

Project: Fashion Search AI on Myntra Dataset

Introduction:

To Create a generative search system capable of searching a plethora of product descriptions to find and recommend appropriate choices against a user query. You may use this [Myntra dataset](#) on Kaggle to build such a system.

Objectives:

The main objective of this project is to develop a fashion query response system that utilizes AI models to provide detailed and user-friendly responses to fashion-related queries. The system aims to enhance user experience by generating informative and contextually relevant answers, thereby assisting users in finding fashion items based on their preferences.

Design:

The project involves two main layers: the search layer and the generation layer. The search layer is responsible for retrieving relevant fashion items from the dataset based on keyword matching or predefined criteria. The generation layer utilizes advanced AI models like GPT-3.5 to generate detailed responses to user queries, incorporating context and generating natural language responses.

Implementation:

The implementation involved several steps, including data preprocessing, model integration, and query response generation. Data Preprocessing: The CSV dataset was formatted to enhance data quality and readability. Blank entries were replaced, decimal points were standardized, and unnecessary columns were removed. Text columns were cleaned to remove HTML tags and extra characters.

2. Read, Process the csv Files

```
# Define the file path to the Fashion Dataset CSV file
file_path = '/content/drive/MyDrive/Gen_AI_Fashion_AI/Fashion Dataset v2.csv'

# Read the Fashion Dataset CSV file into a Pandas DataFrame
fashion_data = pd.read_csv(file_path)

# Display the first few rows of the DataFrame to verify the data has been loaded correctly
fashion_data.head()
```

	p_id	name	products	price	colour	brand	img	ratingCount	avg_rating	description	p_attributes
0	17048614	Khushal K Women Black Ethnic Motifs Printed Ku...	Kurta, Palazzos, Dupatta	5099.0	Black	Khushal K	http://assets.myntassets.com/assets/images/170...	4522.0	4.418399	Black printed Kurta with Palazzos with dupatta...	{'Add-Ons': 'NA', 'Body Shape ID': '443,333,32...
1	16524740	InWeave Women Orange Solid Kurta with Palazzos...	Kurta, Palazzos, Floral Print Dupatta	5899.0	Orange	InWeave	http://assets.myntassets.com/assets/images/165...	1081.0	4.119334	Orange solid Kurta with Palazzos with dupatta<...	{'Add-Ons': 'NA', 'Body Shape ID': '443,333,32...
2	16331376	Anubhutee Women Navy Blue Ethnic Motifs Embrol...	Kurta, Trousers, Dupatta	4899.0	Navy Blue	Anubhutee	http://assets.myntassets.com/assets/images/163...	1752.0	4.161530	Navy blue embroidered Kurta with Trousers with...	{'Add-Ons': 'NA', 'Body Shape ID': '333,424', ...
3	14709966	Nayo Women Red Floral Printed Kurta With Trous...	Kurta, Trouser, Dupatta	3699.0	Red	Nayo	http://assets.myntassets.com/assets/images/147...	4113.0	4.088986	Red printed kurta with trouser and dupatta ...	{'Add-Ons': 'NA', 'Body Shape ID': '333,424', ...
4	11056154	AHIKA Women Black & Green Printed Straight Kurta	Kurta	1350.0	Black	AHIKA	http://assets.myntassets.com/assets/images/110...	21274.0	3.978377	Black and green printed straight kurta, has a...	{'Body Shape ID': '424', 'Body or Garment Size...

Model Integration:

Advanced AI models like GPT-3.5 were integrated into the system to generate responses to user queries. Queries were passed through the model to generate detailed and contextually relevant responses.

Query Response Generation:

User queries were processed by both the search layer and the generation layer. The search layer retrieved relevant fashion items from the dataset. The generation layer utilized AI models to generate detailed responses to user queries, incorporating context and generating natural language responses.

Here are the few queries that were used to test the model:

#Query 1:

an orange summer dress or kurta to wear over blue denim jeans.

```
[14] # Prompt the user to enter a fashion-related query with specific requirements
query = input("Please enter your fashion-related query. For example, you can mention the color, size, price range, style, and any specific patterns or designs you prefer: ")

#Query 1: a orange summer dress or kurta to wear over blue denim jeans .
#Query 2: I'm looking for office wear sarees in elegant colors like pink, violet, or green. I prefer sarees with full embroidered designs, suitable for professional settings. There's no specific price
#Query 3: I'm searching for a versatile black leather jacket, suitable for various occasions and effortlessly complementing any outfit. Preferably in size XL.
#Query 4: I'm seeking ethnic attire suitable for adults, with a preference for Ishin brand. These outfits are intended for festival wear. The clothing set I'm looking for should include a Navy Blue a
#Query 5: A silk purple summer dress, for beach party

Please enter your fashion-related query. For example, you can mention the color, size, price range, style, and any specific patterns or designs you prefer: Pink, No price limit, XL, ETHNIC
```

Search Layer:

Query:
Pink, No price limit, XL, ETHNIC

Top 3 Search Layer

ID: 16556024

ID: 15816452

ID: 15980892

ID: 11362372

ID: 18301324

	Metadata	Documents	Distances	IDs
0	{'Brand': 'Global Desi', 'Colour': 'Pink', 'De...	Pink and white ethnic motifs embroidered tunic...	0.300063	16556024
1	{'Brand': 'Fabindia', 'Colour': 'Pink', 'Descr...	Pink and golden regular topEt...	0.311516	15816452

Generation layer:

These items offer various options in the pink color with ethnic influences, providing a mix of embroidered details, prints, and crop styles. For further details, you can visit our website to explore th

ID: 11362372

ID: 16556024

ID: 15980892

#Query 2:


I'm looking for office wear sarees in elegant colors like pink, violet, or green. I prefer sarees with full embroidered designs, suitable for professional settings. There's no specific price range, as quality is my priority. These sarees will be worn for office meetings and formal events.

Search Layer:

Query:
office wear Saree, green


Top 3 Search Layer

ID: 14121912




Metadatas

ID: 13771272




Documents

ID: 17193238




Distances

ID: 11642362



IDs


ID: 14960498




Generation layer:

Unfortunately, none of the items specifically match the query for a green Saree. However, if you are open to exploring different options for office wear in green, these fashion items could be a stylis


ID: 11642362



ID: 13771272



ID: 14960498



#Query 3:

I'm searching for a versatile black leather jacket, suitable for various occasions and effortlessly complementing any outfit. Preferably in size XL.

4. Semantic Search with Cache

In this section, we will perform a semantic search of a query in the collections embeddings to get several top semantically similar results.

Prompt the user to enter a fashion-related query with specific requirements

query = input("Please enter your fashion-related query. For example, you can mention the color, size, price range, style, and any specific patterns or designs you prefer: ")

#Query 1: a orange summer dress or kurta to wear over blue denim jeans .

#Query 2: I'm looking for office wear sarees in elegant colors like pink, violet, or green. I prefer sarees with full embroidered designs, suitable for professional settings. There's no specific price

#Query 3: I'm searching for a versatile black leather jacket, suitable for various occasions and effortlessly complementing any outfit. Preferably in size XL.

#Query 4: "I'm seeking ethnic attire suitable for adults, with a preference for Ishin brand. These outfits are intended for festival wear. The clothing set I'm looking for should include a Navy Blue a

#Query 5: A silk purple summer dress, for beach party

Please enter your fashion-related query. For example, you can mention the color, size, price range, style, and any specific patterns or designs you prefer: XL size, versatile black leather jacket

Search Layer:

Query:
XL size, versatile black leather jacket

Top 3 Search Layer


ID: 14846678


ID: 17968486


ID: 18163884


ID: 18161132


ID: 14218480











Metadatas

Documents

Distances

IDs


Generation layer:


Unfortunately, based on the available information, there doesn't seem to be a specific black leather jacket in XL size listed in the dataset. For further details on black leather jackets, particularly if you have any specific preferences or requirements for the black leather jacket, feel free to provide more details for better assistance in finding the right product.


ID: 14846678

ID: 17968486

ID: 18161132







#Query 4:

"I'm seeking ethnic attire suitable for adults, with a preference for Ishin brand. These outfits are intended for festival wear. The clothing set I'm looking for should include a Navy Blue and golden foil printed top with a round neck and three-quarter sleeves, along with a matching skirt featuring a drawstring closure. Both pieces should be made of pure cotton.

4. Semantic Search with Cache

In this section, we will perform a semantic search of a query in the collections embeddings to get several top semantically similar results.

Prompt the user to enter a fashion-related query with specific requirements
query = input("Please enter your fashion-related query. For example, you can mention the color, size, price range, style, and any specific patterns or designs you prefer: ")

#Query 1: a orange summer dress or kurta to wear over blue denim jeans .
#Query 2: I'm looking for office wear sarees in elegant colors like pink, violet, or green. I prefer sarees with full embroidered designs, suitable for professional settings. There's no specific price
#Query 3: I'm searching for a versatile black leather jacket, suitable for various occasions and effortlessly complementing any outfit. Preferably in size XL.
#Query 4: "I'm seeking ethnic attire suitable for adults, with a preference for Ishin brand. These outfits are intended for festival wear. The clothing set I'm looking for should include a Navy Blue and golden foil printed top with a round neck and three-quarter sleeves, along with a matching skirt featuring a drawstring closure. Both pieces should be made of pure cotton."
#Query 5: A silk purple summer dress, for beach party

you prefer: seeking ethnic attire suitable for adults, with a preference for Ishin brand. These outfits are intended for festival wear. The clothing set I'm looking for should include a Navy Blue and golden foil printed top with a round neck and three-quarter sleeves, along with a matching skirt featuring a drawstring closure. Both pieces should be made of pure cotton.

Search Layer:

Top 3 Search Layer



Generation layer:



There were many more queries that was used, for project purpose I have listed only few here as well as at the code terminal.

Challenges:

Several challenges were encountered during the implementation process:

Metadata Processing:

Initially, there were issues with processing metadata, which affected the display of results. This was resolved through debugging and refining the metadata processing code.

Dataset Chunking:

Loading the entire dataset without chunking posed a challenge due to memory constraints. However, chunking was not implemented due to time constraints and the manageable size of the dataset.

Lessons Learned:

Proper data preprocessing is crucial for ensuring data quality and readability. Integrating advanced AI models can significantly enhance the system's capabilities in generating detailed

and contextually relevant responses. Handling large datasets requires careful consideration of memory constraints and implementation of efficient data processing techniques.

Future Scope:

As a future scope, the project can be extended by implementing it as a Flask web application and hosting it on a website. This would provide better interaction and aesthetics for users. Additionally, rephrasing prompts and introducing interactive sessions with criteria-based filters could further enhance the system's performance and user experience.

Conclusion:

Upon comparison of the search query outcomes from both the search and generation layers, it becomes apparent that the generation layer produces more detailed and comprehensible results. While the search layer retrieves relevant information through keyword matching or set criteria, the generation layer utilizes advanced AI models like GPT-3.5 to dynamically generate responses, incorporating context and furnishing tailored and informative answers.

The generation layer's capability to comprehend query context and furnish responses in natural language enhances overall readability and user engagement. Its flexibility allows it to adapt to various query formats and provide customized responses, catering to diverse user preferences. This versatility makes it a valuable tool for tasks necessitating comprehensive understanding and communication of information.

In summary, while the search layer efficiently retrieves relevant information, the generation layer significantly enhances output quality and readability, making it a preferred choice for tasks requiring detailed and user-friendly responses.