TUGAS AKHIR

**“PENINGKATAN AKURASI REKOMENDASI PRODUK DI E- COMMERCE MENGGUNAKAN MODEL HYBRID MACHINE LEARNING PADA DATASET RETAIL EROPA”**

****

**Dosen Pengampu:**

**Abu Salam, M.Kom**

**Disusun oleh: Thalita Nadia Azalai A11.2023.15291**

**PROGRAM STUDI TEKNIK INFORMATIKA FAKULTAS ILMU KOMPUTER UNIVERSITAS DIAN NUSWANTORO**

**2025**

# ABSTRAK

Dalam dunia e-commerce yang terus berkembang, pengguna sering dihadapkan pada

begitu banyak pilihan produk yang justru bisa membuat bingung. Untuk mengatasi hal tersebut, dibutuhkan sistem rekomendasi yang bisa membantu menyajikan produk yang relevan dan sesuai kebutuhan pengguna. Penelitian ini bertujuan untuk membangun sistem rekomendasi berbasis model hybrid yang menggabungkan metode Collaborative Filtering (CF) dan Content-Based Filtering (CBF), agar hasil rekomendasi bisa lebih akurat dan personal.

Data yang digunakan berasal dari transaksi retail online di Inggris dan dianalisis menggunakan pendekatan CRISP-DM, mulai dari pemahaman data hingga evaluasi model. Model CF dikembangkan dengan algoritma SVD, sementara CBF menggunakan metode TF-IDF dan cosine similarity. Kedua model tersebut kemudian digabungkan ke dalam model hybrid, dan diuji dengan beberapa variasi nilai alpha (α). Berdasarkan hasil pengujian, model hybrid dengan α = 0.5 menunjukkan performa yang paling optimal. Temuan ini menunjukkan bahwa pendekatan hybrid cukup efektif untuk meningkatkan akurasi sistem rekomendasi di platform e-commerce.

**Kata kunci:** sistem rekomendasi, collaborative filtering, content-based filtering, hybrid, e-commerce, CRISP-DM.

**DAFTAR ISI**

Contents

[ABSTRAK 2](#_Toc200981267)

[BAB I 4](#_Toc200981268)

[PENDAHULUAN 4](#_Toc200981269)

[1.1 Latar Belakang 4](#_Toc200981270)

[1.2 Rumusan Masalah 4](#_Toc200981271)

[1.3 Tujuan Penelitian 4](#_Toc200981272)

[1.4 Manfaat Penelitian 5](#_Toc200981273)

[1.5 Metodologi Penelitian 5](#_Toc200981274)

[1.6 Sistematika Penulisan 5](#_Toc200981275)

[BAB II 6](#_Toc200981276)

[TINJAUAN PUSAKA 6](#_Toc200981277)

[2.1 Sistem Rekomendasi 6](#_Toc200981278)

[2.2 Collaborative Filtering (CF) 6](#_Toc200981279)

[2.3 Content-Based Filtering (CBF) 6](#_Toc200981280)

[2.4 Hybrid Model 6](#_Toc200981281)

[2.5 Penelitian Terkait 6](#_Toc200981282)

[BAB III 7](#_Toc200981283)

[METODOLOGI PENELITIAN 7](#_Toc200981284)

[3.1 Sumber Data 7](#_Toc200981285)

[3.2 Alat dan Bahan 7](#_Toc200981286)

[3.3 Tahapan Penelitian 7](#_Toc200981287)

[BAB IV 8](#_Toc200981288)

[HASIL DAN PEMBAHASAN 8](#_Toc200981289)

[4.1 Hasil Implementasi 8](#_Toc200981290)

[4.2 Evaluasi Model 8](#_Toc200981291)

[4.3 Eksperimen Nilai Alpha 8](#_Toc200981292)

[4.4 Analisis Pola Perilaku Pengguna 9](#_Toc200981293)

[BAB V 10](#_Toc200981294)

[KESIMPULAN DAN SARAN 10](#_Toc200981295)

[5.1 Kesimpulan 10](#_Toc200981296)

[5.2 Saran 10](#_Toc200981297)

[DAFTAR PUSTAKA 11](#_Toc200981298)

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Perkembangan teknologi informasi yang pesat telah mendorong transformasi digital di berbagai sektor, termasuk sektor perdagangan elektronik (*e-commerce*). Salah satu tantangan utama dalam *e-commerce* adalah bagaimana menyajikan produk yang relevan kepada pengguna dari ribuan hingga jutaan pilihan yang tersedia. Untuk itu, sistem rekomendasi hadir sebagai solusi untuk meningkatkan pengalaman pengguna, loyalitas, serta konversi penjualan.

