

**IMPLEMENTASI *POWER TRANSFORMER* DAN *PRINCIPAL COMPONENT ANALYSIS* UNTUK OPTIMALISASI *CLUSTERING* PADA
SEGMENTASI PELANGGAN USIA DEWASA MENGGUNAKAN
ALGORITMA K-MEANS**

SKRIPSI

Diajukan Untuk Memenuhi Sebagian Syarat Guna
Memperoleh Gelar Sarjana Komputer (S.Kom.)
Pada Program Studi Sistem Informasi



OLEH :

**CINDY AVITASELLY BAMBANG SAPUTRI
2213030057**

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK DAN ILMU KOMPUTER
UNIVERSITAS NUSANTARA PERSATUAN GURU REPUBLIK
INDONESIA
UN PGRI KEDIRI
2026**

Skripsi oleh :

CINDY AVITASELLY BAMBANG SAPUTRI

NPM : 2213030057

Judul

**IMPLEMENTASI *POWER TRANSFORMER* DAN *PRINCIPAL COMPONENT ANALYSIS* UNTUK OPTIMALISASI *CLUSTERING* PADA
SEGMENTASI PELANGGAN USIA DEWASA MENGGUNAKAN
ALGORITMA K-MEANS**

Telah disetujui untuk diajukan Kepada
Panitia Ujian/Sidang Skripsi Program Studi Sistem Informasi
FTIK UN PGRI Kediri

Tanggal: Kamis, 10 Juni 2026

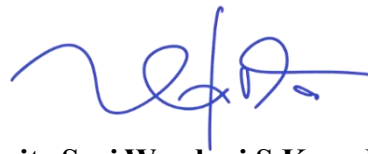
Pembimbing I



Arie Nugroho, S.Kom.MM., M.Kom

NIDN. 0712108103

Pembimbing II



Anita Sari Wardani, S.Kom., M.Kom

NIDN. 0713018402

Skripsi oleh :

CINDY AVITASELLY BAMBANG SAPUTRI

NPM : 2213030057

Judul

**IMPLEMENTASI POWER TRANSFORMER DAN PRINCIPAL
COMPONENT ANALYSIS UNTUK OPTIMALISASI CLUSTERING PADA
SEGMENTASI PELANGGAN USIA DEWASA MENGGUNAKAN
ALGORITMA K-MEANS**

Telah dipertahankan di depan Panitia Ujian/Sidang Skripsi

Program Studi Sistem Informasi FTIK UN PGRI Kediri

Tanggal : 25 Juni 2026

Dan Dinyatakan telah Memenuhi Persyaratan

Panitia Penguji :

1. Ketua : Arie Nugroho, S.Kom., M.M., M.Kom [.....]
2. Penguji 1 : Dwi Harini, S.Si., M.M [.....]
3. Penguji 2 : Anita Sari Wardani, S.Kom., M.Kom [.....]



Dr. Sulistiono, M.si
NIP.196807071993031004

PERNYATAAN

Yang bertanda tangan di bawah ini saya,

Nama : Cindy Avitaselly Bambang Saputri
Jenis Kelamin : Perempuan
Tempat/tgl. lahir : Kediri / 02 Agustus 2003
NPM : 2213030057
Fak/Prodi. : FTIK/ S1-Sistem Informasi

menyatakan dengan sebenarnya, bahwa dalam Skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu perguruan tinggi, dan sepanjang pengetahuan saya tidak terdapat karya tulis atau pendapat yang pernah diterbitkan oleh orang lain, kecuali yang secara sengaja dan tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Kediri, 10 Juni 2026

Yang Menyatakan



CINDY AVITASELLY B.S.

NPM: 2213030057

MOTTO

"It does not matter how slowly you go as long as you do not stop."

- Confucius –

"Tidak ada hasil yang mengkhianati proses."

- Pepatah Indonesia –

"Maka sesungguhnya bersama kesulitan ada kemudahan."

- (QS. Al-Insyirah: 6) -

ABSTRAK

Cindy Avitaselly B.S.: Implementasi *Power Transformer* Dan *Principal Component Analysis* Untuk Optimalisasi *Clustering* Pada Segmentasi Pelanggan Usia Dewasa Menggunakan Algoritma K-Means, Skripsi, Sistem Informasi, Fakultas Teknik dan Ilmu Komputer UN PGRI Kediri, 2026.

Kata Kunci: K-Means, Segmentasi Pelanggan, *PowerTransformer*, *Principal Component Analysis* (PCA), CRISP-DM.

Segmentasi pelanggan merupakan salah satu pendekatan yang digunakan untuk mengidentifikasi kelompok pelanggan berdasarkan karakteristik tertentu sehingga dapat mendukung penyusunan strategi bisnis yang lebih efektif. Salah satu metode yang banyak digunakan dalam proses tersebut adalah algoritma K-Means. Namun, performa algoritma ini sangat bergantung pada kondisi data yang digunakan, terutama terkait perbedaan skala atribut, distribusi data yang tidak seimbang, dan keberadaan *noise* yang dapat memengaruhi pembentukan *Cluster*.

Penelitian ini bertujuan untuk mengkaji pengaruh penerapan teknik *Preprocessing* dan reduksi dimensi terhadap kualitas pengelompokan data pelanggan usia dewasa menggunakan algoritma improved K-Means. Penelitian dilaksanakan dengan pendekatan CRISP-DM yang mencakup tahapan pemahaman bisnis, pemahaman data, persiapan data, pemodelan, evaluasi, dan penerapan. Data yang digunakan berasal dari Mall Customer Segmentation *Dataset* dengan atribut Age, Annual Income, dan Spending Score. Beberapa skenario eksperimen diterapkan, meliputi data tanpa *Preprocessing*, penggunaan *MinMaxScaler*, *RobustScaler*, *PowerTransformer*, PCA, serta kombinasi *Preprocessing* dan PCA. Kinerja *model* dievaluasi menggunakan Silhouette Score dan Davies-Bouldin Index (DBI).

