

**ANALISIS DAN PREDIKSI PREFERENSI PENGGUNA
SMARTPHONE BERDASARKAN DEMOGRAFI DAN POLA
PENGGUNAAN APLIKASI MENGGUNAKAN K-MEANS DAN
RANDOM FOREST**

TESIS

Oleh:

**STELA MARIS HAREFA
NIM. 221231041**



**PROGRAM STUDI S-2 TEKNOLOGI INFORMASI
FAKULTAS INFORMATIKA
UNIVERSITAS MIKROSKIL
MEDAN
2025**

**ANALYSIS AND PREDICTION OF SMARTPHONE USER
PREFERENCES BASED ON DEMOGRAPHICS AND APPLICATION
USAGE PATTERNS USING K-MEANS AND RANDOM FOREST**

THESIS

By:

**STELA MARIS HAREFA
ID NUMBER. 221231041**



**MAJOR OF S-2 INFORMATION TECHNOLOGY
FACULTY OF INFORMATICS
UNIVERSITAS MIKROSKIL
MEDAN
2025**

LEMBARAN PENGESAHAN
ANALISIS DAN PREDIKSI PREFERENSI PENGGUNA
SMARTPHONE BERDASARKAN DEMOGRAFI DAN POLA
PENGUNAAN APLIKASI MENGGUNAKAN K-MEANS DAN
RANDOM FOREST

TESIS

Diajukan untuk Melengkapi Persyaratan Guna
Mendapatkan Gelar Magister
Program Studi S-2 Teknologi Informasi

Oleh:

STELA MARIS HAREFA
NIM. 221231041

Disetujui Oleh:

Dosen Pembimbing,


Dr. Ronsen Purba, M.Sc.

Medan, 3 September 2025

Diketahui dan Disahkan Oleh:

Ketua Program Studi
S-2 Teknologi Informasi,


Ir. Erwin Setiawan Panjaitan, M.M.S.I., Ph.D.

HALAMAN PERNYATAAN

Saya yang membuat pernyataan ini adalah mahasiswa Program Studi S-2 Teknologi Informasi Universitas Mikroskil Medan dengan identitas mahasiswa sebagai berikut:

Nama : Stela Maris Harefa
NIM : 221231041

Saya telah melaksanakan penelitian dan penulisan Tesis dengan judul dan tempat penelitian sebagai berikut:

Judul Tesis : Analisis dan Prediksi Preferensi Pengguna Smartphone Berdasarkan Demografi dan Pola Penggunaan Aplikasi Menggunakan K-Means dan Random Forest

Tempat Penelitian : -
Alamat Tempat Penelitian : -
No. Telp. Tempat Penelitian : -

Sehubungan dengan Tesis tersebut, dengan ini saya menyatakan dengan sebenar-benarnya bahwa penelitian dan penulisan Tesis tersebut merupakan hasil karya saya sendiri (tidak menyuruh orang lain yang mengerjakannya) dan semua sumber, baik yang dikutip maupun yang dirujuk, telah saya nyatakan dengan benar. Bila di kemudian hari ternyata terbukti bahwa bukan saya yang mengerjakannya (membuatnya), maka saya bersedia dikenakan sanksi yang telah ditetapkan oleh Universitas Mikroskil Medan, yakni pencabutan ijazah yang telah saya terima dan ijazah tersebut dinyatakan tidak sah.

Selain itu, demi pengembangan ilmu pengetahuan, saya menyetujui untuk memberikan kepada Universitas Mikroskil Medan Hak Bebas Royalti Non-eksklusif (Non-exclusive Royalty Free Right) atas Tesis saya beserta perangkat yang ada (jika diperlukan). Dengan hak ini, Universitas Mikroskil Medan berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (database), merawat, dan mempublikasikan Tesis saya, secara keseluruhan atau hanya sebagian atau hanya ringkasannya saja dalam bentuk format teracak dan/atau elektronik, selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik hak cipta. Menyatakan juga bahwa saya akan mempertahankan hak eksklusif saya untuk menggunakan seluruh atau sebagian isi Tesis saya guna pengembangan karya di masa depan, misalnya dalam bentuk artikel, buku, ataupun perangkat lunak/sistem informasi.

