COLLABORATIVE FILTERING AND CONTENT BASED RECOMMENDATION SYSTEM FOR SCALING BUSINESSES

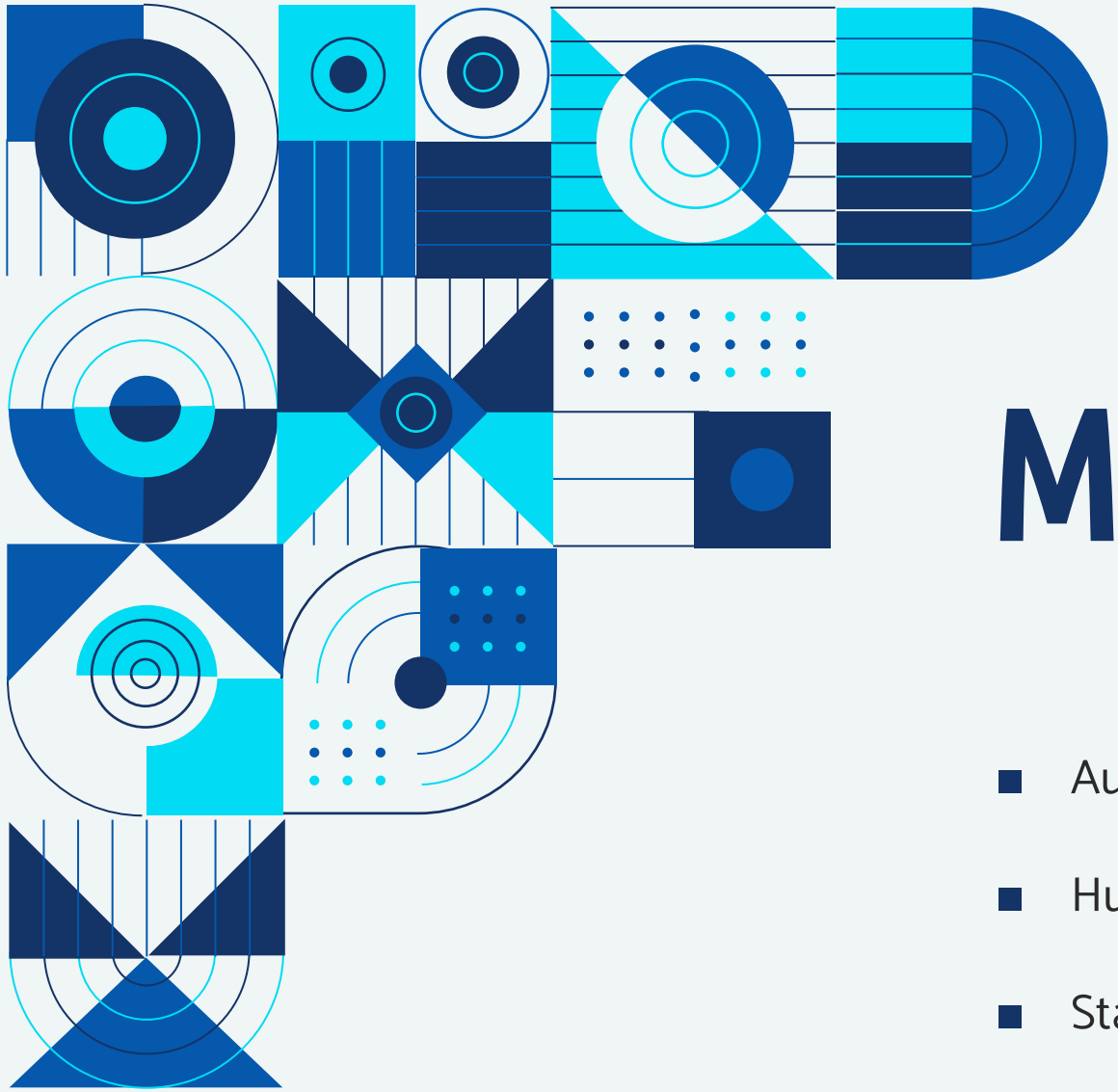
Group Details

Supervisor :
Prof Vrijendra Singh Sir

Group Members:
IIT2018031 - Raushan Raj
IIT2018069 - Suryasen Singh
IIT2018071 - Rahul Kumar Yadav

Problem Statement

- To develop a grocery and product recommendation system
- To develop a system which is majorly based on collaborative filtering technique



Motivation

- Automation
- Human Effort
- Stable recommendation model

LITERATURE SURVEY

S.No.	Title	Authors	Year of Publication
1.	• Movie Recommendation System Using Item Based Collaborative Filtering	Poonam Sharma , Lokesh Yadav	2020
2.	• Deep Learning based product recommendation system.	Akshit Tayade1, Vidhi Sejpal2, Ankit Khivasara3	2021

