PRODUCT RECOMMENDATION SYSYTEM AND ECOMMERCE

Mentor -- ANJANEYULU SIR

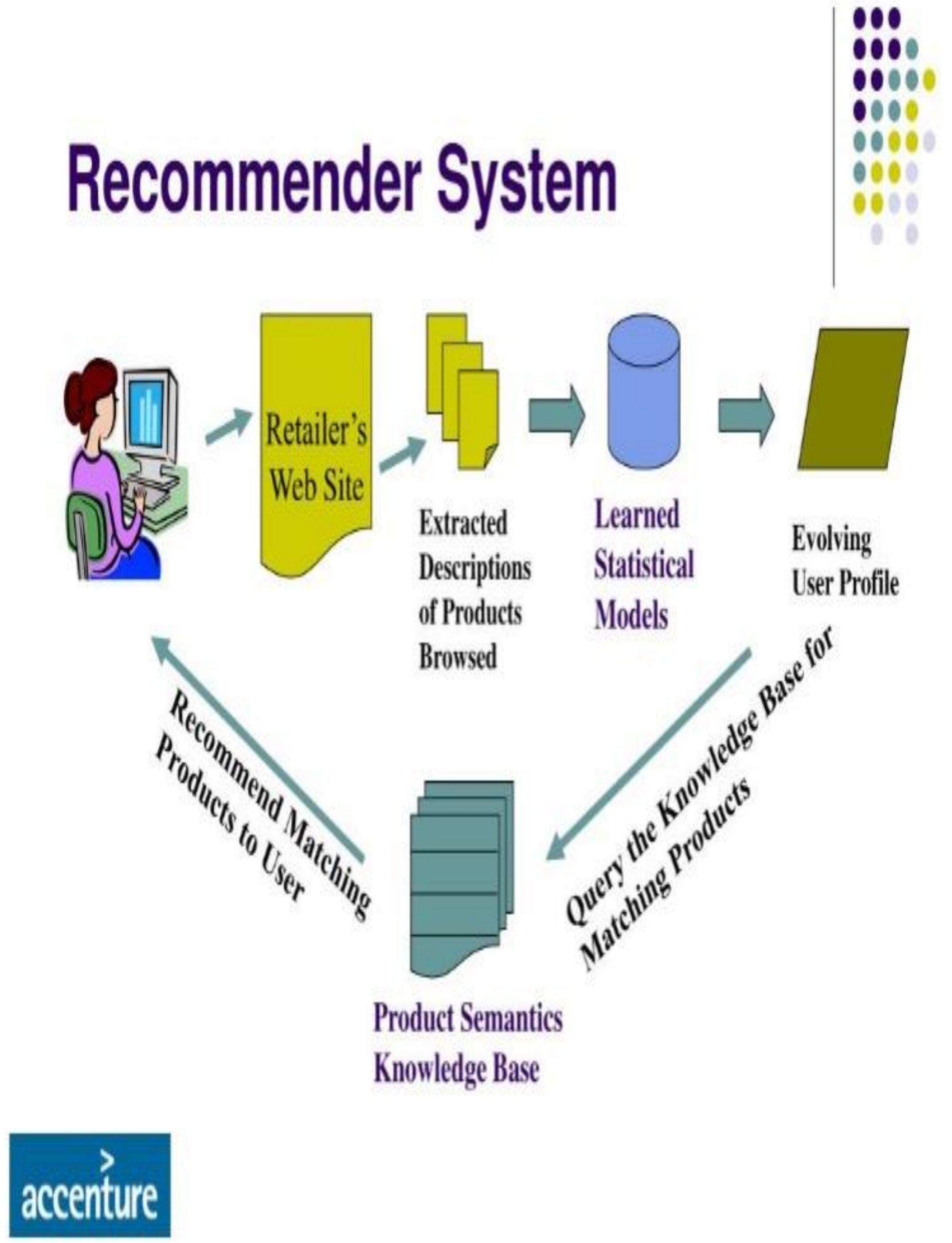
Team Members -- J.FRANKLIN
N.SHIVANI
G.MANISHA
D.SRI LAXMI
K. VAISHNAVI

OBJECTIVE

- TO RECOMMEND TOP 10 BEAUTY PRODUCTS ON AMAZON APP BASED ON CUSTOMER REVIEWS

Abstract

This technical report examines the literature and explains the Recommender System ideas. A recommendation engine is a system that filters information to bring movies , music, books etc to the user. A set of tools to a user , this data has been vetted to ensure that it is likely to pique the user's interest .



