Project Recommendation System

Group 28
Ritika Nagar, Tarun Kumar Gupta, Shubham Sethi, Kartik Gupta, Rachit Gupta, Akshat Tilak

I. Motivation

WITH the wide variety of products and services available on the web, it is difficult for users to choose the product or service that most meets their needs. To reduce or even eliminate this difficulty, recommender systems have emerged. Recommendation systems are a particular type of information filtering system. They are software applications that aim to support users in their decision-making while interacting with a vast amount of information. The current product recommendation system suffers from two significant drawbacks:

- Recommendation redundancy.
- Unpredictability concerning new items (cold start).

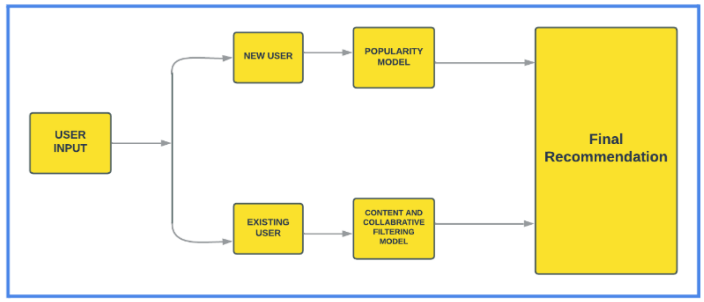
II. PROBLEM STATEMENT

Our product recommendation system aims to provide users with more choices in the ways that products are recommended to them by providing multiple models for users to choose from which can help predict their preferred products depending on their needs. We are providing users with models based on content, collaborative filtering and popularity. To address the cold start problem, the proposed recommendation systems have incorporated a popularity filtering mechanism that recommends popular items to new users. This helps to generate initial recommendations for users who have little or no interaction history with the system. Our future goal was to use reviews from users on certain products to analyze and find a variety of products that the user would find interesting. We have achieved it by merging content-based and collaborative filtering approaches based on their likes and dislikes, as well as those of similar users. By leveraging these approaches, we can provide more accurate and relevant recommendations to each user, improving their overall experience with our product. Using multiple models which can use cosine similarity, Euclidean distance, etc we can uphold a higher measure of effectiveness for our ranking systems by comparing these results to user needs.

III. LITERATURE SURVEY

Some reference papers

- **Reference No: 1**
  Title: Study of an E-commerce recommender system based on Big data Publication: Oxbridge college, kunning university
  Author: Xuesong Zhao
  This paper discussed the era of the web, in which



  a lot of information is overloaded over the Internet. Recommender system algorithms are widely used in e-commerce to provide personalized and more accurate recommendations to online users and enhance the sales and user stickiness of e-commerce. This study aims to build a product recommendation system on ecommerce platform according to user needs.

- **Reference No: 2**
  Title: Collaborative Filtering for Recommender Systems Publication: 2014 Second International Conference on Advanced Cloud and Big Data
  Author: Michael D. Ekstrand, John T. Riedl, and Joseph A. Konstan
  The report highlights the discussion of the types of the recommender systems as general and types of CF such as; memory based, model based, and hybrid model. In addition, this report discusses how to choose an appropriate type of CF. The evaluation methods of the CF systems are also provided throughout the paper.

- **Reference No: 3**
  Title: Content-Based Filtering: Techniques and Applications Publication: 2017 International Conference on Communication, Control, Computing and Electronics Engineering (ICCCCEE)

  Author: Khartoum, Sudan
  Content-based recommender systems make recommendations by analyzing the content of textual information and finding regularities in the content. The major difference between CF and content-based recommender systems is that CF only uses the user-item ratings data to make predictions and recommendations, while content-based recommender systems rely on the features of users and items for predictions. Both content-based recommender

systems and CF systems have limitations. While CF systems do not explicitly incorporate feature information, content-based systems do not necessarily incorporate the information in preference similarity across individuals.

