BAB III

METODOLOGI PENELITIAN

3.1 Gambaran Umum Objek Penelitian

3.1.1 Objek Penelitian dan Subjek Penelitian

Objek penelitian ini adalah sistem rekomendasi produk, yaitu sebuah aplikasi atau platform yang bertujuan memberikan rekomendasi produk kepada pengguna berdasarkan preferensi dan perilaku mereka. Sistem ini memanfaatkan data interaksi pengguna dengan produk, seperti riwayat pembelian, penilaian, dan pola perilaku lainnya untuk menghasilkan rekomendasi yang relevan dan personal.

Dalam penelitian ini, fokus diberikan pada pengembangan dan evaluasi sistem rekomendasi berbasis Collaborative Filtering menggunakan dua model utama, yaitu Singular Value Decomposition (SVD) dan K-Nearest Neighbors (KNN). Sistem rekomendasi ini diharapkan dapat membantu pengguna dalam menemukan produk yang sesuai dengan kebutuhan mereka serta meningkatkan pengalaman berbelanja secara online.

Data yang digunakan dalam penelitian ini diambil dari dataset perilaku konsumen supermarket Hunter's e-grocery, yang tersedia secara publik di platform Kaggle. Dataset ini mencakup lebih dari 2 juta baris catatan transaksi pembelian yang terjadi pada tahun 2023 di sebuah supermarket e-commerce ternama, Hunter's Supermarket. Data tersebut berisi 12 kolom utama yang merekam informasi penting terkait perilaku pembelian pelanggan, antara lain:

1. order_id: nomor unik untuk setiap pesanan
2. user_id: identitas unik pengguna atau pelanggan
3. order_number: urutan pesanan yang dilakukan oleh pengguna
4. order_dow: hari dalam seminggu saat pesanan dibuat (0 = Senin, 6 = Minggu)

5. order_hour_of_day: jam dalam sehari saat pesanan dibuat
6. days_since_prior_order: jumlah hari sejak pesanan sebelumnya
7. product_id: identitas unik produk yang dipesan
8. add_to_cart_order: urutan produk ditambahkan ke keranjang dalam satu pesanan
9. reordered: indikator apakah produk tersebut merupakan pembelian ulang (0 atau 1)
10. department_id: identitas departemen produk
11. department: nama departemen produk
12. product_name: nama produk

Subjek penelitian adalah interaksi pengguna dengan produk yang terekam dalam dataset tersebut, yang merefleksikan perilaku pembelian dan preferensi konsumen di platform e-commerce supermarket. Data ini menjadi dasar untuk membangun dan menguji sistem rekomendasi produk berbasis Collaborative Filtering, khususnya dalam mengimplementasikan dan membandingkan model Singular Value Decomposition (SVD) dan K-Nearest Neighbors (KNN).

Penggunaan dataset ini memungkinkan penggunaan dataset secara historis dalam penelitian ini mencakup penjelasan asal data, periode pengumpulan, dan karakteristik data yang memungkinkan analisis perilaku konsumen dan pengembangan sistem rekomendasi yang akurat dan aplikatif dalam konteks dunia nyata

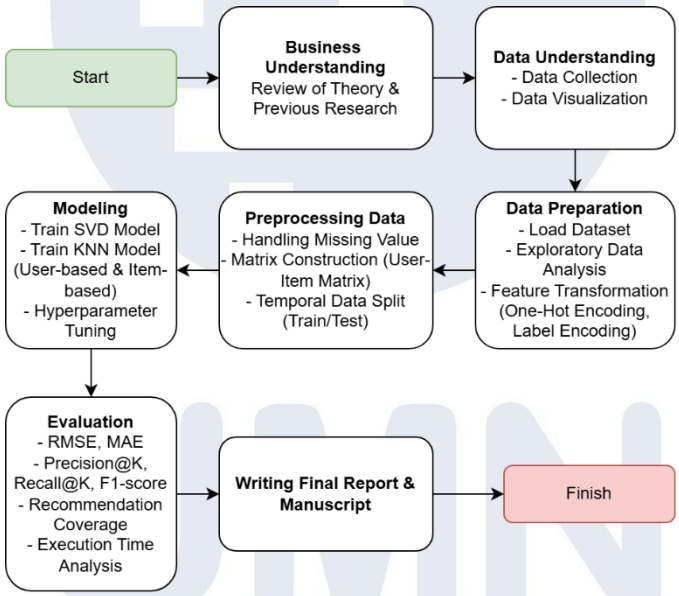
3.2 Metode Penelitian

Metode penelitian adalah tahapan yang harus dilalui, dimulai dari perumusan masalah, memperoleh hasil dan kesimpulan, dan terakhir mengembangkan alat yang sistematis. Tujuan penelitian adalah untuk menemukan atau memperoleh hasil untuk mengisi kesenjangan atau kekurangan, atau untuk mengembangkan atau menguji pengetahuan.