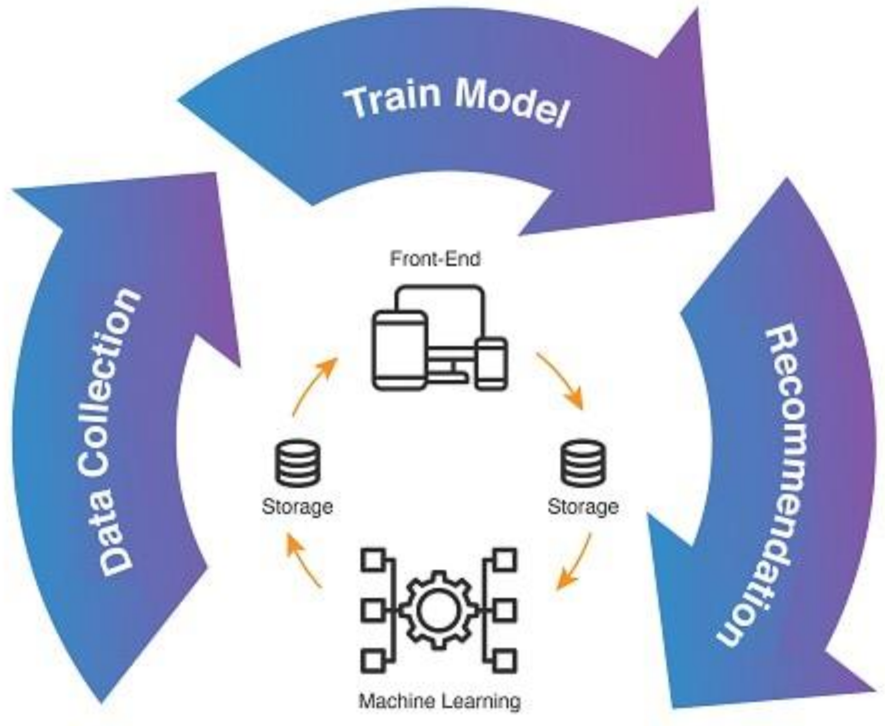
Introduction to Recommender Systems and E-Commerce

RECOMMENDER SYSTEMS

- Recommendation system is a technology which is used for filter and retrieval the data .

- In the methods of unsupervised machine learning ,the task of machine or model is to group the uncategorized data , according to similarities, patterns and differences without given any kind of training to the machine

- These can be based on various criteria ,including past puchases ,search history, demographic information , and other factors



E-COMMERCE

- “E-commerce" or "electronic commerce" is the trading of goods and services on the internet. It is your bustling city center or brick-and-mortar shop translated into zeroes and ones on the internet superhighway.

- Ecommerce works by connecting buyers and sellers using various electronic channels.

For example, you need a channel, such as a website or social media, so customers can find products and services to purchase. Then a payment processor enables the exchange of the goods or services. Once the transaction succeeds, the customer receives a confirmation email or SMS, and a printable receipt.

Literature Review

SIVAPALAN

- The significant chanllanges of collaborative filtering -- sparsity
  -- scalability
- Collaborative filtering measures -- cosine metric
  -- jaccard coefficient
  -- personalized data
- Due to data sparsity , sales volumes on large e-commerce are decreasing.
- Suggested that the **developers of RSs** should stop using **the nearest neighbouring**

G . GUPTA & R . KATARYA

- In UBCF techniques ,the recommendations are generated acc to the like or dislike of the neighbours node
- In IBCF technique, the similarity between the items are calculated

ZHAO , ZHI-DAN
&
MING

The drawback of UBCF   ---- if user U like an item I but his/her neighbours not gave good ratings good ratings to that item I will not be recommended

IBCF calculates similarity between two items

Source of the data

Collected the data of amazon beauty products from Kaggle website https://www.kaggle.com/skillsmuggler/amazon-ratings

