

**Implementasi K-Prototypes dengan Ensemble Learning untuk
Segmentasi Pelanggan E-Commerce PT. XYZ**



Tugas Akhir

Cianando Pautrisio Cendranadi

00000057710

**PROGRAM STUDI TEKNIK KOMPUTER
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG**

2025

**Implementasi K-Prototypes dengan Ensemble Learning untuk
Segmentasi Pelanggan E-Commerce PT. XYZ**



Tugas Akhir

Diajukan sebagai Salah Satu Syarat untuk Memperoleh
Gelar Sarjana Teknik

Cianando Pautrisio Cendranadi
00000057710

PROGRAM STUDI TEKNIK KOMPUTER
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG
2025

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Cianando Pautrisio Cendranadi

Nomor Induk Mahasiswa : 00000057710

Program Studi : Teknik Komputer

Skripsi dengan judul:

Implementasi K-Prototypes dengan Ensemble Learning untuk Segmentasi Pelanggan E-Commerce PT. XYZ

Merupakan hasil karya saya sendiri bukan plagiat dari laporan karya tulis ilmiah yang ditulis oleh orang lain, dan semua sumber, baik yang dikutip maupun dirujuk, telah saya nyatakan dengan benar serta dicantumkan di Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan/penyimpangan, baik dalam pelaksanaan maupun dalam penulisan laporan karya tulis ilmiah, saya bersedia menerima konsekuensi dinyatakan TIDAK LULUS untuk mata kuliah yang telah saya tempuh.

Tangerang, 27 November 2025



Cianando Pautrisio Cendranadi

HALAMAN PENGESAHAN

Skripsi dengan judul

Implementasi K-Prototypes dengan Ensemble Learning untuk Segmentasi
Pelanggan E-Commerce PT. XYZ

Oleh

Nama : Cianando Pautrisio Cendranadi

NIM : 00000057710

Program Studi : Teknik Komputer

Fakultas : Teknik dan Informatika

Telah diujikan pada hari Rabu, 10 Desember 2025
Pukul 13.00 s.d 15.00 dan dinyatakan
LULUS

Dengan susunan penguji sebagai berikut.

Ketua Sidang



Samuel Hutagalung, M. T. I.
0304038902

Pembimbing



Monica Pratiwi, S.ST., M.T.
0325059601

Ketua Program Studi Teknik Komputer

Penguji



Dareen Kusuma Halim, S.Kom.,
M.Eng.Sc.
NIDN. 0317129202



Samuel Hutagalung, M. T. I.
0304038902

HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH

Yang bertanda tangan di bawah ini:

Nama : Cianando Pautrisio Cendranadi
NIM : 00000057710
Program Studi : Teknik Komputer
Jenjang : S1
Judul Karya Ilmiah : Implementasi K-Prototypes dengan
Ensemble Learning untuk Segmentasi Pelanggan E-Commerce

Menyatakan dengan sesungguhnya bahwa saya bersedia (**pilih salah satu**):

- ☒ Saya bersedia memberikan izin sepenuhnya kepada Universitas Multimedia Nusantara untuk mempublikasikan hasil karya ilmiah saya ke dalam repositori Knowledge Center sehingga dapat diakses oleh Sivitas Akademika UMN/Publik. Saya menyatakan bahwa karya ilmiah yang saya buat tidak mengandung data yang bersifat konfidensial.
- ☐ Saya tidak bersedia mempublikasikan hasil karya ilmiah ini ke dalam repositori Knowledge Center, dikarenakan: dalam proses pengajuan publikasi ke jurnal/konferensi nasional/internasional (dibuktikan dengan *letter of acceptance*) **.
- ☐ Lainnya, pilih salah satu:
 - ☐ Hanya dapat diakses secara internal Universitas Multimedia Nusantara
 - ☐ Embargo publikasi karya ilmiah dalam kurun waktu 3 tahun.

Tangerang, 27 November 2025



Cianando Pautrisio Cendranadi

KATA PENGANTAR

Puji Syukur atas selesainya Penulisan Laporan Tugas Akhir ini dengan judul: Implementasi K-Prototypes dengan Ensemble Learning untuk Segmentasi Pelanggan E-Commerce PT. XYZ yang dilakukan untuk memenuhi salah satu syarat untuk mencapai gelar Strata 1 Jurusan Teknik Komputer pada Fakultas Teknik dan Informatika Universitas Multimedia Nusantara. Saya menyadari bahwa tanpa bantuan dari berbagai pihak, sangatlah sulit bagi saya untuk menyelesaikan tugas akhir ini, oleh karena itu saya ingin mengucapkan terima kasih kepada

Mengucapkan terima kasih

1. Dr. Andrey Andoko., selaku Rektor Universitas Multimedia Nusantara.
2. Bapak Dr. Eng. Niki Prastomo, S.T., M.Sc., selaku Dekan Fakultas Universitas Multimedia Nusantara.
3. Bapak Samuel Hutagalung, M.T.I., selaku Ketua Program Studi Teknik Komputer Universitas Multimedia Nusantara.
4. Ibu Monica Pratiwi, S.ST., M.T., sebagai Pembimbing pertama yang telah memberikan bimbingan, arahan, dan motivasi atas terselesainya tugas akhir ini.
5. Keluarga saya yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan tugas akhir ini.

