Report

Question 1,2

In First question I have converted the csv input file in such a way that a dictionary is created where the key is product ID and for each key/product ID we have a tuple which contains the image URL converted to PIL object and review text in a string format.Total number of IDs were coming out to be 994 and total number of images are to be 1640.Product ID 3073 had a nan review therefore i converted it to an empty string to retain the image and not deleting the id for better RGB features during imagenet model evaluation.

In Second question I have preprocessed the image by resizing to a size which is 224x224 which is compatible with VGG16 model which takes the input features as (224x224x3) which takes 3 RGB channels.For further preprocessing I have augmented to gain more information about the dataset for better feature extraction.In General models like VGG16 which are based on CNN architecture are quite powerful and within its nodes it pre processes the input therefore I have also used their preprocessing input for best feature extraction possible.I have further normalised the features for all images in every product ID.For text preprocessing i have used the basic preprocessing techniques and used Lemmatization which is better for searching which is ideal purpose of the assignment.I have calculated the TF-idf scores from scratch and adjusted the dictionary such that every product ID as a key has a value which is itself a dictionary where each term and its respective TF-idf score is printed.

Question 3

I have saved all the features and earlier dictionary in pickles and loaded them for image based and text based retrieval.I have parsed the input link and review and which also checks if the link is legit by using RE and regular expression.It raises error if the url link is wrong.
For image based retrieval I have loaded the same base model for finding the input features of the input image and normalised it.I have also preprocessed the input text by following the similar techniques.Then I have calculated the cosine similarities.One assumption I have made is that I have taken the similarity score with every product ID image in case of multiple images and returned the max cosine similarity image.However the other images can be printed because we need to deal with the product id if we use the reference of Amazon search for example.Taking average could lead to ambiguous results because one image could be very relevant and the other could not be relevant(I checked the dataset there are few cases) so the overall similarity score would decrease so its better to print max and other subsequent product id images for reference.I have printed the corresponding review for that ID and calculate the cosine similarity score and printed the reviews. Similar process is done for text based retrieval. First I have calculated the tf idf of input and calculated the similarity score. Now for the Kids retrieved I have calculated the cosine similarity for image based and printed the image and score.

Question 4

In Combined retrieval I have printed all the features list and have calculated the combined composite similarity score pairwise which do match as per the ranking and Rank the pairs based on the composite similarity score.

Question 5

After calculating the combined similarity scores of Image based retrieval and Text based retrieval we can see the the upper rankings are from the image based retrieval whereas the text based retrieval are somewhat in lower rankings.We can therefore conclude that image based retrieval could be more relevant and more effective for the retrieval system based on input image,review pair.If we look at the list of tuple for all such cases we can see that image based retrieval have a better relevancy based on the score.
There could be few reasons why is that so,
Images do contain a lot of information and features with it.It consists of several attributes suchy as RGB colour features, texture, and more important features in case of image matching.The image arrays contain pixel by pixel features that can cover a larger corpus of context with respect to the text based retrieval.the images provide a more additional information to review text which can improve the overall accuracy and similarities as per the question.

Reviews often contain the wording,tones,perspective and subjectiveness along with varied biassed-unbiased reviews which may not accurately describe the same product.One assumption is that I have included the tone and perspective tokens in the text based retrieval which may help to better extract the relevant rankings.The images are more definitive and consistent depiction of product giving a more clear relationship between images and review texts.

images provide a more intuitive and immediate understanding of the product compared to text. consumers often rely on visual cues to make purchasing decisions, and images can convey aspects such as product appearance, design, and functionality more effectively than text alone. This visual representation can lead to a stronger emotional connection with the product, influencing the perception of its quality and desirability. Therefore, in a retrieval system, prioritising image-based features may lead to more accurate and relevant results for users seeking information or recommendations based on visual characteristics.

Some additional challenges and potential improvements in the retrieval process:
incorporating multiple related images: many products have multiple related images showcasing different angles, features, or variations. However, current retrieval systems often consider only one representative image per product id, potentially missing out on valuable information. improvement: enhance the retrieval process to consider all related images associated with a product id. This can involve incorporating techniques such as ensemble learning or attention

mechanisms to aggregate information from multiple images effectively.risk of misleading similarity based on visual features: visual similarity alone may not always indicate relevance, especially if two images share similar visual features but belong to different product categories or have different contexts. For instance, products with similar colour schemes may be mistakenly ranked higher. improvement: develop more sophisticated models that not only analyse visual features but also consider contextual information and semantic understanding to ensure the relevance of retrieved results.user-generated content and informal language: user-generated content, such as reviews, can contain informal language, misspellings, abbreviations, and slang, which pose challenges for text-based retrieval systems. improvement: implement robust preprocessing techniques to handle noisy and informal text, including spell checking, normalisation, and sentiment analysis. Additionally, leverage natural language processing models trained on user-generated content to better understand and extract meaningful information from reviews.addressing concept drift and evolving trends: products, preferences, and trends evolve over time, leading to concept drift in the dataset. However, retrieval models trained on historical data may become less effective in capturing current user preferences and trends. improvement: implement mechanisms for continuous learning and adaptation, such as online learning techniques and active learning strategies. regularly update the model with new data and monitor performance metrics to identify and adapt to evolving trends and user preferences.balancing interpretability and complexity: more complex models may achieve higher accuracy but at the cost of interpretability, making it challenging for users to understand and trust the retrieval results. improvement: strike a balance between model complexity and interpretability by incorporating explainable ai techniques, such as attention mechanisms and feature visualisation. provide users with insights into how the model generates recommendations and allow for user feedback to refine the retrieval process iteratively Assigning equal weights to both the top three images and top three text may not always be optimal. in multimodal retrieval systems, it's feasible to prioritise one modality over the other based on its significance or relevance for the specific task. This can be achieved by assigning different weights to the textual and visual components, thereby influencing the total similarity score accordingly.Consideration of temporal dynamics: product preferences, trends, and user behaviours can change over time, leading to shifts in the relevance of retrieval results. improvement: implement mechanisms to capture temporal dynamics in the retrieval process, such as time-sensitive relevance scoring or incorporating time-stamped data into the model. This allows the system to adapt and provide more accurate and up-to-date recommendations based on the current context and trends.