A Major Project Phase - I(A)

E-Commerce Products Recommendation System using Machine Learning

Submitted in Partial fulfilment for the
degree of Bachelor of Technology in
Computer Science and Technology

Submitted by

Vaishnavi Agawane(01)

Pooja Rathod(48)

Neha Tembhe(61)

Under the guidance of

Prof. Prajakta Gotarne

Usha Mittal Institute Of Technology

S.N.D.T. WOMEN'S UNIVERSITY

MUMBAI-400049

2024-2025

CERTIFICATE

This is to certify that **Vaishnavi Agawane, Pooja Rathod** and **Neha Tembhe** has successfully completed major project phase-I(A) work on **E-Commerce Products Recommendation System using Machine Learning** in the partial fulfillment for the bachelor's degree in **Computer Science and Technology** during the year 2024-25 as prescribed by SNDT Women's University.

GUIDE
PROF. PRAJAKTA GOTARNE

H.O.D
PROF. KUMUD WASNIK

Principle
DR.YOGESH NERKAR

Examiner 1

Examiner 2

ACKNOWLEDGEMENT

We are deeply grateful to Prof. Prajakta Gotarne for their invaluable guidance, continuous support, and encouragement throughout this project. We would also like to thank our Head of Department, Prof. Kumud D. Wasnik., for providing us with this opportunity and necessary resources to work on this project.

Our heartfelt thanks go to our classmates and family members for their unwavering support during the course of this project. Their motivation helped us immensely in completing this project successfully.

Vaishnavi Agawane

Pooja Rathod

Neha Tembhe

ABSTRACT

This project focuses on developing an e-commerce recommendation system using Flask and machine learning to enhance user experience by providing personalized product suggestions. The system employs various recommendation techniques including content-based,collaborative filtering, hybrid, and multi-model approaches.

The development involves collecting and preprocessing an ecommerce dataset, implementing diverse recommendation algorithms, and integrating the system with a Flask web application. The Flask application features user registration, product browsing, search functionality, and personalized recommendations,ensuring a secure and seamless shopping experience.This solution aims to boost user engagement, sales, and customer satisfaction in the digital marketplace.

Contents

List of Figures

0.1 Introduction

In the digital era, e-commerce platforms have revolutionized the way people shop, offering a vast array of products from around the globe. However, the sheer volume of available products can overwhelm consumers, making it difficult to find items that match their preferences. To tackle this issue, e-commerce platforms increasingly rely on recommendation systems, which use algorithms to suggest products tailored to individual user tastes and behaviors. These systems enhance the shopping experience by streamlining product discovery, thereby improving customer satisfaction and driving sales. The integration of machine learning in recommendation systems has significantly advanced their accuracy and effectiveness, making them an essential tool in the competitive e-commerce landscape. This project aims to develop a sophisticated recommendation system using Flask and machine learning, employing various techniques to deliver highly personalized product suggestions.

0.1.1 Purpose and Objectives

1. **Develop a Content-Based Recommendation Algorithm:** Utilize product attributes and user preferences to suggest similar items.
2. **Implement Collaborative Filtering Models:** Analyze user behavior, such as ratings and interactions, to predict preferences.
3. **Create Hybrid Recommendation Models:** Combine content-based and collaborative filtering approaches to improve accuracy and diversity of recommendations.
4. **Explore Multi-Model Strategies:** Integrate various machine learning models to cater to different user preferences and product characteristics.
5. **Integrate the Recommendation System with a Flask Web Application:** Develop a user-friendly interface for seamless interaction, including features such as user registration, product browsing, search functionality, and feedback mechanisms.
6. **Evaluate System Performance:** Use metrics such as accuracy, precision, recall, and F1-score to measure the effectiveness of the recommendation system.
7. **Ensure Security and Ethical Standards:** Implement user authentication, session management, and maintain data privacy and confidentiality.

0.2 Problem Statement

In today's fast-paced and dynamic e-commerce landscape, the vast array of product choices can overwhelm customers, making it difficult to discover relevant items efficiently. This challenge not only hampers user experience but also affects customer retention and business growth. The proposed system seeks to address this issue by leveraging advanced machine learning algorithms to deliver highly personalized and accurate product recommendations. By analyzing user preferences, browsing behavior, and purchase history, the system aims to streamline the shopping experience, increase customer satisfaction, and optimize conversion rates.

