E - Commerce Recommendation System

Student Name: PRISHA GABBAD

Reg. No.: 25BCY10139

Subject: CSA2001 - Fundamentals in AI and ML

College: Vellore Institute of Technology BHOPAL

Year: 2025

Submitted To: Dr. E. Nirmala ma'am

Slot: A21+A22+A23

Introduction

The Flipkart Product Recommendation System is designed to help users discover relevant products by analyzing product metadata, ratings, descriptions, and categories. This project uses content-based filtering, rating-based trending analysis, and collaborative filtering to recommend items.

The system is fully implemented in Python using Pandas, NumPy, TF-IDF, cosine similarity, visualization libraries, and basic ML concepts.

Problem Statement

E-commerce platforms contain millions of products, making it difficult for users to find relevant items quickly.

The project aims to build a hybrid recommendation system that provides personalized and meaningful product suggestions based on:

Ratings

Product metadata

User behavior

Content similarity

Functional Requirements

The system supports 3 major functional modules (as required):

Module 1 — Rating-Based Recommendation System

Calculates average rating per product

Shows top 10 trending/highest rated product

Module 2 — Content-Based Recommendation System

Cleans product description, brand, and categories

Extracts tags using NLP

Generates TF-IDF vectors

Computes cosine similarity

Recommends similar products

Module 3 — Collaborative Filtering System

Simulates 500 users

Builds a user–item matrix

Calculates user similarity

Recommends products liked by similar users

Non-Functional Requirements

1. Performance:

Efficient vector computations using sparse TF-IDF matrices.

2. Scalability:

Can handle thousands of products; suitable for more with optimization.

3. Usability:

Simple input — just enter a product name or user ID.

4. Maintainability:

Modular code: cleaning, tagging, recommendation, collaborative logic separated.

5. Reliability:

Handles missing values, invalid inputs, and inconsistent data formats.

System Architecture

Architecture Layers:

1. Data Layer

Flipkart CSV dataset

Preprocessing, feature extraction

2. Logic Layer

Rating-based module

Content-based module

Collaborative filtering module

3. Recommendation Engine Layer

Similarity computation

Ranking & scoring

4. Output Layer

Tables / DataFrames

Visualization graphs (heatmaps, bar charts)

Design Diagrams

Actors: User

Use cases: Get top products, get similar products, get personalized recommendations

Workflow Diagram

1. Load dataset

2. Clean missing values

3. Extract tags using NLP

4. Build TF-IDF matrix

5. Compute similarity

6. Generate recommendations

Implementation Details

code implements:

  Data cleaning

  Null handling

  Tag creation using spaCy

  TF-IDF vectorization

  Cosine similarity

  User-item matrix creation

  Collaborative filtering logic

  Visualizations:

Testing Approach

Tested for null values

Tested invalid product names

Tested missing user IDs

Verified similarity output correctness

Checked visualization accuracy

Challenges Faced

Missing or corrupted dataset values

Inconsistent product descriptions

NLP cleaning complexity

Slow TF-IDF on large text fields

Learnings & Key Takeaways

Learned handling real e-commerce data

Understood TF-IDF vectorization deeply

Learned collaborative filtering basics

Learned system design and documentation

Future Enhancements

Real user authentication

Real-time recommendation engine

Deep learning based models (BERT, neural CF)

API deployment & front-end interface