

Assignment 2 Report

Step 1: Load the data with columns as ['Product ID.', 'Image', 'Review text'],

Step 2: Pre process the Image column to convert the links from string to list form

Step 3: Convert all the image links their respective images by downloading the images.

Step 4: Pre process the images by resizing them, adjusting the contrast and the brightness of the image

Step 5: Extract the features of all the images using VGG16 model and normalize them.

Step 6: Preprocess the review text using the same techniques used in assignment 1

Step 7: Calculate the tf, idf and tfidf scores of all the preprocessed texts

Step 8: To get the most similar images Calculate the the cosine similarity of the extracted images features and the input image features and then sort them to get the top 3 most similar images.

Step 9: To get the most similar reviews calculate the cosine similarity of the tfidf scores of the reviews and the tfidf scores of the input review features and then sort them to get the top 3 most similar reviews.

Step 10: To get the composite similarity score, get the average of the cosine similarity score of the image and the review.