0.3 Literature Review

Sr.No.	Paper Title	Author Name	Methodology	Observation	Limitations
1.	Techniques, benefits, and challenges of recommendation system in e-commerce: A literature review Publish Year:2021	Thanh Vu Ngoc and Huong Tran Thi	Collaborative filtering, content-based filtering, hybrid methods, and deep learning-based methods.	Improved user experience, increased sales, and reduced customer churn.	Data sparsity, cold-start problem, scalability, and privacy concerns.
2.	E-Commerce Product Recommendation System based on ML Algorithms. Publish Year:2020	Md. Zahurul Haque, Department of CSE, Manarat International University, Dhaka, Bangladesh.	Used various machine learning algorithms, including collaborative filtering, content-based filtering, and hybrid methods.	The hybrid approach combining collaborative filtering and content-based filtering achieved the best performance.	The system was not evaluated on a large-scale dataset.
3.	The Power of Personalization: Enhancing User Experience in E-commerce Publish Year:2023	Shreyas Ghansawant1 , Gaurang Kumbhar2 , Prof. Nikita kawase3 , Prof. Deepak K. Sharma4	Discussed the importance of personalization in e-commerce and how it can improve user experience.	Personalised recommendations can significantly increase customer engagement and satisfaction.	The paper did not provide specific examples or case studies.
4.	Neural Collaborative Filtering Publish Year:2017	Xiangnan He, Lizi Liao, Hanwang Zhang	Introduced a new neural collaborative filtering model that combines collaborative filtering and neural networks.	The proposed model outperformed existing collaborative filtering methods on various datasets.	The model may be computationally expensive for large-scale datasets.
5.	Matrix factorization Techniques for Recommender system Publish Year:2019	Yehuda Koren, Yahoo Research Robert Bell and Chris Volinsky, AT&T Labs—Research	Reviewed various matrix factorization techniques used in recommender systems, including singular value decomposition (SVD) and its variants.	Matrix factorization is a popular approach for recommendation systems due to its simplicity and effectiveness.	The paper did not discuss the limitations of matrix factorization techniques or compare them with other approaches.

Figure 1: Literature Review

0.4 Exiting System

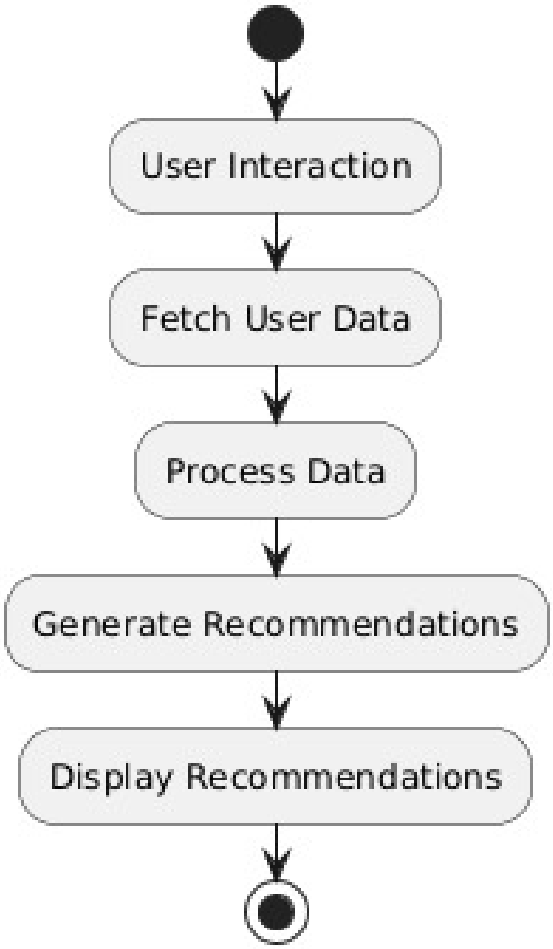

Figure 2: Data flow diagram of Exiting System

Drawbacks of the Existing E-commerce Product Recommendation System:

While the described system provides a solid foundation for product recommendations, there are several potential drawbacks that can be addressed to improve its effectiveness:

- **Data Quality and Quantity:** The accuracy and comprehensiveness of the user data collected can significantly impact the quality of recommendations. If the data is incomplete, inaccurate, or biased, the system may generate irrelevant or misleading suggestions.
- **Algorithm Limitations:** The chosen recommendation algorithm may have inherent limitations. For example, collaborative filtering can struggle with cold-start problems (recommending products to new users with limited interaction history) and the sparsity of user-item interactions.
- **Over-reliance on Historical Data:** The system may become overly reliant on historical

data, failing to adapt to changing user preferences or emerging trends. This can lead to stale and outdated recommendations.