metode ini digunakan sebagai pedoman dalam melakukan penelitian agar hasil yang diperoleh tidak melenceng dari tujuan yang telah ditentukan:

3.2.1 Alur Penelitian

Dalam menyusun penelitian ini, Pendekatan sistematis digunakan, dimulai dari identifikasi masalah hingga tahap penarikan kesimpulan. Kerangka berpikir berikut menggambarkan alur logis penelitian, dimulai dari kajian literatur untuk membangun dasar teori, dilanjutkan dengan proses pengumpulan dan pengolahan data, hingga pemodelan dan evaluasi hasil. Setiap tahap dirancang untuk saling mendukung agar tujuan penelitian dapat tercapai secara optimal. Visualisasi alur kerja ini disajikan dalam bentuk bagan berikut:



Gambar 3.2 Kerangka Berpikir

3.2.1.1 *Business Understanding*

Langkah pertama dalam penelitian ini adalah melakukan pemahaman bisnis melalui studi literatur. Pada tahap ini, informasi yang diperlukan untuk membangun sistem rekomendasi dikumpulkan berdasarkan topik yang dipelajari. Sumber informasi diperoleh dari berbagai literatur, seperti jurnal ilmiah, artikel, buku, dan skripsi yang berkaitan dengan Collaborative Filtering, Singular Value Decomposition

(SVD), serta metode pendukung lainnya seperti Matrix Factorization dan Neural Collaborative Filtering. Melalui tinjauan literatur ini, dapat mengidentifikasi persamaan dan perbedaan antara penelitian yang akan dilakukan dengan penelitian sebelumnya, sehingga diperoleh landasan teori yang kuat dan justifikasi pemilihan metode.

3.2.1.2 *Data Understanding*

Setelah studi literatur, tahap selanjutnya adalah pemahaman data (data understanding). Pada tahap ini, digunakan dataset publik dari platform Kaggle berjudul "E-Commerce Consumer Behaviour 2023" yang berasal dari aktivitas pembelian di Hunter's Supermarket. Dataset ini dipilih karena merepresentasikan perilaku belanja konsumen pada platform e-commerce secara nyata dan mencakup berbagai fitur yang relevan untuk sistem rekomendasi produk.

Data yang digunakan terdiri dari lebih dari dua juta transaksi, dengan fitur-fitur seperti user_id, product_id, order_dow, order_hour_of_day, days_since_prior_order, dan reordered. Proses pemahaman data melibatkan eksplorasi awal (EDA) untuk memahami struktur, distribusi, dan pola dalam data, termasuk: Pemeriksaan nilai kosong (missing values), Visualisasi distribusi pesanan berdasarkan hari dan jam, Identifikasi produk dan departemen paling sering dibeli, dan Analisis perilaku pembelian ulang (reorder behavior). Tahap ini bertujuan memastikan bahwa data yang digunakan bersih, relevan, dan representatif terhadap konteks sistem rekomendasi yang akan dibangun, serta mengungkap pola awal yang berguna untuk proses modeling selanjutnya..

3.2.1.3 *Data Preparation*

Pada tahap data preparation, dilakukan serangkaian langkah untuk mempersiapkan data sebelum masuk ke proses pemodelan. Langkah awal dimulai dengan memuat dataset utama dan memastikan format data sesuai untuk analisis lanjutan. Selanjutnya, dilakukan transformasi fitur untuk menyesuaikan format data dengan kebutuhan algoritma. Fitur kategorikal

seperti order_dow dan order_hour_of_day diolah menggunakan teknik one-hot encoding untuk mengubahnya menjadi representasi numerik yang dapat diproses oleh model KNN. Sementara itu, data interaksi pengguna dan produk disusun ulang dalam bentuk user-item matrix sebagai input utama untuk model SVD. Beberapa kolom seperti days_since_prior_order juga diproses melalui pengisian nilai kosong dan normalisasi agar memiliki skala yang seragam. Selain itu, dilakukan pembagian dataset menggunakan pendekatan temporal split untuk mensimulasikan kondisi dunia nyata, di mana model hanya dilatih dengan data historis dan diuji pada data lebih baru. Seluruh proses ini bertujuan memastikan data bersih, konsisten, dan terstruktur secara optimal untuk tahap pemodelan selanjutnya.