Semoga karya ilmiah ini dapat bermanfaat bagi seluruh civitas akademika UMN dan pembaca lainnya.

Tangerang, 27 November 2025



Cianando Pautrisio Cendranadi

Implementasi K-Prototypes dengan Ensemble Learning untuk

Segmentasi Pelanggan E-Commerce PT. XYZ

Cianando Pautrisio Cendranadi

ABSTRAK

Segmentasi pelanggan merupakan aspek vital untuk kesuksesan platform e-commerce dalam memberikan personalisasi yang tepat sasaran. Proses clustering manual yang saat ini dijalankan PT. XYZ memerlukan waktu 1-3 hari kerja per siklus dan rentan terhadap inkonsistensi, sehingga diperlukan pendekatan analitik otomatis yang lebih efisien. Penelitian ini mengimplementasikan metode ensemble K-Prototypes untuk clustering pelanggan PT. XYZ berbasis data transaksi dan demografi campuran. Dataset mencakup 75.232 transaksi dari tiga periode bulanan (Juni-Agustus 2025) dengan tiga fitur utama: usia (numerik ternormalisasi), gender, dan kategori produk (kategorikal). Proses clustering dilakukan melalui bootstrap ensemble clustering untuk meningkatkan robustness, dilanjutkan dengan meta-clustering untuk memperoleh konsensus cluster terbaik. Evaluasi kualitas menggunakan metrik internal (Silhouette Score, Davies-Bouldin Index, Calinski-Harabasz Index) sebagai validasi utama, dilengkapi AdaBoost classifier untuk mengukur separability cluster. Hasil penelitian menunjukkan bahwa metode ensemble K-Prototypes berhasil menghasilkan empat segmen pelanggan dengan profil demografis yang jelas dan terpisah, serta konsisten di periode 3 bulan. Metrik evaluasi internal menunjukkan kualitas clustering yang cukup baik (Silhouette Score > 0.65 , Davies-Bouldin Index < 1.0 , Calinski-Harabasz Index > 100), dan evaluasi separability mencapai accuracy $> 85\%$. Waktu eksekusi total pipeline adalah sekitar 15 menit (920 detik), lebih efisien dibandingkan proses manual yang memerlukan 1-3 hari kerja. Segmen yang dihasilkan juga dapat digunakan sebagai bahan pertimbangan dalam strategi pemasaran terpersonalisasi, dengan setiap cluster menunjukkan karakteristik unik dalam profil demografis, preferensi produk, dan dinamika pertumbuhan.

Kata kunci: e-commerce, clustering, K-Prototypes, clustering

Implementation of K-Prototypes with Ensemble Learning for PT.

XYZ E-Commerce Customer Segmentation

Cianando Pautrisio Cendranadi

ABSTRACT (English)

Customer segmentation is a vital aspect for the success of e-commerce platforms in delivering targeted personalization. The manual segmentation process currently used by PT. XYZ takes 1–3 working days per cycle and is prone to inconsistency, so a more efficient, automated analytical approach is needed. This study implements an ensemble K-Prototypes method for segmenting PT. XYZ's customers, based on mixed transaction and demographic data. The dataset consists of 75,232 transactions from three monthly periods (June–August 2025) with three main features: age (normalized numeric), gender, and product category (categorical). The segmentation process is carried out using bootstrap ensemble clustering to improve robustness, followed by meta-clustering to obtain the best consensus clusters. Quality evaluation uses internal metrics (Silhouette Score, Davies-Bouldin Index, Calinski-Harabasz Index) as the primary validation, complemented by an AdaBoost classifier to measure cluster separability. The results show that the ensemble K-Prototypes method successfully produces four customer segments with clear and well-separated demographic profiles that remain consistent over the three-month period. The internal evaluation metrics indicate reasonably good clustering quality (Silhouette Score > 0.65, Davies-Bouldin Index < 1.0, Calinski-Harabasz Index > 100), and the separability evaluation reaches accuracy > 85%. The total pipeline execution time is about 15 minutes (920 seconds), which is more efficient than the manual process that requires 1–3 working days. The resulting segments can also be used as a basis for personalized marketing strategies, with each cluster showing unique characteristics in demographic profiles, product preferences, and growth dynamics.

