Skincare Products Recommendations

Final project by Anastasia Gorina

Overview

- Since **Sephora** doesn't have a **recommendation system** at the moment, I have decided to build one (or three).
- All data was scraped from **Sephora.com**: 17,000 **user reviews** for 1,400 unique **skincare products**
- Ended up with three models: filtering based on **user features**, **content-based** recommender, and **collaborative filtering**.

Pit stop for fun (review generator with Markov chain):

“AMAZING I originally purchased the Balance and Clarify set which contained a 2 step peel and I broke out”

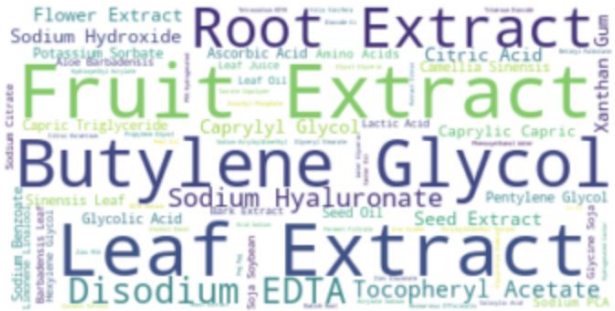
“The only thing it seemed to work because I’m self conscious of the work getting rid of it”

“It smells like bad BO and I love the gel cream”

“If your eyes it remove makeup i love it but the more expensive items in my midlate 20s normal skin”

EDA: are the products different?

Yes. Logistic regression classified them with 97% accuracy



Treatments

Masks

Cleansers

Moisturizers

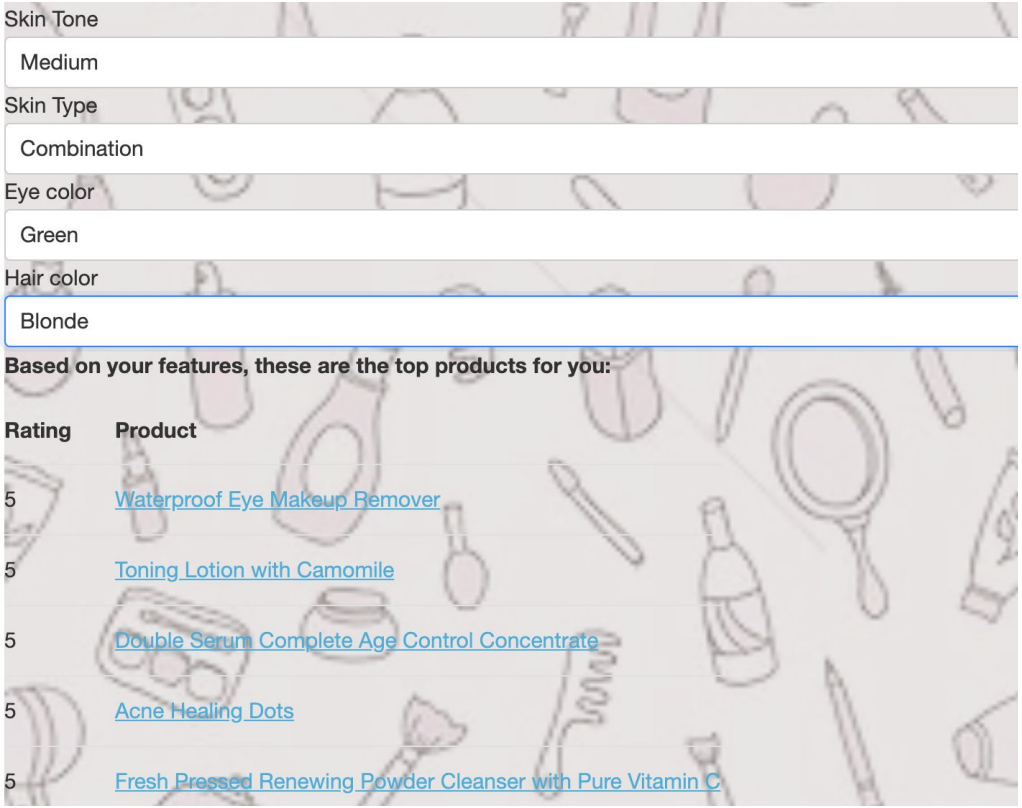
EDA: Can we predict if a product gets 5-stars?

Not really…

accuracy 0.5991678224687933

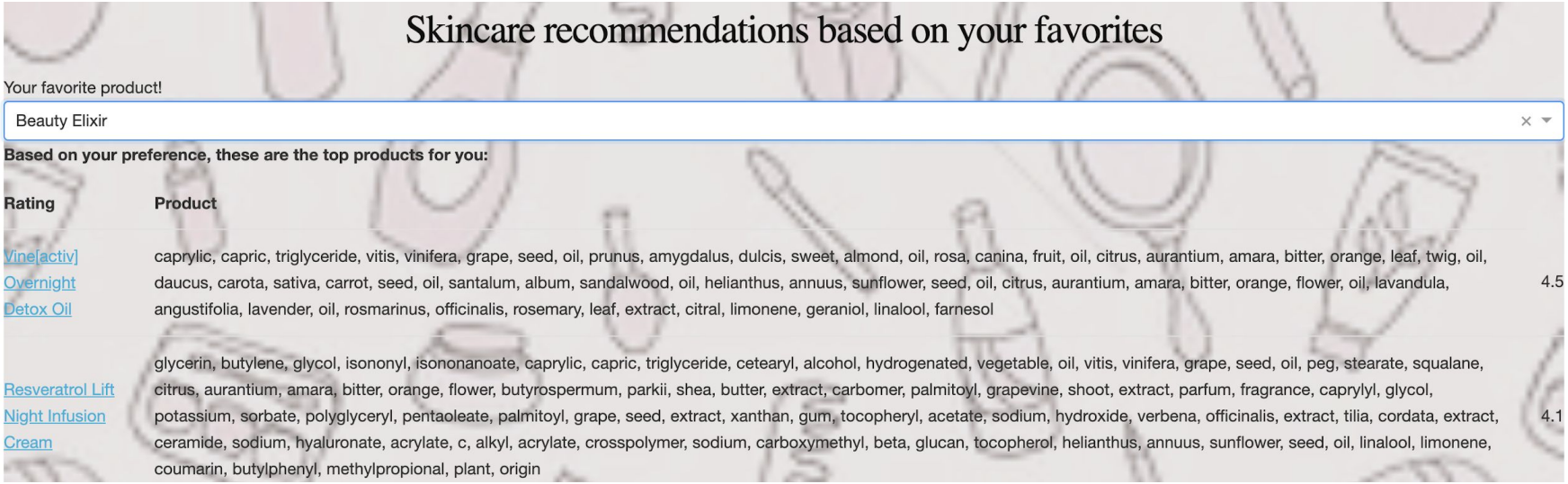
	precision	recall	f1-score	support
1	0.49	0.26	0.34	856
0	0.63	0.82	0.71	1307
micro avg	0.60	0.60	0.60	2163
macro avg	0.56	0.54	0.53	2163
weighted avg	0.57	0.60	0.56	2163

Model #1: By user Features



- Good for cold start problem: a user just needs to look in the mirror
- It takes user features, and gives top rated products from the customers who look like the person who is providing input

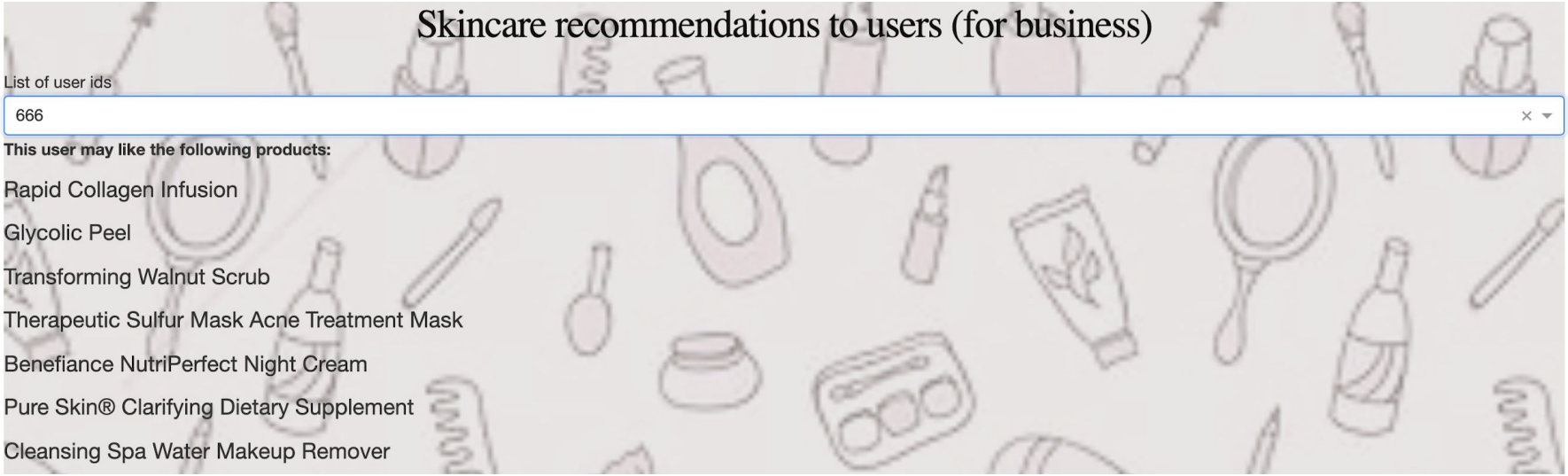
Model #2: By similarity of the ingredients



Good for opinionated and experienced users.

Tested on my favorite products, and it gives meaningful recommendations!

Model #3: By user similarity



Designed for business use. For example, customer service or marketing department representatives can type in a user id and email them a list of recommended products.

Future work

- Sell the app to Sephora for a million dollars
- Collaborate with **biochemistry experts**, **dermatologists**, and **geneticists** to collect more data about the users (e.g. skin **PH-level**, predisposition for **acne/photoageing**, etc.) and products, and use it to maximize the **personalized** approach to recommendations.