- **Lack of Contextual Understanding:** The system may not fully consider the context of a user's interaction, such as the time of day, location, or specific shopping scenario. This can result in irrelevant or untimely recommendations.
- **Privacy Concerns:** Collecting and analyzing user data raises privacy concerns. Ensuring that user data is handled securely and ethically is crucial to maintain trust and compliance with regulations.
- **Limited Diversity:** The system may generate recommendations that are too similar to the user's existing preferences, limiting their exposure to new and diverse products.

0.5 Proposed System

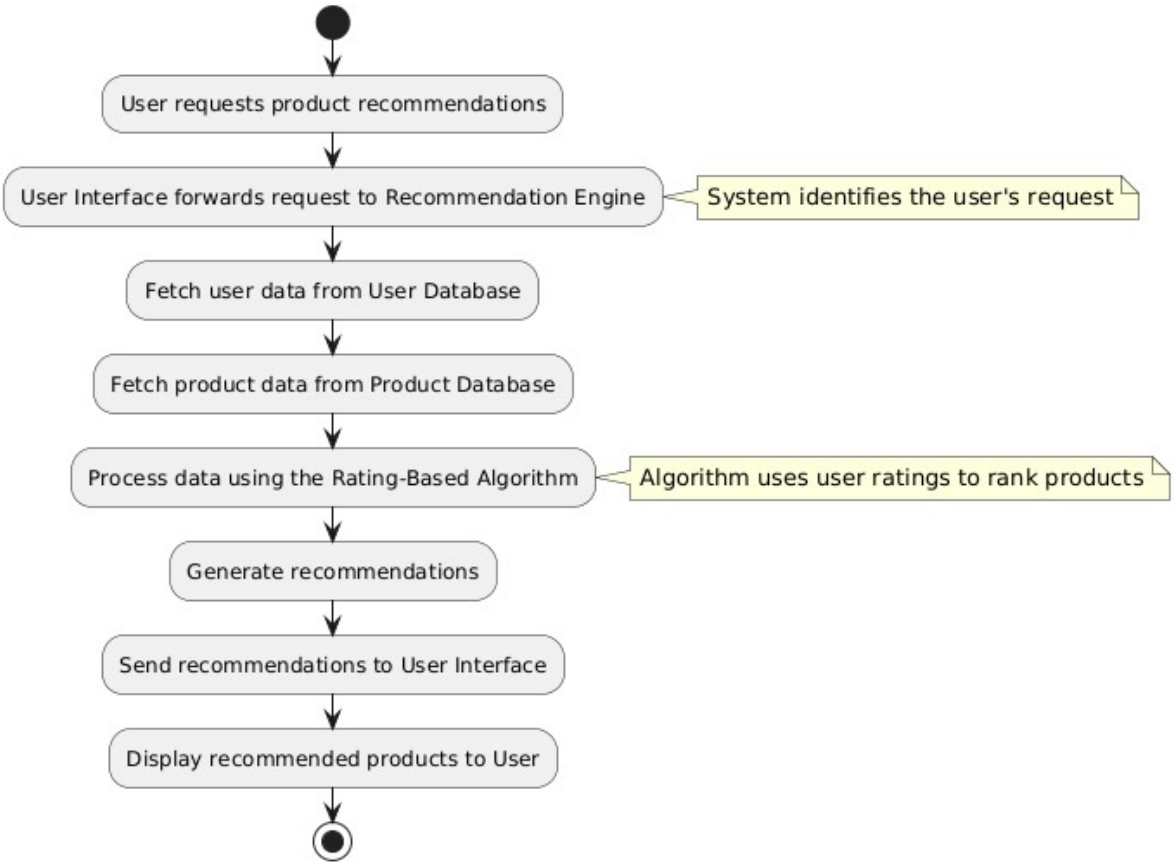

Figure 3: Data flow diagram of Proposed System

The proposed system is an E-Commerce Product Recommendation System that leverages machine learning algorithms to provide personalized product recommendations. The system integrates content-based filtering, collaborative filtering, hybrid models, multi-model approaches, and rating-based recommendations to offer a highly tailored shopping experience for users.