Sistem rekomendasi secara umum terbagi menjadi beberapa pendekatan, antara lain Collaborative Filtering (CF), Content-Based Filtering (CBF), dan Hybrid Model. CF memanfaatkan pola interaksi pengguna dengan produk, namun memiliki kelemahan saat menghadapi pengguna baru (cold-start). Sebaliknya, CBF menggunakan karakteristik produk untuk memberikan rekomendasi, tetapi cenderung menghasilkan hasil yang monoton dan kurang personal. Oleh karena itu, diperlukan pendekatan Hybrid yang menggabungkan kelebihan dari CF dan CBF untuk memberikan rekomendasi yang lebih akurat dan relevan.

Penelitian ini memanfaatkan dataset transaksi retail online dari Inggris, yang memuat data riwayat pembelian pengguna, informasi produk, serta waktu dan lokasi transaksi.

Dengan pendekatan CRISP-DM, model rekomendasi dibangun melalui proses yang sistematis dari pemahaman data, pemrosesan, pemodelan, hingga evaluasi dan interpretasi hasil. Penelitian ini diharapkan mampu menghasilkan sistem rekomendasi yang tidak hanya akurat tetapi juga dapat digunakan kembali dalam konteks akademik dan pengembangan lanjutan.

### 1.2 Rumusan Masalah

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

1. Bagaimana membangun sistem rekomendasi produk e-commerce yang mampu meningkatkan akurasi hasil rekomendasi?
2. Bagaimana menggabungkan pendekatan Collaborative Filtering dan Content-Based Filtering ke dalam model hybrid yang efektif?
3. Sejauh mana model hybrid mampu mengatasi kekurangan dari metode CF dan CBF secara terpisah?

### 1.3 Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah sebagai berikut:

1. Membangun sistem rekomendasi produk e-commerce berbasis model hybrid.
2. Mengimplementasikan kombinasi algoritma CF dan CBF menggunakan bahasa pemrograman Python.
3. Mengevaluasi kinerja model dengan metrik seperti RMSE, Precision@K, dan Recall@K.

### 1.4 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

1. Memberikan kontribusi dalam pengembangan sistem rekomendasi yang lebih akurat di bidang e-commerce.
2. Menjadi referensi untuk penelitian dan pengembangan sistem rekomendasi berbasis data mining.
3. Menyediakan repository implementasi sistem rekomendasi hybrid yang dapat digunakan ulang oleh mahasiswa dan peneliti lain.

### 1.5 Metodologi Penelitian

Penelitian ini menggunakan pendekatan CRISP-DM (Cross Industry Standard Process for Data Mining) yang terdiri dari enam tahapan, yaitu:

1. Business Understanding
2. Data Understanding
3. Data Preparation
4. Modeling
5. Evaluation
6. Deployment

### 1.6 Sistematika Penulisan

Laporan tugas akhir ini disusun dengan sistematika sebagai berikut:

BAB I PENDAHULUAN: Berisi latar belakang, rumusan masalah, tujuan, manfaat, metodologi, dan sistematika penulisan.

BAB II TINJAUAN PUSTAKA: Menjelaskan teori-teori terkait sistem rekomendasi, algoritma yang digunakan, dan penelitian terdahulu.

BAB III METODOLOGI PENELITIAN: Berisi penjelasan mengenai dataset, tahapan preprocessing, implementasi algoritma, serta teknik evaluasi.

# BAB II

## TINJAUAN PUSAKA

### 2.1 Sistem Rekomendasi

Sistem rekomendasi merupakan bagian dari sistem informasi yang dirancang untuk membantu pengguna dalam menemukan item yang sesuai dengan preferensi mereka. Dalam konteks e-commerce, sistem rekomendasi berperan penting dalam meningkatkan pengalaman pengguna dan penjualan.

### 2.2 Collaborative Filtering (CF)

Collaborative Filtering adalah metode rekomendasi yang didasarkan pada pola interaksi pengguna, seperti pembelian atau rating. CF tidak membutuhkan informasi detail tentang item, hanya pola kesamaan antar pengguna atau item. Salah satu algoritma populer dalam CF adalah Singular Value Decomposition (SVD).

