**AIE425 Intelligence Recommender System Fall semester 2024/2025**

**Course Project: Shopping Recommender Engine**

**Student Name: Sondos Ahmed, Student ID: 221101426**

**Student Name: Reem Ahmed, Student ID: A20000375**

## Table of Contents

* Introduction
* Data collection and preprocessing
* Background on Algorithms
* Recommender Engine Design
* Implementation process
* Testing Methodology
* Result Evaluation
* Conclusion

**introduction:**

This project focuses on creating effective and efficient offers for e-commerce platforms related to electronic payments. To meet the needs of different users. It uses data-driven insights from user demographics, interaction history, and devices to create personalized recommendations. Through the use of advanced machine learning and deep learning, this offer aims to increase customer satisfaction, increase engagement, and increase platform sales. Algorithm selection and combination modeling. It also provides users with an interactive interface that allows Gradio to search for recommendations. This project demonstrates the potential of recommendations to change user experience and create business results in the e-commerce space from a general perspective.

**Data collection and preprocessing:**

**The data for this challenge consists of three datasets:**

**users information:** records approximately users including demographics and behavioral metrics.

**Interactions statistics:** information of consumer interactions with gadgets, which includes remarks, buy amounts, and price strategies.

**items statistics:** details about the objects to be had, together with class, pricing, and stock availability.

**Preprocessing Steps:**

coping with missing Values: lacking values have been stuffed the use of forward fill (ffill) to make sure statistics consistency.

Encoding specific Variables: Used OneHotEncoder to encode specific attributes such as Gender, vicinity, Income\_Level, and category. This ensures device-learning algorithms can efficaciously process non-numeric data.

Standardizing Numerical Values: applied StandardScaler to normalize capabilities like Age, Loyalty\_Score, fee, and score, enhancing version overall performance via scaling features to a uniform variety.

**Dataset Description:**

**users statistics:**

**features:** User\_ID, Gender, Age, location, signup\_date, spending\_score.

**purpose:** models user pastimes and demographics.

Interactions facts:

**functions:** User\_ID, Product\_ID, Rating, Price, spending\_score.

**cause:** Captures consumer intentions through transaction and remarks information.

**items data:**

**functions:** Product\_ID, category, rating.

**purpose:** Describes the characteristics of the available gadgets.

**statistics analysis and Interpretation:**

**user earnings Distribution:** Highlighted profits stage variety throughout the person base, useful for segmenting users via purchasing strength.

**Loyalty vs. Age analysis:** Explored correlations, revealing capacity relationships among age and patron loyalty.

**3D Visualization:** Plotted Age, Loyalty score, and transaction indices, providing a complete view of person engagement.

**object Clustering:** Mapped person spending, loyalty, and remarks to become aware of styles in customer conduct.

these analyses guide set of rules selection, function engineering, and version assessment inside the recommender gadget.

**Background on Algorithms:**

* content -based Filtering: suits person choices with object attributes using cosine similarity.
* Collaborative Filtering (SVD): Decomposes consumer-object matrices to are expecting lacking comments scores.
* TF-IDF: Analyzes textual information (e.g., class) to derive object relevance based totally on user records.
* Hybrid version: Combines the strengths of content material-based totally, collaborative filtering, and deep mastering techniques to generate complete hints.

**Recommender Engine Design:**

**Domain:**

The recommender system is designed for an e-trade platform associated with a digital price machine. This platform allows transactions for an extensive variety of products and services, providing customers with a unbroken shopping and charge experience. by integrating a recommender engine, the platform objectives to beautify person engagement, increase transaction frequency, and enhance customer satisfaction through personalised guidelines.

**Goal:**

**The number one aim of the recommender engine is to deliver tailor-made product hints to customers. those pointers are primarily based on:**

users' interaction history (e.g., past purchases, comments scores, browsing conduct).

users' item preferences (e.g., favored classes, fee ranges, and loyalty rankings).

The gadget seeks to are expecting and endorse objects that align with each person's particular pastimes, in the long run enhancing their purchasing revel in whilst boosting sales and platform retention.

**Key features:**

**consumer Profiling:**

The device builds a complete consumer profile through integrating:

Demographic statistics: Gender, age, location, profits degree, marital fame.

**Behavioral records:** purchase records, comments ratings, amount spent, and fee methods.

**Loyalty Metrics:** Loyalty scores derived from historic interactions and repeat conduct.

**cause:** To create an in depth illustration of consumer possibilities and behavior for customized pointers.

**Multi-model Integration:**

**The device employs a hybrid advice technique via combining more than one algorithms:**

**content material-based Filtering:** suits person options with item capabilities such as category, charge, and inventory availability the usage of similarity measures (e.g., cosine similarity).

