# BAB II

# LANDASAN TEORI

## 2.1 Penelitian Terdahulu

Tabel 2.1 Tabel Penelitian Terdahulu

| No | Nama Jurnal | Jurnal (vol no brp) | Tahun | Penulis | Hasil |
|---|---|---|---|---|---|
| 1 | Analisis Metode Collaborative Filtering menggunakan KNN dan SVD++ untuk Rekomendasi Produk E-commerce Tokopedia | *Edumatic: Jurnal Pendidikan Informatika*, Vol. 8, No. 1 | 2024 | Chalista Yulia Hazizah, Triyanna Widiyaningtyas | SVD++ lebih unggul dari KNN dalam akurasi (MAE: 0,161176 vs 0,163964; RMSE: 0,185252 vs 0,197045). SVD++ lebih efektif menangani data sparsity dan feedback implisit, serta memberikan rekomendasi yang lebih relevan pada e-commerce.[1][2] |
| 2 | Sistem Rekomendasi Dengan Teknik Collaborative Filtering Menggunakan Faktorisasi Matriks dan KNN Untuk Produk Elektronik | Institut Teknologi Sepuluh Nopember (ITS) | 2023 | Muhammad Fadli | SVD sebagai teknik matrix factorization menghasilkan prediksi rating dan rekomendasi produk elektronik yang lebih baik dibanding KNN, terutama pada dataset besar. KNN kurang efisien untuk data besar.[5] |
| 3 | Extended Collaborative Filtering Recommendation System with Adaptive KNN and SVD | SSRN Electronic Journal | 2023 | Y. Gu, Z. Ding, S. Wang, D. Yin | Sistem rekomendasi berbasis SVD dan adaptive KNN meningkatkan akurasi dan personalisasi pada e-commerce. SVD efektif untuk data sparse, sementara adaptive KNN membantu pada data padat.[8] |

| | | | | | |
|---|---|---|---|---|---|
| 4 | Pemanfaatan Metode Collaborative Filtering dengan Algoritma KNN dalam Sistem Rekomendasi Produk E-commerce | *Mind Journal: Jurnal Psikologi dan Informatika*, Vol. 9, No. 2 | 2023 | Rahmat Hidayat | KNN dapat digunakan untuk sistem rekomendasi produk e-commerce, namun performanya menurun pada data besar dan sparse, sehingga disarankan penggunaan matrix factorization seperti SVD untuk hasil lebih baik. [3] |
| 5 | Performance Comparison and Analysis of SVD and ALS in Recommendation System | Applied and Computational Engineering, Vol. 49 | 2024 | Tianyi Zhao | SVD menunjukkan akurasi lebih tinggi daripada ALS pada data padat dan kecil. Penelitian menekankan pentingnya pemilihan algoritma berdasarkan karakteristik dataset. |
| 6 | Hybrid-based Food Recommender System Utilizing KNN and SVD Approaches | Cogent Engineering, Vol. 11 | 2024 | Yap, Zhi-Toung; Haw, Su-Cheng; Ruslan, Nur | Sistem rekomendasi makanan berbasis hybrid yang menggabungkan pendekatan KNN dan SVD berhasil meningkatkan akurasi serta keberagaman rekomendasi. Dengan menggabungkan Collaborative Filtering dan Content-Based Filtering, serta penerapan mekanisme pembelajaran berkelanjutan, |

| | | | | | |
|---|---|---|---|---|---|
| | | | | | sistem mampu menyesuaikan preferensi pengguna secara dinamis dari waktu ke waktu, meningkatkan relevansi rekomendasi dan kepuasan pengguna. |

Berdasarkan Tabel 2.1, enam penelitian yang disebutkan menunjukkan bahwa berbagai studi telah dilakukan untuk membandingkan efektivitas algoritma Singular Value Decomposition (SVD) dan K-Nearest Neighbors (KNN) dalam sistem rekomendasi berbasis Collaborative Filtering.