Key Features:

- **Content-Based Filtering:** This technique suggests products by analyzing their attributes (e.g., product category, description) and matching them to users' past preferences. Text-based methods such as TF-IDF and word embeddings will be used to extract features from product descriptions.
- **Collaborative Filtering:** By analyzing historical user behavior, such as ratings, clicks, and purchase history, this method predicts user preferences. Techniques like matrix factorization (e.g., SVD) will be used to uncover hidden relationships between users and products.

- **Rating-Based Recommendations:** This feature focuses on user-generated ratings to improve recommendation accuracy. The system will analyze ratings provided by users for different products and use algorithms such as neighborhood-based collaborative filtering (user-user or item-item) to recommend products with similar ratings. It will prioritize products that are highly rated by users with similar tastes.

- **Hybrid Recommendation:** A hybrid model will combine content-based filtering, collaborative filtering, and rating-based recommendations to increase accuracy. This method capitalizes on the strengths of each approach to deliver more personalized and relevant product suggestions.

- **Multi-Model Recommendation:** This approach integrates various machine learning models, such as neural collaborative filtering and deep learning models, to account for a variety of user behaviors and preferences.

0.5.1 Advantages of Proposed System

The proposed e-commerce product recommendation system, as depicted in the flowchart, incorporates several key improvements to address the limitations of the existing system:

- **Real-Time Data Processing:**

  Overcomes the limitation: By processing user data in real-time, the system can adapt more quickly to changing user preferences and behaviors. This ensures that recommendations remain relevant and up-to-date.

- **User Feedback Loop:**

  Overcomes the limitation: The inclusion of a user feedback loop allows the system to continuously learn and improve its recommendations. By collecting and analyzing user feedback, the system can fine-tune its algorithms and models to better align with user preferences.

- **Data Privacy and Security:**

  Overcomes the limitation: Explicitly addressing data privacy and security throughout the process ensures that user information is handled responsibly and in compliance with relevant regulations. This helps build trust with users and mitigates potential risks.

- **Scalable Cloud Storage:**

  Overcomes the limitation: Storing data in scalable cloud storage allows the system to handle large volumes of data efficiently and cost-effectively. This is particularly important as the system grows and the amount of user data increases.

- **Potential for Enhanced Recommendation Algorithms:**

  Overcomes the limitation: While not explicitly shown in the flowchart, the proposed system leaves room for the implementation of more advanced recommendation algorithms. This could include hybrid approaches that combine the strengths of different techniques to address the limitations of individual methods.

- **Contextual Understanding:**

  Overcomes the limitation: Although not explicitly depicted, the system could be enhanced to incorporate contextual information such as time of day, location, and user mood. This would allow for more personalized and relevant recommendations.

  By incorporating these enhancements, the proposed system aims to provide a more effective, efficient, and user-centric product recommendation experience. It addresses the key drawbacks of the existing system while ensuring data privacy and scalability.