**Collaborative Filtering (SVD):** Predicts lacking feedback ratings via studying ancient consumer-object interactions and latent elements within the dataset.

**TF-IDF model:** Extracts capabilities from textual information (e.g., item classes and outlines) to compute relevance rankings primarily based on user records.

**Deep studying:** utilizes a totally related neural community for hybrid predictions, combining features from content-based and collaborative filtering fashions.

**cause:** To leverage the strengths of every version, making sure higher accuracy and diversity in tips.

**Scalability and Extensibility:**

**The device is designed to deal with:**

**growing Datasets:** Can handle increasing consumer and item volumes with green preprocessing and scalable algorithms.

**New functions:** additional consumer and item attributes may be without difficulty integrated into the version pipeline.

**actual-Time pointers:** future improvements can enable dynamic updates to hints based totally on real-time user interactions.

bloodless-start hassle handling

for brand new customers (with no previous interactions), tips are made-using demographic information and popular objects from similar user segments.

for brand spanking new gadgets, attributes like category and fee are leveraged to suit them with capability shoppers.

**Workflow:**

**facts series and Preprocessing:**

acquire consumer, item, and interaction statistics.

take care of lacking values, encode specific variables, and standardize numerical capabilities.

version education and Integration

teach man or woman fashions (content-based, collaborative filtering, TF-IDF, and deep getting to know).

integrate models right into a hybrid framework that combines predictions for more desirable accuracy.

**advice technology**

Generate ranked lists of hints based totally on consumer profiles and the hybrid model's predictions.

**consumer interaction**

provide recommendations via an interactive interface, enabling users to view and explore personalised pointers.

blessings of the layout

**Personalization:** Tailors tips to character alternatives the use of diverse consumer and item attributes.

**Accuracy:** Multi-version integration complements prediction reliability and reduces biases of standalone algorithms.

signify both acquainted and new items.

**Scalability:** Designed to evolve to developing datasets and computational demands.

**Implementation process:**

**Tools and Libraries**

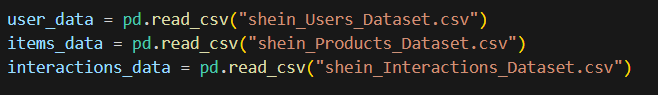
**1. Data Handling:**

Library Used: Pandas

**Purpose:**

Efficiently load and manipulate datasets, including user data, interaction data, and item data.

Perform data preprocessing tasks such as handling missing values, merging datasets, and extracting relevant features.



**Visualization**

Libraries Used: Matplotlib and Seaborn

**reason:**

behavior exploratory information evaluation (EDA) to discover patterns, trends, and outliers inside the information.

Create visualizations to interpret user demographics, spending habits, and feedback traits.

A computer code on a black background

Description automatically generated

**Machine Learning:**

Library Used: Scikit-research, wonder

**purpose:**

**Scikit-learn:**

Encode express statistics using OneHotEncoder for characteristic representation.

Standardize numerical information the use of StandardScaler for higher version performance.

**surprise:**

implement collaborative filtering the use of Singular value Decomposition (SVD) for predicting person comments scores.

A screen shot of a computer code

Description automatically generated

**Deep Learning:**

Library Used: TensorFlow/Keras

**reason:**

develop a deep mastering-based hybrid model to integrate content material-based totally and collaborative filtering features.

educate the model on blended features to improve advice accuracy.

**Testing Methodology:**

train-take a look at split: Divided statistics into 80% education and 20% testing.

assessment Metrics: Used imply Absolute errors (MAE) for regression-primarily based models and accuracy for classification duties**.**