Secara umum, hasil-hasil penelitian tersebut menunjukkan konsistensi dalam mengidentifikasi keunggulan SVD dalam hal akurasi, efisiensi komputasi, serta kemampuannya menangani data sparse, terutama pada platform e-commerce berskala besar. Hazizah dan Widiyaningtyas menemukan bahwa algoritma SVD++ secara signifikan lebih unggul dibanding KNN dalam hal MAE dan RMSE pada dataset Tokopedia. Temuan ini diperkuat oleh penelitian Fadli , yang menunjukkan bahwa pendekatan matrix factorization seperti SVD memberikan hasil prediksi lebih akurat pada data produk elektronik dalam skala besar. Gu et al. memperluas pendekatan ini dengan menggabungkan adaptive KNN dan SVD, menghasilkan sistem rekomendasi yang lebih adaptif dan personal. Hidayat menyoroti performa KNN yang menurun pada dataset besar dan sparse, dan merekomendasikan penggunaan SVD sebagai alternatif yang lebih stabil. Penelitian Zhao menambahkan perspektif dari konteks internasional, menunjukkan bahwa SVD secara signifikan mengungguli algoritma Alternating Least Squares (ALS) pada dataset Amazon Electronics, terutama pada data yang kecil dan padat. Terakhir, penelitian Yap et al. mengusulkan sistem hybrid berbasis KNN dan SVD dalam konteks rekomendasi makanan, dengan pendekatan pembelajaran adaptif

yang terbukti mampu menyesuaikan rekomendasi terhadap perubahan preferensi pengguna secara berkelanjutan.

Gap penelitian terletak pada terbatasnya studi komparatif yang secara langsung membandingkan SVD dan KNN dalam konteks real-world dataset yang bersifat terbuka, kompleks, dan relevan dengan sistem rekomendasi produk berbasis personalisasi. Meskipun berbagai penelitian terdahulu telah menunjukkan keunggulan masing-masing algoritma, sebagian besar hanya menguji model pada satu jenis domain atau dengan pendekatan non-tuning. Selain itu, masih sedikit penelitian yang membahas secara simultan metrik evaluasi regresi (RMSE, MAE) dan klasifikasi (precision, recall, F1-score) dalam membandingkan kedua model secara menyeluruh. Oleh karena itu, penelitian ini bertujuan untuk mengisi kekosongan tersebut dengan melakukan evaluasi komprehensif terhadap model SVD dan KNN yang telah dituning pada sistem rekomendasi produk, sehingga dapat memberikan panduan yang lebih praktis dan terukur dalam pengembangan Personalization Recommender System.

Secara keseluruhan, keterkaitan antar penelitian tersebut mengindikasikan bahwa SVD menunjukkan keunggulan yang konsisten dalam hal akurasi prediksi dan efisiensi komputasi, terutama ketika diterapkan pada data yang besar dan bersifat sparse. Di sisi lain, algoritma KNN tetap relevan untuk digunakan pada skenario dengan skala data kecil hingga menengah, namun performanya cenderung menurun secara signifikan ketika diterapkan pada dataset yang lebih besar atau kompleks. Menariknya, pendekatan hybrid yang menggabungkan kekuatan SVD dan adaptive KNN telah terbukti menjadi strategi yang menjanjikan untuk meningkatkan personalisasi serta ketahanan model terhadap variasi karakteristik data. Berdasarkan sintesis dari berbagai studi terdahulu, dapat disimpulkan bahwa evaluasi komparatif antara SVD dan KNN dengan mempertimbangkan berbagai metrik performa pada dataset nyata e-commerce merupakan langkah krusial dalam mendukung pemilihan model

Collaborative Filtering yang paling tepat guna dalam pengembangan sistem rekomendasi berbasis personalisasi.

## 2.2 Teori tentang Topik Skripsi

### 2.2.1 Sistem Rekomendasi

Personal Recommender System (PRS) adalah sistem yang digunakan untuk merekomendasikan objek-objek yang mungkin menarik bagi pengguna berdasarkan preferensi dan kebiasaan mereka. PRS berperan penting dalam kehidupan sosial, ekonomi, dan analisis ilmiah. Sistem ini membangun jembatan antara pengguna dan objek-objek yang ada dengan membangun kesamaan antara mereka. Dengan demikian, PRS dapat membantu pengguna menemukan objek-objek yang sesuai dengan minat mereka [1].

