



SKRIPSI

SEGMENTASI POLA PEMBAYARAN PELANGGAN PROPERTI PADA PT XYZ MENGUNAKAN METODE *X-MEANS*

EDELIN FORTUNA
NPM 21083010087

DOSEN PEMBIMBING

Dr. Eng. Ir. Dwi Arman Prasetya, S.T., M.T., IPU., Asean, Eng.
Kartika Maulida Hindrayani, S.Kom., M.Kom.

**KEMENTERIAN PENDIDIKAN TINGGI, SAINS, DAN TEKNOLOGI
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI SAINS DATA
SURABAYA
2025**



SKRIPSI

**SEGMENTASI POLA PEMBAYARAN
PELANGGAN PROPERTI PADA PT XYZ
MENGUNAKAN METODE *X-MEANS***

EDELIN FORTUNA
NPM 21083010087

DOSEN PEMBIMBING

Dr. Eng. Ir. Dwi Arman Prasetya, S.T., M.T., IPU., Asean, Eng.
Kartika Maulida Hindrayani, S.Kom., M.Kom.

**KEMENTERIAN PENDIDIKAN TINGGI, SAINS, DAN TEKNOLOGI
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI SAINS DATA
SURABAYA
2025**



SKRIPSI

SEGMENTASI POLA PEMBAYARAN PELANGGAN PROPERTI PADA PT XYZ MENGUNAKAN METODE X-MEANS

EDELIN FORTUNA
NPM 21083010087

DOSEN PEMBIMBING

Dr. Eng. Ir. Dwi Arman Prasetya, S.T., M.T., IPU., Asean, Eng.
Kartika Maulida Hindrayani, S.Kom., M.Kom.

**KEMENTERIAN PENDIDIKAN TINGGI, SAINS, DAN TEKNOLOGI
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI SAINS DATA
SURABAYA
2025**

LEMBAR PENGESAHAN

SEGMENTASI POLA PEMBAYARAN PELANGGAN PROPERTI PADA PT XYZ MENGGUNAKAN METODE X-MEANS

Oleh:
EDELIN FORTUNA
NPM. 21083010087

Telah dipertahankan di hadapan dan diterima oleh Tim Penguji Sidang Skripsi Program Studi Sains Data Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jawa Timur Pada tanggal 10 September 2025.

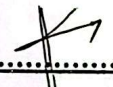
Menyetujui,

Dr. Eng. Ir. Dwi Arman Prasetya, S.T., M.T.,
IPU., Asean, Eng.
NIP. 19801205 200501 1 002


.....

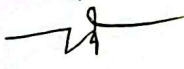
(Pembimbing I)

Kartika Maulida Hindrayani, S.Kom., M.Kom.
NIP. 19920909 202203 2 009


.....

(Pembimbing II)

Dr. Ir. Mohammad Idhom, S.P., S.Kom., M.T.
NIP. 19830310 202121 1 006


.....

(Ketua Penguji)

Alfan Rizaldy Pratama, S.Tr.T., M.Tr.Kom.
NIP. 19990606 202406 1 001


.....

(Penguji I)

Mengetahui,

Dekan Fakultas Ilmu Komputer


.....

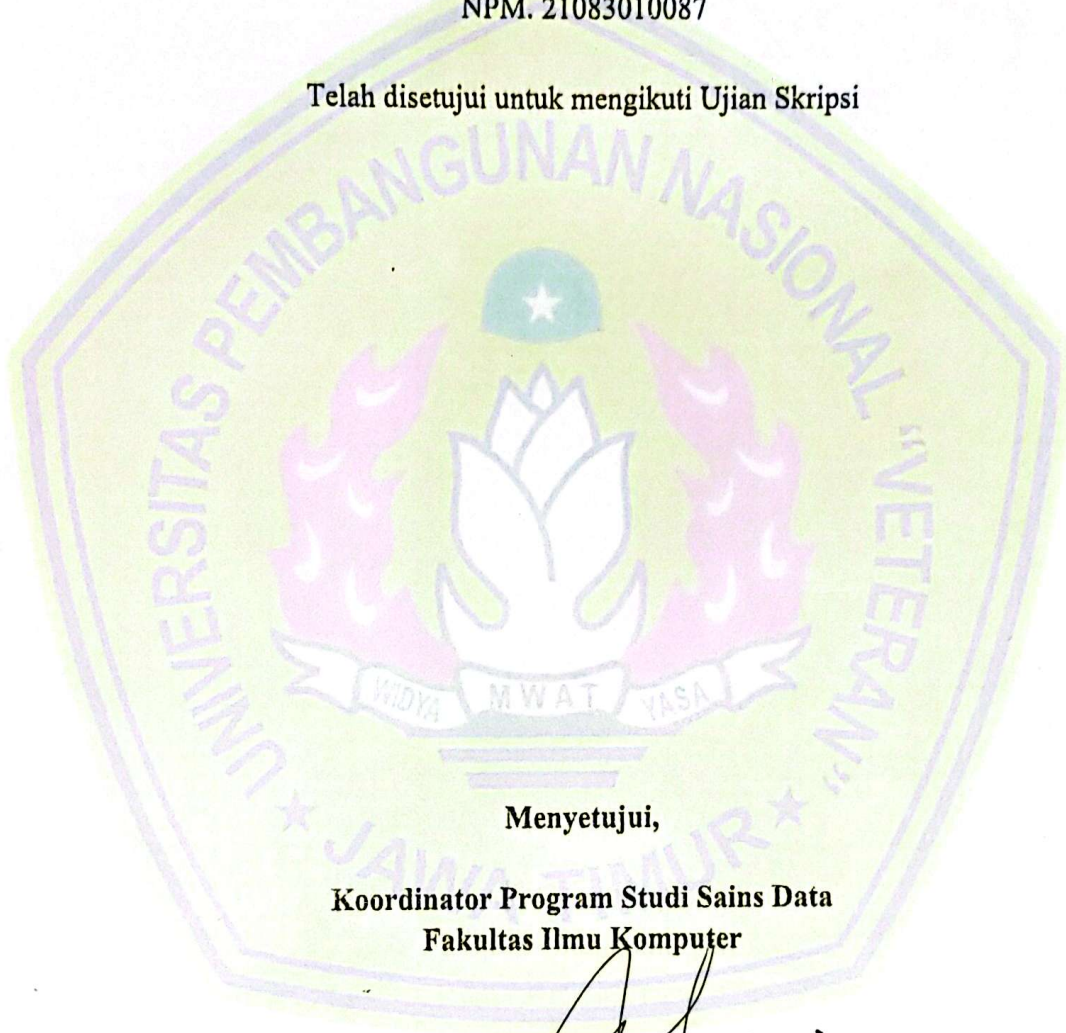
Prof. Dr. Ir. Novirina Hendrasarie, MT.
NIP. 19681126 199403 2 001