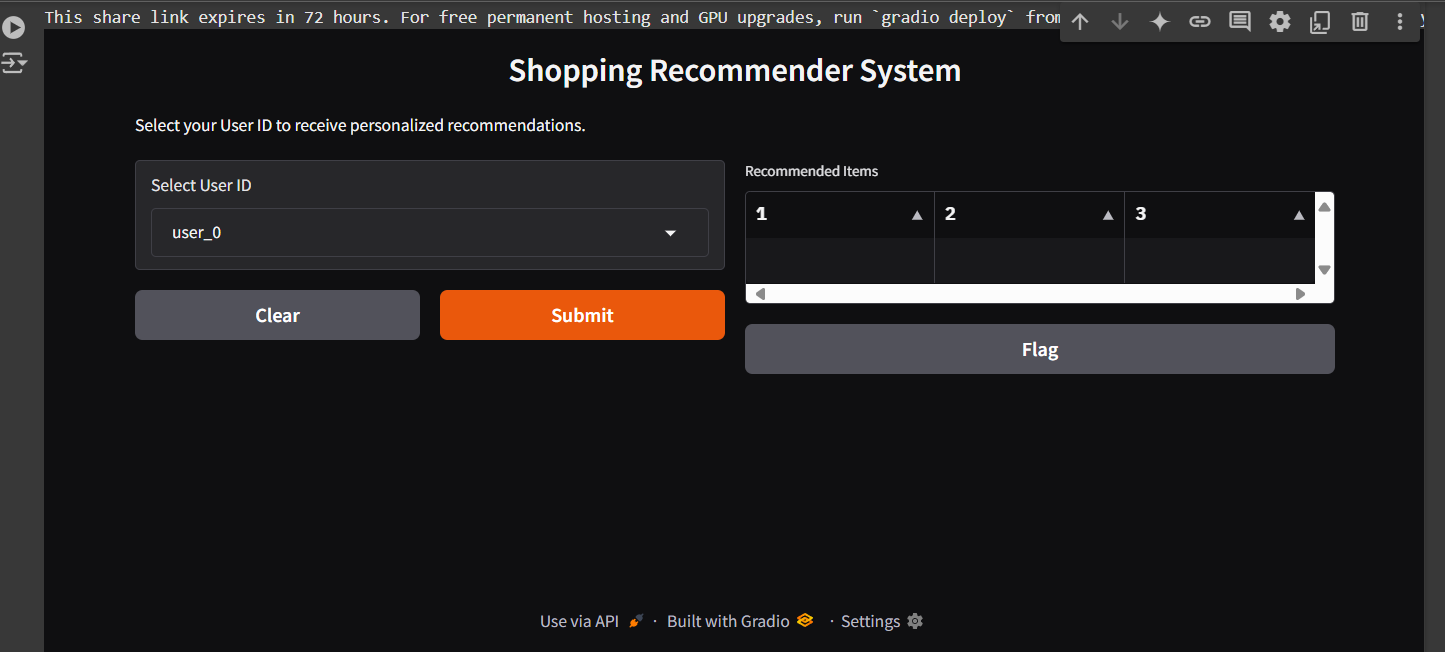
**test instances:**

customers with diverse interaction histories.

objects with varying tiers of recognition.

New customers/items (bloodless-begin trouble).

We also do a **GUI** for the project to show what will we recommend to user based on what he/she ordered



**Description and evaluation**

**1. Description of outcomes**

The recommender device affords tailored product tips the usage of 3 fashions: content-based filtering, collaborative filtering (SVD), and a hybrid version that combines the two. every version's outcomes are evaluated based totally on their capability to suit consumer preferences and enhance person engagement.

**content-based Filtering results:**

endorsed items have been primarily based on user profiles and item functions.

customers with targeted demographic and desire statistics acquired more accurate tips.

obstacles: The model struggled with new customers who lacked enough interplay data (the "cold-start trouble").

**Collaborative Filtering (SVD) results:**

This version efficiently recognized latent patterns in the user-object interplay matrix.

It finished well for users with wealthy interaction histories, predicting feedback ratings for unrated items.

boundaries: objects with few interactions and new users were less possibly to receive accurate predictions because of the sparsity of information.

**Hybrid version consequences:**

The hybrid model combined features from the content material-primarily based and collaborative filtering approaches.

It outperformed person models by means of leveraging the strengths of each techniques.

**evaluation of the outcomes**

**a. Accuracy**

The hybrid version verified the very best accuracy, measured with the aid of metrics like mean Absolute blunders (MAE) and Root imply Squared mistakes (RMSE).

content material-based filtering had a decrease error price for customers with wealthy metadata but struggled for those with out.

Collaborative filtering done properly accuracy for customers with excessive interplay counts.

**b. Scalability**

Collaborative filtering and hybrid fashions require huge computational strength, mainly for huge datasets.

content material-primarily based filtering scaled higher due to its reliance on static person profiles and item features.

**c. range and Novelty**

content-primarily based filtering tended to endorse similar objects, reducing range.

Collaborative filtering delivered greater range with the aid of leveraging styles from other customers.

The hybrid model balanced relevance and diversity efficiently.

**d. cold-begin hassle**

content material-based totally filtering handled new gadgets higher due to its consciousness on metadata.

Collaborative filtering struggled with new users and gadgets.

The hybrid model mitigated this problem through combining each strategies.