Personalized Review Ranking System for [Enhanced Shopper Decision-Making in E-Commerce]

Movies & TV Recommender

Our Demo

https://drive.google.com/file/d/1LYBNFM0oa_05AQ0Sx6t2Bd_aAze-m7su/view?usp=sharing

Team Members

Asmaa Fares
1210084

Ghasssan Qandeel
1212397

Alhasan Manasrah
1211705

Suhaib Badaha
1210472

Introduction

This project focuses on the development of a **personalized movie recommendation system** using **Ridge Regression**, a machine learning technique that ranks and personalizes movie reviews for users. The goal is to provide users with **tailored movie reviews based on their preferences** and **past interactions** with the system. The recommendation process leverages a combination of **text analysis**, **sentiment detection**, **user profiling**, and **machine learning** algorithms to enhance the accuracy and relevance of the reviews displayed to the user.

Data Collection & Preprocessing

1-Data Collection and Import:

We collected the dataset containing movie and TV reviews , The dataset includes columns such as **reviewText**, **asin**, and **reviewText_ar**

2-Data Cleaning & Preprocessing:

We performed cleaning on the dataset by **removing unnecessary columns** and ensuring that the review data was in a usable format. We also **handled missing values** and **applied text preprocessing**

3-Text Translation:

we used Google Translate API to translate the original reviews

5-Data Merging and Indexing:

We **merged** the translated Arabic reviews (**reviewText_ar**) back into the **same dataset** with a matching index.

4-Feature Engineering:

After preprocessing the reviews, we **created new features** for better analysis and machine learning purposes, such as **sentiment** features, **domain-specific** features (**like mentions of war/history**), and other **text-based** features

Personalization

User Profile Creation

- Collect **user preferences** and **interests** (e.g., history, emotional impact) from their **queries and past behaviors.**

Feature Extraction & Sentiment Analysis

- Extract review features like **length, word count**, and **sentiment**, along with **domain-specific** keywords (e.g., "history," "war") to assess review relevance.

Implementation Plan

- **Calculate** a personalized **score** for each review, **prioritizing those most relevant** to the user's interests, ensuring tailored recommendations.

Machine Learning Model

Ridge Regression (ML-based Ranking System)

- **Purpose:** Used to rank reviews based on personalized relevance scores and other features.

- **Key Functions:**

1- train_ranking_models_ridge_only: Trains a Ridge Regression model for ranking.

2- phase3_ml_ranking_system_ridge_only: Applies the trained Ridge model to rank reviews for each user.

- **Advantages**: Regularized model to prevent **overfitting** and effectively **rank** reviews **based on multiple features.**

Evaluation Metrics

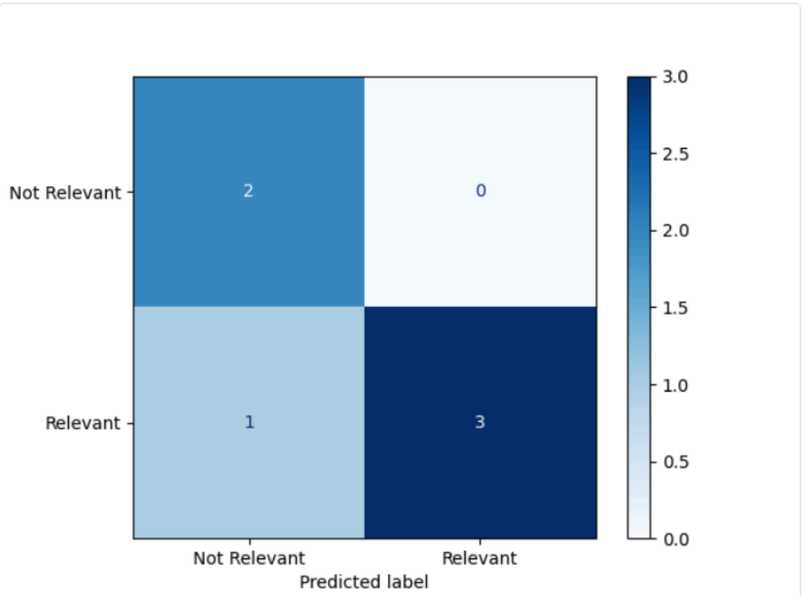
- **Precision**: A precision of 1 means that all of the recommendations made by the system were relevant.