LEMBAR PERSETUJUAN

**SEGMENTASI POLA PEMBAYARAN PELANGGAN PROPERTI PADA
PT XYZ MENGGUNAKAN METODE X-MEANS**

Oleh:
EDELIN FORTUNA
NPM. 21083010087

Telah disetujui untuk mengikuti Ujian Skripsi



Menyetujui,

**Koordinator Program Studi Sains Data
Fakultas Ilmu Komputer**

Dr. Eng. Ir. Dwi Arman Prasetya, S.T., M.T., IPU., Asean, Eng.
NIP. 19801205 200501 1 002

SURAT PERNYATAAN BEBAS PLAGIASI

Saya yang bertanda tangan di bawah ini:

Nama Mahasiswa : Edelin Fortuna
NPM : 21083010087
Program : Sarjana (S1)
Program Studi : Sains Data
Fakultas : Fakultas Ilmu Komputer

Menyatakan bahwa dalam dokumen ilmiah Skripsi ini tidak terdapat bagian dari karya ilmiah lain yang telah diajukan untuk memperoleh gelar akademik di suatu lembaga Pendidikan Tinggi, dan juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang/lembaga lain, kecuali yang secara tertulis disitasi dalam dokumen ini dan disebutkan secara lengkap dalam daftar pustaka.

Dan saya menyatakan bahwa dokumen ilmiah ini bebas dari unsur-unsur plagiasi. Apabila di kemudian hari ditemukan indikasi plagiat pada Skripsi ini, saya bersedia menerima sanksi sesuai dengan peraturan perundang-undangan yang berlaku.

Demikian surat pernyataan ini saya buat dengan sesungguhnya tanpa ada paksaan dari siapapun juga dan untuk dipergunakan sebagaimana mestinya



Surabaya, 11 September 2025
Yang Membuat Pernyataan,



EDELIN FORTUNA
NPM. 21083010087

ABSTRAK

Nama Mahasiswa / NPM : Edelin Fortuna / 21083010087
Judul Proposal Skripsi : Segmentasi Pola Pembayaran Pelanggan Properti
Pada PT XYZ Menggunakan Metode *X-Means*
Dosen Pembimbing : 1. Dr. Eng. Ir. Dwi Arman Prasetya, S.T., M.T.,
IPU., Asean, Eng.
2. Kartika Maulida Hindrayani, S.Kom., M.Kom.

Memahami perilaku pelanggan sangat penting untuk mendukung keberlangsungan bisnis properti. Penelitian ini bertujuan untuk mengelompokkan pelanggan PT XYZ berdasarkan pola pembayaran angsuran mereka dengan menggunakan algoritma klasterisasi *X-Means*. Sebanyak 9.615 catatan transaksi di integrasi menjadi 386 profil pelanggan dan dianalisis menggunakan fitur yang telah dipilih karena menggambarkan karakteristik individual pembeli, seperti jumlah keterlambatan pembayaran dan status pembayaran. *X-Means* dipilih karena memiliki efisiensi dalam komputasi membentuk klaster. Proses klasterisasi menghasilkan 3 klaster. Dilakukan evaluasi klaster dengan menggunakan metrik evaluasi *Silhouette Score* diperoleh 0,606, *Davies Bouldin Index* (DBI) sebesar 0,538, dan *Calinski Harabsz* sebesar 472.924 yang menunjukkan pemisahan klaster dapat dikategorikan cukup baik. Hasil klasterisasi ini memiliki pola pelanggan yang melakukan transaksi pembayaran sekali langsung lunas, pelanggan yang masih menunggu, pelanggan yang telat bayar. Beragam karakteristik pelanggan, dapat dimanfaatkan untuk merancang strategi komunikatif kepada pelanggan, mengoptimalkan proses penagihan, serta meningkatkan pengelolaan hubungan pelanggan secara keseluruhan guna mendukung pertumbuhan bisnis perusahaan.