SAMPLE DATA

This is the sample data of size 20

UserId	ProductId	Rating	Timestamp
A39HTATAQ9V7YF	205616461	5	1369699200
A3JM6GV9MNOF9X	558925278	3	1355443200
A1Z513UWSAAO0F	558925278	5	1404691200
A1WMRR494NWEWV	733001998	4	1382572800
A3IAAVS479H7M7	737104473	1	1274227200
AKJHHD5VEH7VG	762451459	5	1404518400
A1BG8QW55XHN6U	1304139212	5	1371945600
A22VW0P4VZHDE3	1304139220	5	1373068800
A3V3RE4132GKRO	130414089X	5	1401840000
A327B0I7CYTEJC	130414643X	4	1389052800
A1BG8QW55XHN6U	130414643X	5	1372032000
AIFAAVTUYHEHB	130414643X	4	1378252800
AVOGV98AYOFG2	1304146537	5	1372118400
A22VW0P4VZHDE3	130414674X	5	1371686400
AVOGV98AYOFG2	1304168522	5	1372118400
A6R426V4J7AOM	1304168522	5	1373414400
A22VW0P4VZHDE3	1304174778	5	1372896000
AKGB62WGF35J8	1304174778	5	1372896000
A22VW0P4VZHDE3	1304174867	5	1373068800

Data Description

This dataset contains product reviews and metadata from Amazon, including 142.8 million reviews spanning May 1996 - July 2014.

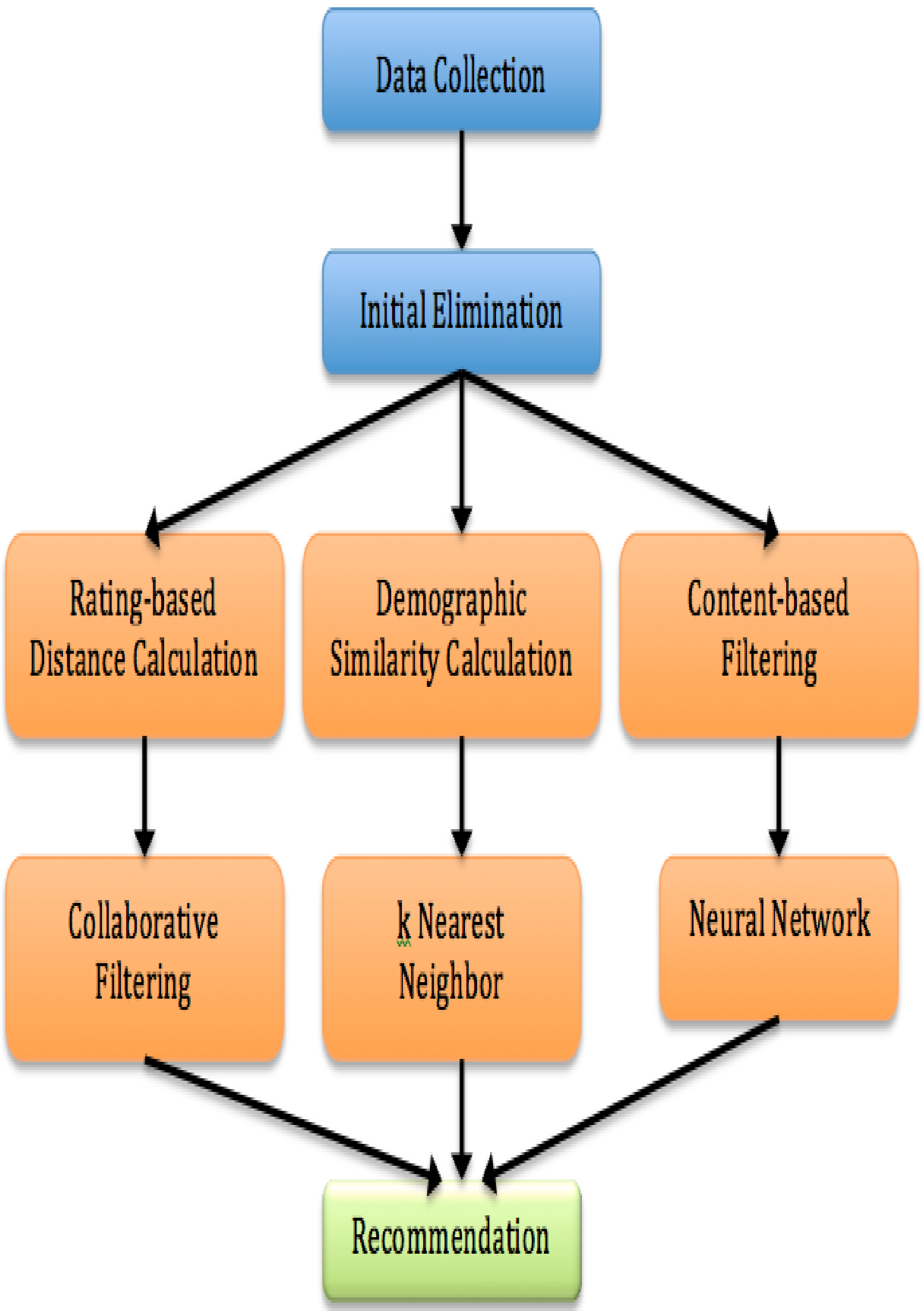
- It contains
  - the unique **UserId** (Customer Identification),
  - the product **ASIN** (Amazon's unique product identification code for each product),
  - **Ratings** (ranging from 1-5 based on customer satisfaction) and
  - of the rating (in UNIX time) the **Timestamp**