LITERATURE SURVEY

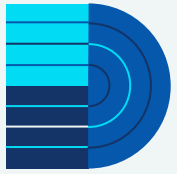
S.No.	Title	Authors	Year of Publication
3.	Personalized Book Recommendation System using Machine Learning Algorithm	Dhiman Sarma, Tanni Mittra, and Mohammad Shahadat Hossain	2021
4.	Music Recommender System Based on Collaborative Filtering	AnuPrabha P S, HarsithaN, Vaishnavi K, Dr.P.Velvadivu and Dr.M.Sujithra	2020
5.	A Movie Recommender System: MOVREC	Manoj Kumar, D.K. Yadav, Ankur Singh, Vijay Kr. Gupta	2015

Tools and Libraries

Google Colaboratory

Python

Pandas



TF-IDF Vectorizer

Cosine Similarity

Pearson Correlation

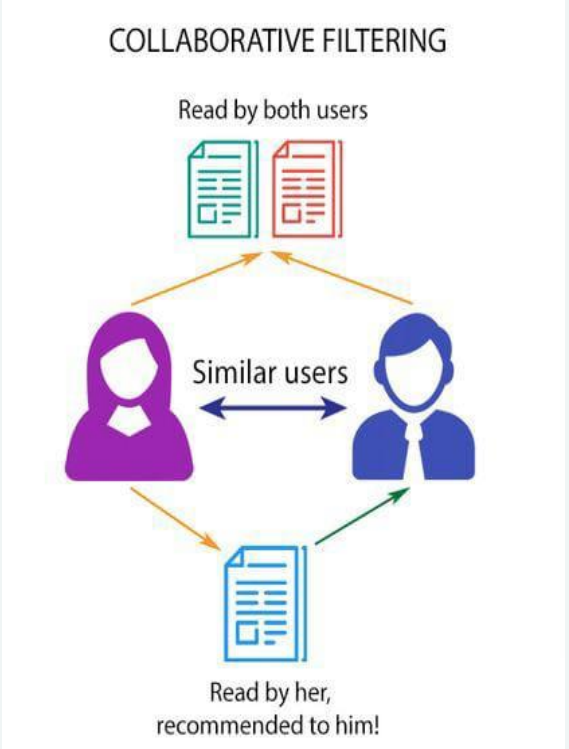
METHODOLOGY

Techniques Used

1. Collaborative Filtering

- Identify relationships between pieces of data.
- Frequently used in recommendation system.

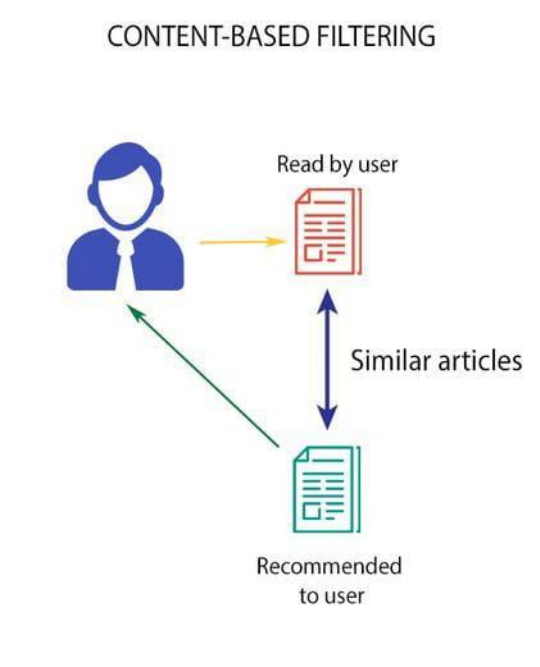
Users	Item 1	Item 2	Item 3	Item 4
User 1	5	4		5
User 2	4		3	
User 3		1		2
User 4	1	2		



➢ *User 1 and 2 have similar interest likewise User 3 and 4 are also showing similar kind of interest.*

2. Content Based Filtering

- Based on user preferences for product features
- It doesn’t need any data about other users, since the recommendations are specific to this user



Data Set Overview :

Two kinds of Dataset has been used :

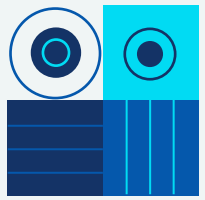
- ❑ Amazon product ratings dataset.
- ❑ Flipkart product dataset.