Kata kunci: Klasterisasi, Pelanggan, Pola Pembayaran, Segmentasi *Silhouette Score*, *X-Means*

Halaman ini sengaja dikosongkan

ABSTRACT

Student Name / NPM : Edelin Fortuna / 21083010087
Thesis Title : *Segmentation of Customer Payment Patterns in Property at PT XYZ Using the X-Means Method*
Advisor : 1. Dr. Eng. Ir. Dwi Arman Prasetya, S.T., M.T.,
IPU., Asean, Eng.
2. Kartika Maulida Hindrayani, S.Kom., M.Kom.

Understanding customer behavior is essential to support the sustainability of the property business. This study aims to segment the customers of PT XYZ based on their installment payment patterns using the X-Means Clustering algorithm. A total of 9,615 transaction records were integrated into 386 customer profiles and analyzed using selected features that represent individual buyer characteristics, such as the number of late payments and payment status. X-Means was chosen due to its computational efficiency in forming clusters. The Clustering process resulted in third clusters. Cluster evaluation was conducted using three metrics: the Silhouette Score 0.606, Davies-Bouldin Index 0,538, and Calinski-Harabasz score 472.924, indicating that the cluster separation is reasonably good. The Clustering results revealed distinct customer patterns, including those who make a single full payment, those with outstanding balances, and those with frequent payment delays. These diverse customer characteristics can be leveraged to design targeted communication strategies, optimize the billing process, and enhance overall customer relationship management, ultimately supporting the company's business growth.

Keywords: *Clustering, Payment Pattern, Segmentation, Silhouette Score, X-Means Clustering*

Halaman ini sengaja dikosongkan

KATA PENGANTAR

Puji syukur kehadirat Allah SWT atas segala rahmat, hidayah dan karunia-Nya kepada penulis sehingga skripsi berjudul **“Segmentasi Pola Pembayaran Pelanggan Properti Pada PT XYZ Menggunakan Metode X-Means”** dapat terselesaikan dengan baik. Penulis mengucapkan terima kasih kepada Bapak Dr. Eng. Ir. Dwi Arman Prasetya., ST., MT., IPU., Asean. Eng dan juga Bu Kartika Maulida Hindrayani, S.Kom., M.Kom., selaku dosen pembimbing saya yang telah meluangkan waktunya untuk membimbing dan memberikan kritik serta saran yang membangun, nasihat serta motivasi kepada penulis. Dan penulis juga banyak menerima bantuan dari berbagai pihak, baik itu berupa moril, spiritual maupun materiil. Untuk itu penulis mengucapkan terima kasih kepada:

1. Bapak Prof. Dr. Ir. Akhmad Fauzi, MMT., IPU., ASEAN. Eng., selaku Rektor Universitas Pembangunan Nasional “Veteran” Jawa Timur.
2. Ibu Prof. Dr. Ir. Novirina Hendrasarie, MT., beserta jajarannya selaku Dekan dan Wakil Dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur.
3. Bapak Dr. Eng. Ir. Dwi Arman Prasetya., ST., MT., IPU., Asean. Eng., selaku Koordinator Program Studi Sains Data Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur.
4. Dosen-dosen Program Studi Sains Data yang telah membimbing serta memberikan arahan selama masa perkuliahan.
5. Kedua orang tua penulis yang selalu senantiasa mendoakan dan mendukung penulis, memberikan nasihat-nasihat yang membuat penulis bisa menyelesaikan penelitian dan buku skripsi ini.
6. Kedua kakak penulis Rani dan Ame serta adik penulis Aurel yang selalu membeikan dukungan dalam bentuk apapun demi menyelesaikan penelitian dan buku skripsi ini.
7. Kepada sahabat-sahabat penulis, Aisyah, Putri, Dzakiyyah, Bella, Dona, Azrel, Made Ayu, Lila, sahabat kecil penulis Syifa, Alvira, dan Diana yang selalu memberikan pertolongan apapun saat penulis kesusahaan dan menemani penulis saat mengerjakan buku skripsi ini.

8. Kepada teman-teman kuliah penulis, Nayya, Awal, Cesaria, Firly, Novita, dan Radya yang telah memberikan dukungan serta selalu mengajak penulis untuk melakukan bimbingan dan memberikan kritik masukan kepada penulis dikala kebingungan dalam menyelesaikan penelitian ini.
9. Himpunan Mahasiswa Sains Data yang telah memberikan pengalaman baru yang berkesan dan tak terlupakan kepada penulis saat masa perkuliahan berlangsung.
10. Seluruh angkatan 2021 Program Studi Sains Data Universitas Pembangunan Nasional “Veteran” Jawa Timur.
11. Dan yang terakhir, kepada diri saya sendiri, apresiasi sebesar-besarnya karena telah bertanggung jawab atas pilihan yang diambil untuk menyelesaikan apa yang telah dimulai. Terima kasih telah bertahan, mau terus berusaha dan berjuang sehingga tidak pantang menyerah dalam situasi dan kondisi apapun.

Penulis menyadari bahwa di dalam penyusunan skripsi ini masih terdapat kekurangan. Untuk itu kritik dan saran yang membangun dari semua pihak sangat diharapkan demi kesempurnaan penulisan buku skripsi ini. Dengan segala keterbatasan yang penulis miliki semoga penelitian ini dapat bermanfaat bagi seluruh pihak yang membaca buku skripsi ini.