0.6 System Architecture

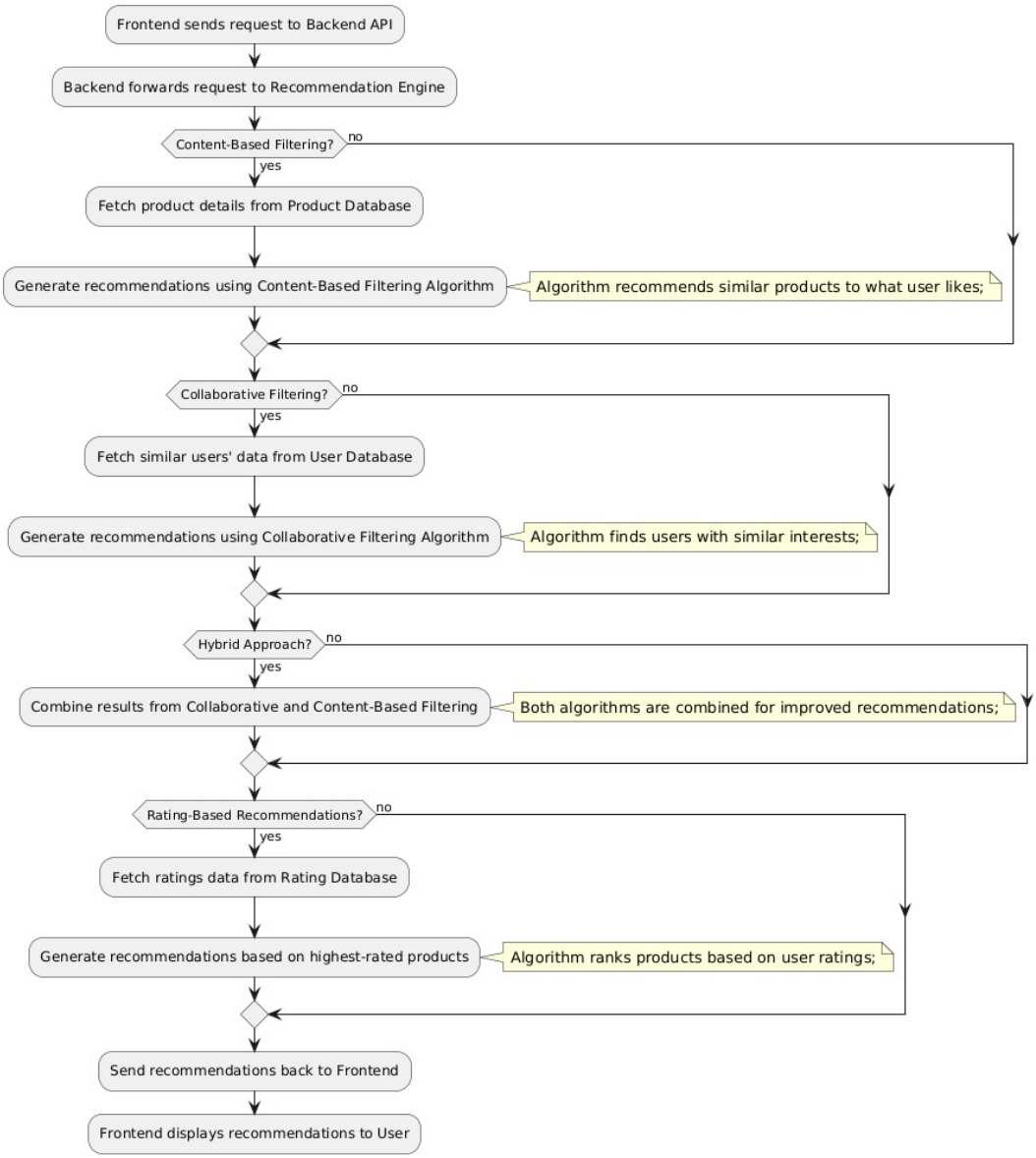

Figure 4: Sysytem Architecture

The system is designed with multiple recommendation algorithms such as content-based filtering,Collaborative Filtering, Hybrid Approach and rating-based recommendations. Flask serves as the web interface, allowing users to browse recommended products based on personalized suggestions.

0.7 Algorithm

In this project, each algorithm plays a crucial role in the E-Commerce Products Recommendation System by suggesting products to users based on different methodologies. Here's a brief explanation of each algorithm and its purpose in the project:

1. **Content-Based Filtering**

    - Content-based filtering recommends products by analyzing the attributes or characteristics of items that the user has interacted with in the past. It focuses on the similarities between products rather than user preferences.
    - **Purpose in Project:**In your e-commerce recommendation system, content-based filtering can recommend products based on attributes like the product category, brand, and descriptions. If a user interacts with or buys a particular product, the system suggests products with similar attributes, helping users discover related items.

2. **Collaborative Filtering**

    - Collaborative filtering recommends products by identifying patterns of user behavior. It uses historical data of user interactions, such as ratings, clicks, or purchases, and compares users with similar preferences.
    - **Purpose in Project:**This approach is useful for recommending products that other users with similar purchasing patterns have liked. Collaborative filtering helps when there's no direct connection between products but a strong connection between users' behavior and preferences.

3. **Rating-Based Recommendations**

    - Rating-based recommendations rank and suggest products based on user-generated ratings. Products with higher ratings are prioritized for recommendations.
    - **Purpose in Project:**This algorithm ensures that highly rated products are recommended to users, enhancing their trust in the recommendation system. It can be used as a complementary method to filter out products that are well-received by the community.

