BAB I
PENDAHULUAN

1.1 Latar Belakang

Dalam era digital saat ini, *smartphone* telah menjadi bagian integral dari kehidupan sehari-hari. Menurut laporan dari Statista, jumlah pengguna *smartphone* di seluruh dunia diperkirakan mencapai 6,8 miliar pada tahun 2023, menunjukkan pertumbuhan yang signifikan dalam penggunaan perangkat ini (Statista, 2023) [1]. *Smartphone* tidak hanya berfungsi sebagai alat komunikasi, tetapi juga sebagai platform untuk berbagai aplikasi yang mendukung aktivitas sehari-hari, mulai dari pendidikan, hiburan, hingga transaksi keuangan.

Proyeksi tahun 2025 menunjukkan bahwa jumlah unduhan aplikasi secara global diperkirakan melebihi 300 miliar, dengan pengguna *smartphone* mencapai lebih dari 6,3 miliar orang di seluruh dunia. Secara rata-rata, pengguna menghabiskan waktu lebih dari empat jam per hari untuk berinteraksi dengan berbagai aplikasi, dimana 88% dari total waktu penggunaan dialokasikan khusus untuk aktivitas dalam aplikasi. Adapun kategori aplikasi yang mendominasi jumlah unduhan mencakup aplikasi permainan, platform media sosial, serta aplikasi produktivitas (Appinventiv, 2025) [2]. Menurut laporan *Pew Research Center* (2023), secara kolektif, pengguna aplikasi mobile di Amerika Serikat menghabiskan waktu hingga ratusan miliar jam per tahun (akumulasi seluruh waktu penggunaan dari jutaan pengguna aktif), dengan kategori aplikasi yang paling dominan adalah *game, video streaming, social media*, dan aplikasi hiburan [3]. Di sisi lain, rata-rata pengguna *smartphone* di Indonesia menghabiskan waktu lebih dari 5,7 jam per hari di dalam aplikasi *mobile*, menjadikan Indonesia sebagai negara dengan *screen time* tertinggi di dunia sepanjang tahun 2023 [4]. Menurut *Business of Apps* (2025), jumlah pengguna aktif *smartphone* di Indonesia mencapai sekitar 209 juta orang. Aplikasi seperti TikTok, YouTube, WhatsApp, serta kategori game, utilitas, dan fintech menjadi yang paling sering digunakan oleh pengguna di Indonesia [5].

Setiap pengguna memiliki preferensi unik terhadap aplikasi yang digunakan, sehingga analisis terhadap data dapat memberikan wawasan berharga tentang perilaku pengguna yang dapat dimanfaatkan untuk pengembangan produk, layanan, atau strategi pemasaran yang lebih tepat sasaran, dan memberikan manfaat yang signifikan di berbagai sektor.

© Karya Dilindungi UU Hak Cipta
1. Dilarang menyebarluaskan dokumen tanpa izin.
2. Dilarang melakukan plagiasi.
3. Pelanggaran diberlakukan sanksi sesuai peraturan UU Hak Cipta.

Meningkatnya ketergantungan pengguna terhadap aplikasi *smartphone* menciptakan tantangan dalam memahami preferensi dan perilaku pengguna, yang mana jika tidak dipahami dengan baik, dapat menyebabkan strategi pemasaran yang tidak efektif, pengembangan produk yang tidak tepat sasaran, hingga meningkatnya risiko *churn* (pengguna berhenti menggunakan layanan) yang tinggi. Meskipun banyak penelitian telah dilakukan untuk menganalisis pola penggunaan aplikasi, masih terdapat kekurangan dalam mengidentifikasi faktor-faktor yang mempengaruhi preferensi pengguna. Penelitian terkini menunjukkan bahwa rata-rata aplikasi kehilangan sekitar 77% pengguna aktif harian dalam 3 hari pertama setelah instalasi, dan hanya tersisa sekitar 5–6% pengguna aktif setelah 30 hari [6]. Hal ini mengindikasikan bahwa personalisasi yang kurang efektif merupakan salah satu faktor penyebab *churn* yang tinggi dan potensi kerugian bisnis. Sebaliknya, data menunjukkan bahwa personalisasi yang tepat dapat meningkatkan retensi hingga sekitar 26–30% [7]. Oleh karena itu, penelitian ini mengembangkan metodologi yang menggabungkan segmentasi dan prediksi preferensi pengguna untuk menciptakan pengalaman digital yang lebih relevan.

