SEGMENTASI PERILAKU PENGGUNA KARTU KREDIT DAN PREDIKSI DEFAULT MENGGUNAKAN KMEANS CLUSTERING

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika

disusun oleh

FAIZ DAFFA KURNIA

22.11.4627

Kepada

FAKULTAS ILMU KOMPUTER

UNIVERSITAS AMIKOM YOGYAKARTA

YOGYAKARTA

2026

SEGMENTASI PERILAKU PENGGUNA KARTU KREDIT DAN PREDIKSI DEFAULT MENGGUNAKAN KMEANS CLUSTERING

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika

disusun oleh

FAIZ DAFFA KURNIA

22.11.4627

Kepada

FAKULTAS ILMU KOMPUTER

UNIVERSITAS AMIKOM YOGYAKARTA

YOGYAKARTA

2026

HALAMAN PERSETUJUAN

SKRIPSI

SEGMENTASI PERILAKU PENGGUNA KARTU KREDIT DAN PREDIKSI DEFAULT MENGGUNAKAN KMEANS CLUSTERING

yang disusun dan diajukan oleh

Faiz Daffa Kurnia

22.11.4627

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 27 Februari 2026

Dosen Pembimbing,

Supriatin, A.Md., S.Kom., M.Kom.
NIK. 190302239

HALAMAN PENGESAHAN

SKRIPSI

SEGMENTASI PERILAKU PENGGUNA KARTU KREDIT DAN PREDIKSI DEFAULT MENGGUNAKAN KMEANS CLUSTERING

yang disusun dan diajukan oleh

Faiz Daffa Kurnia

22.11.4627

Telah dipertahankan di depan Dewan Penguji
pada tanggal 27 Februari 2026

Susunan Dewan Penguji

Nama Penguji	Tanda Tangan
Rizqi Sukma Kharisma, S.Kom, M.Kom **NIK. 190302215**	
Sharazita Dyah Anggita, S.Kom., M.Kom. **NIK. 190302285**	
Supriatin, A.Md., S.Kom., M.Kom. **NIK. 190302239**	

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 27 Februari 2026

DEKAN FAKULTAS ILMU KOMPUTER

Prof. Dr. Kusrini, M.Kom.
NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Faiz Daffa Kurnia
NIM : 22.11.4627

Menyatakan bahwa Skripsi dengan judul berikut:

Segmentasi Perilaku Pengguna Kartu Kredit dan Prediksi Default Menggunakan K-Means Clustering

Dosen Pembimbing : Supriatin, A.Md., S.Kom., M.Kom.

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 27 Februari 2026

Yang Menyatakan,

Faiz Daffa Kurnia

HALAMAN PERSEMBAHAN

Skripsi ini penulis persembahkan dengan sepenuh hati kepada Allah SWT atas segala rahmat, hidayah, dan karunia-Nya sehingga penulis diberikan kekuatan dan kemampuan untuk menyelesaikan skripsi ini.

Kemudian kepada kedua orang tua tercinta, yang tidak pernah berhenti mendoakan, menyayangi, dan berkorban demi keberhasilan penulis. Juga kepada Bapak/Ibu Dosen Universitas Amikom Yogyakarta, atas bimbingan, arahan, dan ilmu yang diberikan selama masa studi hingga menyelesaikan skipsi ini.Dan tak lupa untuk sahabat-sahabat terbaik yang selalu hadir di saat susah maupun senang, yang menemani begadang, berbagi tawa dan keluh kesah sepanjang perjalanan kuliah ini. Kalian adalah bagian dari cerita terbaik penulis.

KATA PENGANTAR

Puji syukur penulis panjatkan kehadirat Allah SWT yang telah melimpahkan rahmat dan hidayah-Nya, sehingga penulis dapat menyelesaikan skripsi ini dengan lancar dan tepat waktu. Skripsi ini disusun untuk memenuhi salah satu syarat mencapai derajat Sarjana Program Studi Informatika Fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta. Dalam penyusunan skripsi ini, penulis menyadari bahwa keberhasilan ini tidak terlepas dari bantuan, bimbingan, dan dukungan berbagai pihak. Oleh karena itu, penulis mengucapkan terima kasih yang sebesar-besarnya kepada:

