**Recommendations:**

The recommendations given by the model for 5 different examples are as follows ,

Top 5 recommendations of Radiant Bay Floral Flakes Yellow Gold Diamond 18 K Ring are:

1 Radiant Bay Floral Flakes White Gold Diamond 18 K Ring

2 Radiant Bay Floral Fusion Yellow Gold Diamond 18 K Ring

3 Radiant Bay Floral Fusion White Gold Diamond 18 K Ring

4 Radiant Bay Blend Yellow Gold Diamond 18 K Ring

5 Radiant Bay For Forever Yellow Gold Diamond 18 K Ring

Top 5 recommendations of Vishudh Printed Women's Straight Kurta are:

1 Vishudh Printed Women's Anarkali Kurta

2 MASARA Solid Women's Straight Kurta

3 Jaipurkurti Striped Women's Straight Kurta

4 Noor Embroidered Women's Straight Kurta

5 GO LUCKNOW Embroidered Women's Straight Kurta

Top 5 recommendations of Abr Casual Shoes are:

1 People Casual Shoes

2 Zobello Casual Shoes

3 Duke Casual Shoes

4 Clincher Casual Shoes

5 Kujaparis Casual Shoes

Top 5 recommendations of INDRICKA Casual Roll-up Sleeve Solid Women's Top are:

1 INDRICKA Casual Full Sleeve Solid Women's Top

2 INDRICKA Casual Butterfly Sleeve Solid Women's Top

3 PURYS Casual Roll-up Sleeve Printed Women's White Top

4 F&S Casual, Party Short Sleeve Self Design Women's Top

5 Tops And Tunics Casual Butterfly Sleeve Solid Women's Top

Top 5 recommendations of Rorlig RR-028 Expedition Analog Watch - For Men, Boys are:

1 Rorlig RR-030 Essentials Analog Watch - For Men, Boys

2 R.S D&G16 Analog Watch - For Men

3 Q&Q DA37J504Y Analog Watch - For Women

4 Q&Q VQ13-008 Analog Watch - For Girls, Boys

5 Sonata Everyday Analog Watch - For Men

Recommendations using cosine\_similarity and Euclidean distance:

**Cosine-Similarity :**

Top 5 recommendations of Alisha Solid Women's Cycling Shorts are:

1 Mynte Solid Women's Cycling Shorts, Gym Shorts, Swim Shorts

2 Nuteez Printed Women's Basic Shorts

3 Provalley Solid Boy's Basic Shorts

4 Ashdan Solid Women's Basic Shorts

5 UFO Printed Girl's Basic Shorts

**Euclidean Distance:**

Top 5 recommendations of Alisha Solid Women's Cycling Shorts are:

1 Mynte Solid Women's Cycling Shorts, Gym Shorts, Swim Shorts

2 Nuteez Printed Women's Basic Shorts

3 Provalley Solid Boy's Basic Shorts

4 Ashdan Solid Women's Basic Shorts

5 UFO Printed Girl's Basic Shorts

**INSIGHTS :**

Since no user data/preference were given , instead of collaborative based filtering using any complex machine learning/deep learning model , its convenient to opt for item-item content based filtering.

The Data mining technique/algorithm used in this question is of TF-IDF (Term frequency – Inverse Document Frequency) vectorization. The weight or importance of a word is found out by multiplication of term frequency and inverse document frequency and the words in the categorical feature is replaced by a weight vector.

Our aim is to tf-idf vectorize some important features and based on the occurrence and importance of the words in a query , build a recommendation model finding products having similar words.

After scorching through the data and making models only on one features , it showed that its better to combine two features namely product\_name and product\_category\_tree to properly extract the information of the product as product\_category\_tree gives out terms like furniture , clothing which makes it easier to differentiate and product\_name gives the exact details of the product like ‘18k gold ring’ in which product\_category\_tree will only give ‘jewelry ,rings’ . Thus , a new product NameTree was used to combine both features and then it was tf-idf vectorized.

Cosine-similarity and Euclidean Distance gave the similarity measure of a query and other products comparing their respective tf-idf vectors.

These similarity scores were taken , sorted from highest to lowest and the duplicates were removed , finally to give the indices of the closest 5 products to the query .

The five products were printed and hence recommended.

**Observations :**

1. Cold start was prevented as no prior information is required of any new product because tf-df vector just need the frequency of words and calculates weights based on it .
2. Sparsity problems didn’t arise as content based filtering were done on features like name and category tree instead of using features like overall rating or product rating
3. Cosine Similarity and Euclidean distance gave the same results and thus either one of the two can be used. This assignment uses cosine\_similarity