0.8 Wrokflow of Proposed Model

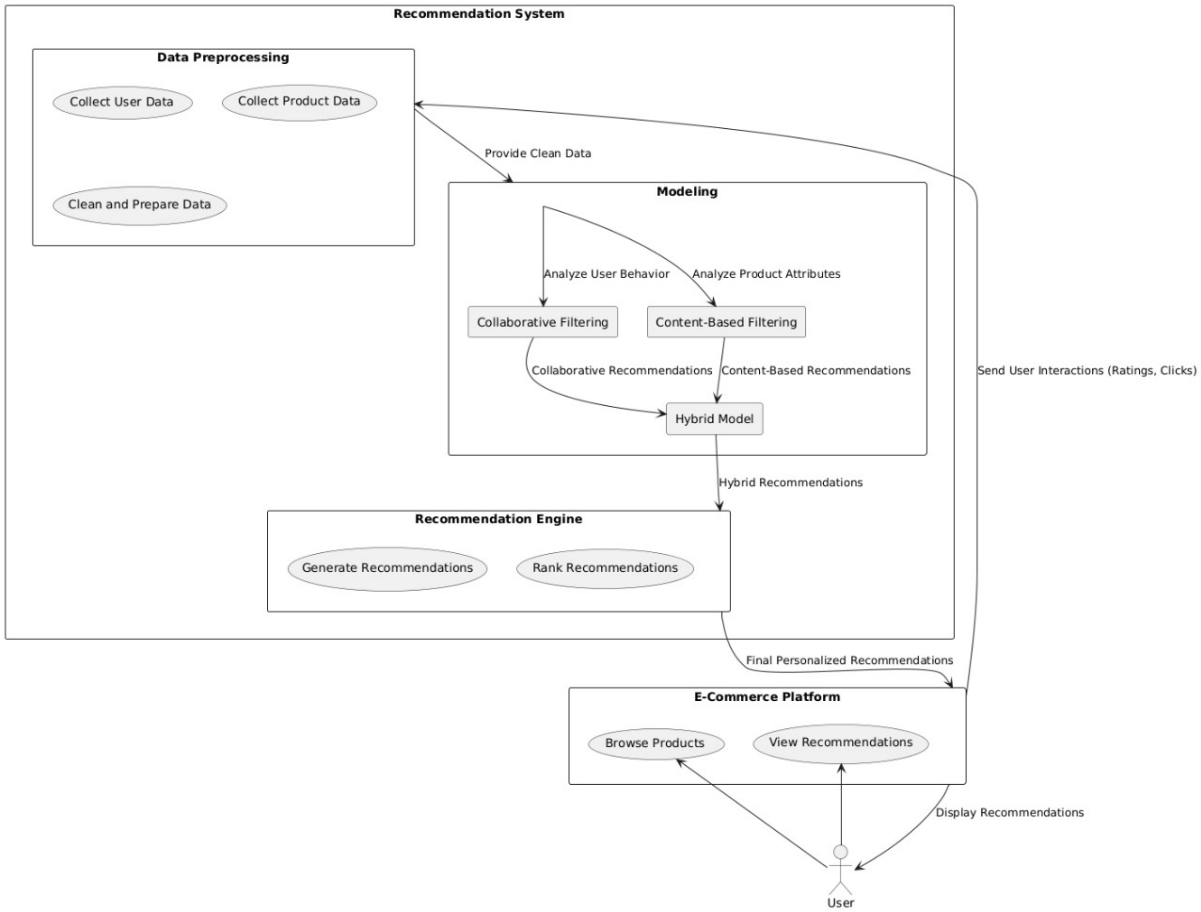

Figure 5: Workflow of Proposed Model

0.9 Hardware and Software Requirements

0.9.1 Hardware Requirements:

1. **CPU:** Multi-core processor
2. **RAM:** 8 GB minimum
3. **Storage:** 10 GB (SSD)

0.9.2 Software Requirements:

1. **OS:** Windows
2. **Python:** 3.7
3. **Libraries:**
    - **Pandas:** Data manipulation and analysis
    - **NumPy:** Numerical operations
    - **Matplotlib:** Plotting graphs
    - **Seaborn:** Statistical data visualization
    - **Scikit-Learn:** Machine learning algorithms
    - **SciPy:** Scientific computations
    - **SpaCy:** Natural language processing
    - **Text Editor:** Jupyter Notebook