1. Prof. Dr. Kusrini, M.Kom., selaku Dekan Fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta.
2. Ibu Supriatin, M.Kom., selaku Dosen Pembimbing Skripsi yang telah meluangkan waktu, tenaga, dan pikirannya untuk memberikan bimbingan serta arahan selama penyusunan skripsi ini.
3. Seluruh Dosen dan Staf Program Studi Informatika Universitas AMIKOM Yogyakarta atas ilmu, fasilitas, dan dukungan yang diberikan selama penulis menempuh studi.
4. Kedua orang tua tercinta yang selalu menjadi sumber kekuatan penulis. Terima kasih atas doa yang tidak pernah putus, kasih sayang yang tulus, dan segala pengorbanan yang telah diberikan demi keberhasilan penulis.
5. Teman-teman seperjuangan yang telah berbagi semangat, waktu, dan kenangan selama masa studi.
6. Semua pihak yang tidak dapat penulis sebutkan satu persatu, yang telah memberikan bantuan dan dukungan dalam penyelesaian skripsi ini.

Yogyakarta, 24 Februari 2026

Penulis

DAFTAR ISI

HALAMAN JUDUL i
HALAMAN PERSETUJUAN ii
HALAMAN PENGESAHAN iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI iv
HALAMAN PERSEMBAHAN v
KATA PENGANTAR vi
DAFTAR ISI vii
DAFTAR TABEL ix
DAFTAR GAMBAR x
DAFTAR LAMPIRAN xi
DAFTAR LAMBANG DAN SINGKATAN xii
DAFTAR ISTILAH xiii
INTISARI xv
ABSTRACT xvi
BAB I PENDAHULUAN 1
1.1 Latar Belakang 1
1.2 Rumusan Masalah 3
1.3 Batasan Masalah 3
1.4 Tujuan Penelitian 3
1.5 Manfaat Penelitian 4
1.6 Sistematika Penulisan 4

BAB II TINJAUAN PUSTAKA 5
2.1 Studi Literatur 5
2.2 Dasar Teori 9
BAB III METODE PENELITIAN 16
3.1 Objek Penelitian 16
3.2 Alur Penelitian 16
3.3 Alat dan Bahan 20
3.4 Rancangan antarmuka 24
BAB IV HASIL DAN PEMBAHASAN 26
4.1 Akuisisi Data 26
4.2 Preprocessing Data & EDA 27
4.3 Pengembangan Model dan Evaluasi 33
4.4 Deployment 40
BAB V PENUTUP 42
5.1 Kesimpulan 42
5.2 Saran 42
REFERENSI 44
LAMPIRAN 47

DAFTAR TABEL

Tabel 2.1 Tabel Keaslian Penelitian 7
Tabel 3.1 Variabel Dataset 20
Tabel 3.2 *Hardware* Penelitian 22
Tabel 3.3 *Software* Penelitian 23

DAFTAR GAMBAR

Gambar 2.2 Alur K-Means Clustering 13
Gambar 3.2 Diagram Alur Penelitian 17
Gambar 3.4 Rancangan Antarmuka Sistem 25
Gambar 4.1 Informasi Dataset 26
Gambar 4.2 Cek Missing Values 27
Gambar 4.3 Pelabelan fitur kategorikal 28
Gambar 4.4 Standarisasi fitur numerik 28
Gambar 4.5 Visualisasi Distribusi Usia Pelanggan 29
Gambar 4.6 Visualisasi Distribusi Annual Income 29
Gambar 4.7 Visualisasi Distribusi Total Spend 30
Gambar 4.8 Visualisasi Distribusi Credit Score 31
Gambar 4.9 Visualisasi Average Income by Age Group 31
Gambar 4.10 Visualisasi Average Spending by Education Level 32
Gambar 4.11 Visualisasi Tabel Correlation Matrix 33
Gambar 4.12 Seleksi Fitur Utama 34
Gambar 4.13 Kode Implementasi Persamaan WCSS dan Plot Elbow 35
Gambar 4.14 Visualisasi Grafik Elbow Plot 35
Gambar 4.15 Kode Silhouette Score Setelah Seleksi Fitur 36
Gambar 4.16 Kode Train Model 36
Gambar 4.17 Visualisasi Cluster K = 2 37
Gambar 4.18 Visualisasi Cluster K = 3 38
Gambar 4.19 Visualisasi Cluster K = 4 39
Gambar 4.20 Kode Pengembangan Tampilan Streamlit 40
Gambar 4.21 Tampilan Akhir Sistem 41