Surabaya, 10 September 2025

Penulis

DAFTAR ISI

HALAMAN SAMPUL.....	i
LEMBAR PENGESAHAN	ii
LEMBAR PERSETUJUAN	iii
SURAT PERNYATAAN BEBAS PLAGIASI.....	iv
ABSTRAK	v
<i>ABSTRACT</i>	vii
KATA PENGANTAR.....	ix
DAFTAR ISI.....	xi
DAFTAR GAMBAR.....	xv
DAFTAR TABEL	xvii
DAFTAR LAMPIRAN	xix
DAFTAR NOTASI.....	xxi
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	5
1.3. Batasan Masalah.....	5
1.4. Tujuan Penelitian	6
1.5. Manfaat Penelitian	6
BAB II TINJAUAN PUSTAKA.....	9
2.1 Penelitian Terdahulu	9
2.2 Dasar Teori.....	14
2.2.1. Properti.....	14
2.2.2. <i>Inhouse</i> Syariah.....	15
2.2.3. Segmentasi	15
2.2.4. Klasterisasi	17
2.2.5. <i>K-Means</i>	18
2.2.6. <i>X-Means</i>	19
2.2.7. <i>RobustScaler</i>	21
2.2.8. <i>Sillhouette Score</i>	22
2.2.9. <i>Davies Bouldin Index (DBI)</i>	23

2.2.10.	<i>Calinski Harabasz</i>	24
BAB III DESAIN DAN IMPLEMENTASI SISTEM		27
3.1	Variabel Penelitian dan Sumber Data	27
3.2.	Langkah Analisis.....	28
3.2.1	Identifikasi Masalah	28
3.2.2	Persiapan Data.....	29
3.2.3	<i>Preprocessing</i> Data	31
3.3.3	Explorasi Data Analisis.....	32
3.3.4	Implementasi <i>X-Means</i>	34
3.3.5	Evaluasi Model	35
3.3.6	Interpretasi Klaster	37
3.3.7	Visualisasi Klaster.....	37
3.3.8	Rekomendasi Strategi.....	39
3.3.	Desain Sistem.....	39
BAB IV HASIL DAN PEMBAHASAN		43
4.1.	Persiapan Data.....	43
4.1.1.	Membaca Data	43
4.1.2.	Seleksi Fitur	44
4.1.3.	<i>Missing Value</i>	46
4.1.4.	Konversi Tipe Data	47
4.1.5.	Agregasi dan Penggabungan Data	48
4.1.6.	<i>Missing Value</i>	53
4.2.	Data <i>Preprocessing</i>	55
4.2.1.	Transformasi Data.....	55
4.2.2.	Standarisasi Data.....	57
4.3.	Explorasi Data Analisis.....	60
4.3.1.	Analisis Distribusi.....	60
4.3.2.	Identifikasi <i>Outlier</i>	63
4.3.3.	Korelasi <i>Pearson</i>	66
4.3.4.	Menghitung <i>Variance Inflation Factor (VIF)</i>	68
4.4.	Implementasi <i>X-Means</i>	70
4.5.	Evaluasi Model <i>X-Means</i>	73

4.6.	Perbandingan Waktu Komputasi <i>X-Means</i> dan <i>K-Means</i>	75
4.7.	Interpretasi Klaster	75
4.8.	Visualisasi Klaster.....	77
4.9.	Kontribusi dan Pemanfaatan, Rekomendasi, serta Implementasi Strategi	80
4.10.	<i>Deployment</i>	82
BAB V PENUTUP		85
5.1.	Kesimpulan	85
5.2.	Saran Pengembangan	86
DAFTAR PUSTAKA		87
LAMPIRAN.....		91

Halaman ini sengaja dikosongkan

DAFTAR GAMBAR

Gambar 2.1. Ilustrasi Segementasi Pelanggan[23]	16
Gambar 2.2. Ilustrasi <i>Clustering</i> [24]	18
Gambar 3.1. Diagram Alir Penelitian	28
Gambar 3.2. Diagram alir Desain Sistem.....	41
Gambar 4.1. Distribusi Jumlah Transaksi.....	61
Gambar 4.2. Distribusi Total Pembayaran.....	61
Gambar 4.3. Distribusi Harga	62
Gambar 4.4. Distribusi Selisih	62
Gambar 4.5. Distribusi Jumlah Terlambat	63
Gambar 4.6. <i>Box Plot</i> Jumlah Transaksi.....	64
Gambar 4.7. <i>Box Plot</i> Total Pembayaran.....	64
Gambar 4.8. <i>Box Plot</i> Harga	65
Gambar 4.9. <i>Box Plot</i> Selisih	65
Gambar 4.10. <i>Box Plot</i> Jumlah terlambat	66
Gambar 4.11. Korelasi <i>Pearson</i>	68
Gambar 4.12. Anggota Klaster.....	77
Gambar 4.13. Jumlah Pelanggan Berdasarkan Klaster dan Status Pembayaran...	78
Gambar 4.14. t-SNE Persebaran Klaster.....	79
Gambar 4.15. Halaman Utama.....	82
Gambar 4.16. Fitur Web	83
Gambar 4.17. Halaman Fitur Data	84
Gambar 4.18. Halaman Fitur Analisis & Klasterisasi.....	84

