K-Means* untuk mengelompokkan 102 pelanggan PDAM Tirta Muare Ulakan Sambas ke dalam tiga *cluster*, yaitu tingkat tunggakan rendah, sedang, dan tinggi. Hasil *clustering* menunjukkan bahwa *Cluster* 0 (tingkat tunggakan sedang) memiliki jumlah anggota terbanyak, yaitu 44 pelanggan (43,14%), diikuti *Cluster* 1 (tingkat tunggakan tinggi) sebanyak 32 pelanggan (31,37%), dan *Cluster* 2 (tingkat tunggakan rendah) sebanyak 26 pelanggan (25,49%). Pengelompokan tersebut memberikan informasi mengenai tingkat risiko tunggakan pelanggan sehingga dapat digunakan sebagai dasar dalam menentukan prioritas penagihan.

Analisis berdasarkan ID_Golongan menunjukkan bahwa golongan B3 merupakan golongan dengan jumlah pelanggan menunggak terbanyak, sedangkan analisis berdasarkan alamat menunjukkan bahwa Jl. Syafiudin, Jl. Sei Rambi, dan Jl. Merdeka merupakan wilayah dengan jumlah pelanggan menunggak yang paling banyak. Hasil *clustering* yang diimplementasikan ke dalam aplikasi berbasis Python menggunakan Streamlit mampu membantu PDAM Tirta Muare Ulakan Sambas dalam mengidentifikasi pelanggan dan wilayah yang menjadi prioritas penagihan sehingga proses penagihan dapat dilakukan secara lebih efektif, efisien, dan tepat sasaran.

DAFTAR PUSTAKA

[1] Rahayu, Prastyadi Wibawa, et al. *Buku Ajar Data Mining*. PT Sonpedia Publishing Indonesia; 2024.
[2] Putra RF, et al. *Data Mining: Algoritma dan Penerapannya*. PT Sonpedia Publishing Indonesia; 2023.

[3] Syahfitri N. "Pengelompokan Produk Berdasarkan Data Persediaan Barang Menggunakan Metode Elbow dan K-Medoid." *KLIK: Kajian Ilmiah Informatika dan Komputer*. 2023;4(3):1668–1675.
[4] Hutagalung J. "Pemetaan Siswa Kelas Unggulan Menggunakan Algoritma K-Means Clustering." *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*. 2022;9(1):606–620.
[5] Dewi FP, Aryni PS, Umaidah Y. "Implementasi Algoritma K-Means Clustering Seleksi Siswa Berprestasi Berdasarkan Keaktifan dalam Proses Pembelajaran." *JISKA (Jurnal Informatika Sunan Kalijaga)*. 2022;7(2):111–121.
[6] Putra A, Saputra D. "Tahapan Data Mining dalam Proses Knowledge Discovery in Database (KDD)." *Indonesian Journal Computer Science*. 2025;10(1):1–10.
[7] Nazihah F, et al. "Analisis Segmentasi Pelanggan Menggunakan Algoritma K-Means pada Data Penjualan." *Jurnal Algoritma*. 2025;22(1):1–10.
[8] Maesaroh S, et al. *Bahasa Pemrograman Python*. Sada Kurnia Pustaka; 2024.
[9] Rahman S, et al. *Python: Dasar dan Pemrograman Berorientasi Objek*. Penerbit Tahta Media; 2023.
[10] Amal I, et al. *Aplikasi Pendeteksi Berita Palsu Bahasa Indonesia Menggunakan Framework Flask dan Streamlit serta Algoritma Machine Learning*. Skripsi. Universitas Muhammadiyah Surakarta; 2023.
[11] Surya J, Aminuddin FH. *Pemrograman MySQL Database with Streamlit Python*. PT Sonpedia Publishing Indonesia; 2024.
[12] Wenny. "Normalisasi Data Kependudukan dengan Model Min Max dan Algoritma K-Means untuk Pengelompokan Tingkat Ekonomi Masyarakat." *Bulletin of Information System Research*. 2024;2(2):53–63.
[13] Akbar SCD, Defit S, Hendrik B. "Segmentasi Tunggakan Pelanggan Menggunakan Algoritma K-Means Cluster pada Perusahaan Air Minum Daerah." *Jurnal Pustaka AI (Pusat Akses Kajian Teknologi Artificial Intelligence)*. 2025;5(2):349–355.
[14] Perdana SA, Florentin SF, Santoso A. "Analisis Segmentasi Pelanggan Menggunakan K-Means Clustering Studi Kasus Aplikasi Alfagift." *Sebatik*. 2022;26(2):446–457.
[15] Handayani F. "Aplikasi Data Mining Menggunakan Algoritma K-Means Clustering untuk Mengelompokkan Mahasiswa Berdasarkan Gaya Belajar." *Jurnal Teknologi dan Informasi*. 2022;12(1):46–63.
[16] Mutiara M, Siregar S. "Analisis Penyebab Tunggakan Pembayaran Rekening Air Pelanggan pada Perusahaan Daerah Air Minum (PDAM) Tirta Sari Binjai." *Ekonomi Bisnis Manajemen dan Akuntansi (EBMA)*. 2025;6(2):233–237.
[17] Rohana S, Buaton R, Annatasia K. "Cluster Pelanggan Listrik PLN UP3 Binjai dengan Metode K-Means." *Jurnal Nasional Teknologi Komputer*. 2025;5(3):495–507.
[18] [18] Sari P. "Algoritma K-Means Clustering: Sebuah Studi Literatur." *Jurnal Informatika*. 2024;1(1):1–7.
[19] Hendriyani Y. *Dasar-Dasar Data Mining: Konsep, Teknik, dan Aplikasi*. 2025.
[20] Ocktavia S, Atmojo WT. "Analisis Segmentasi Pelanggan Pembiayaan Berdasarkan Demografi untuk Memprediksi Tingkat Kredit Menggunakan Algoritma K-Means." *Jurnal Informatika Teknologi dan Sains*. 2025;7(1):394–401.