Sistem rekomendasi yang modern saat ini semakin mengandalkan informasi kontekstual seperti lokasi, waktu, dan aktivitas pengguna untuk menghasilkan rekomendasi yang lebih relevan dan sesuai dengan kebutuhan individu [10]. Sebagai contoh konkret, ketika mahasiswa sarjana mencari informasi tentang "Metode fuzzy" untuk tugas kelasnya, rekomendasi yang akan mereka terima mungkin berbeda dengan apa yang direkomendasikan kepada mahasiswa pascasarjana yang sedang menulis makalah penelitian dengan topik yang sama. Perbedaan ini dapat terjadi karena persyaratan tugas yang berbeda serta perbedaan dalam jenjang pendidikan formal yang menjadi faktor kontekstual yang signifikan [11].

Penting untuk dicatat bahwa informasi kontekstual menjadi aspek kunci dalam menjaga keakuratan rekomendasi yang diberikan oleh sistem [12][13]. Sistem rekomendasi yang dianggap handal sering kali menerapkan pendekatan yang mempertimbangkan konteks dalam memberikan rekomendasi kepada pengguna berdasarkan situasi mereka saat ini atau faktor-faktor tertentu yang berlaku [14][15] Selama dekade terakhir, terdapat peningkatan signifikan dalam penelitian dan implementasi sistem

rekomendasi, yang telah berhasil diterapkan dalam berbagai bidang aplikasi seperti manajemen pengetahuan, e-commerce, e-learning, dan e-health [16].

Sistem rekomendasi ini membuktikan nilai pentingnya dalam memberikan saran kepada pengguna untuk membantu mereka mengambil keputusan yang lebih baik, seperti menentukan tempat yang ingin dikunjungi, memilih film untuk ditonton, atau menambahkan teman ke dalam jejaring sosial mereka. Dalam lingkungan perangkat bergerak, terdapat banyak faktor kontekstual yang dapat diintegrasikan ke dalam proses rekomendasi, seperti faktor-faktor pribadi, sosial, dan lingkungan, sehingga memungkinkan rekomendasi yang lebih akurat dan sesuai dengan situasi emosional atau aktivitas saat ini, serta sejarah pengguna [17] Dengan mempertimbangkan semua faktor ini, sistem rekomendasi dapat memberikan rekomendasi yang lebih relevan pada waktu yang tepat dan di lokasi yang sesuai untuk memenuhi kebutuhan pengguna.

Sistem rekomendasi terdiri dari perangkat lunak atau alat dan teknik perangkat lunak yang menyarankan item yang berguna bagi pengguna [18]. Menurut Prasetya, sistem rekomendasi juga dapat menginstruksikan pengguna untuk memilih produk sesuai dengan kebutuhan pengguna [19]. Sistem rekomendasi adalah jenis aplikasi yang menyelidiki situasi dan keinginan pengguna. Oleh karena itu, sistem pemberi rekomendasi memerlukan jenis rekomendasi yang tepat, akurat dan memenuhi keinginan pengguna [20]. Berkaitan dengan sistem rekomendasi yang umum digunakan. Sistem rekomendasi dikategorikan menjadi beberapa jenis yaitu [21]:

1. Berbasis konten: Pembuatan profil pengguna
2. Pemfilteran kolaboratif: menggunakan peringkat
3. Berbasis pengetahuan: menggunakan pola pengetahuan
4. Berbasis hibrid: kombinasi dua metode

### 2.2.3 Accuracy

Accuracy dalam konteks Personal Recommender System (PRS) mengacu pada sejauh mana sistem dapat memberikan rekomendasi yang sesuai dengan preferensi pengguna. Landasan teori ini penting karena tujuan utama dari PRS adalah untuk memberikan rekomendasi yang akurat dan relevan kepada pengguna.

Dalam penelitian yang dilakukan oleh Zhou et al. [1], mereka mengukur akurasi PRS dengan menggunakan skor peringkat (ranking score). Skor peringkat ini menggambarkan sejauh mana rekomendasi yang diberikan oleh sistem sesuai dengan preferensi pengguna. Hasil penelitian menunjukkan bahwa dengan mempertimbangkan korelasi tingkat tinggi antara objek-objek yang direkomendasikan, akurasi PRS dapat ditingkatkan hingga 23% dalam kasus optimal.