Keywords: *e-commerce, segmentation, K-Prototypes, clustering*

DAFTAR ISI

HALAMAN PERSETUJUAN	3
HALAMAN PENGESAHAN	4
HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH	5
ABSTRAK	7
<i>ABSTRACT (English)</i>	8
DAFTAR ISI	9
DAFTAR TABEL	11
DAFTAR GAMBAR	12
DAFTAR LAMPIRAN	13
BAB I	
PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Pertanyaan Penelitian	3
1.3 Batasan Penelitian	3
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	3
1.6 Sistematika Penulisan	4
BAB II	
TINJAUAN PUSTAKA	5
2.1 Justifikasi Solusi	5
2.1.1 Extensions to the K-Means Algorithm for Clustering Large Data Sets with Categorical Values [6]	5
2.1.2 Bagging to improve the accuracy of a clustering procedure [4]	6
2.1.3 Intuitive-K-prototypes: A mixed data clustering algorithm with intuitionistic distribution centroid [10]	7
2.1.4 Accuracy Measure of Customer Churn Prediction in Telecom Industry using Adaboost over Random Forest Algorithm [7]	7
2.1.5 Customer Segmentation and Classification Using K-Modes Clustering with Ensemble Learning [8]	8
2.2 Tinjauan Teori	9
2.2.1 Customer Segmentation in E-Commerce	9
2.2.2 K-Prototypes Clustering Algorithm	11
2.2.3 Bootstrap Ensemble Principles	12
2.2.4 Meta-Clustering dan Cluster Ensemble	14
2.2.5 Adaboost Classification	15
BAB III	
METODE PENELITIAN	18

3.1 Flow Existing	18
3.2 Gambaran Umum Metodologi	18
3.3 Dataset	19
3.4 Perancangan Solusi	19
3.4.1 Data Preprocessing	20
3.4.2 Cari K terbaik, dari Elbow Method , Cari Metode Init terbaik	23
Tabel 3.1 Perbandingan metrik evaluasi clustering	24
3.4.2.1 Elbow Method	24
3.4.2.2 Silhouette Score	25
3.4.2.3 Davies Bouldin Index	26
3.4.2.4 Calinski Harabazs Index	27
3.4.3 Implementasi Algoritma Ensemble K-Prototypes	28
3.4.4 Evaluasi Konsistensi Label Cluster Antar Periode	28
3.4.5 Evaluasi Hasil Clustering dan Customer Segment Profiling	29
3.4.6 Cluster Analysis	30
3.5 Metode Pengujian	31
3.5.1 Pengujian Efisiensi Waktu	31
3.5.2 Pengujian Kualitas Clustering	33
3.5.3 Pengujian Validasi Cluster (AdaBoost Classification)	34
BAB IV	
IMPLEMENTASI DAN PENGUJIAN	36
4.1 Spesifikasi Sistem	36
4.1.1 Hardware	36
4.1.2 Software	36
4.2 Limitasi dan Scope Solusi	37
4.3 Implementasi Solusi	37
4.3.1 Data Preprocessing	39
4.3.2 Cari K dan Init terbaik	41
4.3.3 Implementasi Algoritma	46
4.3.4 Dashboard	54
4.4 Pengujian & Analisis Solusi	60
Tabel 4.1 Hasil internal metrik clustering	61
Tabel 4.2 Hasil Adaboost classifier	62
4.4.1 Analisis Clustering 3 Bulan	63
4.4.2 Hasil Cluster dan Insight Bisnis	65
4.4.3 Interpretasi Efisiensi Waktu	68
BAB V	
SIMPULAN DAN SARAN	69

5.1 Simpulan	69
5.2 Saran	69
DAFTAR PUSTAKA	71
LAMPIRAN	73



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR TABEL

Tabel 3.1 Perbandingan metrik evaluasi clustering	23
Tabel 4.1 Hasil internal metrik clustering	59
Tabel 4.2 Hasil Adaboost classifier	60



DAFTAR GAMBAR

Gambar 1.1 Chart prioritas tujuan penelitian	2
Gambar 3.1 Flow Existing	18
Gambar 3.2 Alur Penelitian	20
Gambar 4.1 Hasil dari Elbow Method	43
Gambar 4.2 Silhouette Score, Davies Bouldin Index, dan Calinski Harabasz Index	45
Gambar 4.3 Halaman awal dashboard	55
Gambar 4.4 Tampilan ketika load dataset	56
Gambar 4.5 Tampilan ketika data preprocessing	57
Gambar 4.6 Tampilan metrik setelah clustering	57
Gambar 4.7 Chart Cluster Distributions	58
Gambar 4.8 Chart Gender Distribution per Cluster	59
Gambar 4.9 Chart Age Distribution per Cluster	59
Gambar 4.10 Main Category Distribution per Cluster	60
Gambar 4.11 Total Waktu	60
Gambar 4.12 Distribusi masing-masing cluster	63
Gambar 4.13 Distribusi gender pada masing-masing cluster	64
Gambar 4.14 Distribusi umur pada masing-masing cluster	65
Gambar 4.15 Distribusi kategori produk pada masing-masing cluster	66
Gambar 4.16 Timing breakdown	68



DAFTAR LAMPIRAN

Lampiran A - Turnitin	72
Lampiran B - Konsultasi Bimbingan	73
Lampiran C - Daftar Penggunaan AI	74
Lampiran D - Hasil Kuesioner	75
Lampiran E - Dataset	77



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA