Personalized Recommendation System

DEPI Graduation Project

Ministry of Communications and Information Technology

Supervisor Name: Mohammed Agoor

Group name: ALX_AIS5_S3e

	الاسم
1	ميار محمود خليل
2	ميار مصطفى حسن
3	سناء محمد مصطفى
4	حاتم محمود محمد عبد الهادي
5	محمد أمير فتحي الحوفي
6	محمد صبري الطيبي منصور

Table of contents:

- Abstract:

This project presents a personalized recommendation system developed using the Amazon Fine Food Reviews dataset. The primary objective is to provide users with tailored product suggestions that reflect their preferences. The system leverages ensemble learning techniques combined with collaborative filtering to generate recommendations. Specifically, it outputs a diverse array of products, including those based on user preferences, items predicted to align with user tastes, and trending products related to the user's interests. The approach aims to enhance user experience by providing relevant recommendations that cater to individual tastes, making it a valuable tool for enhancing customer engagement in the food industry. Future enhancements include the categorization of data for more precise recommendations and the integration of user feedback to continually refine the model.

- Introduction:

In an era where consumers are inundated with choices, the ability to provide personalized recommendations has become a critical component of successful online retail platforms. Personalized recommendation systems enhance user experience by filtering through vast product offerings to present users with items tailored to their tastes and preferences. This project focuses on developing a sophisticated recommendation system using the Amazon Fine Food Reviews dataset, which consists of rich information regarding consumer feedback on various food products.

The recommendation system employs a hybrid approach, integrating ensemble learning techniques with collaborative filtering methods. This combination allows the system to leverage both the collective behavior of users and the unique preferences of individuals, resulting in more accurate and meaningful product suggestions. By analyzing user interactions and product ratings, the system identifies patterns that inform its recommendations.

The output of this system includes a balanced mix of products: five items specifically reflecting user preferences based on past ratings, five predicted products expected to resonate with the user's taste, and three hot products that are trending in the food market. This structure not only caters to individual user preferences but also introduces novelty through trending items, making the shopping experience more engaging.

As online shopping continues to evolve, the importance of effective recommendation systems cannot be overstated. They are essential for improving customer satisfaction, increasing conversion rates, and fostering brand loyalty. This project aims to contribute to this field by creating a robust recommendation system that aligns closely with user expectations and market trends.

• Methodology:

The methodology employed in this project for developing the personalized recommendation system consists of several key steps: data preparation, model selection, implementation of ensemble learning techniques, and evaluation of the recommendation performance.

1. Data Preparation:

The foundation of any recommendation system lies in the quality and relevance of the data used. This project utilizes the Amazon Fine Food Reviews dataset, which contains over 500,000 reviews spanning more than a decade. The dataset includes user IDs, product IDs, ratings, and textual reviews, providing a comprehensive view of user interactions with food products.

Data preparation involved the following steps:

- **Data Cleaning**: Remove duplicates, handle missing values, and ensure that the data is in the correct format for analysis.
- **Feature Selection**: Focus on essential features such as user ID, product ID, and ratings while discarding irrelevant data.
- **Data Filtering:** Using users that made at least 10 reviews.

2. Model Selection:

The primary approaches were employed in the recommendation system are: collaborative filtering, SVD and ensemble learning.

- **Singular Value Decomposition (SVD)**: SVD is a powerful matrix factorization technique used in collaborative filtering. It decomposes the user-item interaction matrix into three matrices: one representing user preferences, one representing item attributes, and a diagonal matrix containing singular values that capture the importance of latent factors. By retaining only the top singular values, SVD reduces noise and enhances the ability of the model to generalize from the data. This results in improved prediction accuracy, particularly for users or items with sparse data.
- **Collaborative Filtering**: This method analyzes user behavior and identifies patterns based on the interactions of similar users. Both user-based and item-based collaborative filtering techniques were utilized, allowing the system to make recommendations based on the preferences of like-minded users or similar products.
- **Ensemble Learning**: This approach combines multiple models to improve prediction accuracy. By integrating different algorithms, the system enhances its ability to capture various aspects of user preferences and item characteristics.

3. Implementation:

The implementation phase involved coding the recommendation algorithms and incorporating them into a system. The following steps were taken:

- **Model Training**: Train the recommendation models on the user-item interaction data. For collaborative filtering, models such as KNNWithMeans and SVD were utilized to capture similarities between users and items.
- **Prediction Generation**: The trained models were used to generate recommendations for users. This included calculating scores for potential products based on historical user behavior and predicted preferences.

- **Combining Results**: The final recommendations were generated by combining the outputs from user-based and item-based collaborative filtering with ensemble learning techniques, including SVD, to produce a well-rounded set of suggestions.

4. Evaluation:

To evaluate the performance of the recommendation system, various metrics were employed, including:

- **Root Mean Square Error (RMSE)**: To measure the differences between predicted ratings and actual ratings, providing insight into the accuracy of the recommendations.
- **Mean Absolute Error (MAE)**: This metric offers an alternative perspective on prediction accuracy by calculating the average magnitude of errors in a set of predictions.

5. MLOps :

MLOps, or Machine Learning Operations, is an emerging discipline that combines machine learning, DevOps, and data engineering practices to streamline and automate the end-to-end lifecycle of machine learning models. In the context of this project, MLOps principles are essential for ensuring that the personalized recommendation system is not only developed effectively but also deployed, monitored, and maintained efficiently in a production environment.

- ## Results:

Firslt MLOps experiments:

The following are the results of some experiments that have been done using SVD , user-based & item-based collabotive filtering:

KNNWithMeans approach:

User-based:

	Model Type	K	Similarity	RMSE	MAE
1	KNNWithMeans	30	cosine	0.921888	0.592654
2	KNNWithMeans	50	cosine	0.922226	0.592230
3	KNNWithMeans	30	pearson	0.922576	0.592345
4	KNNWithMeans	40	cosine	0.923451	0.593613
5	KNNWithMeans	40	pearson	0.923671	0.593612
6	KNNWithMeans	10	cosine	0.925180	0.594892
7	KNNWithMeans	10	pearson	0.925403	0.595703
8	KNNWithMeans	10	pearson	0.925449	0.595118

Item_based:

	Model Type	K	Similarity	RMSE	MAE
0	KNNWithMeans	10	pearson	0.899115	0.505366
1	KNNWithMeans	40	cosine	0.908886	0.544615
2	KNNWithMeans	10	cosine	0.911373	0.542600

SVD approach:

	Model Type	Learning Rate	RMSE	MAE
0	SVD	0.1	0.798877	0.477719
1	SVD	0.05	0.803777	0.489880
2	SVD	0.01	0.831311	0.544022
3	SVD	0.005	0.873406	0.612804
4	SVD	0.3	1.455030	0.818055

Comment: As we notice that the best approach is SVD with RMSE equals to 0.799.

SVD is better with lr = 0.1

User_based is better with cosine similarity and k = 30.

Item_based is better with pearson similarity and k = 10.

We will use these parameters in the Ensamble Learning with different weights.

Second using Ensemble Learning:

We employed ensemble learning to enhance the recommendation system's performance. This method integrates multiple collaborative filtering techniques, leveraging their strengths to improve prediction accuracy. Specifically, we utilized the Singular Value Decomposition (SVD) method with a weight of 0.7, which allows the model to capture the underlying structure of the data effectively.

Additionally, we incorporated the user-based K-Nearest Neighbors (KNN) algorithm with a weight of 0.15, enabling personalized recommendations by considering similar users' preferences. The item-based KNN method was also included, assigned a weight of 0.15, which focuses on the relationships between items to suggest products that align with the user's tastes.

SVD weight = 0.7

knn_user weight = 0.15

knn_item weight = 0.15

The following are the metrics produced:

```
RMSE: 0.7962
MAE:  0.4914
```

Comment: This weighted ensemble approach helps balance the influence of each model, providing a robust and nuanced recommendation output that can cater to diverse user preferences.