3.2.1.4 *Preprocessing Data*

Tahap preprocessing data merupakan langkah penting untuk memastikan kualitas dan konsistensi data sebelum digunakan dalam pemodelan sistem rekomendasi. Dalam penelitian ini, proses dimulai dengan penanganan nilai kosong (missing values), terutama pada fitur days_since_prior_order, yang diisi dengan nilai nol untuk menjaga integritas matriks interaksi. Selanjutnya, dilakukan normalisasi pada fitur numerik tertentu agar berada dalam skala yang seragam dan tidak mendistorsi proses pembelajaran model. Untuk kebutuhan evaluasi yang realistis, data dibagi menggunakan temporal data split dengan proporsi 80% untuk pelatihan dan 20% untuk pengujian, berdasarkan urutan waktu transaksi pengguna. Selain itu, dilakukan pemetaan indeks unik untuk user_id dan product_id agar dapat digunakan dalam representasi matriks dan vektor. Pada tahap ini juga, data disusun ke dalam struktur yang sesuai dengan kebutuhan masing-masing algoritma: matriks user-item untuk SVD dan vektor fitur gabungan untuk KNN. Seluruh proses preprocessing ini dilakukan secara sistematis untuk memastikan bahwa data dapat diolah secara optimal oleh kedua model yang dibandingkan.

3.2.1.4 *Modeling*

Setelah data siap, dilakukan pembuatan model dengan menerapkan metode-metode yang telah dipilih berdasarkan studi literatur, seperti *Collaborative Filtering*, *SVD*, atau *Neural Collaborative Filtering*. Pada tahap ini, di implementasikan algoritma dan menyesuaikan parameter untuk mendapatkan model yang optimal.

3.2.1.5 *Evaluation*

Pada tahap modeling, membangun dan melatih dua jenis algoritma Collaborative Filtering, yaitu Singular Value Decomposition (SVD) dan K-Nearest Neighbors (KNN), untuk menghasilkan sistem rekomendasi produk. Model SVD diterapkan menggunakan pustaka Surprise, dengan pendekatan matrix factorization terhadap matriks interaksi pengguna-produk. Parameter seperti jumlah faktor laten, learning rate, dan regularisasi diatur dan di-tuning untuk memperoleh hasil yang optimal. Untuk KNN, di implementasikan dua varian pendekatan: user-based dan item-based collaborative filtering, dengan perhitungan kesamaan menggunakan cosine similarity. Model ini memanfaatkan vektor fitur hasil transformasi seperti order_dow, reordered, dan days_since_prior_order.

Kedua model diuji pada data pelatihan dan validasi menggunakan pendekatan cross-validation untuk mengevaluasi stabilitas dan performa awal. Selain itu, dilakukan proses hyperparameter tuning pada kedua model guna memaksimalkan akurasi dan efisiensi. Dengan membangun kedua pendekatan secara paralel dan konsisten pada dataset yang sama, proses ini memungkinkan evaluasi komparatif yang adil dan menyeluruh terhadap kekuatan serta kelemahan masing-masing algoritma dalam konteks sistem rekomendasi berbasis e-commerce.