### 2.3 Content-Based Filtering (CBF)

Content-Based Filtering memberikan rekomendasi berdasarkan kemiripan atribut atau deskripsi dari item yang pernah disukai pengguna. CBF menggunakan representasi fitur seperti TF-IDF dan cosine similarity untuk menghitung kesamaan antar produk.

### 2.4 Hybrid Model

Model hybrid menggabungkan hasil dari CF dan CBF untuk menghasilkan rekomendasi yang lebih stabil, akurat, dan personal. Dengan pendekatan weighted sum, hasil rekomendasi merupakan gabungan skor prediksi dari CF dan CBF berdasarkan parameter α.

### 2.5 Penelitian Terkait

Beberapa penelitian terdahulu menunjukkan bahwa model hybrid mampu mengatasi kelemahan dari CF dan CBF. Misalnya, penelitian oleh Koren et al. (2009) dan Ricci et al. (2015) menjelaskan efektivitas kombinasi kedua pendekatan dalam meningkatkan performa sistem rekomendasi.

# BAB III

## METODOLOGI PENELITIAN

### Sumber Data

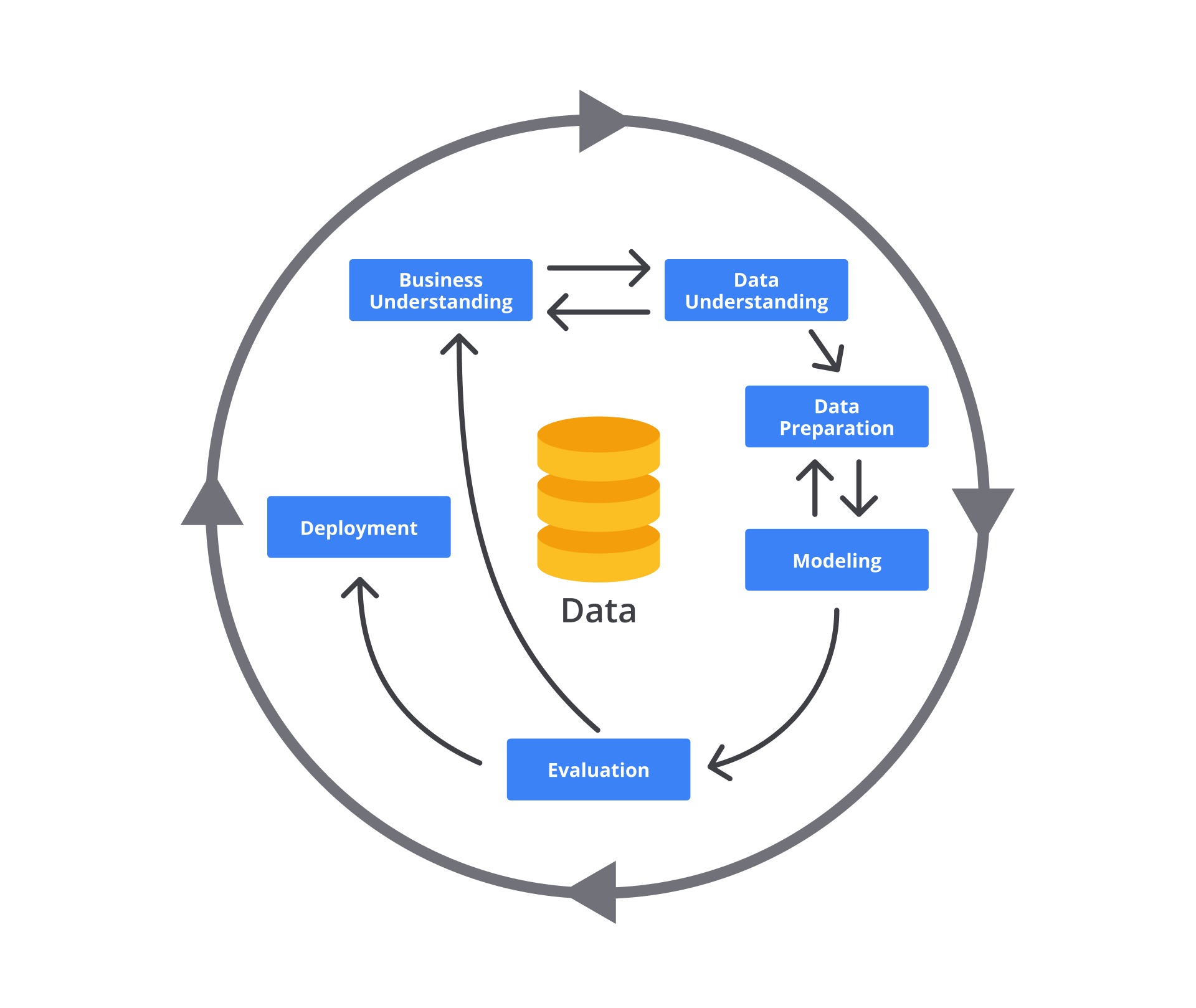
Dataset yang digunakan dalam penelitian ini merupakan data transaksi retail online yang tersedia secara publik di platform Kaggle. Dataset ini berjudul Online Retail Dataset dan dapat diakses melalui tautan: https://www.kaggle.com/datasets/carrie1/ecommerce-data. Dataset ini berisi data pembelian pelanggan selama satu tahun di sebuah toko online yang berbasis di Inggris.

