**Recommendation systems based on product similarity**

**Introduction**

In the dynamic landscape of the commercial web, recommendation systems have emerged as indispensable tools that shape our online shopping experiences. These intelligent algorithms leverage vast amounts of data and sophisticated techniques to deliver personalized and relevant suggestions to users, guiding them through the overwhelming sea of choices. However, despite their remarkable capabilities, recommendation systems face a significant hurdle known as the cold start problem. This challenge arises when new users or items enter the system, lacking sufficient historical data for accurate recommendations. The cold start problem poses a formidable obstacle, as traditional algorithms heavily rely on past interactions to predict future preferences. Addressing this issue has become a critical focus for researchers and businesses alike. In this article, we delve into the intricacies of recommendation systems in the commercial web, exploring the nuances of the cold start problem and examining innovative approaches that strive to overcome this challenge.

Today , algorithms like market basket analysis that based on purchase history or collaborative filtering that combining purchase history and consumer characteristics don’t achieve a good result with a new product in a recommendation system , so let’s begin and see what we did .

**Data**

Our databases are from h&m company :

1. Includes all products (104,500) on the company's online shopping site.

2. Photos of the products, there are about 103,400 photos.

3. The purchases made on the site and each purchase is characterized by the buyer's name, the product purchased, date, product price.

Our goal is to recommend a new product that has entered the site in a good way.

First, we will define a "shopping basket" - all the products that a certain person purchased on a certain day . To check the strength of the relationship between each pair of products we will check how many times each 2 products appear together in all the shopping baskets.

For example: our products are - 1,2,3,4,5.

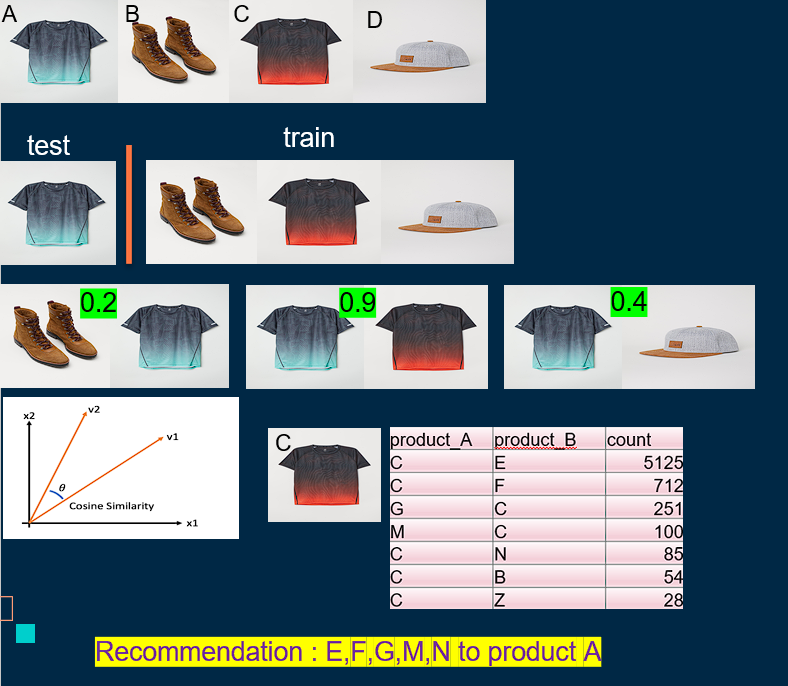


**Methodology**

**Solution 1 : image similarity**

**First algorithm**

1. Gathering Products: The algorithm starts by collecting all the products available in the system.
2. Identifying New Products: Approximately 15% of the products are classified as new, indicating they have no purchase history or limited data.
3. Calculating Similarity: To determine the similarity between each new product and existing products, the algorithm employs a technique called "embedding" using a model called "img2Vec." This model converts the product images into numerical vectors that represent their visual features.
4. Comparing Image Embeddings: The algorithm calculates the vector distance or similarity between the embedding of each new product's image and the embeddings of the product images in the training data. This step helps identify the existing product with the most similar image to the new product.
5. Recommending Related Products: Once the most visually similar existing product is identified, the algorithm recommends the five products that are most frequently purchased together with that existing product. These associated products are likely to be complementary or related to the new product, providing valuable recommendations to users.



**Second algorithm**

The first three steps are the same as in the first algorithm .