0.10 Implementation

0.10.1 Load Packages and Libraries

Step 0: Load Packages and Libraries

[54]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

from sklearn.metrics.pairwise import cosine_similarity
from sklearn.feature_extraction.text import TfidfVectorizer

import os
from scipy.sparse import coo_matrix
```

Figure 6: Load Packages and Libraries

0.10.2 Data Loading and Preprocessing

[56]:
```
# Read your dataset
train_data = pd.read_csv(r'C:\Users\247386\OneDrive\Desktop\E-Commerece-Recommendation-System-Machine-Learning-Product-Recommendation-system--main\market

train_data.columns
```

[56]:
```
Index(['Uniq Id', 'Crawl Timestamp', 'Dataset Origin', 'Product Id',
       'Product Barcode', 'Product Company Type Source',
       'Product Brand Source', 'Product Brand Normalised Source',
       'Product Name Source', 'Match Rank', 'Match Score', 'Match Type',
       'Retailer', 'Product Category', 'Product Brand', 'Product Name',
       'Product Price', 'Sku', 'Upc', 'Product Url', 'Market',
       'Product Description', 'Product Currency',
       'Product Available Inventory', 'Product Image Url',
       'Product Model Number', 'Product Tags', 'Product Contents',
       'Product Rating', 'Product Reviews Count', 'Bsr', 'Joining Key'],
      dtype='object')
```

[57]:
```
train_data = train_data[['Uniq Id','Product Id', 'Product Rating', 'Product Reviews Count', 'Product Category', 'Product Brand', 'Product Name', 'Product
train_data.head(3)
```

[57]:

	Uniq Id	Product Id	Product Rating	Product Reviews Count	Product Category	Product Brand	Product Name	Product Image Url
0	1705736792d82aa2f2d3caf1c07c53f4	2e17bf4acecdece67fc00f07ad62c910	NaN	NaN	Premium Beauty > Premium Makeup > Premium Nail...	OPI	OPI Infinite Shine, Nail Lacquer Nail Polish, ...	https://i5.walmartimages.com/asr/0e1f4c51-c1a4...
1	95a9fe6f4810fcfc7ff244fd06784f11	076e5854a62dd283c253d6bae415af1f	NaN	NaN	Beauty > Hair Care > Hair Color > Auburn Hair ...	Nice'n Easy	Nice n Easy Permanent Color, 111 Natural Mediu...	https://i5.walmartimages.com/asr/9c8e42e4-13a5...
2	8d4d0330178d3ed181b15a4102b287f2	8a4fe5d9c7a6ed26cc44d785a454b124	4.5	29221.0	Beauty > Hair Care > Hair Color > Permanent Ha...	Clairol	Clairol Nice N Easy Permanent Color 7/106A Nat...	https://i5.walmartimages.com/asr/e3a601c2-6a2b...

Figure 7: Data Loading and Preprocessing

0.10.3 Data Cleaning and Tags Creation

Step 3: Data Cleaning and Tags Creations

[134]:
```
import pandas as pd
import spacy
from spacy.lang.en.stop_words import STOP_WORDS

train_data = pd.read_csv(r'C:\Users\247386\OneDrive\Desktop\E-Commerece-Recommendation-System-Machine-Learning-Product-Recommendation-system--main\market
# Initialize spaCy model
nlp = spacy.load("en_core_web_sm")

def clean_and_extract_tags(text):
    if pd.isna(text):
        return ''
    doc = nlp(str(text).lower())
    tags = [token.text for token in doc if token.text.isalnum() and token.text not in STOP_WORDS]
    return ', '.join(tags)

columns_to_extract_tags_from = ['Product Category', 'Product Brand', 'Product Description']

for column in columns_to_extract_tags_from:
    if column in train_data.columns:
        train_data[column] = train_data[column].astype(str).apply(clean_and_extract_tags)
    else:
        print(f"Column '{column}' does not exist in the dataset.")