### 2.2.4 E-commerce

*E-commerce* telah mengubah cara orang berbelanja dan berbisnis. Dengan adanya *e-commerce*, konsumen dapat dengan mudah mencari dan membeli produk atau layanan dari berbagai penjual di seluruh dunia tanpa harus pergi ke toko fisik. Di sisi lain, perusahaan dapat memperluas jangkauan pasar mereka dan menjual produk mereka kepada pelanggan di berbagai lokasi.

*E-commerce* juga memberikan kemudahan dan kenyamanan bagi konsumen dengan adanya fitur seperti pembayaran *online*, pengiriman barang, dan layanan pelanggan yang tersedia 24 jam. Selain itu, *e-commerce* juga memberikan peluang bagi pelaku bisnis untuk mengumpulkan data pelanggan dan menerapkan strategi pemasaran yang lebih personal melalui penggunaan *Personal Recommender System* (PRS) [1].

### 2.2.5 Sampel Dan Populasi

Pernyataan Sugiyono [28] tentang populasi adalah bahwa populasi adalah suatu wilayah yang digeneralisasikan yang meliputi obyek atau

subjek yang mempunyai nilai dan ciri tertentu, dan digunakan untuk mempelajari dan menarik kesimpulan. Saya tekankan. Lebih lanjut Anshori & Iswati (2011) menegaskan bahwa populasi tidak hanya berarti jumlah subjek penelitian saja, namun mencakup seluruh ciri dan sifat subjek tersebut.

Dalam konteks penelitian ini, populasi yang dimaksud adalah pengguna e-commerce pada platform tokopedia. Dengan mengambil pandangan dari Sugiyono dan Anshori & Iswati, populasi penelitian ini mencakup tidak hanya jumlah objek penelitian, tetapi juga semua karakteristik atau kualitas yang dimiliki oleh objek tersebut.

Dalam konteks penelitian ini, sampel yang diambil juga mencerminkan representasi dari populasi yang sedang diteliti. Pengambilan jumlah sampel yang layak harus mempertimbangkan keadaan dan karakteristik dari populasi tersebut agar hasil penelitian memiliki tingkat keakuratan dan keberlanjutan yang optimal. Oleh karena itu, penetapan ukuran sampel dalam penelitian ini juga akan didasarkan pada pertimbangan tersebut untuk memastikan hasil yang dapat diandalkan dan mewakili populasi secara menyeluruh

## 2.3 Teori tentang Framework/Algoritma yang digunakan

### 2.3.1 Collaborative filtering

Collaborative filtering adalah pendekatan yang umum digunakan dalam PRS. Pendekatan ini menggunakan informasi dari pengguna lain untuk merekomendasikan objek yang mungkin menarik bagi pengguna tersebut. Metode ini didasarkan pada asumsi bahwa pengguna yang memiliki preferensi yang serupa dalam masa lalu cenderung memiliki preferensi yang serupa di masa depan [1].

Collaborative filtering adalah salah satu metode yang umum digunakan dalam sistem rekomendasi. Metode ini mengumpulkan preferensi pengguna dari sejumlah pengguna yang serupa untuk membuat rekomendasi. Dalam collaborative filtering, rekomendasi didasarkan pada

kesamaan preferensi antara pengguna. Metode ini dapat digunakan untuk menghasilkan rekomendasi yang personal dan efektif [2].

Collaborative filtering pada sistem rekomendasi merupakan metode yang memanfaatkan informasi dari pengguna berupa skor peringkat produk dan pilihan sesuai keinginan dan preferensi pengguna [4]. Metode kolaboratif menciptakan spesifikasi yang direkomendasikan pengguna untuk suatu produk berdasarkan evaluasi atau nilai guna, seperti pembelian produk atau penilaian pengguna terhadap produk terkait [6]. Pemfilteran kolaboratif dapat dibagi menjadi dua metode untuk memperoleh informasi dalam pemeringkatan pengolahan data. Kedua metode tersebut adalah kolaborasi berbasis pengguna (user-based Collaborative Filtering) dan kolaborasi berbasis produk (item-based Collaborative Filtering) [3].

