Skincare Recommendation System Data Report

1. Business Understanding

Introduction

In the realm of machine learning-based recommender systems, the field of cosmetics and skin care products have often been overlooked. However, we firmly believe that by developing a comprehensive recommender system, we can greatly simplify the process of finding the perfect products tailored to each individual's personalised needs.

Our goal is to create a one-stop shop experience where users receive recommendations for a complete skin care product based on their personalised factors (such as skin type, skin tone), similar user choices, and budget.

By leveraging these insights, we aim to provide users with a curated selection of products that will enable them to achieve their desired look effortlessly. Through this project, we aspire to revolutionise the way cosmetics are recommended and empower users to make informed choices.

Problem statement

One of the key challenges in e-commerce is effectively matching customers with products that align with their interests especially in the beauty and cosmetics industry. Many Users have relied on Influencers, advertisements and promotions to experiment on the best products to use for their skincare.

There is a wealth of information hidden in past purchase history and reviews that can provide valuable insights into their tastes and preferences. However, existing recommender systems often overlook this historical data, limiting their ability to provide accurate and personalised recommendations.

The problem at hand is to leverage the on the data to build a data-driven recommender system that can accurately predict products that are likely to be of interest to the users.

Objectives

The primary objective of this project is to develop a robust skincare recommender system that offers a one-stop shop experience for users seeking beauty and cosmetics products.

- Developing an intuitive user interface that allows users to input their preferences, such as skin type, skin tone and budget preferences.
- Implementing collaborative filtering algorithms to analyse user data and generate personalised recommendations.
- Providing users with a curated selection of products from different categories to create a cohesive and personalised skincare.

Success metrics

- RMSE(Root Mean Squared Error) of less than 0.05
- MAE(Mean Absolute Error) of less than 0.05
- Precision, recall and accuracy of higher than 90

2. Data understanding

This dataset was collected from Kaggle which was scraped on March 2023 and was complimented by another dataset that we scraped from sephora website and contains:

- information about all beauty products (over 9,000) from the Sephora online store, including product and brand names, prices, ingredients, ratings, and all features.
- user reviews (over 1 million on over 2,000 products) of all products from the Skincare category, including user appearances, and review ratings by other users

3.Data preparation

Dealing with missing values

- 10 columns have no missing values, while another 4 have less than 10 percent of their data missing.
- The "Helpfulness" column, which is the ratio of the number of positive feedback divided by the total positive feedback, has approximately 51 percent of its data missing. We assume this is because it either represents negative feedback or no feedback at all, so we will fill the missing values with zero.
- User feature columns such as user skin tone, skin type, eye colour, and haircolor have approximately 20 percent of their data missing. Since these columns are very sensitive (as guessing users' features is not possible) and crucial for our modelling, we will have to drop the missing values.

Dealing with duplicates

- Duplicated rows constitute of 15.91 % of our dataset
- All duplicated rows were dropped.

4. Modelling

The problem at hand is a recommendation task.

The following models were developed for this task:

Memory based collaborative filtering

Model 1: Nearest Neighbors Model:

- .Create a sparse matrix from user ratings on products: First, create mappings from unique author IDs and product names to indices to construct the sparse matrix.
- Fit a Nearest Neighbors model to the sparse matrix: Utilise the Nearest Neighbors algorithm to find similar products based on cosine similarity.
- Finally create a function that takes a product name as input and returns a list of recommended products that are similar to the input product based on user ratings.

Model 2: KNNBasic Model:

When fitting the model, we need to create a sample of the data because the dataset is huge.

- An RMSE of 0.2724 means that, on average, the predicted ratings deviate from the actual ratings by approximately 0.2724 units.

User based collaborative filtering

Model 3: SVD Model:

The Singular Value Decomposition (SVD) model captures underlying relationships and generates recommendations based on these latent features.

- An RMSE of 0.0343 indicates that the SVD recommender model has a low average deviation between predicted and actual ratings. The MAE value of 0.0158 further confirms the model's accuracy. Additionally, with precision and recall values close to 1 and an accuracy score of 0.9998, the SVD model provides highly accurate and relevant personalized recommendations.

Model 4: SVDpp Model:

Building upon the SVD model, the SVD++ algorithm further considers implicit feedback signals, such as user interactions and implicit preferences, to enhance the recommendation process.

- With an RMSE of 0.0290 and an MAE of 0.0113, the model achieves a remarkably low average deviation between predicted and actual ratings, indicating its high accuracy. The Precision value of 0.9997 suggests that the majority of recommendations made by the system are relevant. The Recall value of 1.0 indicates that the model successfully identified all relevant recommendations. The Accuracy score of 0.9997 further demonstrates the overall correctness of the recommendations provided by the SVD++ model.