train_data['Tags'] = train_data[columns_to_extract_tags_from].fillna('').apply(lambda row: ', '.join(row), axis=1)
```

[73]:
```
# Concatenate the cleaned tags from all relevant columns
train_data['Tags'] = train_data[columns_to_extract_tags_from].apply(lambda row: ', '.join(row), axis=1)
```

Figure 8: Data Cleaning and Tags Creation

0.10.4 Rating Based Recommendation System

Rating Base Recommendations System

[75]:
```
train_data = pd.read_csv(r'C:\Users\247386\OneDrive\Desktop\E-Commerece-Recommendation-System-Machine-Learning-Product-Recommendation-system--main\market
average_ratings = train_data.groupby(['Product Name', 'Product Reviews Count', 'Product Brand', 'Product Image Url'])['Product Rating'].mean().reset_inde
```

[76]:
```
top_rated_items = average_ratings.sort_values(by='Product Rating', ascending=False)

rating_base_recommendation = top_rated_items.head(10)
```

[77]:
```
rating_base_recommendation.loc[:, 'Product Rating'] = rating_base_recommendation['Product Rating'].astype(float)
rating_base_recommendation.loc[:, 'Product Reviews Count'] = rating_base_recommendation['Product Reviews Count'].astype(int)
```

[132]:
```
rating_base_recommendation.loc[:, ['Product Name', 'Product Rating', 'Product Reviews Count', 'Product Brand', 'Product Image Url']] = rating_base_recomm
rating_base_recommendation
```

[132]:

	Product Name	Product Reviews Count	Product Brand	Product Image Url	Product Rating
2466	Prell Shampoo, Classic Clean 13.50 oz (Pack of 4)	8.0	Prell	https://i5.walmartimages.com/asr/591617a5-4b84...	5.0
2840	Shea Soft & Smooth Body Milk Spray 6.8 oz	3.0	OGX	https://i5.walmartimages.com/asr/a055cdd5-695b...	5.0
2398	Philips Sonicare HX9023/65 C2 Optimal Plaque C...	1.0	Sonicare	https://i5.walmartimages.com/asr/767db079-1563...	5.0
2393	Peter Thomas Roth Water Drench Hyaluronic Clou...	2.0	Peter Thomas Roth	https://i5.walmartimages.com/asr/2178c851-e4ae...	5.0
862	Colgate Total Whole Mouth Health Gel Toothpast...	1.0	Colgate	https://i5.walmartimages.com/asr/51016094-008e...	5.0
2385	Perricone MD Face Finishing Moisturizer 118ml/...	2.0	Perricone MD	https://i5.walmartimages.com/asr/c958cad5-7996...	5.0
2381	Perfecting Neck and Decollete Cream	1.0	PCA Skin	https://i5.walmartimages.com/asr/7d49107b-85a2...	5.0
869	Colorescience Sunforgettable Brush On Sunscree...	3.0	ColoreScience	https://i5.walmartimages.com/asr/e4582fe9-7bc7...	5.0
2379	Perfect Bar, Dark Chocolate Chip Peanut Butter...	2.0	Perfect Bar	https://i5.walmartimages.com/asr/8a3dfffc-0f47...	5.0
874	Comfort Bath Heavyweight Aloe Cleansing Washcl...	2.0	Sage	https://i5.walmartimages.com/asr/c2a4f9f1-0e3e...	5.0

[ ]:

Figure 9: Rating Based Recommendation System

0.11 Future Scope

The future work will focus on improving recommendation accuracy by incorporating advanced machine learning models and user feedback. Key future developments include:

- Content Base Filtering
- Collaborative Filtering
- Multi-Model Approaches

0.12 References

1. Thanh Vu Ngoc and Huong Tran Thi.Techniques, benefits, and challenges of recommendation system in e-commerce: A literature review.Publish Year:2021

2. Md. Zahurul Haque, Department of CSE, Manarat International University, Dhaka. Bangladesh.E-Commerce Product Recommendation System based on ML Algorithms.Publish Year:2020

3. Shreyas Ghansawant1 , Gaurang Kumbhar2 , Prof. Nikita kawase3 ,Prof. Deepak K. Sharma4.The Power of Personalization: Enhancing User Experience in E-commerce.Publish Year:2023

4. Xiangnan He, Lizi Liao, Hanwang Zhang.Neural Collaborative Filtering Publish Year:2017

5. Yehuda Koren, Yahoo Research Robert Bell and Chris Volinsky, ATT Labs-Resear ch.Matrix factorization Techniques for Recommender system.Publish Year:2019