Implementation :

Our model has been implemented using three different techniques :-

- Popularity based recommendation system
- Content based recommendation system
- Item to item based collaborative filtering

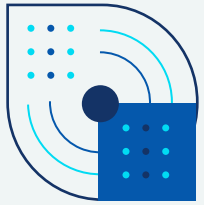
Procedural Steps :



Importing Libraries

Loading of Dataset

Data preprocessing



Generation of similarity score

Recommendation

Popularity based recommendation

1. Grouping of product based on rating
2. Reverse sorting
3. Recommendation

ProductId	occurences
B001MA0QY2	7533
B0009V1YR8	2869
B0043OYFKU	2477
B0000YUXI0	2143
B003V265QW	2088
B000ZMBSPE	2041
B003BQ6QXK	1918
B004OHQR1Q	1885
B00121UVU0	1838
B000FS05VG	1589

Top 10 most popular products

Content based recommendation :

1. Preparation of documents
2. Text vectorization

```
from sklearn.feature_extraction.text import TfidfVectorizer
documents = df2['description']
tf = TfidfVectorizer(stop_words='english')
tfidf_matrix = tf.fit_transform(documents)
```

3. Finding similarities between products

```
from sklearn.metrics.pairwise import cosine_similarity
cosine_sim = cosine_similarity(tfidf_matrix, tfidf_matrix)
```

4. Recommendation

Below are the recommendations for the product - HRS ULTIMATE MEN Chest Pads

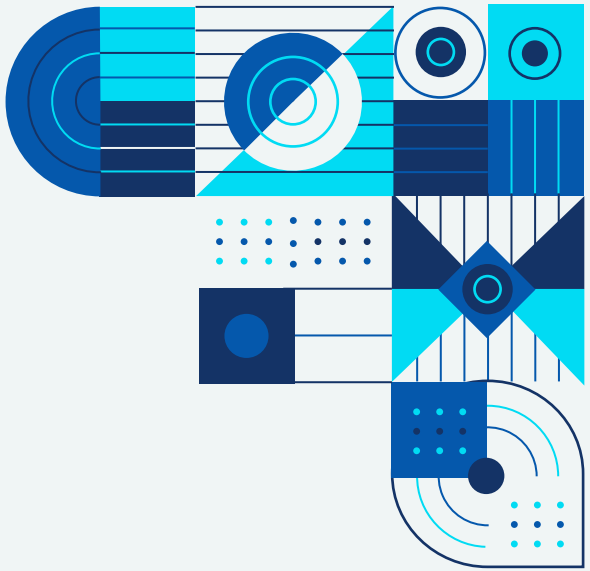
	Product name	Similarity score
0	HRS ULTIMATE MEN Chest Pads	1.000000
1	HRS ULTIMATE BOY Chest Pads	0.891930
2	HRS ULTIMATE BOY Elbow Pads	0.740572
3	HRS CLUB BOY Thigh Pads	0.721257
4	Parth Collection English Y-pad	0.171612
5	India Inc Women's Solid Casual Shirt	0.159057
6	TeeMoods Casual Full Sleeve Striped Women's Top	0.154496
7	TeeMoods Casual Full Sleeve Striped Women's Top	0.154496
8	TeeMoods Casual Full Sleeve Striped Women's Top	0.154496
9	TeeMoods Casual Full Sleeve Striped Women's Top	0.154496
10	TeeMoods Casual Full Sleeve Striped Women's Top	0.154496

Top 10 similar products

Item - Item based collaborative filtering

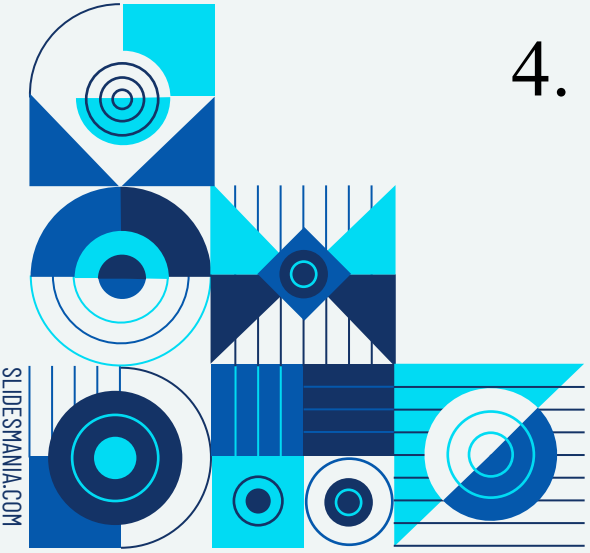
1. Preparation of Utility matrix.
2. Removal of insignificant values and handling null values

```
utility_matrix = df3.pivot_table(values='Rating', index='UserId', columns='ProductId')
utility_matrix = utilty_matrix.dropna(thresh=10, axis=1).fillna(0)
```



3. Finding relationship between products

```
df_item_similarity = utility_matrix.corr(method='pearson')
df_item_similarity.head(10)
```



4. Preparation of user history and recommendation

Top 10 recommendations

```
Below are the recommendation based on the User history -  [('1403790965', 4), ('B0002VHBTU', 5), ('535795531X', 2)]

B0002VHBTU    2.499806
1403790965    1.499566
B000052YKM    0.075251
B00028OSI0    0.041307
B00028MLG6    0.034251
B00014DMQE    0.031249
B0000C0XL8    0.030753
B000142P12    0.023551
B0000DD8VH   -0.000462
B0000589YB   -0.000480
dtype: float64
```

Thank you!