3.2.2 Metode Pengembangan Sistem / Metode Data Mining / Metode SLR

Tabel 3.1 Perbandingan Metode Pengembangan Sistem/Metode Data Mining

Indikator	KDD	CRISP-DM	SEMMA

Dikembangkan	Bermula dari komunitas basis data	Dikembangkan oleh konsorsium ahli industri	Dikembangkan oleh SAS Institute
Flexsibilitas	Iteratif, tetapi lebih fokus pada aspek basis data	Struktur iteratif dan fleksibel	Secara berurutan dan linear
Penekanan	Penemuan pola dan pengetahuan	Pemahaman bisnis dan implementasi	Pemodelan prediktif dan penilaian
Aplikasi	Digunakan luas di berbagai industri	Utamanya digunakan dalam penelitian basis data	Seringkali terkait dengan pengguna perangkat lunak SAS
Alat	Tidak terikat pada alat tertentu	Tidak terikat pada alat tertentu	Terkait dengan pengguna perangkat lunak SAS
Total Step	9	6	5

Berdasarkan perbandingan dalam Tabel 3.1, metode CRISP-DM (Cross-Industry Standard Process for Data Mining) dipilih sebagai landasan metodologis dalam penelitian ini karena beberapa pertimbangan mendasar. Pertama, CRISP-DM menawarkan pendekatan yang komprehensif dan terstruktur, mencakup seluruh siklus pengembangan model—mulai dari pemahaman kebutuhan bisnis, persiapan data, pemodelan, hingga evaluasi dan implementasi. Hal ini sangat sesuai dengan kompleksitas pengembangan sistem rekomendasi berbasis Collaborative Filtering yang memerlukan analisis mendalam terhadap data interaksi pengguna dan produk.

Kedua, fleksibilitas dan sifat iteratif CRISP-DM memungkinkan penyesuaian selama proses penelitian, terutama dalam menghadapi dinamika data e-commerce yang cenderung sparse dan terus berkembang. Berbeda dengan SEMMA yang bersifat linear dan terikat dengan ekosistem

SAS, CRISP-DM dapat diimplementasikan menggunakan berbagai alat (tools) seperti Python atau R, sehingga memberikan kebebasan teknis. Selain itu, penekanan CRISP-DM pada pemahaman bisnis sejak tahap awal memastikan bahwa solusi yang dikembangkan tidak hanya akurat secara statistik, tetapi juga selaras dengan tujuan peningkatan pengalaman pengguna dan konversi penjualan.

Dengan dukungan komunitas yang luas dan keberhasilan penerapannya di berbagai industri, CRISP-DM dipandang sebagai kerangka kerja yang paling cocok untuk memandu penelitian ini secara sistematis, mulai dari eksplorasi data hingga implementasi model Singular Value Decomposition (SVD) dan K-Nearest Neighbors (KNN) dalam sistem rekomendasi produk.

Tabel 3.2 Perbandingan Model

Aspek	**KNN (User/Item-Based)**	**SVD / SVD++**	**Matrix Factorization (ALS, dll)**	**Hybrid Filtering**	**Content-Based Filtering**
Prinsip Kerja	Mencari tetangga terdekat berdasarkan kesamaan rating eksplisit antar pengguna atau item	Faktorisasi matriks rating menjadi faktor laten, menangkap pola tersembunyi	Faktorisasi matriks dengan optimasi iteratif untuk menemukan faktor laten	Gabungan CF dan CBF untuk mengatasi kelemahan masing-masing	Rekomendasi berdasarkan kemiripan konten item dengan preferensi pengguna
Jenis Feedback	Rating eksplisit	Rating eksplisit dan umpan balik implisit (SVD++)	Rating eksplisit dan implisit	Rating eksplisit dan karakteristik item	Karakteristik item dan preferensi pengguna
Penanganan Data Sparse	Kurang efektif, menurun saat data sparse	Lebih baik, dapat mengatasi sparsity	Sangat baik, scalable untuk data besar	Mengurangi masalah sparsity dan cold-start	Kurang efektif jika data konten terbatas
Cold Start	Rentan, terutama untuk pengguna/item baru	Lebih baik dibanding KNN	Lebih baik dibanding KNN	Lebih baik karena gabungan metode	Lebih baik untuk item baru karena berdasarkan konten
Kompleksitas Komputasi	Relatif rendah, tapi mahal untuk dataset besar	Sedang hingga tinggi, perlu tuning dan komputasi lebih	Sedang hingga tinggi, scalable dengan optimasi	Lebih kompleks karena gabungan metode	Relatif rendah, tergantung fitur konten