Model 5 : Tuned SVDpp Model

Best parameters from the GridSearchCV :

{ n_epochs = 50, n_factors = 50, lr_all = 0.01, reg_all = 0.02}

- This tuned SVDpp is our final model, demonstrating outstanding performance with the lowest RMSE of 0.0283 and an MAE of 0.0099. The model's ability to minimise the deviation between actual and predicted ratings signifies its exceptional accuracy. With a precision value of 0.9998 and recall value of 1.0, the model generates highly relevant recommendations. Furthermore, the accuracy score of 0.9998 solidifies its effectiveness in delivering accurate and personalised recommendations.

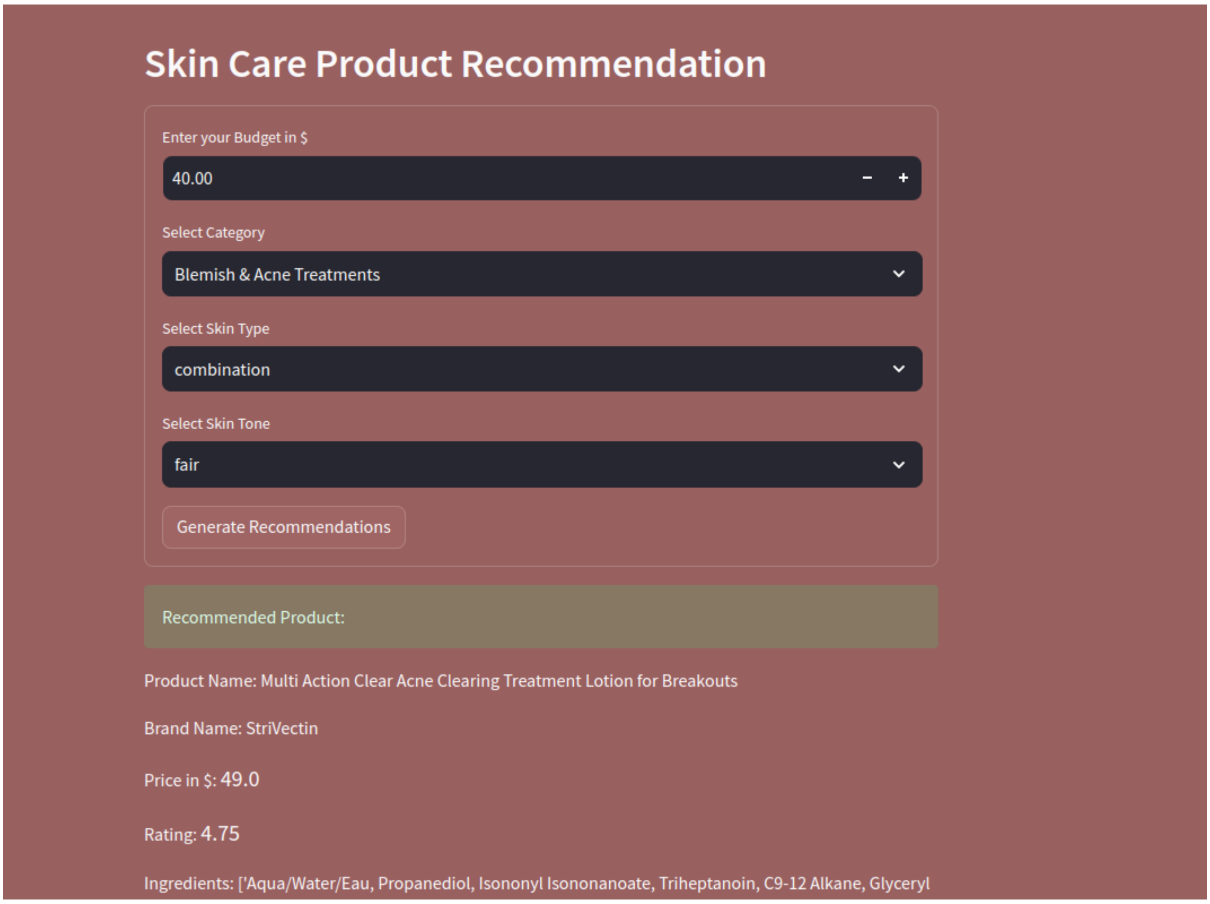
5.Evaluation

- The Tuned SVDpp model has met the objectives of the project.
- The Tuned SVDpp model is able to recommend a skin care product to the user with an accuracy of 99% and an RMSE of 0.0283.
- The model is able to take into account the user input and recommend the best product.

6.Deployment

The Tuned SVDpp was deployed using streamlit.

Below is the deployed interface



Recommendations

Based on our experience in developing the skincare recommender system, we would like to provide the following recommendations:

- The business should provide more products such as face serums since they were rated as the most helpful.
- They should collect more reviews from all races to improve the accuracy of the recommendation system.
- They should Collaborate with skincare experts, dermatologists, or industry professionals who can provide valuable insights and expertise on the quality and safety of skincare products.
- Alongside the product recommendations, the business can provide educational resources such as skincare guides, tutorials, and tips. This can help users make informed decisions about their skincare routines, understand the benefits of different ingredients, and address specific skin concerns