1. Selecting Similar "Old" Products: Instead of choosing just one existing product with the most similar image, this algorithm selects the five "old" products (products with purchase history) that have the closest image resemblance to the new product. This step expands the options for recommendation based on similar visual characteristics.
2. Analyzing Purchase Behavior: For each of the five similar "old" products, the algorithm examines the purchase history to determine which product was most frequently bought together with them. This analysis helps identify the products that have the strongest association or correlation with the selected "old" products.
3. תמונה שמכילה טקסט, צילום מסך, תוכנה, סמל מחשב

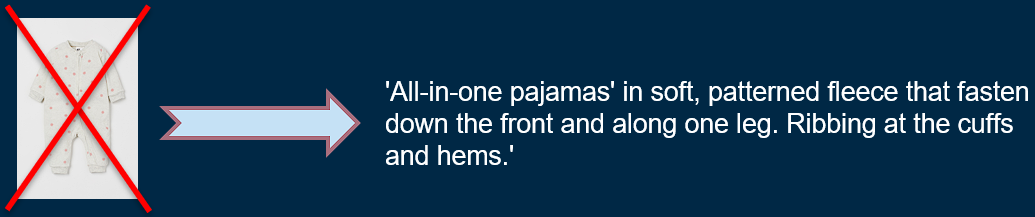
   התיאור נוצר באופן אוטומטיRecommending Related Products: Finally, the algorithm recommends the five products that were most purchased together with the selected "old" products. These recommended products are likely to have a high likelihood of being relevant or complementary to the new product.

**Solution 2 : description similarity**

Just like solution **1** with the **2** algorithms .

The difference is that we use the textual description of the products to look for similarities between the new product and the old products, instead of the image.

We extracted for each textual description a vector representation with the help of one of the Bert models.



**Naive Solution**

To simulate the situation today when trying to solve the problem of a new product entering the system, we built a model based on market basket analysis combined with clustering.

1. Clustering with Kmodes Algorithm: The first step involves performing clustering using the Kmodes algorithm on all products. The clustering is based on the characteristics of the products, such as the product name, category, product color, and other relevant attributes. This process groups similar products together, creating distinct clusters.
2. Division into New and Old Products: After clustering, the products are divided into two categories: new products and "old" products with a purchase history. This division is performed based on the available historical data, with products without purchase history classified as new products and products with purchase history classified as "old" products.
3. Finding Popular "Old" Products per Cluster: For each cluster, the algorithm identifies the five most popular "old" products within that cluster. The popularity index is determined by the frequency of purchases or the occurrence of these products in the historical data. This step helps to understand the preferences and trends within each cluster.

* In each cluster can be new and old products but for the new products we don’t have an early knowledge .

1. For each product we recommend the five most popular old products according to the cluster to which it is associated, this recommendation assumes that new products within a certain cluster share similar characteristics or appeal to customers with similar preferences.
2. The number of clusters chosen by elbow measure :

תמונה שמכילה טקסט, תרשים, קו, עלילה

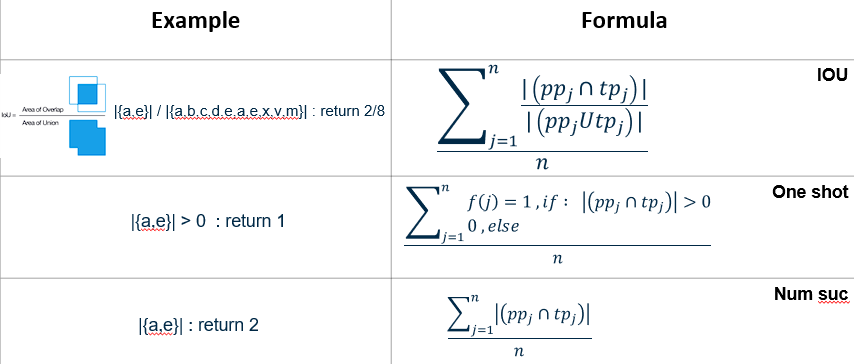
התיאור נוצר באופן אוטומטי

**Measures**

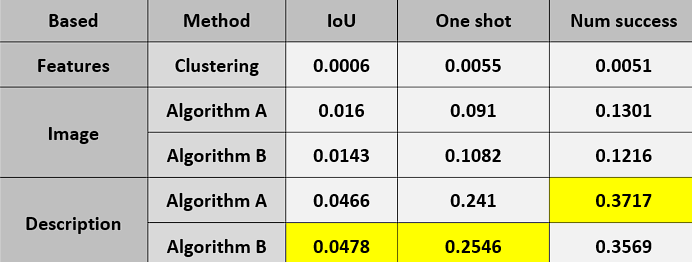
For each product J, we need to compare the 5 products that we predict will be bought the most with the new product, and the 5 products that were bought the most with the new product .

Example :

To new J product : predict – {a,b,c,d,e} , true –{ a,e,x,v,m}



**Results**



**Conclusion**

As we can see the description gives us the best solution and we recommend using it to find similarities between a new product and existing products in recommendation systems, although we recommend testing a hybrid model that combines both the image and the textual description in the recommendation.