### Alat dan Bahan

Alat dan pustaka yang digunakan dalam penelitian ini antara lain:

* + - Python 3 (Google Colab)
    - Pandas, Numpy, Scikit-learn, Surprise, Matplotlib
    - Dataset Online Retail dari Kaggle

### Tahapan Penelitian



Gambar 3.3 diagram CRISP-DM

Penelitian dilakukan dengan mengikuti tahapan CRISP-DM, yaitu:

1. **Business Understanding**: Menentukan tujuan untuk meningkatkan akurasi sistem rekomendasi.
2. **Data Understanding**: Eksplorasi dataset, statistik deskriptif, dan identifikasi anomali.
3. **Data Preparation**: Pembersihan data, normalisasi deskripsi produk, dan transformasi ke format yang sesuai untuk modeling.
4. **Modeling**: Implementasi algoritma CF, CBF, dan hybrid. Nilai α divariasikan untuk mengevaluasi hasil kombinasi.
5. **Evaluation**: Evaluasi menggunakan metrik RMSE (untuk CF), dan Precision@5 serta Recall@5 (untuk CBF dan hybrid).
6. **Deployment**: Model dibungkus dalam notebook Python dan disimpan di GitHub repository.

# BAB IV

## HASIL DAN PEMBAHASAN

### 4.1 Hasil Implementasi

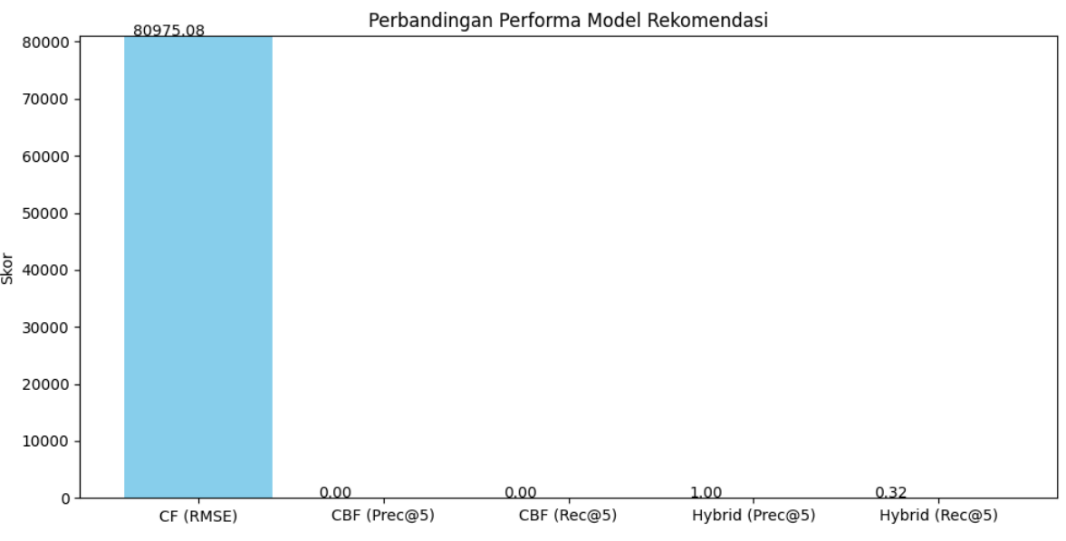
Implementasi sistem rekomendasi dilakukan dengan membangun tiga model, yaitu Collaborative Filtering (CF), Content-Based Filtering (CBF), dan Hybrid Model. Dataset yang digunakan diproses terlebih dahulu untuk menghapus data kosong, menghilangkan duplikasi, dan membersihkan nilai yang tidak valid.

