

**ANALISIS PERILAKU PELANGGAN E-COMMERCE
MENGUNAKAN MODEL KLUSTERING DENGAN
ALGORITMA K-MEANS**

SKRIPSI SARJANA INFORMATIKA

Oleh

Faris Helmi

207064516114



**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNOLOGI KOMUNIKASI DAN
INFORMATIKA**

UNIVERSITAS NASIONAL

2023

ANALISIS PERILAKU PELANGGAN E-COMMERCE MENGUNAKAN MODEL KLUSTERING DENGAN ALGORITMA K-MEANS

SKRIPSI SARJANA

Karya ilmiah sebagai salah satu syarat untuk memperoleh gelar
Sarjana Teknik Teknologi Informatika dari Fakultas Teknologi Komunikasi dan Informatika

Oleh

Faris Helmi

207064516114



**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNOLOGI KOMUNIKASI DAN
INFORMATIKA
UNIVERSITAS NASIONAL
2023**

HALAMAN PENGESAHAN
TUGAS AKHIR

Analisis Perilaku Pelanggan E-Commerce Menggunakan Model Klustering
Dengan Algoritma K-Means



Dosen Pembimbing 1

A handwritten signature in black ink, consisting of several loops and a long vertical stroke.

Dr. Agung Triayudi, S.Kom., M.Kom

Dosen Pembimbing 2

A handwritten signature in black ink, featuring a large, stylized 'R' and several horizontal strokes.

Ratih Titi Komala Sari S.T., M.M

HALAMAN PENGESAHAN
TUGAS SARJANA

ANALISIS PERILAKU PELANGGAN E-COMMERCE
MENGGUNAKAN MODEL KLUSTERING DENGAN
ALGORITMA K-MEANS



Pembimbing I

(Dr. Agung Triayudi, S.Kom., M.Kom)

NIP. 0107019009

Pembimbing II

(Ratih Titi Komala Sari S.T., M.M)

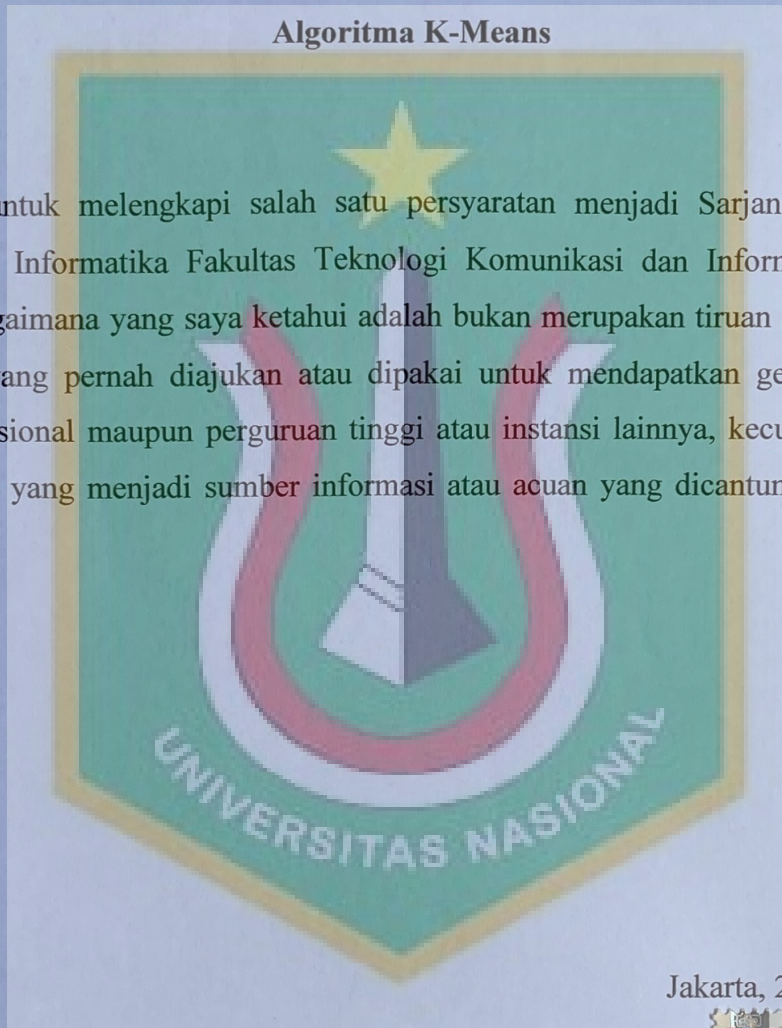
NIP. 0103150850

PERNYATAAN KEASLIAN TUGAS AKHIR

Saya menyatakan dengan sesungguhnya bahwa Tugas Akhir dengan judul :