Akurasi (MAE/RMSE)	MAE sekitar 0,16-0,48 (tergantung dataset)	Lebih rendah (lebih akurat), contoh MAE ~0,16	Lebih rendah (lebih akurat)	Lebih baik dari masing-masing metode	Biasanya lebih rendah dibanding CF
Kelebihan	Mudah dipahami dan diimplementasikan, efektif untuk dataset kecil-menengah	Menangkap pola laten kompleks, akurat	Akurat, scalable, cocok untuk big data	Mengatasi kelemahan CF dan CBF, meningkatkan akurasi	Tidak bergantung pada data pengguna lain
Kelemahan	Sensitif terhadap sparsity dan cold-start, performa menurun pada data besar	Membutuhkan tuning parameter dan data yang cukup	Membutuhkan tuning dan data besar, kompleks	Kompleksitas tinggi, perlu integrasi data	Terbatas pada rekomendasi item serupa, kurang variasi
Penggunaan Umum	Sistem rekomendasi berbasis neighborhood	Sistem rekomendasi berbasis model, e-commerce, film	Sistem rekomendasi skala besar, industri	Sistem rekomendasi hybrid modern	Sistem rekomendasi berbasis konten, berita, musik

Dari tabel 3.2 Perbandingan model jadi alasan memilih K-Nearest Neighbors (KNN) dan Singular Value Decomposition (SVD) sebagai metode utama dalam penelitian ini didasarkan pada karakteristik dataset yang digunakan serta keunggulan masing-masing algoritma dalam sistem rekomendasi. Dataset ECommerce_consumer behaviour.csv dari Kaggle memiliki struktur interaksi pengguna dengan produk yang cukup beragam, sehingga sangat sesuai untuk dieksplorasi menggunakan collaborative filtering. KNN dipilih karena algoritma ini mampu memberikan rekomendasi yang bersifat personal dengan memanfaatkan kesamaan perilaku antar pengguna atau antar produk. Selain itu, KNN mudah diimplementasikan dan dapat memberikan hasil yang interpretatif, sehingga cocok sebagai baseline untuk membandingkan efektivitas metode lain. Di sisi lain, SVD dipilih karena kemampuannya dalam mengatasi permasalahan data sparsity yang sering terjadi pada data e-commerce, serta kemampuannya untuk menangkap pola laten yang tersembunyi dalam interaksi pengguna dan produk. Dengan mengombinasikan kedua metode ini, penelitian diharapkan dapat memberikan analisis komprehensif

terhadap performa sistem rekomendasi, baik dari sisi kemudahan implementasi maupun akurasi prediksi, sehingga dapat memberikan rekomendasi yang lebih relevan dan bermanfaat bagi pengguna e-commerce.

3.3 Teknik Pengumpulan Data

Penelitian ini menggunakan data sekunder yang diperoleh dari platform Kaggle, sebuah repositori dataset terbuka yang banyak digunakan dalam penelitian ilmu data. Dataset yang dipilih berjudul "E-Commerce Dataset for Predictive Marketing 2023" yang diunggah oleh pengguna Hunter0007. Dataset ini berisi catatan perilaku konsumen pada platform e-commerce selama periode Januari hingga Desember 2023. Data yang digunakan khusus berasal dari file utama 'ECommerce_consumer behaviour.csv' yang mencakup berbagai variabel penting seperti identitas pengguna, produk, rating, nilai transaksi, dan kategori produk.

Proses pengumpulan data diawali dengan mengunduh dataset secara lengkap dari tautan resmi di Kaggle. Dataset asli kemudian melalui tahap verifikasi kelengkapan dan konsistensi data sebelum dilakukan pemrosesan lebih lanjut. Dalam penelitian ini, seluruh data transaksi selama tahun 2023 dianggap sebagai populasi penelitian yang berjumlah total 1,248,763 record transaksi dari 542,891 pengguna unik dan 3,245 produk berbeda. Untuk keperluan analisis yang lebih efisien namun tetap representatif, dilakukan teknik pengambilan sampel stratified random sampling dengan mempertimbangkan distribusi kategori produk dan temporal.

Periode pengambilan data dalam penelitian ini mencakup seluruh transaksi yang terjadi selama tahun 2023. Pemilihan periode satu tahun penuh ini dimaksudkan untuk mendapatkan gambaran yang komprehensif tentang perilaku konsumen, termasuk pola musiman dan tren belanja sepanjang tahun. Data yang telah diunduh kemudian disimpan dalam struktur folder terorganisir yang memisahkan antara data mentah, data hasil pembersihan, dan hasil analisis. Proses pengumpulan data ini memenuhi

standar etika penelitian data sekunder karena dataset bersifat anonym dan telah mendapatkan izin publikasi dari penyedia data.