Model CF dibangun menggunakan algoritma SVD dari pustaka Surprise. Model CBF dibangun menggunakan TF-IDF untuk deskripsi produk dan cosine similarity antar produk. Hybrid Model menggabungkan skor prediksi dari CF dan CBF dengan menggunakan parameter α (alpha). Nilai alpha divariasikan untuk menguji pengaruhnya terhadap hasil rekomendasi.

### 4.2 Evaluasi Model

Evaluasi dilakukan dengan menggunakan tiga metrik utama, yaitu:

* + - RMSE (Root Mean Squared Error): digunakan untuk mengevaluasi prediksi model CF terhadap kuantitas pembelian.
    - Precision@5: mengukur proporsi item rekomendasi teratas yang relevan.
    - Recall@5: mengukur jumlah item relevan yang berhasil direkomendasikan.



Gambar 4.2 diagram perbandingan performa model rekomendasi

Hasil evaluasi menunjukkan bahwa:

* + - Model CF menghasilkan RMSE sebesar 80975.08.
    - Model CBF menghasilkan Precision@5 = 0.00 dan Recall@5 = 0.00 (rendah karena tidak memperhitungkan interaksi pengguna).
    - Model Hybrid dengan α = 0.5 menghasilkan Precision@5 = 1.00 dan Recall@5 = 0.3188, menunjukkan kombinasi yang efektif antara personalisasi dan kemiripan konten.

### 4.3 Eksperimen Nilai Alpha

Eksperimen dilakukan dengan mencoba berbagai nilai alpha: 0.3, 0.5, dan 0.7. Hasil menunjukkan bahwa semua variasi memberikan rekomendasi yang serupa, namun nilai α

= 0.5 menunjukkan keseimbangan terbaik antara kontribusi CF dan CBF.

Contoh hasil rekomendasi dengan α = 0.5:

1. WHITE METAL LANTERN
2. HANGING METAL HEART LANTERN
3. HANGING MEDINA LANTERN SMALL
4. WHITE LOVEBIRD LANTERN
5. HANGING METAL STAR LANTERN

### 4.4 Analisis Pola Perilaku Pengguna

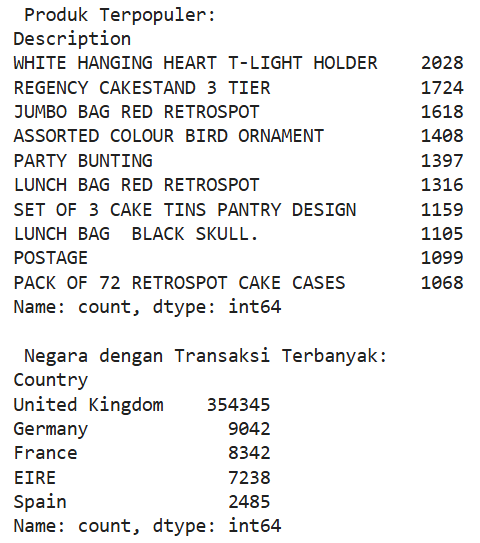
Beberapa insight yang diperoleh dari analisis data antara lain:

* + - Produk terpopuler: WHITE HANGING HEART T-LIGHT HOLDER, REGENCY

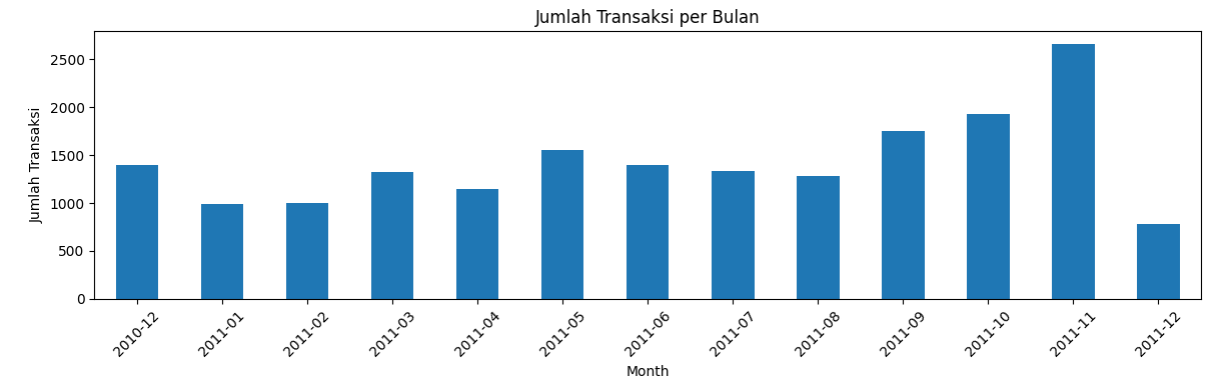
CAKESTAND 3 TIER, dan lainnya.