### 2.3.2 Singular Value Decomposition (SVD) dalam Collaborative Filtering

Singular Value Decomposition (SVD) adalah teknik matrix factorization yang memecah matriks user-item menjadi tiga matriks: U, Σ, dan V^T, sehingga. Dalam konteks collaborative filtering, matriks user-item biasanya berisi rating atau interaksi pengguna terhadap item. SVD bertujuan menemukan representasi laten baik untuk pengguna maupun item, sehingga dapat memprediksi rating yang hilang atau merekomendasikan produk baru. [3][4][7]

Proses SVD pada Collaborative Filtering:

- Matriks User-Item: Data interaksi pengguna dan item diubah ke dalam bentuk matriks.
- Decomposition: Matriks ini difaktorkan menjadi tiga matriks (U, Σ, V^T) menggunakan teknik SVD.
- Latent Factor Extraction: Faktor laten dari pengguna dan item diidentifikasi, yang merepresentasikan pola tersembunyi dalam preferensi pengguna.

- Prediksi: Dengan merekonstruksi matriks menggunakan sejumlah faktor utama (k), sistem dapat memprediksi rating yang belum diberikan oleh pengguna terhadap item tertentu.

Keunggulan SVD:

- Akurasi tinggi pada data yang sparse karena mampu menangkap pola laten yang tidak terlihat secara eksplisit.
- Scalability: Efektif untuk dataset besar dengan banyak missing value.
- Reduksi Dimensi: Mengurangi kompleksitas data dengan hanya mempertahankan faktor utama.

Keterbatasan SVD:

- Kompleksitas komputasi yang tinggi pada dataset sangat besar.
- Membutuhkan penanganan khusus untuk data yang sangat sparse atau memiliki missing value.

Hasil penelitian menunjukkan bahwa implementasi SVD pada sistem rekomendasi film menghasilkan nilai RMSE 0.4002 dan MAE 0.1186, menandakan tingkat akurasi prediksi yang tinggi dan error yang rendah.[8]

### 2.3.3 KNN dalam collaborative filtering

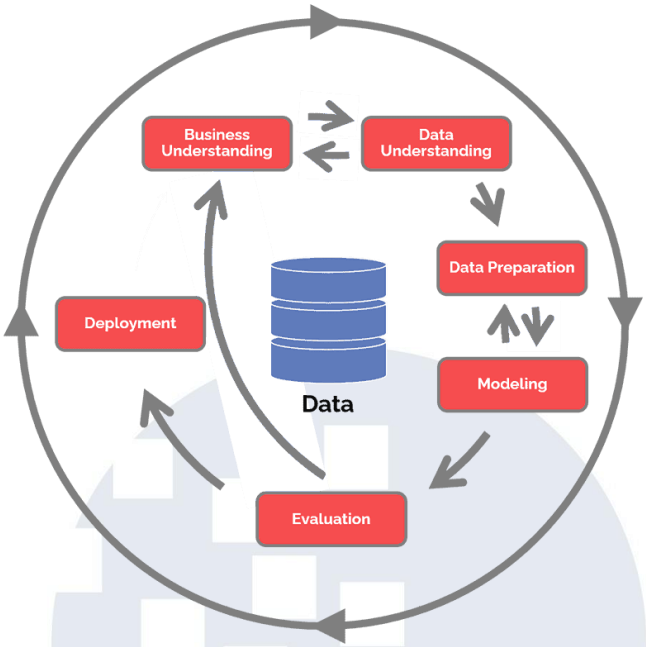
Pada sistem rekomendasi, collaborative filtering merupakan salah satu pendekatan yang banyak digunakan untuk memberikan rekomendasi kepada pengguna berdasarkan pola perilaku atau preferensi pengguna lain yang serupa. Salah satu algoritma yang sering diterapkan dalam collaborative filtering adalah K-Nearest Neighbors (KNN). Algoritma KNN bekerja dengan prinsip mencari sejumlah tetangga terdekat (K) yang memiliki kesamaan preferensi atau perilaku dengan pengguna atau item yang sedang dianalisis. Kesamaan ini biasanya diukur menggunakan metode seperti cosine similarity, Pearson correlation, atau Euclidean distance. Dalam pendekatan user-based collaborative filtering, sistem akan

mencari pengguna-pengguna lain yang memiliki pola rating atau interaksi yang mirip dengan pengguna target, kemudian merekomendasikan item yang disukai oleh tetangga terdekat tersebut. Sementara itu, pada item-based collaborative filtering, sistem akan mencari item-item yang mirip berdasarkan pola rating dari para pengguna, sehingga item yang sering dinilai serupa akan direkomendasikan satu sama lain.