Methodology

Model based collaborative system

POPULARITY BASED REOMMENDATION SYSTEM

- Popularity based are a great strategy to target the new customers with the most popular products sold on a business's website and is very useful to cold start a recommendation engine.



STATISTICAL TECHNIQUES

CONTENT BASED RECOMMENDATION SYSTEM

Comparing user interests to product features . The products that have the most overlapping features with user interests are what's recommended

Collaborative Filtering

Measures or techniques used are -

- neighbour(cosine, correlation)
- clustering
- Graph theory

Commonly used techniques -

- Bayesian approach
- clustering
- Artificial nueral networks
- Linear regression
- Probabilistic models

HYBRID RECOMMENDATION SYSTEM

Uses both collaborative data and content based . Simultaneously which helps to suggest a similar or close item to the users

- Methods we used are **content based & collaborative filtering** and techniques we used is **correlation matrix based on data**

DATA VISUALIZATION

Popularity based Recommendation System

In [1]:
```python
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt


plt.style.use("ggplot")

import sklearn
from sklearn.decomposition import TruncatedSVD
```

In [3]:
```python
amazon_ratings = pd.read_csv('C:\\Users\\KISHORE\\Desktop\\kkkkk\\ratings_Beauty.csv')
amazon_ratings = amazon_ratings.dropna()
amazon_ratings.head()
```

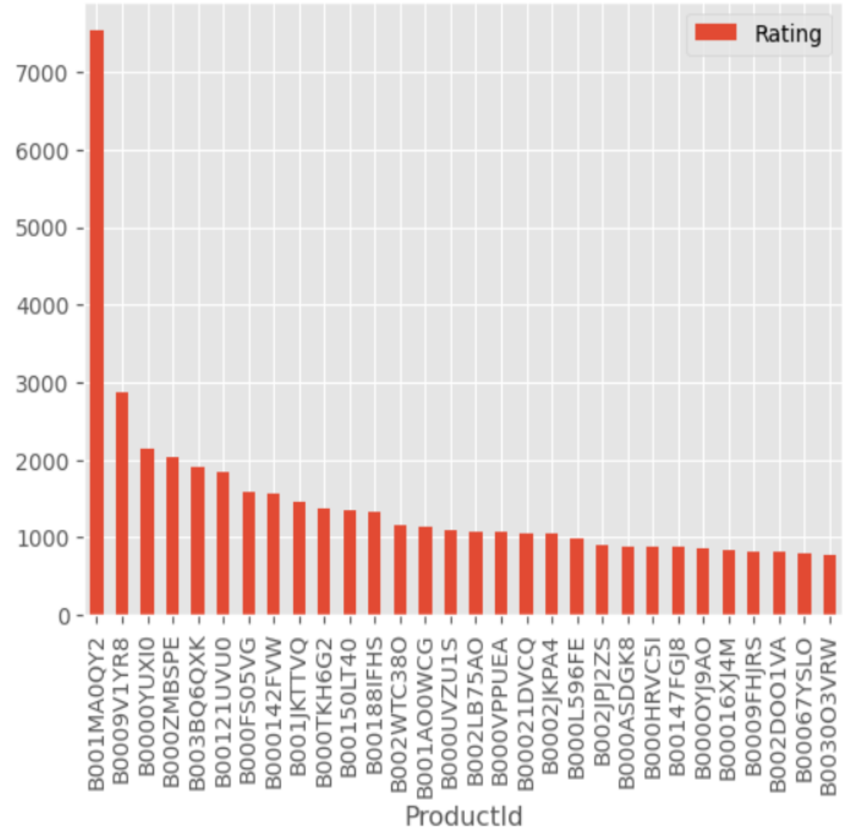
Out[3]:

	UserId	ProductId	Rating	Timestamp
0	A39HTATAQ9V7YF	205616461	5	1369699200
1	A3JM6GV9MNOF9X	558925278	3	1355443200
2	A1Z513UWSAAO0F	558925278	5	1404691200
3	A1WMRR494NWEWV	733001998	4	1382572800
4	A3IAAVS479H7M7	737104473	1	1274227200

In [4]:
```
amazon_ratings.shape
```
Out[4]:
```
(1048575, 4)
```
In [5]:
```
popular_products = pd.DataFrame(amazon_ratings.groupby('ProductId')['Rating'].count())
most_popular = popular_products.sort_values('Rating', ascending=False)
most_popular.head(10)
```
Out[5]:

ProductId	Rating
B001MA0QY2	7533
B0009V1YR8	2869
B0000YUXI0	2143
B000ZMBSPE	2041
B003BQ6QXK	1918
B00121UVU0	1838
B000FS05VG	1589
B000142FVW	1558
B001JKTTVQ	1468
B000TKH6G2	1379

In [6]:
```
most_popular.head(30).plot(kind = "bar")
```
Out[6]: <Axes: xlabel='ProductId'>



Analysis:

- The above graph gives us the most popular products (arranged in descending order) sold by the business.
- For eaxmple, product, ID # B001MA0QY2 (Maybelline New York Sky High washable Masacara)has sales of over 7000, the next most popular product, ID # B0009V1YR8(PanOxyl Acne Foaming Wash) has sales of 3000, etc.

Item based collaborative using correlation matrix

In [7]:
```
amazon_ratings1 = amazon_ratings.head(10000)
```

In [8]:
```
ratings_utility_matrix = amazon_ratings1.pivot_table(values='Rating', index='UserId', columns='ProductId', fill_value=0)
ratings_utility_matrix.head()
```

Out[8]:

ProductId	1304139212	1304139220	130414089X	130414643X	1304146537	130414674X	1304168522	1304174778	1304174867	1304174905	...
UserId											
A00205921JHJK5X9LNP42	0	0	0	0	0	0	0	0	0	0	...
A024581134CV80ZBLIZTZ	0	0	0	0	0	0	0	0	0	0	...
A03056581JJIOL5FSKJY7	0	0	0	0	0	0	0	0	0	0	...
A03099101ZRK4K607JVHH	0	0	0	0	0	0	0	0	0	0	...
A0505229A7NSH3FRXRR4	0	0	0	0	0	0	0	0	0	0	...

5 rows × 886 columns

In [9]:
```
ratings_utility_matrix.shape
```

Out[9]: (9697, 886)

In [10]:
```
X = ratings_utility_matrix.T
X.head()
```

Out[10]:

UserId ProductId	A00205921JHJK5X9LNP42	A024581134CV80ZBLIZTZ	A03056581JJIOL5FSKJY7	A03099101ZRK4K607JVHH	A0505229A7NSH3FRXRR4	A05492663T95K\
1304139212	0	0	0	0	0	
1304139220	0	0	0	0	0	
130414089X	0	0	0	0	0	
130414643X	0	0	0	0	0	
1304146537	0	0	0	0	0	

5 rows × 9697 columns

```
In [11]: X.shape

Out[11]: (886, 9697)

In [12]: X1 = X

In [14]: SVD = TruncatedSVD(n_components=10)
         decomposed_matrix = SVD.fit_transform(X)
         decomposed_matrix.shape

Out[14]: (886, 10)

In [15]: correlation_matrix = np.corrcoef(decomposed_matrix)
         correlation_matrix.shape

Out[15]: (886, 886)

In [16]: X.index[99]

Out[16]: '6162071103'

In [17]: i = "6117036094"

         product_names = list(X.index)
         product_ID = product_names.index(i)
         product_ID

Out[17]: 96
```

```
In [18]: correlation_product_ID = correlation_matrix[product_ID]
         correlation_product_ID.shape

Out[18]: (886,)
```

```
In [24]: Recommend = list(X.index[correlation_product_ID > 0.90])
         Recommend.remove(i)

         Recommend[0:9]
```