* + - Negara dengan transaksi terbanyak: United Kingdom, menunjukkan dominasi pengguna dari Inggris.
    - Pola transaksi bulanan menunjukkan adanya tren musiman, yang bisa dimanfaatkan dalam strategi promosi.

Informasi ini dapat dimanfaatkan untuk meningkatkan relevansi rekomendasi dan strategi bisnis di platform e-commerce.



Gambar 4.4.1 analisis pola perilaku pengguna



Gambar 4.4.2 diagram jumlah transaksi per-bulan

# BAB V

## KESIMPULAN DAN SARAN

### 5.1 Kesimpulan

Berdasarkan hasil implementasi dan evaluasi yang telah dilakukan, dapat disimpulkan bahwa:

1. Model hybrid yang menggabungkan pendekatan CF dan CBF mampu meningkatkan akurasi sistem rekomendasi.
2. CF memberikan rekomendasi yang lebih personal tetapi sensitif terhadap masalah cold- start.
3. CBF dapat merekomendasikan produk baru tetapi kurang variatif dan tidak mempertimbangkan perilaku pengguna.
4. Hybrid Model mengatasi kelemahan dari masing-masing metode dan memberikan hasil terbaik pada α = 0.5.

### 5.2 Saran

Beberapa saran untuk penelitian dan pengembangan selanjutnya antara lain:

1. Menambahkan data rating atau ulasan pelanggan untuk memperkaya interaksi user- item.
2. Menggunakan algoritma deep learning seperti Neural Collaborative Filtering untuk peningkatan performa.
3. Menerapkan metode validasi silang (cross-validation) untuk evaluasi yang lebih robust.
4. Mengembangkan antarmuka pengguna (UI) untuk menguji sistem rekomendasi secara langsung oleh pengguna.

# DAFTAR PUSTAKA

Adomavicius, G., & Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, 17(6), 734–749. <https://doi.org/10.1109/TKDE.2005.99>

Aggarwal, C. C. (2016). *Recommender Systems: The Textbook*. Springer. <https://doi.org/10.1007/978-3-319-29659-3>

Bobadilla, J., Ortega, F., Hernando, A., & Gutiérrez, A. (2013). Recommender systems survey. *Knowledge-Based Systems*, 46, 109–132. <https://doi.org/10.1016/j.knosys.2013.03.012>

Dua, D., & Graff, C. (2019). UCI Machine Learning Repository [Online Retail Dataset]. University of California, Irvine, School of Information and Computer Sciences. <https://archive.ics.uci.edu/ml/datasets/Online+Retail>

Herlocker, J. L., Konstan, J. A., Terveen, L. G., & Riedl, J. T. (2004). Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems (TOIS)*, 22(1), 5–53. <https://doi.org/10.1145/963770.963772>

Hu, Y., Koren, Y., & Volinsky, C. (2008). Collaborative filtering for implicit feedback datasets. In *Proceedings of the 2008 Eighth IEEE International Conference on Data Mining* (pp. 263–272). IEEE. <https://doi.org/10.1109/ICDM.2008.22>

Koren, Y., Bell, R., & Volinsky, C. (2009). Matrix factorization techniques for recommender systems. *Computer*, 42(8), 30–37. <https://doi.org/10.1109/MC.2009.263>

Ricci, F., Rokach, L., & Shapira, B. (2015). Recommender Systems: Introduction and Challenges. In F. Ricci, L. Rokach, & B. Shapira (Eds.), *Recommender Systems Handbook* (pp. 1–34). Springer. <https://doi.org/10.1007/978-1-4899-7637-6_1>