Usage on a specific user:

Let us try our ensemble learning approach on user (user_id = 'A383XURHVF8ON6') for example. The following result are 3 trendy products related to the user preference, and the other 10 are 5 of the top rated products by the user and 5 of the unrated products which user is expected to like them.

```
Recommendations for user A383XURHVF8ON6:

Hot Products (Trending):
- Product ID: B002IEVJRY
- Product ID: B002IEZJMA
- Product ID: B0026KPDG8

The 10 Recommended Products:
- Product ID: B00135XQCK
- Product ID: B001FSK3PI
- Product ID: B0009JOX2S
- Product ID: B002T0EX7U
- Product ID: B000E158ES
- Product ID: B0002Z9BF8
- Product ID: B001DY6TWU
- Product ID: B000EMOD3O
- Product ID: B000VVT9IM
- Product ID: B001HN5Z4K
```

For more details about the trendy top products score:

The score of the product = Number Of Ratings * Average Rating

```
Top 3 Recommended Hot Products for user: A383XURHVF8ON6
Item ID: B002IEVJRY, Popularity Score: 1433.0, Predicted Rating: 4.2342337655058415, Number of Ratings: 373, A
verage Rating: 3.841823056300268
Item ID: B002IEZJMA, Popularity Score: 1379.0, Predicted Rating: 4.519109450255137, Number of Ratings: 390, Av
erage Rating: 3.5358974358974358
Item ID: B0026KPDG8, Popularity Score: 1299.0, Predicted Rating: 5.0, Number of Ratings: 304, Average Rating:
4.2730263157894735
```

The 10 products rating:

```
Top 5 Actual Products for user A383XURHVF8ON6:
Item ID: B0002Z9BF8, Actual Rating: 5
Item ID: B001DY6TWU, Actual Rating: 5
Item ID: B000EMOD3O, Actual Rating: 5
Item ID: B000VVT9IM, Actual Rating: 5
Item ID: B001HN5Z4K, Actual Rating: 5

Top 5 Unrated Product Predictions for user A383XURHVF8ON6:
Item ID: B000EDGBDI, Combined Prediction: 5.0
Item ID: B001J8H1E0, Combined Prediction: 5.0
Item ID: B000EDG3UE, Combined Prediction: 5.0
Item ID: B0029NTQ1K, Combined Prediction: 5.0
Item ID: B0083QJU5O, Combined Prediction: 5.0
```

- Conclusion:

In this project, we developed a personalized recommendation system utilizing the Amazon Fine Food Reviews dataset. By combining advanced techniques such as ensemble learning and collaborative filtering, we were able to create a robust model that provides tailored product suggestions to users. The integration of multiple methods, including Singular Value Decomposition (SVD) and K-Nearest Neighbors (KNN), allowed us to leverage the strengths of each approach, resulting in improved accuracy and user satisfaction.

Our findings demonstrated that the system effectively recommends products based on user preferences, alongside trending items that enhance user engagement. The structured approach to implementing ensemble learning, with thoughtfully assigned weights to each algorithm, contributed significantly to the overall performance of the recommendation engine.

Looking ahead, there are numerous avenues for enhancing the system further, such as incorporating user feedback mechanisms and exploring additional recommendation algorithms. These improvements will enable the model to adapt dynamically to changing user preferences, ultimately leading to a more personalized and satisfying user experience.

Overall, this project highlights the potential of machine learning in revolutionizing the e-commerce landscape by providing users with intelligent and relevant recommendations, thereby enhancing their shopping experience.

- ## Future Work:

As we look to improve and expand the capabilities of the personalized recommendation system, several opportunities for future work have been identified:

1. **Category-Specific Recommendations**: By categorizing the dataset into distinct product categories, we can enhance the recommendation process. This will allow for more tailored suggestions that align with user interests within specific domains.
2. **Incorporation of User Feedback**: Implementing a feedback mechanism will enable users to rate the recommendations they receive. This feedback can be utilized to refine the model and improve future predictions, ensuring that the system evolves with changing user preferences.
3. **Exploration of Additional Recommendation Algorithms**: While the current system employs SVD and KNN methods, further exploration of alternative algorithms, such as matrix factorization techniques or deep learning approaches, could yield even better performance and accuracy in recommendations.