Halaman ini sengaja dikosongkan

DAFTAR TABEL

Tabel 2.1. Tabel Penelitian Terdahulu	9
Tabel 3.1. Metadata.....	27
Tabel 4.1. Data Angsuran.....	44
Tabel 4.2. Data Utama	44
Tabel 4.3. Pemilihan Data Angsuran pertama	45
Tabel 4.4. Pemilihan Data Utama	45
Tabel 4.5. Pemilihan Data Angsuran Kedua.....	46
Tabel 4.6. <i>Output Missing Value</i>	47
Tabel 4.7. Agregasi df1	49
Tabel 4.8. Hasil Gabungan df1 dan df2	50
Tabel 4.9. Hasil Perhitungan Selisih Pembayaran	50
Tabel 4.10. Penambahan Status Pembayaran.....	51
Tabel 4.11. Perhitungan Keterlambatan Pembayaran	52
Tabel 4.12. Jumlah Terlambat Setiap Unit.....	52
Tabel 4.13. Dataset Analisis.....	53
Tabel 4.14. Melihat <i>Missing Value</i>	54
Tabel 4.15. Dataset Penelitian.....	55
Tabel 4.16. Data Hasil <i>Encoding</i>	56
Tabel 4.17. Penjelasan Fitur.....	57
Tabel 4.18. Dataset Standarisasi	59
Tabel 4.19. Perhitungan VIF	69
Tabel 4.20. Perhitungan VIF Setelah Diatasi.....	70
Tabel 4.21. Dataset.....	71
Tabel 4.22. Perbandingan Waktu Komputasi	75
Tabel 4.23. Kontribusi dan Pemanfaatan Strategi.....	80
Tabel 4.24. Rekomendasi Strategi	81
Tabel 4.25. Contoh Implementasi Nyata	81

Halaman ini sengaja dikosongkan

DAFTAR LAMPIRAN

Lampiran 1. Data Penelitian.....	91
Lampiran 2. Kode <i>Script</i>	92
Lampiran 3. <i>Script</i> Pembuatan <i>Interface</i>	93
Lampiran 4. LoA Artikel.....	94

Halaman ini sengaja dikosongkan

DAFTAR NOTASI

k	: Jumlah klaster
c_i	: Himpunan data diklaster i
x	: Vektor data
μ_i	: <i>Centroid</i> klaster
$C_{m(q)}$	: <i>Centroid</i> dari klaster m pada perhitungan ke q
n_m	: Banyaknya data pada klaster m
$x_{i(q)}$	: Data ke i pada iterasi ke q
BIC	: Metrik statistik <i>Bayesian Information Criterion</i>
L	: Faktor bobot
$\frac{p}{2}$	: Jumlah klaster
$\log N$	: Jumlah data titik
$\frac{n}{2}$	: Jumlah data
d	: Dimensi data
σ^2	: Variansi dalam klaster
μ	: Mean (rata-rata) dari distribusi data (pusat)
x_i	: Data ke i
$\ x_i - \mu\ ^2$	: Jarak kuadrat (<i>squared Euclidean distance</i>) antara data ke i dengan mean (rata-rata)
$\sum_{i=1}^n \ x_i - \mu\ ^2$	: Total jarak kuadrat semua data terhadap mean (rata-rata)
$\frac{n}{2} \log(2\pi)$	: Konstanta normalisasi Gaussian
SSE	: Total error kuadrat dari semua titik ke <i>centroid</i>
n	: Jumlah titik dalam klaster
x'	: <i>Scaling robustscaller</i>
x	: Nilai asli data
IQR	: Rentang antar kuartil
$Q3$	: Kuartil 3, nilai persentil ke-75 (75% data terbawah)
$Q1$	: Kuartil 1, nilai persentil ke-25 (25% data terbawah)

$s(i)$	: <i>Silhouette score</i> untuk data ke i
$b(i)$	: Rata-rata jarak antara data i dengan klaster terdekat lainnya (klaster tetangga terdekat)
$a(i)$	: Rata-rata jarak antara data i dengan semua data lain dalam klaster yang sama
$\max(a(i), b(i))$	: Normalisasi supaya nilai $s(i)$ selalu berada antara -1 sampai +1
DBI	: Metrik statistik <i>Davies Bouldin Index</i>
$\frac{1}{k}$	: Jumlah klaster
s_i	: Rata-rata jarak antara setiap titik data dalam klaster i terhadap pusat klaster
s_j	: Rata-rata jarak antara setiap titik data dalam klaster j terhadap pusat klaster
M_{ij}	: Jarak antara pusat klaster
$\frac{s_i + s_j}{M_{ij}}$	: Perhitungan rasio kemiripan antara klaster i dan j
CH	: Calinski Harabasz
$Tr(B_k)$	: <i>Trace</i> atau jumlah diagonal dari matriks disperse antar klaster
$Tr(W_k)$	: <i>Trace</i> dari matriks disperse dalam klaster
n	: Jumlah total data
k	: Jumlah klaster

ABSTRAK

Nama Mahasiswa / NPM : Edelin Fortuna / 21083010087
Judul Proposal Skripsi : Segmentasi Pola Pembayaran Pelanggan Properti
Pada PT XYZ Menggunakan Metode *X-Means*
Dosen Pembimbing : 1. Dr. Eng. Ir. Dwi Arman Prasetya, S.T., M.T.,
IPU., Asean, Eng.
2. Kartika Maulida Hindrayani, S.Kom., M.Kom.