Analisis Perilaku Pelanggan E-Commerce Menggunakan Model Klustering Dengan Algoritma K-Means

Yang dibuat untuk melengkapi salah satu persyaratan menjadi Sarjana Komputer pada Program Studi Informatika Fakultas Teknologi Komunikasi dan Informatika Universitas Nasional, sebagaimana yang saya ketahui adalah bukan merupakan tiruan atau publikasi dari Tugas Akhir yang pernah diajukan atau dipakai untuk mendapatkan gelar di lingkungan Universitas Nasional maupun perguruan tinggi atau instansi lainnya, kecuali pada bagian – bagian tertentu yang menjadi sumber informasi atau acuan yang dicantumkan sebagaimana mestinya.



Jakarta, 26 Februari 2024



Faris Helmi

207064516114

LEMBAR PERSETUJUAN TUGAS AKHIR

Tugas Akhir dengan judul :

Analisis Perilaku Pelanggan E-Commerce Menggunakan Model Klustering Dengan Algoritma K-Means

Dibuat untuk melengkapi salah satu persyaratan menjadi Sarjana Komputer pada Program Studi Informatika, Fakultas Teknologi Komunikasi dan Informatika Universitas Nasional. Tugas Akhir ini diujikan pada Sidang Akhir Semester Ganjil 2023-2024 pada tanggal 19 Februari Tahun 2024



Dosen Pembimbing 1

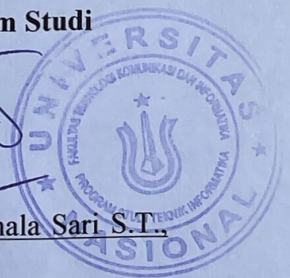
Dr. Agung Triayudi,
S.Kom., M.Kom
NID 0419068604

Ketua Program Studi

Ratih Titi Komala Sari S.T.

M.M

NID 0301038302



HALAMAN PERNYATAAN ORISINALITAS

Skripsi ini adalah hasil karya saya sendiri, dan semua sumber baik yang dikutip maupun dirujuk telah saya nyatakan dengan benar. Bilamana di kemudian hari ditemukan bahwa karya tulis ini menyalahi peraturan yang ada berkaitan etika dan kaidah penulisan karya ilmiah yang berlaku, maka saya bersedia dituntut dan diproses sesuai dengan ketentuan yang berlaku.

Yang menyatakan,

Nama : Faris Helmi
NIM : 207064516114
Tanda Tangan :
Tanggal : 21 Februari 2024

Mengetahui

Pembimbing I : Dr. Agung Triayudi, S.Kom., M.Kom (.....)

Pembimbing II : Ratih Titi Komala Sari S.T., M.M., MMSI (.....)



LEMBAR PERSETUJUAN JUDUL YANG TIDAK ATAU YANG DIREVISI

Nama : Faris Helmi
NPM : 207064516114
Fakultas/Akademi : Fakultas Teknologi Komunikasi dan Informatika
Program Studi : Informatika
Tanggal Sidang : 19 Februari 2024


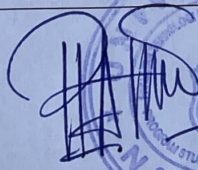
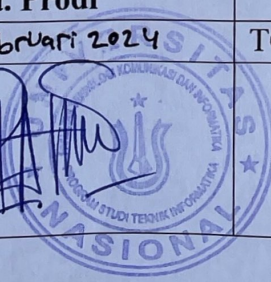
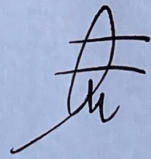
JUDUL DALAM BAHASA INDONESIA :

Analisis Perilaku Pelanggan E-Commerce Menggunakan Model Klustering Dengan Algoritma K-Means

JUDUL DALAM BAHASA INGGRIS :

Analysis of E-Commerce Customer Behavior Using a Clustering Model with K-Means Algorithm

TANDA TANGAN DAN TANGGAL

Pembimbing 1	Ka. Prodi	Mahasiswa
TGL : 26 Februari 2024	TGL : 26 Februari 2024	TGL : 26 Februari 2024
	 	