Secara matematis, prediksi rating atau preferensi dilakukan dengan menghitung rata-rata tertimbang dari rating yang diberikan oleh tetangga terdekat, di mana bobot diberikan berdasarkan tingkat kesamaan antara pengguna atau item. Metrik evaluasi yang umum digunakan untuk menilai kinerja algoritma KNN dalam collaborative filtering antara lain Mean Absolute Error (MAE) dan Root Mean Squared Error (RMSE), yang mengukur seberapa dekat prediksi sistem terhadap data aktual. Kelebihan utama dari metode KNN adalah kemampuannya untuk memberikan rekomendasi tanpa memerlukan pengetahuan khusus tentang domain data, serta kemudahan dalam implementasinya.

Namun, KNN juga memiliki beberapa kelemahan, seperti masalah cold-start pada pengguna atau item baru, serta penurunan performa pada data yang sangat sparse dan kompleksitas komputasi yang meningkat seiring bertambahnya jumlah data. Oleh karena itu, meskipun KNN merupakan metode yang sederhana dan efektif, pengembang sistem rekomendasi sering mengombinasikannya dengan pendekatan lain untuk meningkatkan akurasi dan efisiensi sistem.

### 2.3.3 Cross-Industry Standard Process for Data Mining (CRISP-DM)




Gambar 2.1 CRISP-DM

Sumber: https://www.datascience-pm.com/crisp-dm-2/

Dari gambar 2.1 Bagian ini menjelaskan penerapan kerangka kerja CRISP-DM (*Cross-Industry Standard Process for Data Mining*) sebagai panduan proses penelitian dalam membandingkan algoritma *Singular Value Decomposition* (SVD) dan *K-Nearest Neighbors* (KNN) untuk sistem rekomendasi produk. Enam fase yang ada dalam CRISP-DM menjamin ketelitian metodologis dan keselarasan dengan tujuan bisnis.

### 2.3.3.1 Pemahaman Bisnis (Business Understanding)

Fase ini berfokus pada pemahaman menyeluruh terhadap permasalahan bisnis yang ingin diselesaikan. Dalam konteks penelitian ini, tujuan utamanya adalah membandingkan efektivitas dua algoritma Collaborative Filtering—SVD dan KNN—dalam meningkatkan kualitas rekomendasi produk pada platform e-commerce. Pemahaman terhadap kebutuhan bisnis, seperti peningkatan akurasi rekomendasi dan efisiensi

proses komputasi, menjadi dasar untuk menentukan arah dan fokus analisis data yang akan dilakukan.

#### 2.3.3.2 Pemahaman Data (Data Understanding)

Setelah tujuan bisnis didefinisikan, langkah berikutnya adalah memahami karakteristik dan struktur data yang tersedia. Proses ini melibatkan eksplorasi awal terhadap dataset perilaku konsumen dari platform e-commerce, seperti identifikasi jumlah transaksi, distribusi pembelian ulang (reordered), serta analisis temporal berdasarkan waktu pemesanan (order_hour_of_day dan order_dow). Tujuannya adalah untuk mengetahui pola-pola awal dalam data yang relevan untuk mendukung pengembangan model rekomendasi.

#### 2.3.3.3 Persiapan Data (Data Preparation)

Fase ini mencakup semua aktivitas teknis yang diperlukan untuk menyiapkan data sebelum masuk ke pemodelan. Dalam penelitian ini, proses data preparation meliputi pembersihan nilai kosong, transformasi fitur kategorikal menggunakan one-hot encoding, normalisasi fitur numerik, serta pembentukan struktur user-item matrix untuk model SVD. Data juga dibagi secara temporal untuk memastikan skenario pelatihan dan pengujian lebih realistis. Hasil dari tahap ini adalah dataset yang bersih, terstruktur, dan siap untuk digunakan dalam tahap modeling.