Memahami perilaku pelanggan sangat penting untuk mendukung keberlangsungan bisnis properti. Penelitian ini bertujuan untuk mengelompokkan pelanggan PT XYZ berdasarkan pola pembayaran angsuran mereka dengan menggunakan algoritma klasterisasi *X-Means*. Sebanyak 9.615 catatan transaksi di integrasi menjadi 386 profil pelanggan dan dianalisis menggunakan fitur yang telah dipilih karena menggambarkan karakteristik individual pembeli, seperti jumlah keterlambatan pembayaran dan status pembayaran. *X-Means* dipilih karena memiliki efisiensi dalam komputasi membentuk klaster. Proses klasterisasi menghasilkan 3 klaster. Dilakukan evaluasi klaster dengan menggunakan metrik evaluasi *Silhouette Score* diperoleh 0,606, *Davies Bouldin Index* (DBI) sebesar 0,538, dan *Calinski Harabsz* sebesar 472.924 yang menunjukkan pemisahan klaster dapat dikategorikan cukup baik. Hasil klasterisasi ini memiliki pola pelanggan yang melakukan transaksi pembayaran sekali langsung lunas, pelanggan yang masih menunggu, pelanggan yang telat bayar. Beragam karakteristik pelanggan, dapat dimanfaatkan untuk merancang strategi komunikatif kepada pelanggan, mengoptimalkan proses penagihan, serta meningkatkan pengelolaan hubungan pelanggan secara keseluruhan guna mendukung pertumbuhan bisnis perusahaan.

Kata kunci: Klasterisasi, Pelanggan, Pola Pembayaran, Segmentasi *Silhouette Score*, *X-Means*

Halaman ini sengaja dikosongkan

ABSTRACT

Student Name / NPM : Edelin Fortuna / 21083010087
Thesis Title : *Segmentation of Customer Payment Patterns in Property at PT XYZ Using the X-Means Method*
Advisor : 1. Dr. Eng. Ir. Dwi Arman Prasetya, S.T., M.T.,
IPU., Asean, Eng.
2. Kartika Maulida Hindrayani, S.Kom., M.Kom.

Understanding customer behavior is essential to support the sustainability of the property business. This study aims to segment the customers of PT XYZ based on their installment payment patterns using the X-Means Clustering algorithm. A total of 9,615 transaction records were integrated into 386 customer profiles and analyzed using selected features that represent individual buyer characteristics, such as the number of late payments and payment status. X-Means was chosen due to its computational efficiency in forming clusters. The Clustering process resulted in third clusters. Cluster evaluation was conducted using three metrics: the Silhouette Score 0.606, Davies-Bouldin Index 0,538, and Calinski-Harabasz score 472.924, indicating that the cluster separation is reasonably good. The Clustering results revealed distinct customer patterns, including those who make a single full payment, those with outstanding balances, and those with frequent payment delays. These diverse customer characteristics can be leveraged to design targeted communication strategies, optimize the billing process, and enhance overall customer relationship management, ultimately supporting the company's business growth.

Keywords: *Clustering, Payment Pattern, Segmentation, Silhouette Score, X-Means Clustering*

Halaman ini sengaja dikosongkan

KATA PENGANTAR

Puji syukur kehadirat Allah SWT atas segala rahmat, hidayah dan karunia-Nya kepada penulis sehingga skripsi berjudul **“Segmentasi Pola Pembayaran Pelanggan Properti Pada PT XYZ Menggunakan Metode X-Means”** dapat terselesaikan dengan baik. Penulis mengucapkan terima kasih kepada Bapak Dr. Eng. Ir. Dwi Arman Prasetya., ST., MT., IPU., Asean. Eng dan juga Bu Kartika Maulida Hindrayani, S.Kom., M.Kom., selaku dosen pembimbing saya yang telah meluangkan waktunya untuk membimbing dan memberikan kritik serta saran yang membangun, nasihat serta motivasi kepada penulis. Dan penulis juga banyak menerima bantuan dari berbagai pihak, baik itu berupa moril, spiritual maupun materiil. Untuk itu penulis mengucapkan terima kasih kepada:

1. Bapak Prof. Dr. Ir. Akhmad Fauzi, MMT., IPU., ASEAN. Eng., selaku Rektor Universitas Pembangunan Nasional “Veteran” Jawa Timur.
2. Ibu Prof. Dr. Ir. Novirina Hendrasarie, MT., beserta jajarannya selaku Dekan dan Wakil Dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur.
3. Bapak Dr. Eng. Ir. Dwi Arman Prasetya., ST., MT., IPU., Asean. Eng., selaku Koordinator Program Studi Sains Data Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur.
4. Dosen-dosen Program Studi Sains Data yang telah membimbing serta memberikan arahan selama masa perkuliahan.
5. Kedua orang tua penulis yang selalu senantiasa mendoakan dan mendukung penulis, memberikan nasihat-nasihat yang membuat penulis bisa menyelesaikan penelitian dan buku skripsi ini.
6. Kedua kakak penulis Rani dan Ame serta adik penulis Aurel yang selalu membeikan dukungan dalam bentuk apapun demi menyelesaikan penelitian dan buku skripsi ini.
7. Kepada sahabat-sahabat penulis, Aisyah, Putri, Dzakiyyah, Bella, Dona, Azrel, Made Ayu, Lila, sahabat kecil penulis Syifa, Alvira, dan Diana yang selalu memberikan pertolongan apapun saat penulis kesusahaan dan menemani penulis saat mengerjakan buku skripsi ini.

8. Kepada teman-teman kuliah penulis, Nayya, Awal, Cesaria, Firly, Novita, dan Radya yang telah memberikan dukungan serta selalu mengajak penulis untuk melakukan bimbingan dan memberikan kritik masukan kepada penulis dikala kebingungan dalam menyelesaikan penelitian ini.
9. Himpunan Mahasiswa Sains Data yang telah memberikan pengalaman baru yang berkesan dan tak terlupakan kepada penulis saat masa perkuliahan berlangsung.
10. Seluruh angkatan 2021 Program Studi Sains Data Universitas Pembangunan Nasional “Veteran” Jawa Timur.
11. Dan yang terakhir, kepada diri saya sendiri, apresiasi sebesar-besarnya karena telah bertanggung jawab atas pilihan yang diambil untuk menyelesaikan apa yang telah dimulai. Terima kasih telah bertahan, mau terus berusaha dan berjuang sehingga tidak pantang menyerah dalam situasi dan kondisi apapun.