LEMBAR PERSETUJUAN JUDUL YANG TIDAK ATAU YANG DIREVISI

Nama : Faris Helmi
NPM : 207064516114
Fakultas/Akademi : Fakultas Teknologi Komunikasi dan Informatika
Program Studi : Informatika
Tanggal Sidang : 19 Februari

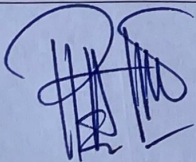

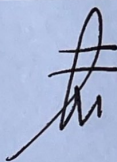
JUDUL DALAM BAHASA INDONESIA :

Analisis Perilaku Pelanggan E-Commerce Menggunakan Model Klustering Dengan Algoritma K-Means

JUDUL DALAM BAHASA INGGRIS :

Analysis of E-Commerce Customer Behavior Using a Clustering Model with K-Means Algorithm

TANDA TANGAN DAN TANGGAL

Pembimbing 2	Ka. Prodi	Mahasiswa
TGL : 26 Februari 2024	TGL :	TGL :
		

HALAMAN PERNYATAAN ORISINALITAS

Skripsi ini adalah hasil karya saya sendiri, dan semua sumber baik yang dikutip maupun dirujuk telah saya nyatakan dengan benar. Bilamana di kemudian hari ditemukan bahwa karya tulis ini menyalahi peraturan yang ada berkaitan etika dan kaidah penulisan karya ilmiah yang berlaku, maka saya bersedia dituntut dan diproses sesuai dengan ketentuan yang berlaku.

Yang menyatakan,

Nama : Faris Helmi

NIM : 207064516114

Tanda Tangan : 

Tanggal : 21 Februari 2024

Mengetahui

Pembimbing I : Dr. Agung Triayudi, S.Kom., M.Kom

(.....)

Pembimbing II : Ratih Titi Komala Sari S.T., M.M., MMSI

(.....)



HALAMAN PENGESAHAN
TUGAS SARJANA

ANALISIS PERILAKU PELANGGAN E-COMMERCE
MENGGUNAKAN MODEL KLUSTERING DENGAN
ALGORITMA K-MEANS



Pembimbing I

(Dr. Agung Triayudi, S.Kom., M.Kom)

NIP. 0107019009

Pembimbing II

(Ratih Titi Komala Sari S.T., M.M)

NIP. 0103150850

KATA PENGANTAR

Puji dan syukur penulis panjatkan kepada Tuhan Yang Maha Esa yang telah memberikan rahmat dan karunia sehingga penulis dapat menyelesaikan skripsi dengan judul "Analisis Perilaku Pengguna E-Commerce Menggunakan Model Klustering Dengan Algoritma K-Means" sebagai salah satu syarat kelulusan Program Studi Sarjana --- Fakultas Teknologi Komunikasi dan Informatika.

Penelitian dan penulisan skripsi ini tidak terlepas dari bantuan berbagai pihak, oleh karena itu penulis menyampaikan banyak terima kasih terutama kepada dosen pembimbing Tugas Akhir, Dr. A yang telah meluangkan banyak waktu, tenaga, pikiran, bimbingan, arahan, motivasi serta memaklumi segala kekurangan penulis selama penelitian tugas akhir dan penyusunan skripsi. Penulis juga mengucapkan banyak terima kasih kepada:

1. Perusahaan Achmad Zaky Foundation (Startup Campus) yang telah memberikan bantuan selama penelitian dalam bentuk dana/data/sarana prasarana.
2. Ayah dan Ibu selaku orangtua penulis yang telah banyak memberi dukungan salam segala bentuk yang tak terhitung.
3. Seluruh dosen pengajar di Program Studi ... FTKI maupun dosen di Program Studi lain yang memberikan banyak ilmu.
4. Teman-teman seangkatan dan sehimpuan berbagai angkatan yang telah membantu dan mendukung.
5. Serta masih banyak pihak yang tidak dapat disebutkan satu per satu yang telah membantu penulis dalam menyelesaikan skripsi dengan baik dan tepat waktu.

Akhir kata, semoga Tuhan Yang Maha Esa membalas kebaikan dan bantuan yang telah diberikan dengan hal yang lebih baik. Penulis mengharapkan kritik dan saran yang bersifat membangun dan semoga skripsi ini dapat memberikan manfaat di bidang Teknologi Informatika.