Demikian pernyataan ini saya perbuat dengan sungguh-sungguh dalam keadaan sadar dan tanpa ada tekanan dari pihak manapun.

Medan, 01 Juli 2025

Saya yang membuat pernyataan,



Stela Maris Harefa

ANALISIS DAN PREDIKSI PREFERENSI PENGGUNA SMARTPHONE BERDASARKAN DEMOGRAFI DAN POLA PENGUNAAN APLIKASI MENGGUNAKAN K-MEANS DAN RANDOM FOREST

Abstrak

Penelitian ini mengatasi keterbatasan studi terdahulu yang hanya menganalisis preferensi pengguna smartphone secara parsial—baik berdasarkan pola penggunaan aplikasi ataupun faktor demografis. Penelitian ini mengintegrasikan kedua dimensi tersebut guna menghasilkan segmentasi dan prediksi preferensi pengguna smartphone. Algoritma K-Means digunakan untuk membuat segmentasi, dengan penentuan jumlah cluster optimal melalui Elbow Method dan Silhouette Score. Karakteristik pengguna dikelompokkan ke dalam empat kelas baru. Kemudian Random Forest digunakan untuk membangun model prediksi preferensi pengguna. Evaluasi model prediksi menggunakan metrik Accuracy, Precision, Recall, F1-Score, Confusion Matrix, dan ROG-AUC menunjukkan Random Forest mencapai kinerja terbaik dengan akurasi rata-rata sebesar 90%. Hasil penelitian ini memberikan gambaran komprehensif mengenai karakteristik pengguna smartphone, sehingga dapat digunakan pengembang aplikasi ataupun pelaku bisnis untuk memahami preferensi pengguna, merancang produk yang lebih sesuai dengan kebutuhan pengguna, serta mengoptimalkan pengembangan produk.

Kata Kunci: Integrasi demografi-pola pengguna, Segmentasi Pengguna, Preferensi Pengguna, K-Means, Elbow Method, Silhouette Score, Random Forest.

Abstract

This study addresses the limitations of previous research, which examined smartphone user preferences only partially—either through application usage patterns or demographic factors. By integrating both dimensions, this research aims to produce more robust segmentation and prediction of smartphone user preferences. The K-Means algorithm is employed to perform segmentation, with the optimal number of clusters determined using the Elbow Method and Silhouette Score. User characteristics are then grouped into four newly defined categories. Subsequently, the Random Forest algorithm is applied to construct the predictive model of user preferences. Model evaluation using Accuracy, Precision, Recall, F1-Score, Confusion Matrix, and ROC-AUC metrics demonstrates that Random Forest achieves the best performance, with an average accuracy of 90%. The findings provide a comprehensive overview of smartphone user characteristics, which can be utilized by application developers and business practitioners to better understand user preferences, design products that are more aligned with user needs, and optimize product development.

Keywords: Demographic-behavioral integration, User preferences, User Segmentation, K-Means, K-Means, Elbow Method, Silhouette Score, Random Forest.

KATA PENGANTAR

Puji dan Syukur penulis panjatkan kepada Tuhan Yang Maha Esa atas berkat dan Rahmat-Nya, penulis dapat menyelesaikan Tesis yang berjudul “Analisis dan Prediksi Preferensi Pengguna *Smartphone* Berdasarkan Demografi dan Pola Penggunaan Aplikasi Menggunakan *K-Means* dan *Random Forest*”. Tesis ini dibuat untuk melengkapi persyaratan kurikulum pada Program Studi S-2 Teknologi Informasi, Universitas Mikroskil Medan. Penulis menyampaikan banyak terima kasih kepada:

1. Bapak Dr. Ronsen Purba, M.Sc., selaku Dosen Pembimbing yang telah membimbing penulis dalam menyelesaikan tesis ini.
2. Bapak Muhammad Fermi Pasha, B.Sc., M.Sc., PhD, selaku Dosen Pendamping Pembimbing yang telah membimbing penulis dalam menyelesaikan tesis ini.
3. Bapak Hardy, S.Kom., M.Sc., Ph.D., selaku Rektor Universitas Mikroskil Medan.
4. Bapak Sunaryo Winardi, S.Kom., M.T., selaku Dekan Fakultas Informatika Universitas Mikroskil Medan.
5. Bapak Ir. Erwin Setiawan Panjaitan, M.M.S.I., Ph.D., selaku Ketua Program Studi S-2 Teknologi Informasi Universitas Mikroskil Medan.
6. Bapak/Ibu Dosen Universitas Mikroskil yang telah membantu proses penulisan tesis ini.
7. Bapak Drs. Nehemia Harefa, M.M. dan Ibu Sari Isa Harefa, S.E., selaku orang tua yang telah memberikan doa, semangat, motivasi untuk menyelesaikan tesis ini.
8. Semua pihak yang telah memberikan dukungan dan motivasi sehingga tesis ini dapat diselesaikan dengan baik.

Penulis menyadari bahwa masih terdapat banyak kekurangan pada tesis ini. Oleh karena itu, kritik dan saran yang bersifat membangun akan sangat diterima. Akhir kata, semoga tesis ini dapat bermanfaat bagi pihak-pihak yang membutuhkan. Terima kasih.

Medan, 01 Juli 2025

Penulis,

Stela Maris Harefa

DAFTAR ISI

ABSTRAK.....	i
KATA PENGANTAR	ii
DAFTAR ISI	iii
DAFTAR GAMBAR.....	v
DAFTAR TABEL.....	vi
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Tujuan Penelitian.....	4
1.4 Manfaat Penelitian	4
1.5 Ruang Lingkup Penelitian.....	5
BAB II KAJIAN LITERATUR.....	6
2.1 Tinjauan Pustaka	6
2.1.1 Preferensi Pengguna.....	6
2.1.2 Segmentasi Pengguna	7
2.1.3 Machine Learning	7
2.1.4 K-Means Clustering.....	8
2.1.5 Random Forest	11
2.1.6 Metrik Evaluasi Model.....	13
2.1.7 Penelitian Terdahulu.....	15
2.2 Kerangka Pikir Pemecahan Masalah.....	16
BAB III METODOLOGI PENELITIAN.....	18
3.1 Analisis Masalah.....	18
3.2 Rancangan Penelitian	18
3.3 Data yang Digunakan	25
3.4 Alat Penelitian	26
3.5 Teknik Analisis Data.....	27

BAB IV HASIL DAN PEMBAHASAN	29
4.1 Hasil	29
4.1.1 Analisis Data Awal	29
4.1.2 Persiapan Data.....	40
4.1.3 Segmentasi Preferensi Pengguna dengan K-Means.....	41
4.1.4 Profiling Cluster untuk Segmentasi	42
4.1.5 Prediksi Preferensi dengan Random Forest	44
4.1.6 Pengujian Hasil Prediksi	47
4.2 Pembahasan.....	50
4.2.1 Evaluasi Hasil Segmentasi K-Means dengan <i>Elbow Method</i>	51
4.2.2 Perbandingan Clustering dengan Beberapa Metrik	51
4.2.3 Evaluasi Hasil Segmentasi K-Means dengan <i>Silhouette Score</i>	53
4.2.4 Perbandingan Hasil Clustering K-Means dengan Algoritma Lain.....	55
4.2.5 Pembahasan Hasil Pengujian Prediksi dengan Random Forest	57
BAB V KESIMPULAN DAN SARAN.....	58
5.1 Kesimpulan	58
5.2 Saran.....	58
DAFTAR PUSTAKA	60
DAFTAR RIWAYAT HIDUP	64