Pada tahun 2021, Tanuwijaya, et al. melakukan penelitian dengan fokus khusus pada perilaku pengguna Netflix menggunakan data *mobile customer behavior*. Dengan menggabungkan metode *K-Means* dan *CatBoost,* mereka mengidentifikasi tiga kelompok pengguna (*Low, Medium, and High consumption*) berdasarkan tingkat konsumsi data dan menemukan kesamaan pola antara pengguna *e-commerce* dengan pelanggan Netflix, dengan akurasi mencapai 86.5% [8]. Namun penelitian ini hanya berfokus pada prediksi Netflix *subscriber* saja. Penelitian berikutnya ditunjukkan oleh Alruban (2022) yang menggunakan pendekatan LSTM untuk menganalisis 76 pengguna terhadap pola urutan penggunaan aplikasi pada *smartphone*. Penelitian ini berhasil mengenali pola temporal penggunaan aplikasi, dimana aplikasi produktifitas dominan digunakan pada pagi hari (08.00-11.00), sementara aplikasi hiburan lebih banyak digunakan pada malam hari (pukul 18.00-22.00), dengan akurasi 80% [9]. Namun penelitian tersebut tidak mempertimbangkan variabel demografis yang dapat mempengaruhi hasil, seperti bagaimana faktor usia dan jenis kelamin mempengaruhi preferensi pengguna terhadap aplikasi produktivitas dan hiburan.

Kurangnya data yang terintegrasi dan analisis yang mendalam menjadi tantangan utama dalam memahami preferensi pengguna. Banyak penelitian sebelumnya hanya fokus pada analisis deskriptif tanpa mempertimbangkan pendekatan prediktif yang dapat memberikan wawasan lebih dalam tentang preferensi pengguna. Penelitian oleh Selvakumar et al. (2023) dan Srinivasa Rao dan Vaishnavi (2024) menyoroti kecanduan *smartphone* dan

© Karya Dilindungi UU Hak Cipta
1. Dilarang menyebarluaskan dokumen tanpa izin.
2. Dilarang melakukan plagiasi.
3. Pelanggaran diberlakukan sanksi sesuai peraturan UU Hak Cipta.

preferensi pembelajaran, tetapi tidak mengeksplorasi dimensi multidimensi yang lebih luas dalam analisis preferensi pengguna, seperti bagaimana faktor usia, jenis kelamin, dan lokasi mempengaruhi preferensi pengguna terhadap aplikasi media sosial dan hiburan [10], [11].

Penelitian ini bertujuan untuk mengisi gap tersebut dengan melakukan segmentasi pengguna *smartphone* berdasarkan pola multidimensi, termasuk usia, gender, dan rasio penggunaan aplikasi. Segmentasi pengguna ini dilakukan dengan membagi pengguna menjadi beberapa kelompok berdasarkan faktor-faktor demografis dan perilaku penggunaan aplikasi, seperti usia, jenis kelamin, dan lokasi. Dengan pendekatan ini, diharapkan dapat memberikan pemahaman yang lebih mendalam tentang preferensi pengguna, yang pada gilirannya dapat membantu pengembang aplikasi dan pemasar dalam merancang produk yang lebih sesuai dengan kebutuhan pengguna.

Dalam penelitian tentang segmentasi pengguna, *K-Means Clustering* telah terbukti menjadi metode yang efektif dan efisien. Jurnal "*Customer Segmentation Using K-Means Clustering*" menunjukkan bahwa *K-Means* mampu mengelompokkan pelanggan berdasarkan karakteristik yang relevan, sehingga memungkinkan perusahaan untuk mengembangkan strategi pemasaran yang lebih terarah dan personal (Krishna Kumar & Priyanka, 2024) [12]. Selain itu, dalam penelitian oleh Gupta et al., 2021, ditemukan bahwa *K-Means* lebih unggul dalam hal kecepatan dan efisiensi ketika menangani *dataset* besar dibandingkan dengan metode *Hierarchical Clustering*, yang sering kali lebih lambat dan kurang efisien dalam pengolahan data yang kompleks [13]. Penelitian oleh Sutramiani et al., 2024, juga menegaskan bahwa *K-Means* memiliki kinerja yang lebih baik dalam hal akurasi dan interpretasi hasil dibandingkan dengan DBSCAN, terutama dalam konteks pengelompokan data yang tidak terstruktur [14]. Oleh karena itu pada penelitian ini akan digunakan algoritma *K-Means* untuk membuat segmentasi pengguna.