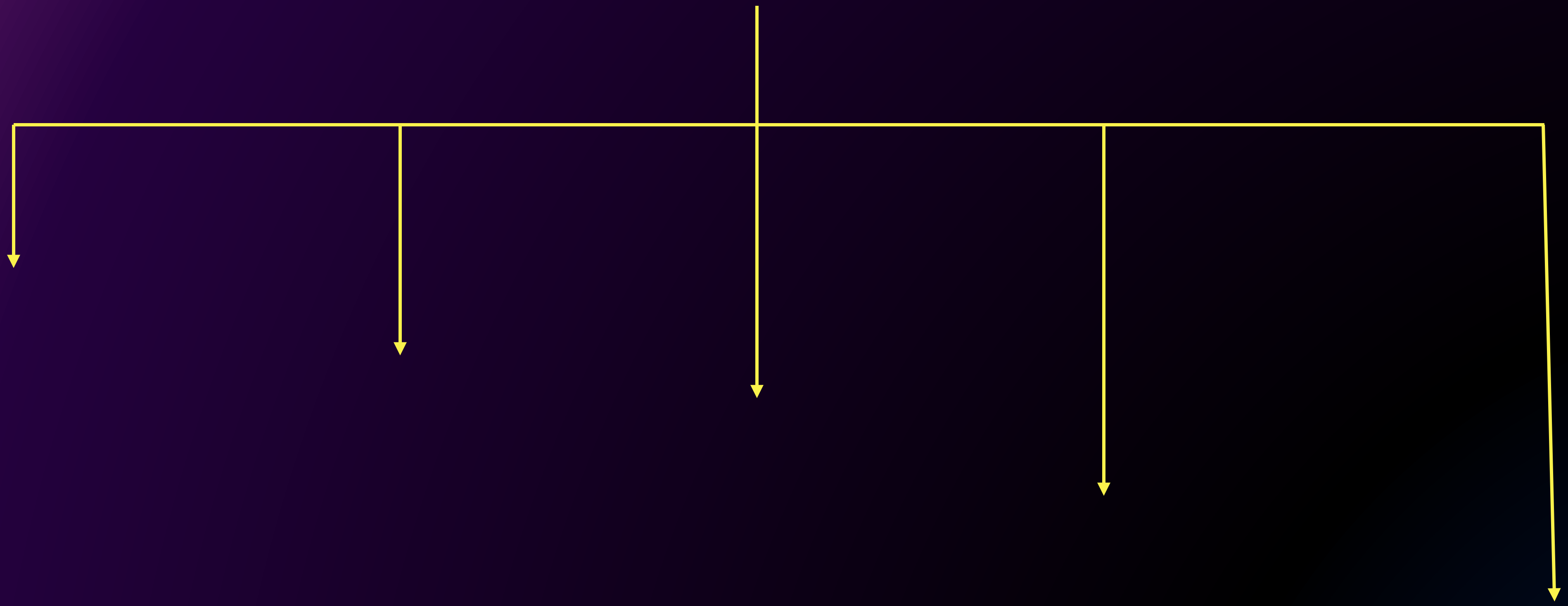
```
['0733001998',
 '1304139212',
 '1304139220',
 '130414089X',
 '130414643X',
 '130414674X',
 '1304174778',
 '1304174867',
 '1304174905']
```

Conclusion:

Here are the top 10 products to be displayed by the recommendation system to the above customer based on the purchase history of other customers in the website. Products named

- -- L'Oreal Paris Voluminous Lash Paradise
- -- Maybelline Instant Age Rewind
- -- Differin GEL
- -- Olaplex No.5 conditioner
- -- Sol De Janeiro body fragrance
- -- Laneige Lip Sleeping Mask
- -- Revlon One-step Volumiser
- -- Maybelline New York Matte ink liquid lipstick
- -- Truskin Naturals vitamin C serum
- -- Heeta scalp MAssager

Work Distribution



REFERENCE

- https://www.kaggle.com/code/shawamar/product-recommendation-system-for-e-commerce
- https://www.google.com/url?sa=t&source=web&rct=j&opi=89978449&url=https://www.investopedia.com/terms/e/ecommerce.asp&ved=2ahUKEwif1enlyfeCAxXW1jgGHbmWDr4QFnoECEUQAQ&usg=AOvVaw3RYZWh2DHXhEBdLbNOdBlS
- https://www.nvidia.com/en-us/glossary/recommendation-system/
- https://www.techtarget.com/searchcio/definition/e-commerce
- https://www.turing.com/kb/collaborative-filtering-in-recommender-system
- https://www.miquido.com/blog/perks-of-recommendation-systems-in-business/

THANK YOU