Penulis menyadari bahwa di dalam penyusunan skripsi ini masih terdapat kekurangan. Untuk itu kritik dan saran yang membangun dari semua pihak sangat diharapkan demi kesempurnaan penulisan buku skripsi ini. Dengan segala keterbatasan yang penulis miliki semoga penelitian ini dapat bermanfaat bagi seluruh pihak yang membaca buku skripsi ini.

Surabaya, 10 September 2025

Penulis

DAFTAR ISI

HALAMAN SAMPUL.....	i
LEMBAR PENGESAHAN	ii
LEMBAR PERSETUJUAN	iii
SURAT PERNYATAAN BEBAS PLAGIASI.....	iv
ABSTRAK	v
<i>ABSTRACT</i>	vii
KATA PENGANTAR.....	ix
DAFTAR ISI.....	xi
DAFTAR GAMBAR.....	xv
DAFTAR TABEL	xvii
DAFTAR LAMPIRAN	xix
DAFTAR NOTASI.....	xxi
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	5
1.3. Batasan Masalah.....	5
1.4. Tujuan Penelitian	6
1.5. Manfaat Penelitian	6
BAB II TINJAUAN PUSTAKA.....	9
2.1 Penelitian Terdahulu	9
2.2 Dasar Teori.....	14
2.2.1. Properti.....	14
2.2.2. <i>Inhouse</i> Syariah.....	15
2.2.3. Segmentasi	15
2.2.4. Klasterisasi	17
2.2.5. <i>K-Means</i>	18
2.2.6. <i>X-Means</i>	19
2.2.7. <i>RobustScaler</i>	21
2.2.8. <i>Sillhouette Score</i>	22
2.2.9. <i>Davies Bouldin Index (DBI)</i>	23

2.2.10. <i>Calinski Harabasz</i>	24
BAB III DESAIN DAN IMPLEMENTASI SISTEM	27
3.1 Variabel Penelitian dan Sumber Data.....	27
3.2. Langkah Analisis	28
3.2.1 Identifikasi Masalah	28
3.2.2 Persiapan Data	29
3.2.3 <i>Preprocessing</i> Data	31
3.3.3 Explorasi Data Analisis	32
3.3.4 Implementasi <i>X-Means</i>	34
3.3.5 Evaluasi Model.....	35
3.3.6 Interpretasi Kluster	37
3.3.7 Visualisasi Kluster	37
3.3.8 Rekomendasi Strategi.....	39
3.3. Desain Sistem	39
BAB IV HASIL DAN PEMBAHASAN	43
4.1. Persiapan Data	43
4.1.1. Membaca Data.....	43
4.1.2. Seleksi Fitur.....	44
4.1.3. <i>Missing Value</i>	46
4.1.4. Konversi Tipe Data.....	47
4.1.5. Agregasi dan Penggabungan Data.....	48
4.1.6. <i>Missing Value</i>	53
4.2. Data <i>Preprocessing</i>	55
4.2.1. Transformasi Data	55
4.2.2. Standarisasi Data	57
4.3. Explorasi Data Analisis	60
4.3.1. Analisis Distribusi	60
4.3.2. Identifikasi <i>Outlier</i>	63
4.3.3. Korelasi <i>Pearson</i>	66
4.3.4. Menghitung <i>Variance Inflation Factor</i> (VIF)	68
4.4. Implementasi <i>X-Means</i>	70
4.5. Evaluasi Model <i>X-Means</i>	73

4.6.	Perbandingan Waktu Komputasi <i>X-Means</i> dan <i>K-Means</i>	75
4.7.	Interpretasi Klaster	75
4.8.	Visualisasi Klaster.....	77
4.9.	Kontribusi dan Pemanfaatan, Rekomendasi, serta Implementasi Strategi	80
4.10.	<i>Deployment</i>	82
BAB V PENUTUP		85
5.1.	Kesimpulan	85
5.2.	Saran Pengembangan	86
DAFTAR PUSTAKA		87
LAMPIRAN.....		91

Halaman ini sengaja dikosongkan

DAFTAR GAMBAR

Gambar 2.1. Ilustrasi Segementasi Pelanggan[23]	16
Gambar 2.2. Ilustrasi <i>Clustering</i> [24]	18
Gambar 3.1. Diagram Alir Penelitian	28
Gambar 3.2. Diagram alir Desain Sistem.....	41
Gambar 4.1. Distribusi Jumlah Transaksi.....	61
Gambar 4.2. Distribusi Total Pembayaran.....	61
Gambar 4.3. Distribusi Harga	62
Gambar 4.4. Distribusi Selisih	62
Gambar 4.5. Distribusi Jumlah Terlambat	63
Gambar 4.6. <i>Box Plot</i> Jumlah Transaksi.....	64
Gambar 4.7. <i>Box Plot</i> Total Pembayaran.....	64
Gambar 4.8. <i>Box Plot</i> Harga	65
Gambar 4.9. <i>Box Plot</i> Selisih	65
Gambar 4.10. <i>Box Plot</i> Jumlah terlambat	66
Gambar 4.11. Korelasi <i>Pearson</i>	68
Gambar 4.12. Anggota Klaster.....	77
Gambar 4.13. Jumlah Pelanggan Berdasarkan Klaster dan Status Pembayaran...	78
Gambar 4.14. t-SNE Persebaran Klaster.....	79
Gambar 4.15. Halaman Utama.....	82
Gambar 4.16. Fitur Web	83
Gambar 4.17. Halaman Fitur Data	84
Gambar 4.18. Halaman Fitur Analisis & Klasterisasi.....	84