- **Recall**: A recall of 0.75 it successfully recommended three out of every four items that were actually relevant.

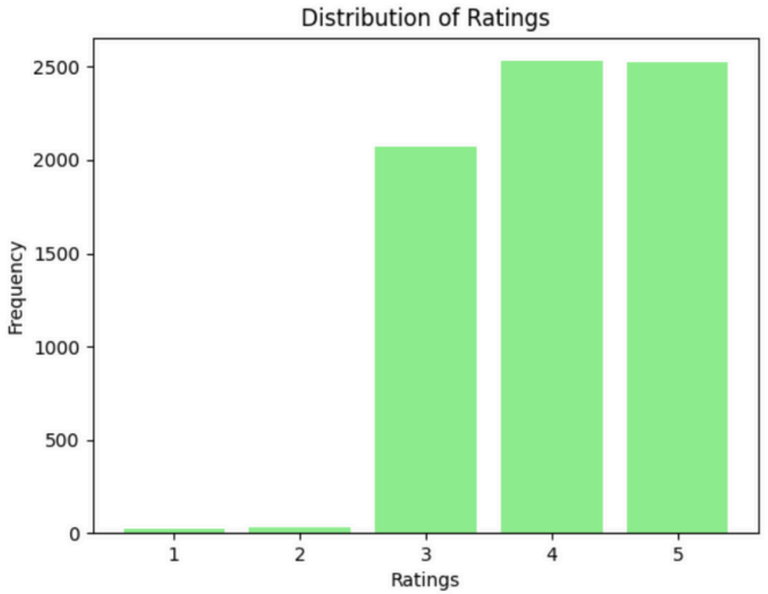
- **F1 Score**: The F1 score of 0.85 confirms that the system is performing well, maintaining a good trade-off between precision and recall.