DAFTAR LAMPIRAN

Lampiran 1. Dataset Credit Card User Behavior and Default 47
Lampiran 2. Source Code Pengembangan Model 47
Lampiran 3. Source Code Pengembangan Aplikasi Streamlit 48
Lampiran 4. Hasil Akhir Tampilan Sistem 48

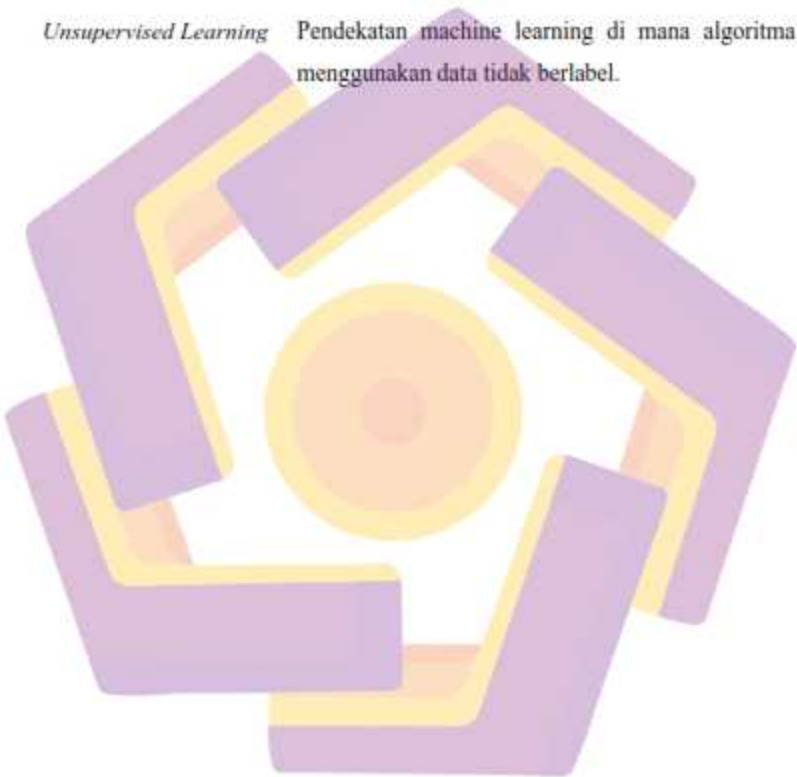
DAFTAR LAMBANG DAN SINGKATAN

J	*Objective function K-Means*
K	Jumlah cluster
n	Jumlah data points
$\|\cdot\|^2$	Jarak *Euclidean* kuadrat
WCSS(K)	*Within-Cluster Sum of Squares* untuk *K cluster*
C_k	Himpunan data points dalam cluster ke-k
AI	*Artificial Intelligence*
CLV	*Customer Lifetime Value*
CRISP-DM	*Cross-Industry Standard Process for Data Mining*
CSV	*Comma-Separated Values*
DBSCAN	*Density-Based Spatial Clustering of Applications with Noise*
EDA	*Exploratory Data Analysis*
GMM	*Gaussian Mixture Model*
PCA	*Principal Component Analysis*
SVM	*Support Vector Machines*
t-SNE	*t-Distributed Stochastic Neighbor Embedding*
WCSS	*Within-Cluster Sum of Squares*

DAFTAR ISTILAH

Centroid	Titik pusat yang merepresentasikan suatu cluster.
Clustering	Teknik mengelompokkan data berdasarkan karakteristik.
Credit Default	Kondisi di mana pemegang kartu kredit gagal memenuhi kewajiban pembayaran.
Data Mining	Proses penemuan pola, korelasi, dan pengetahuan yang bermanfaat dari kumpulan data.
Default Risk	Risiko pemegang kartu kredit tidak mampu memenuhi kewajiban keuangannya.
Deployment	Proses penerapan model ke dalam lingkungan produksi.
Feature Engineering	Proses pemilihan, transformasi, dan pembuatan variabel input yang relevan meningkatkan performa model.
Inertia	Jumlah kuadrat jarak dari setiap data point ke centroid cluster terdekatnya.
One-Hot Encoding	Teknik transformasi variabel kategorikal menjadi representasi biner numerik.
Outlier	Data point yang nilainya menyimpang jauh dari mayoritas data lainnya dalam dataset.
Preprocessing	Proses pembersihan, transformasi, dan persiapan data mentah sebelum digunakan dalam pemodelan machine learning.
Segmentasi	Proses pengelompokan data atau pelanggan ke dalam kelompok yang homogen berdasarkan karakteristik tertentu.