Halaman ini sengaja dikosongkan

DAFTAR TABEL

Tabel 2.1. Tabel Penelitian Terdahulu	9
Tabel 3.1. Metadata.....	27
Tabel 4.1. Data Angsuran.....	44
Tabel 4.2. Data Utama	44
Tabel 4.3. Pemilihan Data Angsuran pertama	45
Tabel 4.4. Pemilihan Data Utama	45
Tabel 4.5. Pemilihan Data Angsuran Kedua.....	46
Tabel 4.6. <i>Output Missing Value</i>	47
Tabel 4.7. Agregasi df1	49
Tabel 4.8. Hasil Gabungan df1 dan df2	50
Tabel 4.9. Hasil Perhitungan Selisih Pembayaran	50
Tabel 4.10. Penambahan Status Pembayaran.....	51
Tabel 4.11. Perhitungan Keterlambatan Pembayaran	52
Tabel 4.12. Jumlah Terlambat Setiap Unit.....	52
Tabel 4.13. Dataset Analisis.....	53
Tabel 4.14. Melihat <i>Missing Value</i>	54
Tabel 4.15. Dataset Penelitian.....	55
Tabel 4.16. Data Hasil <i>Encoding</i>	56
Tabel 4.17. Penjelasan Fitur.....	57
Tabel 4.18. Dataset Standarisasi	59
Tabel 4.19. Perhitungan VIF	69
Tabel 4.20. Perhitungan VIF Setelah Diatasi.....	70
Tabel 4.21. Dataset.....	71
Tabel 4.22. Perbandingan Waktu Komputasi	75
Tabel 4.23. Kontribusi dan Pemanfaatan Strategi.....	80
Tabel 4.24. Rekomendasi Strategi	81
Tabel 4.25. Contoh Implementasi Nyata	81

Halaman ini sengaja dikosongkan

DAFTAR LAMPIRAN

Lampiran 1. Data Penelitian.....	91
Lampiran 2. Kode <i>Script</i>	92
Lampiran 3. <i>Script</i> Pembuatan <i>Interface</i>	93
Lampiran 4. LoA Artikel.....	94

Halaman ini sengaja dikosongkan

DAFTAR NOTASI

k	: Jumlah klaster
c_i	: Himpunan data diklaster i
x	: Vektor data
μ_i	: <i>Centroid</i> klaster
$C_{m(q)}$	: <i>Centroid</i> dari klaster m pada perhitungan ke q
n_m	: Banyaknya data pada klaster m
$x_{i(q)}$	: Data ke i pada iterasi ke q
BIC	: Metrik statistik <i>Bayesian Information Criterion</i>
L	: Faktor bobot
$\frac{p}{2}$	: Jumlah klaster
$\log N$	: Jumlah data titik
$\frac{n}{2}$	: Jumlah data
d	: Dimensi data
σ^2	: Variansi dalam klaster
μ	: Mean (rata-rata) dari distribusi data (pusat)
x_i	: Data ke i
$\ x_i - \mu\ ^2$	: Jarak kuadrat (<i>squared Euclidean distance</i>) antara data ke i dengan mean (rata-rata)
$\sum_{i=1}^n \ x_i - \mu\ ^2$	: Total jarak kuadrat semua data terhadap mean (rata-rata)
$\frac{n}{2} \log(2\pi)$	: Konstanta normalisasi Gaussian
SSE	: Total error kuadrat dari semua titik ke <i>centroid</i>
n	: Jumlah titik dalam klaster
x'	: <i>Scaling robustscaller</i>
x	: Nilai asli data
IQR	: Rentang antar kuartil
$Q3$	: Kuartil 3, nilai persentil ke-75 (75% data terbawah)
$Q1$	: Kuartil 1, nilai persentil ke-25 (25% data terbawah)

$s(i)$	: <i>Silhouette score</i> untuk data ke i
$b(i)$	: Rata-rata jarak antara data i dengan klaster terdekat lainnya (klaster tetangga terdekat)
$a(i)$	: Rata-rata jarak antara data i dengan semua data lain dalam klaster yang sama
$\max(a(i), b(i))$	: Normalisasi supaya nilai $s(i)$ selalu berada antara -1 sampai +1
DBI	: Metrik statistik <i>Davies Bouldin Index</i>
$\frac{1}{k}$	: Jumlah klaster
s_i	: Rata-rata jarak antara setiap titik data dalam klaster i terhadap pusat klaster
s_j	: Rata-rata jarak antara setiap titik data dalam klaster j terhadap pusat klaster
M_{ij}	: Jarak antara pusat klaster
$\frac{s_i + s_j}{M_{ij}}$	: Perhitungan rasio kemiripan antara klaster i dan j
CH	: Calinski Harabasz
$Tr(B_k)$	: <i>Trace</i> atau jumlah diagonal dari matriks disperse antar klaster
$Tr(W_k)$	: <i>Trace</i> dari matriks disperse dalam klaster
n	: Jumlah total data
k	: Jumlah klaster