- **Reference No: 4**
  Title: Popularity-Based Recommendation System: International Journal of Engineering and Advanced Technology (IJEAT)

  Author: Keshetti Sreekala
  The paper highlights how the Popularity based recommendation system works with the current vogue. It explains how it basically uses the items which are in swing at present. Whatever is more popular among the general public that is more likely to be recommended to new customers. The generalized recommendation, not personalized, is based on the count. In this paper, they have used a class which includes the methods to create recommendations and to recommend the item to the user.

IV. NOVELTY

Our product recommendation system have used a noval approach that incorporates a hybrid different filtering mechanisms. It provides a more comprehensive and accurate recommendation system that takes into account the unique preferences and needs of each user. By offering personalized and relevant recommendations, we can enhance the user's shopping experience, increase customer satisfaction.

V. PROPOSED METHOD

- **Popularity-based recommendation system**
  In this product model, we recommend products to the user according to the highest mean average rating of products. In this, we only consider the rating of those users who have rated at least 5(a threshold that can change) different products. It does not suffer from cold start problems.

- **Content-based recommendation system**
  In this product model , we recommend products to the user based on Cosine similarity or a euclidean distance of the vector of two products(item). We find cosine similarity and euclidean distance with the help of pivot table using sklearn library.

- **Collabrative-based recommendation system**
  In this product model, we have used Singluar value decompostion (SVD) for the user's product prediction and have used MSE (Mean Squared Error) in the evaluation matrix.

VI. TABLES USED

- **Cosine similarity table**

```
similarity_score = cosine_similarity(pt)
df2 = pd.DataFrame(similarity_score)
df2
```

	0	1	2	3	4	5	6	7	8	9	...	3081	3082	3083	3084	3085	3086	3087	3088	3089	3090
0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	0.0	0.0	1.0	0.0	1.0	0.0	0.0	0.0	0.0	0.707107	...	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	0.0	0.0	0.0	1.0	0.0	0.0	1.0	0.0	1.0	0.707107	...	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	0.0	0.0	1.0	0.0	1.0	0.0	0.0	0.0	0.0	0.707107	...	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
3086	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.000000	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0
3087	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.131306	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0
3088	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0
3089	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0
3090	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0

3091 rows × 3091 columns

- **Euclidean distance table**

```
similarity_score_2 = euclidean_distances(pt)
df3 = pd.DataFrame(similarity_score_2)
df3
```

Python

	0	1	2	3	4	5	6	7	8	9	...	3081	3082	3083	30
0	0.000000	5.099020	5.099020	5.099020	5.099020	5.099020	5.099020	7.141428	5.099020	7.141428	...	7.681146	3.162278	12.845233	4.1231
1	5.099020	0.000000	7.071068	7.071068	7.071068	7.071068	7.071068	8.660254	7.071068	8.660254	...	9.110434	5.830952	13.747727	6.4031
2	5.099020	7.071068	0.000000	7.071068	0.000000	7.071068	7.071068	8.660254	7.071068	5.000000	...	9.110434	5.830952	13.747727	6.4031
3	5.099020	7.071068	7.071068	0.000000	7.071068	7.071068	0.000000	8.660254	0.000000	5.000000	...	9.110434	5.830952	13.747727	6.4031
4	5.099020	7.071068	0.000000	7.071068	0.000000	7.071068	7.071068	8.660254	7.071068	5.000000	...	9.110434	5.830952	13.747727	6.4031
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
3086	5.000000	7.000000	7.000000	7.000000	7.000000	7.000000	7.000000	8.602325	7.000000	8.602325	...	9.055385	5.744563	13.711309	6.3245
3087	1.414214	5.099020	5.099020	5.099020	5.099020	5.099020	5.099020	7.141428	5.099020	7.141428	...	7.549834	3.162278	12.845233	4.1231
3088	11.313708	12.328828	12.328828	12.328828	12.328828	12.328828	12.328828	13.304135	12.328828	13.304135	...	13.601471	11.661904	17.058722	11.9582
3089	5.099020	7.071068	7.071068	7.071068	7.071068	7.071068	7.071068	8.660254	7.071068	8.660254	...	9.110434	5.830952	13.747727	6.4031
3090	8.185353	9.539392	9.539392	9.539392	9.539392	9.539392	9.539392	10.770330	9.539392	10.770330	...	11.135529	8.660254	15.165751	9.0553