Penelitian oleh Sarker dan Salah (2019) telah menguji beberapa algoritma klasifikasi, termasuk: *ZeroR, Naive Bayes, Decision Tree, Support Vector Machines (SVM),* dan *Logistic Regression*, dalam memprediksi preferensi aplikasi *smartphone* berbasis konteks pengguna, hasil eksperimen menunjukkan bahwa *Random Forest* mencapai akurasi tertinggi (89.7% dalam *F1-Score*). Pada penelitian tersebut, *Random Forest* mengungguli algoritma lainnya karena kemampuannya menangani data multidimensi, mengurangi *overfitting*, dan mengidentifikasi fitur dominan seperti waktu dan lokasi [15]. Berdasarkan temuan ini, peneliti memilih *Random Forest* sebagai algoritma untuk memprediksi preferensi pengguna *smartphone*.

© Karya Dilindungi UU Hak Cipta
1. Dilarang menyebarluaskan dokumen tanpa izin.
2. Dilarang melakukan plagiasi.
3. Pelanggaran diberlakukan sanksi sesuai peraturan UU Hak Cipta.

Berdasarkan latar belakang tersebut, penelitian ini berjudul "**Analisis dan Prediksi Preferensi Pengguna *Smartphone* Berdasarkan Demografi dan Pola Penggunaan Aplikasi Menggunakan K-Means dan Random Forest**".

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, rumusan masalah dalam penelitian ini adalah:

1. Bagaimana melakukan analisis dan segmentasi pengguna *smartphone* berdasarkan pola penggunaan aplikasi dan durasi layar menggunakan metode *K-Means*?
2. Bagaimana membangun model prediksi preferensi pengguna *smartphone* dengan *Random Forest* berdasarkan variabel demografi dan label hasil segmentasi dengan algoritma *K-Means*?
3. Seberapa akurat model klasifikasi dalam memprediksi preferensi pengguna *smartphone* berdasarkan pola penggunaan aplikasi dan atribut demografi?

1.3 Tujuan Penelitian

Tujuan dari penelitian ini adalah:

1. Melakukan segmentasi pengguna *smartphone* berdasarkan pola penggunaan aplikasi dan durasi layar menggunakan algoritma *K-Means*.
2. Membangun model prediksi preferensi pengguna *smartphone* menggunakan algoritma *Random Forest* berdasarkan atribut demografi dan hasil segmentasi.
3. Mengevaluasi performa model prediksi preferensi pengguna menggunakan metrik evaluasi.

1.4 Manfaat Penelitian

Manfaat dari penelitian ini adalah:

1. Memberikan wawasan bagi pengembang aplikasi untuk memahami preferensi pengguna dalam merancang produk yang lebih sesuai.
2. Menyediakan informasi yang berguna bagi pelaku bisnis untuk memahami segmen pengguna secara lebih mendalam berdasarkan demografi dan pola penggunaan aplikasi sehingga personalisasi layanan dapat lebih tepat sasaran.
3. Menambah kontribusi ilmiah dalam bidang analisis data dan *machine learning* yang membahas tentang pengelompokan perilaku pengguna dan sistem rekomendasi berbasis pola penggunaan aplikasi.

© Karya Dilindungi UU Hak Cipta
1. Dilarang menyebarluaskan dokumen tanpa izin.
2. Dilarang melakukan plagiasi.
3. Pelanggaran diberlakukan sanksi sesuai peraturan UU Hak Cipta.

1.5 Ruang Lingkup Penelitian

Ruang lingkup penelitian ini mencakup:

1. Data yang digunakan merupakan *dataset* yang diambil dari situs https://www.kaggle.com/datasets/bhadramohit/smartphone-usage-and-behavioral-*dataset*, 1.000 sampel pengguna dengan 9 atribut.
2. Atribut terdiri dari:
   - Demografi terdiri dari: usia (*age*), jenis kelamin (*gender*), lokasi (*location*)
   - Pola penggunaan: total jam penggunaan aplikasi (*total_app_usage_hours*), waktu layar harian (*daily_screen_time_hours*), jumlah aplikasi (*number_of_apps_used*).
3. *Screen time* berdasarkan kategori aplikasi: penggunaan media sosial (*social_media_usage_hours*), aplikasi produktivitas (*productivity_app_usage_hours*), dan game (*gaming_apps_usage_hours*).
4. Segmentasi pengguna dengan *K-Means* berdasarkan pola penggunaan.
5. Prediksi preferensi menggunakan *Random Forest* dengan *input* hasil segmentasi dari *K-Means* serta variabel demografi.
6. Evaluasi dengan: *Accuracy, F1-Score*, *Precision*, *Recall, Confusion Matrix,* dan *ROC-AUC*.

© Karya Dilindungi UU Hak Cipta
1. Dilarang menyebarluaskan dokumen tanpa izin.
2. Dilarang melakukan plagiasi.
3. Pelanggaran diberlakukan sanksi sesuai peraturan UU Hak Cipta.