#### 2.3.3.4 Model Pembangunan (Modeling):

Tahap modeling adalah saat di mana algoritma machine learning mulai diterapkan. Dalam penelitian ini, dua model Collaborative Filtering digunakan: Singular Value Decomposition (SVD) dan K-Nearest Neighbors (KNN). Masing-masing model dikembangkan dengan konfigurasi parameter yang disesuaikan, serta dilakukan proses tuning untuk meningkatkan performa. Pemodelan dilakukan menggunakan pustaka Python seperti Surprise, dan mencakup baik pendekatan user-based maupun item-based untuk KNN.

#### 2.3.3.5 Evaluasi Model (Evaluation):

Setelah model dibangun, evaluasi dilakukan untuk mengukur efektivitasnya. Dalam penelitian ini, evaluasi dilakukan menggunakan metrik akurasi seperti RMSE, MAE, Precision@K, Recall@K, dan F1-score, serta waktu eksekusi sebagai indikator efisiensi komputasi. Selain itu, dilakukan analisis coverage untuk mengukur sejauh mana model mampu memberikan rekomendasi ke seluruh produk. Hasil evaluasi ini digunakan untuk menentukan model mana yang paling unggul secara keseluruhan dan cocok untuk diterapkan dalam sistem rekomendasi produk.

#### 3.2.2.6 Implementasi (Deployment):

Fase akhir dari CRISP-DM adalah *deployment*, yaitu penerapan hasil model ke dalam konteks bisnis nyata atau simulasi implementasi. Pada penelitian ini, *deployment* dilakukan secara konseptual, di mana rekomendasi disampaikan berdasarkan hasil evaluasi model yang telah dibangun. Pemilihan model juga direkomendasikan untuk berbagai skenario: SVD lebih disarankan untuk sistem berskala besar dengan data yang bersifat *sparse*, sedangkan KNN dianggap lebih sesuai untuk kebutuhan rekomendasi real-time dalam skala yang lebih kecil.

## 2.4 Teori tentang tools/software yang digunakan

### 2.4.1 Jupyter Notebook

Jupyter Notebook adalah alat pengolah data populer menggunakan bahasa pemrograman Python. Notebook Jupyter memungkinkan Anda mengintegrasikan kode dan keluaran dokumentasi secara interaktif. Jupyter (https: //jupyter.org/) sendiri merupakan organisasi nirlaba yang mengembangkan perangkat lunak interaktif dalam berbagai bahasa pemrograman. Notebook, sebaliknya, adalah perangkat lunak Jupyter dalam bentuk aplikasi web sumber terbuka yang memungkinkan Anda membuat dan berbagi dokumen interaktif yang berisi kode langsung, persamaan, visualisasi, dan teks penjelasan yang kaya.

Jupyter adalah aplikasi web gratis untuk membuat dan berbagi dokumen yang berisi kode, perhitungan, visualisasi, dan teks. Jupyter adalah singkatan dari tiga bahasa pemrograman: Julia (Ju), Python (Py), dan R. Ketiga bahasa pemrograman ini penting bagi data scientist. Fungsi Jupyter membantu Anda membuat narasi komputasi yang menjelaskan makna dalam data Anda dan memberikan wawasan tentang data Anda. Selain itu, Jupyter juga memfasilitasi kolaborasi antara insinyur dan ilmuwan data dengan memudahkan pembuatan dan berbagi teks dan kode. Oleh karena itu, Jupyter memudahkan data scientist untuk berkolaborasi dengan data scientist, dan data engineer lainnya [23].

### 2.4.2 Python

Python adalah bahasa pemrograman interpretatif multiguna. Tidak seperti bahasa lain yang susah untuk dibaca dan dipahami, Python lebih menekankan pada keterbacaan kode agar lebih mudah untuk memahami sintaks. Hal ini membuat Python sangat mudah dipelajari baik untuk pemula maupun untuk yang sudah menguasai bahasa pemrograman lain. Bahasa pemrograman Python muncul pertama kali pada tahun 1991, yang dirancang oleh seseorang bernama Guido van Rossum. Sampai saat ini Python masih dikembangkan oleh Python Software Foundation. Bahasa Python mendukung hampir semua sistem operasi, bahkan untuk sistem operasi Linux, dan hampir semua distronya sudah menyertakan Python di dalamnya [24].