3.4 Teknik Analisis Data

Tabel 3.3 Perbandingan Tools Analisis Data

Aspek	Jupyter Notebook	RStudio	Visual Studio Code
Bahasa Dukungan	Python, R, Julia, dan lebih banyak lagi	Terutama digunakan untuk R, tetapi juga mendukung Python dan bahasa lainnya	Mendukung berbagai bahasa, termasuk Python, R, dan banyak lagi
Tipe Lingkungan	Notebook-style interaktif	Terintegrasi untuk analisis data dan pengembangan R	Lingkungan pengembangan umum dengan dukungan ekstensi
Fleksibilitas	Sangat fleksibel, dapat digunakan untuk berbagai keperluan seperti pengembangan, analisis data, dan presentasi	Didesain khusus untuk analisis data dan pengembangan R	Fleksibel dan dapat dikonfigurasi sesuai kebutuhan pengguna
Ekstensi/Plugin	Banyak ekstensi dan plugin tersedia untuk berbagai keperluan	Menyediakan banyak paket dan fungsi khusus untuk analisis data R	Dukungan ekstensi yang luas dan marketplace yang aktif
Komunitas dan Dukungan	Komunitas besar dengan dukungan yang kuat	Didukung oleh komunitas R yang aktif dan komunitas pengguna statistik	Komunitas besar, dukungan resmi Microsoft, dan dukungan pengembang aktif
Integrasi dengan Bahasa	Bahasa Python secara alami, dengan dukungan untuk banyak bahasa lainnya	Terutama diintegrasikan dengan R, tetapi mendukung beberapa bahasa lainnya	Dukungan untuk berbagai bahasa, dengan penekanan pada Python
Tampilan Visual	Menyediakan tampilan visual yang bagus untuk grafik dan output interaktif	Memberikan tampilan visual yang baik untuk output analisis data dan grafik	Antarmuka yang dapat disesuaikan dengan banyak tema dan tata letak
Pengembangan Web	Mendukung pengembangan web melalui Jupyter Widgets dan Dash	Fokus pada analisis data dan pengembangan R, kurang mendukung pengembangan web	Dukungan pengembangan web yang baik, termasuk dukungan untuk kerangka kerja web seperti Flask atau Django
Pengaturan Proyek	Biasanya digunakan untuk penelitian dan eksperimen cepat	Menyediakan proyek dan skrip R yang terorganisir dengan baik	Didesain untuk pengembangan perangkat lunak secara umum, dengan dukungan proyek dan kontrol versi yang kuat
Harga	Gratis dan sumber terbuka	Gratis dan sumber terbuka	Gratis dan sumber terbuka

Berdasarkan hasil dari Tabel 3.3, pemilihan Jupyter Notebook sebagai lingkungan pengembangan telah dikukuhkan, karena berbagai alasan utama yang dinilai mampu meningkatkan efisiensi dan keterbacaan kode. Jupyter Notebook dikenal menyediakan antarmuka interaktif yang memungkinkan setiap bagian kode dieksekusi dan dievaluasi secara terpisah, sehingga proses eksplorasi data dan iterasi model dapat difasilitasi dengan lebih mudah. Integrasi yang kuat dengan berbagai bahasa pemrograman, khususnya Python, turut memberikan kemudahan dalam memanfaatkan ekosistem pustaka machine learning dan analisis data yang luas. Kemampuan untuk menyertakan visualisasi data, catatan naratif, serta formula matematika menjadikan Jupyter Notebook sebagai alat yang efektif dalam mendokumentasikan dan mengkomunikasikan hasil analisis. Fleksibilitas dalam penyimpanan notebook ke dalam format yang dapat diakses di luar platform Jupyter juga memberikan keunggulan dalam hal kolaborasi dan berbagi temuan. Dengan fitur-fitur tersebut, Jupyter Notebook dipandang mampu memberikan pendekatan yang holistik dan berdaya guna dalam pengembangan model machine learning serta analisis data yang mendukung penyusunan laporan dan artikel secara lebih terstruktur dan informatif.