Data Analysis

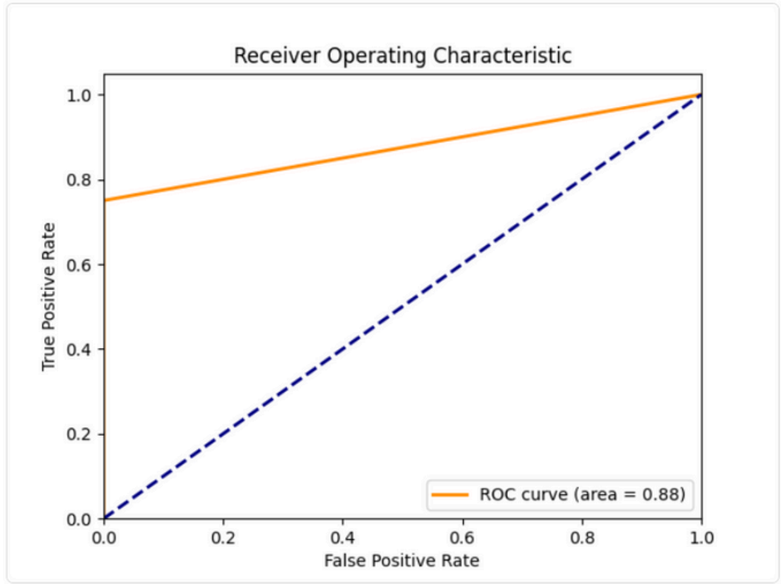
Confusion Matrix



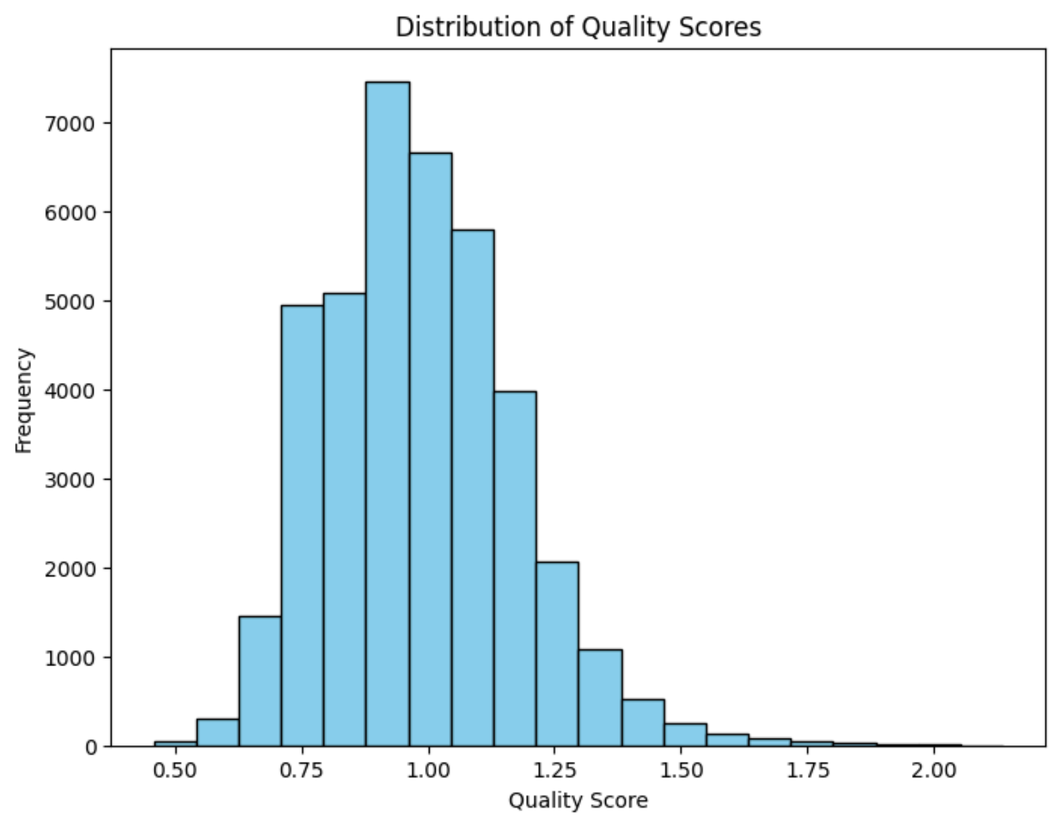
Rating Distribution

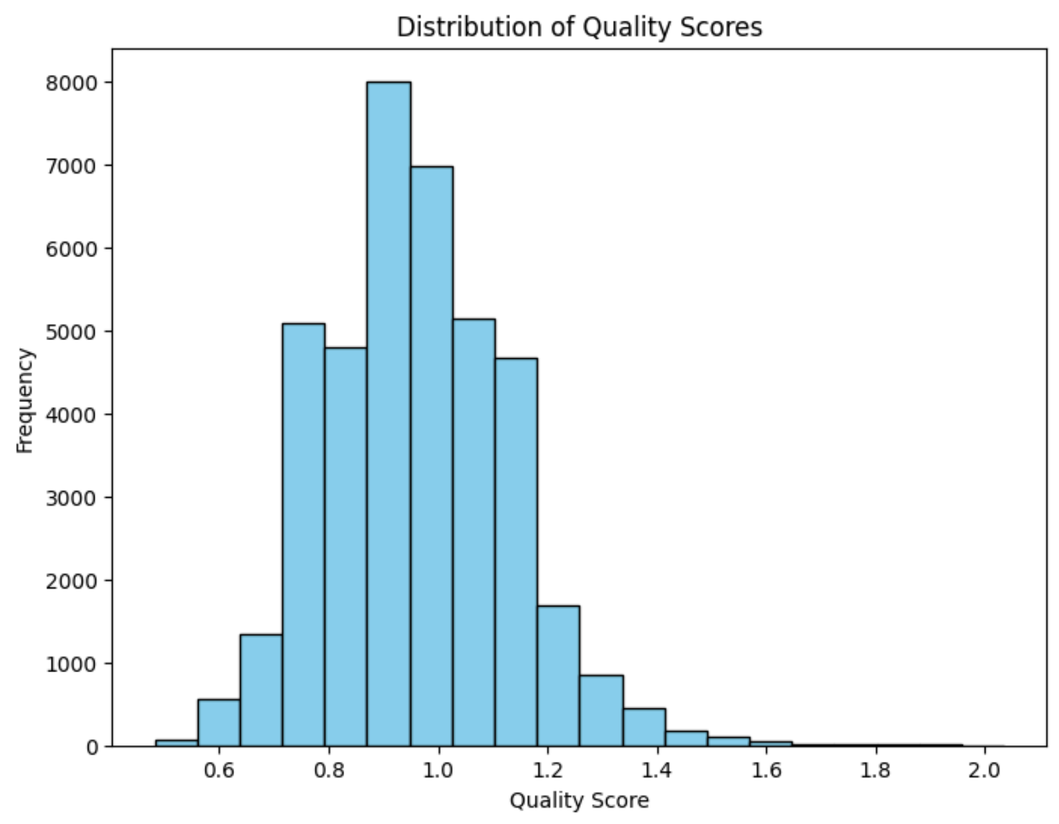


ROC Curve

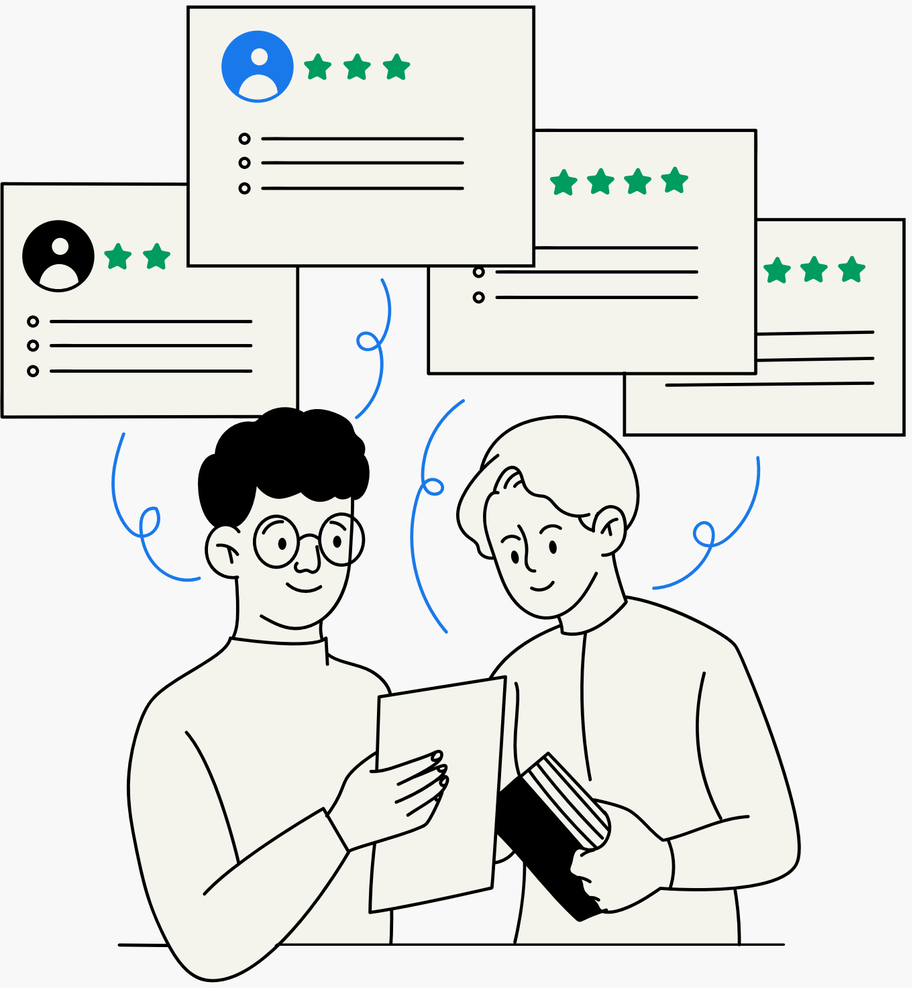


Distribution of Quality Scores for English and Arabic Reviews





Recommendation Generation (to sum up)



1- The Ridge Regression model ranks product reviews based on user preferences and behaviors.

2- Reviews are assigned personalized scores to determine their relevance to the user.

3- The **top k** reviews are selected as personalized recommendations for the user.

4- Each recommendation is accompanied by an **explanation** that justifies why it's relevant to the user .

Conclusion

In this project, we learned how to develop a **personalized recommendation system** by using machine learning techniques like Ridge Regression to rank and suggest products based on **user preferences**.

We also explored the importance of **feature engineering**, particularly focusing on *sentiment analysis*, *content relevance*, and *recency* to enhance the quality of recommendations.

Through **Natural Language Processing** techniques such as **sentiment analysis** and **extracting** domain-specific features , We also handled Arabic reviews by ensuring that both English and Arabic texts were appropriately processed, ensuring the system could effectively work across multiple languages.

Thank you!