3091 rows × 3091 columns

- **SVD Table Results**

Product Name	ASUS ROG Strix G15	ASUS ROG Zephyrus G14	ASUS TUF Gaming A15	Acer Predator Helios 300	Amazon Echo (4th Gen)	Apple AirPods Pro	Apple iPad Air	Apple iPad Pro	Apple iPhone 13 Pro Max	Bo QuietComf 35
user_index										
0	1.340564e-18	7.070102e-17	-5.014076e-17	1.937532e-17	-3.431355e-17	4.596582e-02	4.296188e+00	3.872457e+00	7.635344e-17	-1.916315e-
1	6.901117e-18	4.385073e-17	-4.602211e-17	1.111400e-17	3.317116e-17	-2.436171e-01	3.635722e+00	4.801307e-01	-3.385228e-18	-7.197917e-
2	4.997199e-18	1.044938e-16	-3.963048e-17	9.035241e-18	9.823857e-17	2.413873e-01	-8.023761e-01	1.364786e+00	-5.213814e-17	1.755009e-
3	-5.103911e-18	-5.472515e-17	4.333810e-17	-1.267157e-17	-2.123122e-17	4.707254e+00	5.076368e-01	5.100140e-01	-3.499830e-17	8.490786e-
4	-7.380637e-18	-2.334917e-18	9.545239e-18	9.721361e-19	5.159815e-17	1.414524e-02	4.256497e-01	-1.377486e-01	4.135169e-17	2.691912e-
...	...	...	...	...	...	...	...	...	...	
71	1.186811e-17	-3.190292e-17	2.654830e-17	-9.122255e-18	2.708931e-17	6.646034e-16	2.238983e-17	9.659868e-17	3.342114e-17	1.484686e+
72	2.680383e-34	4.355760e-34	-2.167267e-34	-1.440892e-35	9.620775e-35	-3.078557e-18	1.781660e-18	1.617569e-18	-2.401723e-34	1.212612e-