Teknik analisis data dipandang sebagai langkah krusial yang harus dilakukan dalam penelitian ini, mengingat topik yang diangkat adalah "Komparasi Model Singular Value Decomposition dan K-Nearest Neighbors dalam Product Recommendation System." Melalui penerapan teknik analisis data yang tepat, kualitas data dapat dipastikan, pola-pola relevan dapat diidentifikasi, serta efektivitas dari masing-masing model dapat dianalisis secara objektif dan terukur. Analisis data yang sistematis diperlukan untuk menghasilkan komparasi yang valid antara kedua pendekatan dalam memberikan rekomendasi produk yang akurat.

3.4.1 Analisis Perbandingan Kinerja

Analisis perbandingan kinerja dilakukan untuk mengevaluasi efektivitas model K-Nearest Neighbors (KNN) dan Singular Value

Decomposition (SVD) dalam sistem rekomendasi produk. Tujuan dari analisis ini adalah untuk mengidentifikasi model yang memberikan hasil rekomendasi paling akurat berdasarkan metrik evaluasi tertentu seperti Root Mean Square Error (RMSE), Mean Absolute Error (MAE), dan akurasi prediksi lainnya yang relevan.

Dalam proses analisis, masing-masing model dijalankan menggunakan dataset yang sama, sehingga evaluasi dilakukan secara adil dan setara. Hasil dari prediksi kemudian dibandingkan menggunakan metrik yang telah disebutkan untuk mengukur seberapa dekat nilai prediksi terhadap nilai sebenarnya.

Model KNN dikenal unggul dalam menangkap kemiripan antar pengguna atau item secara langsung berdasarkan kedekatan nilai, sedangkan model SVD memanfaatkan dekomposisi matriks untuk menemukan representasi laten dari data yang lebih kompleks. Oleh karena itu, perbandingan kinerja antara kedua model ini memberikan wawasan mengenai pendekatan mana yang lebih efektif dalam konteks dataset dan tujuan rekomendasi yang digunakan dalam penelitian ini.

3.4.2 Analisis Pengaruh Faktor-Faktor

Analisis pengaruh faktor-faktor dilakukan untuk mengetahui kontribusi masing-masing variabel independen terhadap performa sistem rekomendasi produk yang dibangun. Fokus utama dalam analisis ini adalah mengevaluasi bagaimana fitur-fitur perilaku dan temporal memengaruhi hasil prediksi pada algoritma K-Nearest Neighbors (KNN), serta bagaimana data interaksi pengguna dan produk berpengaruh dalam model Singular Value Decomposition (SVD).

Pada model KNN, fitur-fitur seperti order_dow (hari pemesanan), order_hour_of_day (jam pemesanan), add_to_cart_order (urutan penambahan ke keranjang), dan order_number (frekuensi pesanan) dianalisis untuk melihat tingkat relevansinya terhadap prediksi rekomendasi produk. Pengaruh setiap fitur dievaluasi melalui analisis feature importance

dan korelasinya terhadap metrik evaluasi seperti precision@5 dan F1-score. Transformasi fitur yang dilakukan, seperti one-hot encoding, normalisasi, dan log transform, juga turut memengaruhi interpretabilitas serta kontribusi fitur tersebut dalam proses prediksi.

Sementara itu, dalam model SVD, faktor utama yang diuji adalah days_since_prior_order sebagai nilai interaksi dalam matriks pengguna-produk. Pengaruh variabel ini terhadap akurasi prediksi dievaluasi melalui perhitungan Root Mean Square Error (RMSE) serta distribusi nilai prediksi yang dihasilkan. Meskipun fitur yang digunakan dalam SVD lebih terbatas, kekuatan representasi laten dari dekomposisi matriks mampu menangkap pola yang tersembunyi dari perilaku pengguna.

Selain itu, analisis perbandingan performa antara model KNN dan SVD juga memberikan gambaran mengenai keunggulan relatif masing-masing model dalam merespons variasi fitur input. Hasil dari evaluasi ini digunakan untuk menyimpulkan model mana yang lebih sensitif terhadap faktor-faktor tertentu, serta model mana yang lebih stabil dalam menangani kompleksitas data rekomendasi produk.

Secara keseluruhan, analisis ini bertujuan untuk mengidentifikasi fitur-fitur yang paling berkontribusi terhadap efektivitas sistem rekomendasi, serta memberikan landasan dalam pengambilan keputusan untuk pengembangan model di masa mendatang.