DAFTAR GAMBAR

Gambar 2.1 Flowchart K-Means	10
Gambar 2.2 Ilustrasi Random Forest [38]	12
Gambar 2.3 Kerangka Konseptual.....	17
Gambar 3.1 Tahapan Penelitian	19
Gambar 3.2 Dataset Mobile Usage Behavioral	19
Gambar 3.3 Contoh Elbow Point.....	23
Gambar 3.4 Sampel Dataset	26
Gambar 4.1 Informasi Dataset.....	29
Gambar 4.2 Distribusi Usia	30
Gambar 4.3 Distribusi Gender.....	31
Gambar 4.4 Distribusi Pengguna Berdasarkan Lokasi	32
Gambar 4.5 Durasi Penggunaan Aplikasi.....	33
Gambar 4.6 Distribusi Total Jam Penggunaan Aplikasi dan Jam Layar Harian.....	34
Gambar 4.7 Durasi Media Sosial Berdasarkan Gender	34
Gambar 4.8 Durasi Aplikasi Produktivitas Berdasarkan Gender	35
Gambar 4.9 Durasi Aplikasi Game Berdasarkan Gender	36
Gambar 4.10 Durasi Penggunaan Aplikasi per Lokasi.....	36
Gambar 4.11 Durasi Penggunaan Aplikasi Berdasarkan Lokasi dan Jenis Aplikasi	37
Gambar 4.12 Durasi Rata-rata Penggunaan Aplikasi per Lokasi dan Jenis Aplikasi	38
Gambar 4.13 Heatmap Matriks Korelasi Variabel Numerik	39
Gambar 4.14 Rata-rata Fitur per Cluster	43
Gambar 4.15 Metrik Evaluasi Model Random Forest.....	48
Gambar 4.16 Confusion Matrix.....	49
Gambar 4.17 Hasil Pengujian dengan ROC-AUC	50
Gambar 4.18 Grafik Elbow Point.....	51
Gambar 4.19 Calinski-Harabasz Index.....	52
Gambar 4.20 Davies-Bouldin Index	52
Gambar 4.21 Cross Validation.....	53
Gambar 4.22 Grafik Silhouette Score.....	54
Gambar 4.23 PCA Centroid.....	55

DAFTAR TABEL

Tabel 3.1 Tabel Selisih Kuadrat Nilai ke-i dan Rata-rata	21
Tabel 3.2 Tabel Nilai Z-Score.....	21
Tabel 3.3 Encoding Variable.....	22
Tabel 3.4 Centroid acak	22
Tabel 3.5 Contoh Penentuan Cluster Terdekat	22
Tabel 3.6 Contoh Pemilihan Elbow Point	23
Tabel 3.7 Silhouette Score	23
Tabel 3.8 Centroid Final	24
Tabel 3.9 Hasil Segmentasi K-Means.....	24
Tabel 3.10 Contoh Hasil Prediksi	25
Tabel 3.11 Contoh Data Testing.....	25
Tabel 3.12 Struktur Data Pola Penggunaan Aplikasi.....	25
Tabel 4.1 Tabel Nilai Fitur Awal.....	41
Tabel 4.2 Tabel Nilai Fitur Setelah Standarisasi.....	42
Tabel 4.3 Labeling Data.....	42
Tabel 4.4 Tabel Gender.....	45
Tabel 4.5 One Hot Encoding tanpa Drop First	45
Tabel 4.6 One-Hot Encoding dengan Drop Fisrt.....	46
Tabel 4.7 Encoding Location.....	46
Tabel 4.8 Sampel Data Random Forest	47
Tabel 4.9 Nilai WCSS dan Silhouette Score	54
Tabel 4.10 Metrik Evaluasi.....	56