- **Pivot Table Results**

Product Name	Acetaminophen	Aspirin	Biotin	Calcium	Coenzyme Q10	Cranberry	Echinacea	Fish Oil	Ginkgo Biloba	Ibuprofen	...	Magnesium	Melator
user_index													
0	-1.869857e-16	4.500000e+00	-2.588489e-15	9.969959e-17	-1.874741e-16	2.347696e-16	-9.461999e-16	-5.469887e-16	-1.540118e-16	2.038584e-16	...	1.229272e-15	1.099914e-
1	-3.484190e-17	1.943904e-17	3.020849e-16	-6.094348e-16	-5.206554e-16	6.821755e-16	7.628661e-16	-6.498935e-16	2.742625e-30	-9.298505e-16	...	-4.107825e-16	4.700000e+
2	-7.764292e-32	9.233928e-16	3.444360e-16	-1.422525e-14	-2.646000e-31	5.565453e-16	1.820167e-30	2.912695e-30	-2.867609e-29	-1.656212e-30	...	1.582631e-30	5.953792e-
3	-1.592134e-16	5.645436e-16	9.004009e-16	-6.186871e-17	-7.078284e-16	1.833147e-15	-1.012768e-15	1.183982e-15	-3.693519e-16	1.065814e-15	...	8.215650e-16	2.521924e-
4	6.658841e-32	-3.374533e-16	2.265628e-16	7.657663e-17	4.177411e-31	1.768651e-16	1.771292e-31	1.406635e-31	3.036315e-31	-1.077464e-31	...	-3.155022e-31	-9.78832
5	6.227681e-17	-5.964444e-16	-9.436054e-16	8.386655e-18	1.028285e-16	4.300000e+00	2.847500e-16	5.337533e-16	3.118844e-17	-2.073363e-16	...	1.386391e-15	3.269393e-
6	5.508110e-32	-6.837358e-16	5.932745e-16	-1.196131e-15	9.479124e-32	6.375387e-16	-2.471543e-31	2.002171e-30	-2.185864e-30	3.979531e-31	...	7.399638e-31	-8.48472
7	-2.542636e-16	-5.879033e-17	-6.165718e-16	-3.125805e-16	-4.093376e-16	4.341298e-16	-1.768745e-17	-1.187649e-15	-2.617755e-16	-1.371944e-16	...	1.148907e-15	-1.08115
8	-7.682139e-17	-3.275141e-15	4.600000e+00	5.227481e-17	1.231320e-16	4.928511e-16	-4.508932e-16	5.036813e-16	-2.497146e-16	4.852475e-16	...	1.026956e-15	5.153354e-
9	1.253627e-31	-7.369279e-16	-6.481590e-16	-5.291378e-16	1.413095e-31	3.826328e-16	1.332933e-30	-3.533422e-31	-1.034988e-30	-1.076688e-30	...	-5.314569e-31	1.120460e-
10	5.762991e-32	-1.777059e-16	-2.945351e-16	8.162725e-15	2.185750e-31	3.438725e-17	-8.738635e-31	-1.992001e-30	1.647118e-29	7.282641e-31	...	-8.585120e-31	-2.04235
11	-5.652582e-32	2.866759e-16	6.975433e-16	4.509435e-16	-7.340192e-32	-2.786081e-16	-1.094267e-30	4.157543e-31	8.501548e-31	1.016405e-30	...	3.609872e-31	-1.36571
12	4.146054e-16	-1.872698e-15	-1.541576e-16	1.636488e-16	-9.339045e-31	1.400668e-15	4.410232e-16	-2.726011e-17	2.513346e-17	-1.865268e-16	...	6.161738e-16	6.511407e-
13	1.356122e-17	-9.768983e-16	-4.478717e-17	4.334156e-17	-1.235645e-16	4.637585e-16	4.921059e-17	-1.405292e-15	-6.094116e-16	1.555468e-16	...	-1.283695e-15	-1.27964
14	1.439195e-31	-1.766009e-	-1.682754e-	2.880838e-16	1.159864e-	[illegible]	1.699236e-	4.958250e-	6.702111e-	-6.944377e-	...	-1.058968e-	-7.79795

VII. DATASET USED

Product Name	ProductId	UserId	Rating	Description	category
Advil Liqui-Gels	2001	102001	4.5	Advil Liqui-Gels are a fast-acting pain reliever that provides relief from headaches, muscle a	medicine
Tylenol Extra Strength	2002	102002	4.6	Tylenol Extra Strength is a pain reliever that provides relief from headaches, muscle aches,	medicine
Claritin	2003	10200	4.7	Claritin is an antihistamine that provides relief from allergy symptoms such as sneezing, rur	medicine
Benadryl	2004	102004	4.5	Benadryl is an antihistamine that provides relief from allergy symptoms such as sneezing, r	medicine
Pepto-Bismol	2005	102005	4.5	Pepto-Bismol is an antacid and anti-diarrheal medication that provides relief from heartbur	medicine
Cream Cheese	6001	6	4	Smooth and creamy, great for spreading on bagels	dairy
Chocolate Milk	7001	7	3.7	Tastes good, but a bit too sweet for me	dairy
Butter	8001	8	4.5	Rich and creamy, perfect for baking	dairy
Sour Cream	9001	9	2.8	Has an odd aftertaste, wouldn't buy again	dairy
Strawberry Yogurt	10002	10	4.7	Love the sweet and tangy flavor of this yogurt	dairy
Strawberry Yogurt	10002	1	4.2	Love the sweet and tangy flavor of this yogurt	dairy
Blue Cheese	20002	1	3.9	A bit pungent for my taste, but great for salads	dairy
Blue Cheese	20002	19	3.3	A bit pungent for my taste, but great for salads	dairy
Blue Cheese	20002	11	3.9	A bit pungent for my taste, but great for salads	dairy
Blue Cheese	20002	17	3.7	A bit pungent for my taste, but great for salads	dairy
Heavy Cream	30002	12	4.8	Perfect for whipping into soft peaks, makes great desserts	dairy
Heavy Cream	30002	20	4.6	Perfect for whipping into soft peaks, makes great desserts	dairy
Heavy Cream	30002	19	4.5	Perfect for whipping into soft peaks, makes great desserts	dairy
Heavy Cream	30002	11	3.2	Perfect for whipping into soft peaks, makes great desserts	dairy