Jakarta, 11 November 2023

Faris Helmi

HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademik Program Studi Teknik Informatika, Fakultas Teknologi Komunikasi dan Informatika, saya yang bertanda tangan di bawah ini:

Nama : Faris Helmi
NIM : 207064516114

demikian demi pengembangan ilmu pengetahuan, pengetahuan, menyetujui untuk memberikan kepada Fakultas Teknologi Komunikasi dan Informatika, Hak Bebas Royalti Noneksklusif (Non-exclusive Royalti Free Right) atas karya ilmiah saya yang berjudul:

Analisis Perilaku Pelanggan E-Commerce Menggunakan Model Klustering Dengan Algoritma K-Means

Beserta perangkat yang ada (jika diperlukan). Dengan Hak ini Fakultas Teknologi Komunikasi dan Informatika berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (database), merawat, dan mempublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Jakarta

Pada tanggal : 11 November 2023

Yang menyatakan



(Faris Helmi)

ABSTRAK

E-commerce merupakan salah satu sektor bisnis yang berkembang pesat di era digital. Untuk meningkatkan kinerja dan kepuasan pelanggan perusahaan e-commerce perlu memahami perilaku dan preferensi pelanggan secara mendalam. Salah satu metode yang dapat digunakan adalah analisis RFM, teknik yang digunakan untuk mengidentifikasi pelanggan yang paling berharga berdasarkan kapan pembelian terakhir mereka (*Recency*), seberapa sering mereka melakukan pembelian (*Frequency*), dan berapa total pengeluaran mereka (*Monetary*). Dari analisis RFM tersebut digunakan model *clustering* untuk membuat segmen pelanggan berdasarkan RFM. Penelitian ini juga mengeksplorasi faktor yang menyebabkan pengguna meninggalkan platform e-commerce perusahaan dengan menganalisis aktivitas pelanggan saat menggunakan platform. Model ini diharapkan dapat menghasilkan segmentasi pelanggan yang lebih akurat, stabil, dan interpretabel. Penelitian ini menggunakan data transaksi pelanggan dari salah satu platform e-commerce terkemuka di Indonesia. Hasil penelitian menunjukkan informasi pelanggan dari yang loyal sampai yang hilang berdasarkan perilaku pembelian dan aktivitas penggunaan platform perusahaan. Selain itu, model ini menggunakan empat algoritma, satu algoritma menggunakan k-means untuk membuat segmentasi pelanggan berdasarkan RFM, tiga algoritma lainnya *decision tree*, *random forests*, dan XGBoost digunakan untuk membangun model prediksi faktor yang paling berpengaruh membuat pelanggan meninggalkan platform dan membandingkan kinerja model dari ketiga algoritma.

Kata kunci: e-commerce, perilaku pelanggan, klustering, algoritma K-means, Decision Tree, Random Forests, dan XGBoost.

Analysis of E-Commerce Customer Behavior Using a Clustering Model with K-Means Algorithm

E-commerce is one of the rapidly growing business sectors in the digital era. To improve performance and customer satisfaction e-commerce companies need to understand customer behavior and preferences in depth. One method that can be used is RFM analysis, a technique used to identify the most valuable customers based on when their last purchase was (Recency), how often they make purchases (Frequency), and how much they spend in total (Monetary). From the RFM analysis, a clustering model is used to create customer segments based on RFM. This research also explores the factors that cause users to leave the company's e-commerce platform by analyzing customer activities while using the platform. This model is expected to produce more accurate, stable, and interpretable customer segmentation. This research uses customer transaction data from one of the leading e-commerce platforms in Indonesia. The results show customer information from loyal to lost based on purchase behavior and platform usage activities. In addition, this model uses four algorithms, one algorithm uses k-means to segment customers based on RFM, the other three algorithms decision tree, random forests, and XGBoost are used to build a prediction model of the most influential factors that make customers leave the platform and compare the performance of the model from the three algorithms.

Keywords: e-commerce, customer behavior, clustering, K-means, Decision Tree, Random Forests, and XGBoost algorithm.