Hasil pengujian menunjukkan bahwa perlakuan terhadap data sebelum proses *Clustering* berpengaruh terhadap kualitas *Cluster* yang dihasilkan. Dibandingkan skenario tanpa *Preprocessing*, metode yang melibatkan transformasi data dan reduksi dimensi mampu menghasilkan pengelompokan yang lebih baik. Kombinasi *PowerTransformer* dan PCA pada jumlah *Cluster* lima memberikan hasil paling optimal dengan nilai Silhouette Score sebesar 0,5558 dan Davies-Bouldin Index sebesar 0,5618. Hasil tersebut menunjukkan tingkat kekompakan dan pemisahan *Cluster* yang lebih baik dibandingkan metode lainnya.

Penelitian ini membuktikan bahwa peningkatan kualitas data melalui transformasi distribusi dan penyederhanaan dimensi dapat membantu meningkatkan performa algoritma improved K-Means dalam proses segmentasi pelanggan. Temuan ini diharapkan dapat menjadi referensi dalam pengembangan penelitian *Clustering* serta penerapan analisis pelanggan berbasis data pada berbagai bidang.

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Allah SWT atas segala rahmat, karunia, serta ridho-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul “Implementasi *Power Transformer* dan *Principal Component Analysis* Untuk Optimalisasi *Clustering* Pada Segmentasi Pelanggan Usia Dewasa Menggunakan Algoritma K-Means” dengan baik. Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer pada Program Studi Sistem Informasi, Fakultas Teknik dan Ilmu Komputer, Universitas Nusantara PGRI Kediri.

Proses penyusunan skripsi ini tidak terlepas dari bantuan, dukungan, bimbingan, doa, serta motivasi dari berbagai pihak. Maka dari itu, pada kesempatan ini penulis ingin menyampaikan ucapan terima kasih dan penghargaan yang setulus-tulusnya kepada:

1. Dr. Zainal Afandi, M.Pd., selaku Rektor Universitas Nusantara PGRI Kediri yang telah memberikan kesempatan dan fasilitas kepada penulis untuk menempuh pendidikan hingga tahap penyelesaian skripsi.
2. Dr. Sulistiono, M.Si., selaku Dekan Fakultas Teknik dan Ilmu Komputer Universitas Nusantara PGRI Kediri yang telah memberikan dukungan dan arahan selama proses pendidikan berlangsung.
3. Sucipto, M.Kom., selaku Ketua Program Studi Sistem Informasi Universitas Nusantara PGRI Kediri yang telah memberikan motivasi, arahan, dan dukungan kepada mahasiswa selama menjalani proses perkuliahan.
4. Arie Nugroho, S.Kom., M.M., M.Kom., selaku dosen pembimbing pertama yang telah memberikan bimbingan, arahan, masukan, serta ilmu yang sangat berharga kepada penulis dengan penuh kesabaran selama proses penyusunan skripsi ini.
5. Anita Sari Wardani, S.Kom., M.Kom., selaku dosen pembimbing kedua yang telah memberikan bimbingan, saran, dan motivasi sehingga penulis dapat menyelesaikan penelitian ini dengan lebih baik.
6. Mama tercinta dan Almarhum Papa, yang selalu menjadi sumber kekuatan terbesar dalam hidup penulis. Terima kasih kepada Mama atas kasih sayang,

doa, dukungan, pengorbanan, dan kepercayaan yang selalu diberikan kepada penulis. Untuk Almarhum Papa, segala nilai kehidupan, perjuangan, dan kasih sayang yang telah diberikan akan selalu menjadi inspirasi dalam setiap langkah penulis. Skripsi ini penulis persembahkan sebagai bentuk rasa syukur, cinta, dan penghormatan kepada kedua orang tua atas segala pengorbanan yang telah diberikan.

7. Fariez Firmansyah Putra, yang telah menjadi bagian penting dalam perjalanan penyusunan skripsi ini. Terima kasih atas kesabaran, perhatian, dukungan, dan semangat yang selalu diberikan kepada penulis, terutama pada saat-saat ketika penulis merasa lelah, kehilangan motivasi, dan meragukan kemampuan diri sendiri. Terima kasih karena selalu hadir untuk mendengarkan setiap keluh kesah, memberikan keyakinan ketika penulis merasa tidak mampu, serta tetap mendampingi dalam berbagai proses yang tidak selalu mudah. Dukungan yang diberikan tidak hanya membantu penulis menyelesaikan skripsi ini, tetapi juga membantu penulis untuk tetap percaya pada diri sendiri dan terus melangkah hingga mencapai tahap ini.

Penulis menyadari bahwa skripsi ini masih memiliki keterbatasan dan jauh dari kata sempurna. Maka dari itu, kritik dan saran yang membangun sangat diharapkan demi penyempurnaan penelitian di masa yang akan datang. Penulis berharap skripsi ini dapat memberikan manfaat bagi pengembangan ilmu pengetahuan, khususnya di bidang *data mining* dan *machine learning*, serta dapat menjadi referensi bagi penelitian selanjutnya.

Kediri, 10 Juni 2026



CINDY AVITASELLY B.S.

NPM: 2213030057

DAFTAR ISI

Halaman Sampul.....	i
Halaman Persetujuan	ii
Halaman Pengesahan.....	iii
Halaman Pernyataan	iv
MOTTO	v
ABSTRAK	vi
KATA PENGANTAR.....	vii
DAFTAR ISI.....	ix
DAFTAR GAMBAR.....	xii
DAFTAR TABEL	xiii
DAFTAR LAMPIRAN	xiv
BAB I PENDAHULUAN.....	1
A. Latar Belakang.....	1
B. Batasan Masalah	3
C. Rumusan Masalah	4
D. Tujuan Penelitian.....	4
E. Manfaat Penelitian	5
BAB II KAJIAN PUSTAKA.....	6
A. Kajian Teori	6
1. <i>Data mining</i>	6
2. <i>Clustering</i>	6
3. Algoritma K-Means	7
4. <i>Preprocessing Data</i>	7
5. <i>Principal Component Analysis (PCA)</i>	11
6. <i>Evaluasi Clustering</i>	11
B. Kajian Hasil Penelitian Terdahulu.....	12

C.	Kerangka Berpikir	17
BAB III METODE PENELITIAN		21
A.	<i>Business understanding</i> (Pemahaman Bisnis).....	21
1.	Latar Belakang Masalah Bisnis	21
2.	Tujuan Bisnis.....	22
3.	Kriteria Keberhasilan.....	22
B.	<i>Data understanding</i> (Pemahaman Data)	22
1.	Pengumpulan Data	22
2.	Deskripsi Data	23
3.	Eksplorasi Data Awal	23
4.	Verifikasi Kualitas Data	23
C.	<i>Data preparation</i> (Persiapan Data)	24
1.	Seleksi Data.....	24
2.	Pembersihan Data (<i>Cleaning</i>)	24
3.	Transformasi Data	24
4.	Pembuatan <i>Dataset Modeling</i>	25
D.	<i>Modeling</i>	25
1.	Pemilihan Teknik <i>Model</i>	25
2.	Desain dan Pengujian <i>Model</i>	25
3.	Evaluasi Parameter <i>Model</i>	27
4.	Interpretasi Hasil <i>Modeling</i>	27
E.	<i>Evaluation</i> (Evaluasi)	28
1.	Evaluasi <i>Model</i> Terbaik.....	28
2.	Validasi terhadap Tujuan Bisnis	28
3.	Tinjauan Ulang Proses	28
F.	<i>Deployment</i> (Penerapan).....	29
1.	Rencana Penerapan	29
2.	Implementasi <i>Model</i>	30

3.	Dokumentasi dan Pelaporan	30
BAB IV HASIL DAN PEMBAHASAN		32
A.	Data Hasil Penelitian	32
1.	Implementasi Penelitian	32
2.	Eksplorasi dan <i>Preprocessing</i> Data	34
3.	<i>Hyperparameter tuning</i>	40
4.	Hasil Eksperimen <i>Clustering</i>	41
5.	Perbandingan Hasil Seluruh Metode	49
6.	Analisis Karakteristik <i>Cluster</i> Terbaik	50
B.	Pembahasan	55
1.	Analisis Metode Optimal	55
2.	Evaluasi terhadap Tujuan Penelitian	56
3.	Validitas Hasil Penelitian	57
4.	Rekomendasi Penerapan <i>Model</i>	59
BAB V PENUTUP		61
A.	Kesimpulan	61
B.	Saran	61
DAFTAR PUSTAKA		63
LAMPIRAN		71

DAFTAR GAMBAR

Gambar 2. 1 Kerangka Berpikir.....	17
Gambar 3. 1 Tahapan Metodologi CRISP-DM	21
Gambar 4. 1 <i>Import Library</i>	33
Gambar 4. 2 <i>Load dataset</i>	33
Gambar 4. 3 Seleksi Fitur	35
Gambar 4. 4 <i>Cleaning Data</i>	35
Gambar 4. 5 <i>Outlier removal</i>	36
Gambar 4. 6 Age Binning	37
Gambar 4. 7 <i>Scatter plot</i> Awal	38
Gambar 4. 8 <i>Histogram PowerTransformer</i>	39
Gambar 4. 9 <i>Histogram MinMaxScaler</i>	39
Gambar 4. 10 Elbow Method.....	40
Gambar 4. 11 Tanpa <i>Preprocessing</i>	41
Gambar 4. 12 <i>MinMax Scaler</i>	42
Gambar 4. 13 <i>PowerTransformer</i>	43
Gambar 4. 14 <i>Principal Component Analysis (PCA)</i>	44
Gambar 4. 15 <i>RobustScaler</i>	45
Gambar 4. 16 <i>MinMax Scaler + Principal Component Analysis (PCA)</i>	46
Gambar 4. 17 <i>PowerTransformer + Principal Component Analysis (PCA)</i> .	47
Gambar 4. 18 <i>RobustScaler + Principal Component Analysis (PCA)</i>	48
Gambar 4. 19 Analisis Karakteristik <i>Cluster</i>	50
Gambar 4. 20 <i>Radar chart</i>	53

DAFTAR TABEL

Tabel 2.1. Penelitian Terdahulu.....	15
Tabel 3.1. Deskripsi Data	23
Tabel 3.2. Skenario Eksperimen.....	26
Tabel 3.3. Tools	30
Tabel 4.1. Perbandingan Hasil.....	49

DAFTAR LAMPIRAN

Lampiran 1 Kartu Bimbingan Skripsi	71
Lampiran 2 Surat Keterangan Bebas Similarity	72
Lampiran 3 Bukti Halaman Awal Cek Similarity	73
Lampiran 4 Bukti Submit Artikel/Loa/Terbit.....	74
Lampiran 5 Lembar Berita Acara Ujian	75
Lampiran 6 Lembar Revisi Ujian	76

BAB I

PENDAHULUAN

A. Latar Belakang

Perkembangan teknologi informasi dan meningkatnya volume data dalam berbagai sektor telah mendorong pemanfaatan teknik *data mining* untuk mengekstraksi informasi yang bernilai dalam pengambilan keputusan. Salah satu teknik yang banyak digunakan adalah *Clustering*, yaitu metode pengelompokan data berdasarkan tingkat kemiripan tanpa memerlukan label sebelumnya. *Clustering* telah banyak diterapkan pada berbagai bidang seperti segmentasi pelanggan, analisis perilaku konsumen, pendidikan, serta keuangan karena mampu mengidentifikasi pola tersembunyi dalam data yang tidak dapat diamati secara langsung. *Preprocessing* terhadap kualitas *cluster* Penelitian ini menggunakan algoritma K-Means sebagai metode utama untuk melakukan segmentasi pelanggan usia dewasa berdasarkan atribut Age, Annual Income, dan Spending Score. Hasil **Clustering** selanjutnya digunakan sebagai dasar untuk mengevaluasi pengaruh penera(Djun, Gunadi, & Sariyasa, 2024; Febrianty, Awalina, & Rahayu, 2023; Hendrastuty, 2024; Irawan, Wijaya, & Warisaji, 2025; Marcelina, Kurnia, & Terttiaavini, 2023).

Dalam konteks bisnis, segmentasi pelanggan menjadi salah satu penerapan *Clustering* yang penting, khususnya dalam memahami karakteristik dan perilaku pelanggan. Segmentasi yang tepat dapat membantu perusahaan dalam menentukan strategi pemasaran yang lebih efektif dan tepat sasaran (Kristanti, Junaidi, & Mandyartha, 2024; Surapati & Jannah, 2024). *Dataset* yang digunakan dalam penelitian ini merupakan data pelanggan yang memiliki atribut numerik seperti usia (*Age*), pendapatan tahunan (*Annual Income*), dan skor pengeluaran (*Spending Score*) yang merepresentasikan perilaku konsumsi pelanggan. Variasi nilai pada atribut tersebut menunjukkan adanya perbedaan karakteristik pelanggan yang berpotensi mempengaruhi hasil *Clustering*, terutama dalam hal skala data dan distribusi nilai yang tidak merata (Reddy et al., 2020). *Dataset* Mall Customer Segmentation dipilih karena memiliki karakteristik

data pelanggan yang sesuai untuk proses *Clustering*, khususnya pada atribut numerik yang merepresentasikan perilaku pelanggan seperti usia, pendapatan tahunan, dan skor pengeluaran. *Dataset* ini juga banyak digunakan dalam penelitian *Clustering* sehingga dapat mendukung proses evaluasi metode *Preprocessing* dan reduksi dimensi secara lebih terukur. Fokus penelitian pada pelanggan usia dewasa dilakukan karena kelompok usia tersebut cenderung memiliki pola konsumsi dan perilaku belanja yang lebih stabil dibandingkan kelompok usia remaja, sehingga lebih representatif dalam proses segmentasi pelanggan berbasis *Clustering*. Salah satu algoritma *Clustering* yang paling banyak digunakan adalah K-Means karena memiliki proses komputasi yang sederhana, efisien, dan mampu mengelompokkan data berdasarkan kedekatan terhadap *centroid* (Alhapizi, Nasir, & Effendy, 2020; Kristanti et al., 2024; Marcelina et al., 2023; Surapati & Jannah, 2024). Namun, performa algoritma ini sangat dipengaruhi oleh karakteristik data yang digunakan. Perbedaan skala antar atribut, distribusi data yang tidak normal, serta tingginya dimensi data dapat menyebabkan proses pembentukan *cluster* menjadi kurang optimal (Ghazal et al., 2021; Reddy et al., 2020; Wardani, Ariyanto, Umroh, & Rolliawati, 2023). Maka dari itu, diperlukan tahapan *preprocessing* yang sesuai agar kualitas *Clustering* dapat ditingkatkan. (Sufairoh, Rani, Amalia, & Rolliawati, 2023; Wahyuningtyas, Arafat, Stiawan, & Rolliawati, 2023; Wardani et al., 2023). Selain itu, penelitian lain juga menunjukkan bahwa kualitas *Clustering* sangat dipengaruhi oleh proses *Preprocessing* dan karakteristik distribusi data yang digunakan (Kristanti et al., 2024; Reddy et al., 2020).

Penelitian terdahulu menunjukkan bahwa *preprocessing* dan reduksi dimensi berperan penting dalam meningkatkan kualitas *Clustering*. *Preprocessing* digunakan untuk mengatasi perbedaan skala maupun distribusi data, sedangkan reduksi dimensi bertujuan menyederhanakan representasi data tanpa menghilangkan informasi penting (Ghazal et al., 2021; Hilmy Naufan, Kurniawan, & Suprpti, 2025; Reddy et al., 2020). Namun, sebagian besar penelitian masih menguji metode tersebut secara

terpisah sehingga pengaruh kombinasi berbagai teknik preprocessing dan *Principal Component Analysis* (PCA) terhadap performa algoritma K-Means belum dievaluasi secara komprehensif.

Berdasarkan kondisi tersebut, penelitian ini membandingkan beberapa skenario pengolahan data, yaitu tanpa *preprocessing*, *preprocessing* tunggal, serta kombinasi *preprocessing* dengan *Principal Component Analysis* (PCA) menggunakan algoritma K-Means pada segmentasi pelanggan usia dewasa (Hilmy Naufan et al., 2025). Kualitas *cluster* dievaluasi menggunakan *Silhouette Score* dan *Davies-Bouldin Index* (DBI) karena kedua metrik tersebut mampu mengukur tingkat kekompakan dan pemisahan *cluster* secara kuantitatif (Kristanti et al., 2024; Reddy et al., 2020). Melalui pendekatan tersebut, penelitian ini diharapkan dapat mengidentifikasi metode yang paling efektif dalam meningkatkan kualitas *Clustering* sekaligus memberikan referensi bagi penelitian selanjutnya.

B. Batasan Masalah

1. Penelitian ini menggunakan algoritma K-Means sebagai metode *Clustering* utama.
2. *Dataset* yang digunakan merupakan data pelanggan dengan fokus pada atribut numerik, yaitu Age, Annual Income (k\$), dan Spending Score (1–100).
3. Objek penelitian difokuskan pada segmentasi pelanggan usia dewasa berdasarkan data yang tersedia.
4. Teknik *Preprocessing* yang digunakan meliputi *MinMax Scaler*, *PowerTransformer*, dan *RobustScaler*.
5. Teknik reduksi dimensi yang digunakan adalah *Principal Component Analysis* (PCA).
6. Skenario eksperimen meliputi tanpa *Preprocessing*, *Preprocessing* tunggal, serta kombinasi *Preprocessing* dan PCA.
7. Penentuan jumlah *Cluster* (k) dilakukan melalui proses evaluasi (*hyperparameter tuning*).

8. Pengujian jumlah *Cluster* dilakukan pada rentang nilai $k = 2$ hingga $k = 9$ sesuai kebutuhan eksperimen *Clustering*.
9. Evaluasi kualitas *Clustering* dilakukan menggunakan Silhouette Score dan Davies Bouldin Index (DBI).
10. Penelitian ini tidak melakukan perbandingan langsung dengan metode *Clustering* lain karena fokus pada optimalisasi K-Means.
11. Proses eksperimen dilakukan menggunakan *Google Colaboratory* sebagai lingkungan komputasi.
12. Penelitian ini tidak membahas implementasi hasil *Clustering* dalam sistem bisnis secara langsung, melainkan berfokus pada evaluasi kualitas *Clustering* secara kuantitatif.

C. Rumusan Masalah

1. Bagaimana perbandingan kualitas *Clustering* pada algoritma K-Means antara tanpa *Preprocessing*, *Preprocessing* tunggal, dan kombinasi *Preprocessing* dengan reduksi dimensi yang diukur menggunakan nilai Silhouette Score dan Davies Bouldin Index (DBI)?
2. Bagaimana pengaruh penerapan *PowerTransformer* dan *Principal Component Analysis* (PCA) terhadap peningkatan nilai Silhouette Score dan penurunan nilai Davies Bouldin Index (DBI) pada segmentasi pelanggan usia dewasa?
3. Metode manakah yang menghasilkan kualitas *Clustering* paling optimal berdasarkan nilai Silhouette Score tertinggi dan nilai Davies Bouldin Index (DBI) terendah?

D. Tujuan Penelitian

1. Menganalisis perbandingan kualitas *Clustering* pada algoritma K-Means antara tanpa *Preprocessing*, *Preprocessing* tunggal, dan kombinasi *Preprocessing* dengan reduksi dimensi berdasarkan nilai Silhouette Score dan Davies Bouldin Index (DBI).

2. Mengevaluasi pengaruh penerapan *PowerTransformer* dan *Principal Component Analysis* (PCA) dalam meningkatkan nilai Silhouette Score dan menurunkan nilai Davies Bouldin Index (DBI) pada segmentasi pelanggan usia dewasa.
3. Menentukan metode yang paling optimal dalam menghasilkan kualitas *Clustering* terbaik berdasarkan nilai Silhouette Score tertinggi dan nilai Davies Bouldin Index (DBI) terendah.

E. Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan kontribusi secara teoritis dan praktis. Secara teoritis, penelitian ini dapat memperkaya kajian dalam bidang *data mining*, khususnya terkait optimalisasi algoritma K-Means melalui penerapan teknik *Preprocessing* dan reduksi dimensi. Secara praktis, hasil penelitian ini dapat dimanfaatkan sebagai dasar dalam melakukan segmentasi pelanggan secara lebih efektif, sehingga dapat membantu dalam memahami pola perilaku pelanggan dan mendukung pengambilan keputusan berbasis data, khususnya dalam penentuan strategi pemasaran yang lebih tepat sasaran. Selain itu, penelitian ini diharapkan dapat menjadi referensi dalam penerapan *Preprocessing* dan reduksi dimensi pada penelitian *Clustering* lainnya, khususnya yang menggunakan algoritma K-Means pada data pelanggan.

DAFTAR PUSTAKA

- Aditya, A., Jovian, I., & Sari, B. N. (2020). Implementasi K-Means Clustering Ujian Nasional Sekolah Menengah Pertama di Indonesia Tahun 2018/2019. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 4(1), 51. <https://doi.org/10.30865/mib.v4i1.1784>
- Alhapizi, R., Nasir, M., & Effendy, I. (2020). Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Mahasiswa Baru Universitas Bina Darma Palembang. In *Journal of Software Engineering Ampera* (Vol. 1). <https://doi.org/10.51519/journalsea.v1i1.10>
- Arfan, U., & Paraga, N. (2024). Perbandingan Algoritma K-Means, Naïve Bayes dan Decision Tree Dalam Memprediksi Penjualan Bahan Bakar Minyak. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 4(4), 1379–1389. <https://doi.org/10.57152/malcom.v4i4.1566>
- Azzaria, C., Daniati, E., Ristyawan, A., Nusantara, U., & Kediri, P. (2025). Peningkatan Akurasi Deteksi Liver Disease melalui Hyperparameter Tuning pada Algoritma Random Forest (Vol. 4). <https://doi.org/10.59095/ijcsr.v4i2.198>
- Cahyo Pradikdo, A., & Ristyawan, A. (n.d.). Model klasifikasi abstrak skripsi menggunakan text mining untuk pengkategorian skripsi sesuai bidang kajian. <https://doi.org/10.24176/simet.v9i2.2513>
- Daniati, E., & Nugroho, A. (2016). *K-Means clustering with Decision Support System using SAW: Determining thesis topic*. IEEE. <https://doi.org/10.1109/ICCSCE.2016.7893593>
- Dini Oktaviani, Syarifah Putri Agustini Alkadri, & Sucipto Sucipto. (2025). Klasifikasi Tingkat Kepuasan Pelayanan Pembuatan Paspor Menggunakan Algoritma Naïve Bayes. *JURAL Riset RUMPUN ILMU TEKNIK*, 4(1), 129–144. <https://doi.org/10.55606/jurritek.v4i1.4572>

- Djun, S. F., Gunadi, I. G. A., & Sariyasa, S. (2024). Analisis Segmentasi Pelanggan pada Bisnis dengan Menggunakan Metode K-Means Clustering pada Model Data RFM. *JTIM : Jurnal Teknologi Informasi Dan Multimedia*, 5(4), 354–364. <https://doi.org/10.35746/jtim.v5i4.434>
- Febrianty, E., Awalina, L., & Rahayu, W. I. (2023). Optimalisasi Strategi Pemasaran dengan Segmentasi Pelanggan Menggunakan Penerapan K-Means Clustering pada Transaksi Online Retail Optimizing Marketing Strategies with Customer Segmentation Using K-Means Clustering on Online Retail Transactions. *Jurnal Teknologi Dan Informasi (JATI)*, 13. <https://doi.org/10.34010/jati.v13i2>
- Ghazal, T. M., Hussain, M. Z., Said, R. A., Nadeem, A., Hasan, M. K., Ahmad, M., ... Naseem, M. T. (2021). Performances of k-means clustering algorithm with different distance metrics. *Intelligent Automation and Soft Computing*, 30(2), 735–742. <https://doi.org/10.32604/iasc.2021.019067>
- Guevara-Viejó, F., Valenzuela-Cobos, J. D., Vicente-Galindo, P., & Galindo-Villardón, P. (2021). Application of K-means clustering algorithm to commercial parameters of pleurotus spp. Cultivated on representative agricultural wastes from province of Guayas. *Journal of Fungi*, 7(7). <https://doi.org/10.3390/jof7070537>
- Hendrastuty, N. (2024). Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Dalam Evaluasi Hasil Pembelajaran Siswa. *Jurnal Ilmiah Informatika Dan Ilmu Komputer (JIMA-ILKOM)*, 3(1), 46–56. <https://doi.org/10.58602/jima-ilkom.v3i1.26>
- Herdika, E., Aulia, S., Daniati, E., Muzaki, M. N., & Informasi, S. (2025). Perbandingan Algoritma Machine Learning Dalam Analisis Sentimen Isu Gempa Megathrust (Vol. 4). <https://doi.org/10.59095/ijcsr.v4i2.206>

- Hidayat, T. (2022). Klasifikasi Data Jamaah Umroh Menggunakan Metode K-Means Clustering. *Jurnal Sistim Informasi Dan Teknologi*, 19–24. <https://doi.org/10.37034/jsisfotek.v4i1.115>
- Hilmy Naufan, M., Kurniawan, R., & Suprpti, T. (2025). OPTIMASI NILAI DAVIES BOULDIN INDEX PADA PROGRAM PENDAFTARAN TANAH SISTEMATIS LENGKAP (PTSL) MENGGUNAKAN ALGORITMA K-MEANS DAN PCA. *Jurnal Teknik Elektro Dan Informatika*, 20, 17–28. <https://doi.org/10.30587/e-link.v20i1.9063>
- Ika Murpratiwi, S., Gusti Agung Indrawan, I., & Aranta, A. (2021). ANALISIS PEMILIHAN CLUSTER OPTIMAL DALAM SEGMENTASI PELANGGAN TOKO RETAIL. *Jurnal Pendidikan Teknologi Dan Kejuruan*, 18(2). <https://doi.org/10.23887/jptk-undiksha.v18i2.37426>
- Irawan, D., Wijaya, G., & Warisaji, T. T. (2025). Penerapan Algoritma K-Means Clustering untuk Segmentasi Nasabah Bank. *BIOS: Jurnal Teknologi Informasi Dan Rekayasa Komputer*, 6(1), 47–53. <https://doi.org/10.37148/bios.v6i1.162>
- Jevintya, N. R., Darussalam, U., & Abdullah, S. (2024). APPLICATION OF THE K-MEANS AND DECISION TREE ALGORITHMS IN DETERMINING STUDENT ACHIEVEMENT. *JIKO (Jurnal Informatika Dan Komputer)*, 7(1), 13–18. <https://doi.org/10.33387/jiko.v7i1.7580>
- Kholila, N., Mujiono, M., & Wahyudi, D. (2023). Pemetaan Kondisi Lingkungan Tanam menggunakan K-Means Clustering. *JSITIK: Jurnal Sistem Informasi Dan Teknologi Informasi Komputer*, 1(2), 137–147. <https://doi.org/10.53624/jsitik.v1i2.182>
- Kristanti, B. T., Junaidi, A., & Mandyartha, E. P. (2024). IMPLEMENTASI K-MEANS CLUSTERING DALAM SEGMENTASI PELANGGAN BERDASARKAN USIA, PENDAPATAN, DAN MODEL RFM