VIII. METHODOLOGY

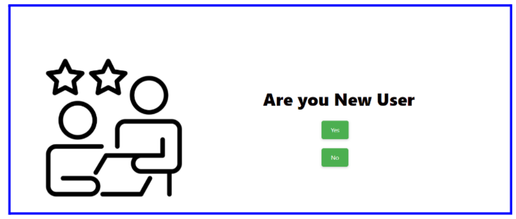
- Firstly,we will extract the a real world and practical dataset for oue model and store them as a CSV file. This will enable us to represent the dataset in a compressed format.
- Then, we will be doing preprocessing of data by doing systematic and comprehensive examination of a dataset using EDA and transform the raw data to a format suitable for further analysis.
- We have converted our data into a pivot table where the rows represent user IDs and the columns represent Product Names. We have also used product ratings from an SVD table, which is a technique for dimensionality reduction that can help to improve the accuracy of collaborative filtering.
- We have then used cosine similarity to find similar users based on their likes and dislikes of the products, and identified the most similar users. Finally, we have used collaborative filtering techniques such as using the likes and dislikes of similar users to recommend products to each user based on their preferences.
- By leveraging pivot tables, cosine similarity, and collaborative filtering techniques, we were able to generate personalized product recommendations for each user based on their past product ratings and the preferences of similar users.
- Currently we are taking inputs from the user based on what all he has bought or reviewed positively in the past and on that basis we provide the user options of different models to recommend a product.

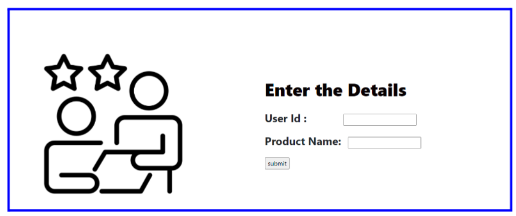
IX. EVALUATION MATRIX

In order to evaluate the accuracy of our collaborative filtering model and its recommendations, we are using a prediction array and a calculation array. The prediction array includes the predicted ratings that a particular user would give to each product. If the user has not given any rating for a product, then the corresponding cell in the prediction array is filled with a 0.

We are then using this prediction array to calculate values for the calculation array, which includes values for each product based on the difference between the actual ratings given by the user and the predicted ratings generated by our model after applying different filtering models.

X. FINAL RESULTS





REFERENCIAS

[1] Xuesong Zhao. " Study of E-commerce recommender system based on Big data Publication.", Publication: Oxbridge college, kunning university, *https://ieeexplore.ieee.org/document/8725694*.

[2] Khartoum and Sudan. "Content-Based Filtering: Techniques and Applications.", Publication: 2017 International Conference on Communication, Control, Computing and Electronics Engineering (ICCCCEE), *https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7867668*.

[3] Parul Aggarwal, Vishal Tomar, Aditya Kathuria, "Comparing content based and collaborative filtering in recommendation systems", *https://www.neliti.com/publications/263309/comparing-content-based-and-collaborative-filtering-in-recommender-systems,*