Silhouette Score	Metrik evaluasi kualitas clustering yang mengukur seberapa baik suatu data point cocok dengan clusternya sendiri
StandardScaler	Teknik normalisasi data yang mengubah fitur numerik sehingga memiliki rata-rata nol dan standar deviasi satu.
Unsupervised Learning	Pendekatan machine learning di mana algoritma dilatih menggunakan data tidak berlabel.



INTISARI

Penggunaan kartu kredit yang meningkat pesat oleh masyarakat membawa risiko *credit default* yang berdampak pada kerugian finansial. Salah satu penyebab tingginya resiko *credit default* antara lain adalah minimnya pemahaman dari pihak lembaga keuangan terhadap pola perilaku. Sebagai sarana pemahaman pola perilaku nasabah, penelitian ini bertujuan untuk mengembangkan model segmentasi perilaku pengguna kartu kredit dan menganalisis potensi *default* dengan menggunakan algoritma *K-Means Clustering*.

Data penelitian ini menggunakan dataset publik *Credit Card User Behavior and Default Dataset* dari Kaggle. Tahapan penelitian meliputi *preprocessing* data, seleksi fitur menggunakan analisis korelasi, penentuan jumlah cluster optimal menggunakan *Elbow Method* dan *Silhouette Analysis*, serta visualisasi hasil clustering menggunakan *Principal Component Analysis* (PCA).

Hasil penelitian menunjukkan bahwa model algoritma *K-Means* berhasil mengelompokkan pengguna kartu kredit ke dalam beberapa segmen dengan 3 model cluster. Fitur utama yang digunakan meliputi usia, pendapatan tahunan, total pengeluaran, skor kredit, dan *Customer Lifetime Value* (CLV). Analisis peluang *default* per cluster memungkinkan identifikasi segmen risiko tinggi secara efektif. Model diimplementasikan dalam aplikasi web berbasis *Streamlit* yang memungkinkan pengguna memasukkan data pelanggan secara manual untuk mendapatkan prediksi segmen beserta estimasi risiko gagal bayar. Penelitian ini diharapkan dapat membantu lembaga keuangan dalam pengambilan keputusan berbasis data terkait manajemen risiko kredit.

Kata kunci: *K-Means Clustering*, Segmentasi Pengguna, Kartu Kredit, *Credit Default*, *Data Mining*.

ABSTRACT

The rapid increase in credit card usage among the public has elevated the risk of credit default, resulting in significant financial losses for issuing institutions. One primary contributing factor to this heightened default risk is the limited understanding of customer behavioral patterns by financial institutions. To address this issue, the present study aims to develop a behavioral segmentation model for credit card users and to analyze default potential through the application of the K-Means Clustering algorithm.

The research utilizes the publicly available Credit Card User Behavior and Default Dataset from Kaggle. The methodology encompasses data preprocessing, feature selection based on correlation analysis, determination of the optimal number of clusters using the Elbow Method and Silhouette Analysis, and visualization of cluxtering outcomes employing Principal Component Analysis (PCA).

The findings demonstrate that the K-Means algorithm effectively segments credit card users into distinct groups, with an optimal configuration of three clusters. The principal features driving the segmentation include age, annual income, total expenditure, credit score, and Customer Lifetime Value (CLV). Subsequent analysis of default probability across clusters enables the precise identification of high-risk segments. The developed model has been implemented in a web-based application using Streamlit, allowing users to input individual customer data manually and obtain the predicted segment along with an estimated default risk assessment. This research is expected to support financial institutions in making data-driven decisions for effective credit risk management.

Keywords: *K-Means Clustering, User Segmentation, Credit Card, Credit Default, Data Mining.*