DAFTAR ISI

HALAMAN JUDUL	1
LEMBAR PERNYATAAN ORISINALITAS.....	2
LEMBAR PENGESAHAN	3
KATA PENGANTAR	4
LEMBAR PERSETUJUAN PUBLIKASI	5
ABSTRAK.....	6
ABSTRACT.....	7
DAFTAR ISI.....	8
DAFTAR GAMBAR	10
DAFTAR TABEL	12
1. PENDAHULUAN.....	13
1.1 Latar Belakang	13
1.2 Rumusan Masalah.....	15
1.3 Tujuan Penelitian	16
1.4 Batasan Masalah	16
1.5 Kontribusi Penelitian.....	16
2. TINJAUAN PUSTAKA.....	18
2.1 Penelitian Terdahulu.....	18
2.2 Landasan Teori	20
2.2.1 Skor Recency, Frequency, dan Monetary Dengan Quantile-Based Discretization dan K-Means Clustering	20
2.2.2 Model Clustering.....	21
2.2.3 Ensemble Learning	22
2.2.4 Decision Tree.....	23
2.2.5 Random Forest	24
2.2.6 XGBoost.....	26
2.2.7 K-Means Clustering	26
3. METODE PENELITIAN.....	28
3.1 Lokasi Penelitian.....	28
3.2 Waktu Penelitian	28

3.3 Penentuan Subjek Penelitian.....	28
3.4 Fokus Penelitian.....	28
3.5 Sumber Data.....	29
3.6 Teknis Pengumpulan Data	29
3.7 Desain Penelitian	30
4. HASIL DAN PEMBAHASAN	32
4.1 Persiapan Data	32
4.2 Rekayasa Fitur	35
4.3 Data Preprocessing.....	36
4.4 Modelling.....	40
4.5 Evaluasi Model dan Hyperparameter Tuning	41
4.6 Hasil	45
4.7 Dashboard Visualisasi.....	47
5. KESIMPULAN DAN SARAN	50
5.1 Kesimpulan	50
5.2 Saran	51
DAFTAR PUSTAKA	



DAFTAR GAMBAR

Gambar 3.1 Relasi Database	30
Gambar 3.2 Alur Penelitian.....	31
Gambar 4.1 Tampilan Google Colab dan persiapan library python yang digunakan	32
Gambar 4.2 tabel data transaksi memiliki kolom berisi tipe data dictionary.....	33
Gambar 4.3 Program parsing tipe data dictionary menjadi kolom dan baris.....	33
Gambar 4.4 Tabel data click stream memiliki kolom berisi tipe data dictionary	34
Gambar 4.5 Membuat fungsi untuk mengecek baris yang kosong dan ekstrak key.....	34
Gambar 4.6 Menggabungkan key yang sudah di ekstrak ke dataframe menggunakan teknik lambda.....	34
Gambar 4.7 Proses rekayasa fitur untuk membuat dataframe baru berisi fitur RFM	35
Gambar 4.8 Dataframe fitur aktivitas pengguna.....	36
Gambar 4.9 Data agregat	36
Gambar 4.10 Distribusi data RFM sebelum preprocessing	37
Gambar 4.11 Program IQR dan Standard Scaler	37
Gambar 4.12 Distribusi data RFM setelah preprocessing	37
Gambar 4.13 Sebelum pembersihan outlier pada dataframe aktivitas pengguna	37
Gambar 4.14 Sesudah pembersihan outlier pada dataframe aktivitas pengguna.....	38
Gambar 4.15 Program segmen pelanggan	39
Gambar 4.16 Pembuatan label untuk model klasifikasi.....	39
Gambar 4.17 Program model clustering menggunakan k-means	41
Gambar 4.18 Splitting data	41
Gambar 4.19 Kode stratified k-fold	41
Gambar 4.20 Evaluasi kinerja model clustering menggunakan Silhouette Score	42
Gambar 4.21 Kode algoritma decision tree menggunakan grid search untuk melakukan hyperparameter tuning otomatis	43
Gambar 4.22 Kode algoritma random forest menggunakan grid search untuk melakukan hyperparameter tuning otomatis	44
Gambar 4.23 Kode algoritma XGBoost menggunakan grid search untuk melakukan hyperparameter tuning otomatis	45
Gambar 4.24 Grafik segmentasi pelanggan	46
Gambar 4.25 Fitur penting yang mempengaruhi prediksi	47

Gambar 4.26 Potongan kode streamlit.....	47
Gambar 4.27 Hasil dashboard yang dibuka melalui browser	48
Gambar 4.28 Tren pengunjung dan penjualan pertahun	48
Gambar 4.29 Tren pendapatan pertahun	49
Gambar 4.30 Kategori produk terlaris	49
Gambar 4.31 Tabel produk yang sudah terjual	49



DAFTAR TABEL

Tabel 2.1 Perbandingan Penelitian Terdahulu.....	18
Tabel 4.1 Segmen skor pelanggan.....	38
Tabel 4.2 Perbandingan kinerja model prediksi tiga algoritma	46