- (STUDI KASUS: LANTIKYA STORE JOMBANG). *Jurnal Informatika Dan Teknik Elektro Terapan*, 12(3).
<https://doi.org/10.23960/jitet.v12i3.4677>
- Kristanto, B., Zy, A. T., & Fatchan, M. (2023). Analisis Penentuan Karyawan Tetap Dengan Algoritma K-Means Dan Davies Bouldin Index. *Bulletin of Information Technology (BIT)*, 4(1), 112–120.
<https://doi.org/10.47065/bit.v3i1>
- Marcelina, D., Kurnia, A., & Terttiaavini, T. (2023). Analisis Kluster Kinerja Usaha Kecil dan Menengah Menggunakan Algoritma K-Means Clustering. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 3(2), 293–301.
<https://doi.org/10.57152/malcom.v3i2.952>
- Miraftabzadeh, S. M., Colombo, C. G., Longo, M., & Foiadelli, F. (2023). K-Means and Alternative Clustering Methods in Modern Power Systems. *IEEE Access*, Vol. 11, pp. 119596–119633. Institute of Electrical and Electronics Engineers Inc.
<https://doi.org/10.1109/ACCESS.2023.3327640>
- Muhammad Alinafiah, A., Ceasar Octariadi, B., Teknik Informatika, J., & Teknik dan Ilmu Komputer, F. (n.d.). *JIP (Jurnal Informatika Polinema) IMPLEMENTAS DATA MINING DALAM PENGELOLAAN STOK OBAT MENGGUNAKAN METODE K-MEANS CLUSTERING DAN ASSOCIATION RULES APRIORI*.
<https://doi.org/10.33795/jip.v10i4.5523>
- Mukhtar, H., Dwi Pramaditya, I., Saputra Weisdiyanto, W., Hardian Putra, S., Trimuawasih, D., Auralia Rilda, A., ... Muhammadiyah Riau, U. (2024). *ALGORITMA K-MEANS UNTUK PENGELOMPOKAN PERILAKU CUSTOMER* (Vol. 4).
<https://doi.org/10.37859/seis.v4i2.7615>
- Mulyadi, I. P. (2022). Klasterisasi Menggunakan Metode Algoritma K-Means dalam Meningkatkan Penjualan Tupperware. *Jurnal*

Informatika Ekonomi Bisnis, 172–179.
<https://doi.org/10.37034/infeb.v4i4.164>

Muningsih, E. (2022). KOMBINASI METODE K-MEANS DAN DECISION TREE DENGAN PERBANDINGAN KRITERIA DAN SPLIT DATA. In *Jurnal TEKNOINFO* (Vol. 16).
<https://doi.org/10.33365/jti.v16i1.1561>

Pasaribu, D. F., Damanik, I. S., Irawan, E., Suhada, & Tambunan, H. S. (2021). Memanfaatkan Algoritma K-Means Dalam Memetakan Potensi Hasil Produksi Kelapa Sawit PTPN IV Marihat. *BIOS : Jurnal Teknologi Informasi Dan Rekayasa Komputer*, 2(1), 11–20.
<https://doi.org/10.37148/bios.v2i1.17>

Putriani, D., Puspa, A., Prayogi, A., Shofyana, A. I., Ristyawan, A., & Daniati, E. (2024). Prediksi Customer Churn Menggunakan Algoritma Decision Tree. In *Agustus* (Vol. 8). Online.
<https://doi.org/10.29407/inotek.v8i1.4914>

Rahmawati, S., & Defit, S. (2024). Prediction of Graduation Accuracy Using the K-Means Clustering Algorithm and Classification Decision Tree. *Jurnal Penelitian Pendidikan IPA*, 10(4), 2007–2013.
<https://doi.org/10.29303/jppipa.v10i4.7073>

Reddy, G. T., Reddy, M. P. K., Lakshmana, K., Kaluri, R., Rajput, D. S., Srivastava, G., & Baker, T. (2020). Analysis of Dimensionality Reduction Techniques on Big Data. *IEEE Access*, 8, 54776–54788.
<https://doi.org/10.1109/ACCESS.2020.2980942>

Regia Raffin, A., Sari Wardani, A., Informasi, S., Teknik, F., & Nusantara PGRI Kediri, U. (2022). Sistem Informasi Penjualan Berbasis Android Pada Outlet Marboba. *JITEKH*, 10(1), 45–51.
<https://doi.org/10.35447/jitekh.v10i1.566>

Risky Agustin, S., Purnamasari, I., & Nurina Sari, B. (2025). Implementasi K-Means Untuk Pengelompokan Kategori Penjualan Barang Berbasis

- Web. *Journal of Informatics Management and Information Technology*, 5(3), 167–176. <https://doi.org/10.47065/jimat.v5i3.610>
- Rizki, A., Pratama, Y., Mirza, A. H., Nasir, M., & Udariansyah, D. (2023). Penerapan Metode Clustering Dalam Analisis Data Produk Laptop di Tokopedia untuk Membuat Rekomendasi Produk: Studi Perbandingan K-Means, Hierarchical, dan DBSCAN. *Media Online*, 99(99), 999–999. <https://doi.org/10.30865/klik.v3i5.665>
- Santoso, H. T., Felmidi, F. A., Fadhila, A. N., Ristyawan, A., & Daniati, E. (2024). Analisis Kinerja Algoritma Data Mining pada Klasifikasi Tingkat Obesitas dengan K-Fold Cross Validation dan AUC. In *Agustus* (Vol. 8). Online. <https://doi.org/10.29407/inotek.v8i1.4917>
- Saputra, A., & Yusuf, R. (2024). Perbandingan Algoritma DBSCAN dan K-MEANS dalam Segmentasi Pelanggan Pengguna Transportasi Publik Transjakarta Menggunakan Metode RFM. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 4(4), 1346–1361. <https://doi.org/10.57152/malcom.v4i4.1516>
- Saputra, D. A., & Andriyanto, T. (2022). Analisis Kualitas Website Sistem Informasi Akademik Universitas Nusantara PGRI Kediri Quality Analysis of Website Academic Information System Universitas Nusantara PGRI Kediri. *Research : Journal of Computer*, 5(1), 17–22. <https://doi.org/10.25273/research.v5i1.9350>
- Saputri, F. W., & Arianto, D. B. (n.d.). *PERBANDINGAN PERFORMA ALGORITMA K-MEANS, K-MEDOIDS, DAN DBSCAN DALAM PENGGEROMBOLAN PROVINSI DI INDONESIA BERDASARKAN INDIKATOR KESEJAHTERAAN MASYARAKAT*. <https://doi.org/10.47111/jti.v7i2.9558>
- Sucipto, S., Al-Mubarak, M. S., & Setiyawan, D. T. (2021). HIERARCHICAL CLUSTERING BAHAN MENU DI KANTIN UNIVERSITAS UNTUK MENUNJANG IMPLEMENTASI SISTEM

- JAMINAN HALAL. *Jurnal Teknologi Industri Pertanian*, 60–69.
<https://doi.org/10.24961/j.tek.ind.pert.2021.31.1.60>
- Sufairoh, I., Rani, A. C., Amalia, K., & Rolliawati, D. (2023). Perbandingan Hasil Analisis Clustering Metode K-Means, DBSCAN Dan Hierarchical Pada Data Marketplace Electronic Phone. *JOINS (Journal of Information System)*, 8(1), 97–105.
<https://doi.org/10.33633/joins.v8i1.8016>
- Sugianto, C. A., Rahayu, A. H., & Gusman, A. (2020). Algoritma K-Means Untuk Pengelompokan Penyakit Pasien Pada Puskesmas Cigugur Tengah. *Journal of Information Technology*.
<https://doi.org/10.47292/joint.v2i2.30>
- Surapati, U., & Jannah, M. (2024). Penerapan Data Mining Menggunakan Metode K-Means Untuk Mengetahui Minat Customer Dalam Pembelian Merchandise Kpop. *Jurnal Sains Dan Teknologi*, 5(3).
<https://doi.org/10.55338/saintek.v5i1.2739>
- Syahputra, H. (2022). Clustering Tingkat Penjualan Menu (Food and Beverage) Menggunakan Algoritma K-Means. *Jurnal KomtekInfo*, 29–33. <https://doi.org/10.35134/komtekinfo.v9i1.274>
- Taha, I. B. M., & Mansour, D. E. A. (2021). Novel power transformer fault diagnosis using optimized machine learning methods. *Intelligent Automation and Soft Computing*, 28(3), 739–752.
<https://doi.org/10.32604/iasc.2021.017703>
- Toto, T. A. (2025). PERBANDINGAN ALGORITMA K-MEANS DAN AGGLOMERATIVE HIERARCHICAL CLUSTERING UNTUK PENGELOMPOKAN DAERAH PENGHASIL PADI DI INDONESIA. *Jurnal Informatika Dan Teknik Elektro Terapan*, 13(3).
<https://doi.org/10.23960/jitet.v13i3.7251>
- Wahyuningtyas, F. D., Arafat, A., Stiawan, A., & Rolliawati, D. (2023). Komparasi Algoritma Hierarchical, K-Means, dan DBSCAN pada Analisis Data Penjualan Melalui Facebook. *Explore: Jurnal Sistem*

Informasi Dan Telematika, 14(1), 7.
<https://doi.org/10.36448/jsit.v14i1.2931>

Wardani, S. D. K., Ariyanto, A. S., Umroh, M., & Rolliawati, D. (2023). PERBANDINGAN HASIL METODE CLUSTERING K-MEANS, DB SCANNER & HIERARCHICAL UNTUK ANALISA SEGMENTASI PASAR. *JIKO (Jurnal Informatika Dan Komputer)*, 7(2), 191. <https://doi.org/10.26798/jiko.v7i2.796>

Widodo, E., & Hadikristanto, W. (2023). Pengelompokan Untuk Penjualan Obat Dengan Menggunakan Algoritma K-Means. *Bulletin of Information Technology (BIT)*, 4(2), 408–413. <https://doi.org/10.47065/bit.v3i1>

Yahya, A., & Kurniawan, R. (2025). Implementasi Algoritma K-Means untuk Pengelompokan Data Penjualan Berdasarkan Pola Penjualan. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 5(1), 350–358. <https://doi.org/10.57152/malcom.v5i1.1773>

Yudistira, A., & Andika, R. (2023). Pengelompokan Data Nilai Siswa Menggunakan Metode K-Means Clustering. *Journal of Artificial Intelligence and Technology Information (JAITI)*, 1(1), 20–28. <https://doi.org/10.58602/jaiti.v1i1.22>

Zahra, A. L., Tiara, S., Ada, R., & Ardini, A. F. (2024). Implementasi Clustering Algoritma K-Means Pada Produksi Beras di Provinsi Jawa Timur Tahun 2022. In *Journal of Computer and Information Systems Ampera* (Vol. 5). <https://doi.org/10.51519/journalcisa.v5i3.485>