Project Report

Semantic Spotter Fashion Recommendor

Your Name

Aparana Bhatt

Table of Contents

1. Ideation

While we explore online women's fashion clothes we get confused as to which option to choose from and find top preferred style.If we have our trained custom recommendation assistant life would be easier and our time will also be saved.

If we know the correct keyword , we can search for good products with less effort but if we have a requirement and our search query is very long then we do not get required suggestions and end up wasting a lot of time searching for the right product.

This is where semantic spotter fashion recommender can be of extreme use because we can write our needs to it and it will understand and return good suggestions in a short time and we would need less time to search for the product online.

It will greatly streamline our choices and help us select our favorite style on the go.

2. All About Data

I have used a publicly available Myntra dataset and sampled it according to the project requirement. The data set is available at [Myntra Fashion Product Dataset (kaggle.com)]

Data consists of a list of women products with description,name,brand and other metadata . Some of which have been used to create a minimal instruction/decision set for the helpmate to take help with.

In [5]:
```
import pandas as pd;

df=pd.read_csv(myntraDataSet)

df.head()
```

Out[5]:

	Unnamed: 0.1	Unnamed: 0	p_id	name	price	colour	brand	img	ratingCount	avg_rating	descript
0	5975	35	12221918.0	SASSAFRAS Women Black Solid Cropped Basic Tube...	1799.0	Black	SASSAFRAS	http://assets.myntassets.com/assets/images/122...	657.0	4.079148	Black s cropp ba jumps with wais
1	5445	495	17365646.0	I Saw It First Women Black & Yellow Floral Pri...	1299.0	Black	I Saw It First	http://assets.myntassets.com/assets/images/173...	NaN	NaN	Black a yellow fl printec line ab
2	2666	686	17570446.0	SHOWOFF Women Blue Boyfriend Fit High-Rise Low...	3720.0	Blue	SHOWOFF	http://assets.myntassets.com/assets/images/175...	NaN	NaN	Dark sha light f blue je
3	5145	195	17829400.0	Forever New Pink Katerina Belted Tiered Skirt	5200.0	Pink	Forever New	http://assets.myntassets.com/assets/images/178...	NaN	NaN	<p>< <l flow roma skir we
4	7754	324	16012536.0	Fabindia Blue Embroidered Tunic	2690.0	Blue	Fabindia	http://assets.myntassets.com/assets/images/160...	NaN	NaN	B embroide tunic ,ha round ne t

The data is processed and sampled to 500 entries only for simplicity.We have also filtered the rows having empty description columns as description is the major source of semantic context in the process. We have also added MetaData column, text length column int the dataset during processing and saved this sampled data in a final csv file, which is being used in the next steps.

In [40]:
```
## filter rows where description is not string type.
df.loc[df['description'].apply(lambda x: isinstance(x, str)), 'Text_Length'] = df['description'].apply(lambda x: len(x.split(' ')
df.head()
```

Out[40]:

	Unnamed: 0	p_id	name	price	colour	brand	img	ratingCount	avg_rating	description	p_attribut
0	0	17048614.0	Khushal K Women Black Ethnic Motifs Printed Ku...	5099.0	Black	Khushal K	http://assets.myntassets.com/assets/images/170...	4522.0	4.418399	Black printed Kurta with Palazzos with dupatta...	{'Add-Or 'NA', 'Bo Shape I '443,333,3
1	1	16524740.0	InWeave Women Orange Solid Kurta with Palazzos...	5899.0	Orange	InWeave	http://assets.myntassets.com/assets/images/165...	1081.0	4.119334	Orange solid Kurta with Palazzos with dupatta<...	{'Add-Or 'NA', 'Bo Shape I '443,333,3
2	2	16331376.0	Anubhutee Women Navy Blue Ethnic Motifs Embroi...	4899.0	Navy Blue	Anubhutee	http://assets.myntassets.com/assets/images/163...	1752.0	4.161530	Navy blue embroidered Kurta with Trousers with...	{'Add-Or 'NA', 'Bo Shape I '333,424',
3	3	14709966.0	Nayo Women Red Floral Printed Kurta With Trous...	3699.0	Red	Nayo	http://assets.myntassets.com/assets/images/147...	4113.0	4.088986	Red printed kurta with trouser and dupatta ...	{'Add-Or 'NA', 'Bo Shape I '333,424',
4	4	11056154.0	AHIKA Women Black & Green Printed Straight 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 Sha ID': '42 'Body Garm Size

3. Implementation Details

The code is written in python and various other components and apis used are mentioned below

1) Python3
2) Openai library
3) LangChain
4) Chroma DB
5) Gpt-3.5 Turbo used for recurring chats
6) Jupyter notebook

4. System Design

System design contains below major layers and steps for a successful response.

1) **Data Layer** : This layer is the first layer and first step where relevant data is sourced and preprocessed.
2) **Text Embedding Layer:** In this layer the data loaded via data loader is converted to embeddings by using OpenAIEmbeddings module and then stored in a vector store called chroma db.
3) **Semantic Search Layer**: We use chroma db as a semantic search store here. This is where the document embeddings are stored pooled from the data layer after preprocessing.
4) **LangChain Chains for prompting**:In this layer we use a stuff chain to take input from a user , convert into prompt template and then pass to LLM for providing a solution for user query.
5) **Generation Layer:** This layer makes use of promptTemplate of lang chain using a simple chain to generate the results.

Code overview and semantic spotter example

Below is the overview of main methods used in the code which are empowering the Movie Assistant .

1) **DataLoader to load data and parse via langchain**: Langchain provides various data loader to support different format. We have used a dataframe data loader to load the fashion dataset data.

```
In [6]: from langchain_community.document_loaders import DataFrameLoader
```

CSV Files

```
In [7]: loader = DataFrameLoader(df, page_content_column="p_id")
```

```
In [8]: loader.load()
```


```
ube-Jumpsuit-9815-1.jpg', 'ratingCount': 657.0, 'avg_rating': 4.079147640791477, 'description': "Black solid cropped basic ju
mpsuit with waist tie-up and side taping detail, strapless, cuffed hemsThe model (height 5'8'') is wearing a size SMaterial:
Polyester<br>  Machine Wash", 'p_attributes': "{'Body or Garment Size': 'To-Fit Denotes Body Measurements in', 'Closure': 'N
A', 'Fabric': 'Polyester', 'Neck': 'Strapless', 'Number of Pockets': 'NA', 'Pattern': 'Solid', 'Sleeve Length': 'Sleeveless',
'Surface Styling': 'Waist Tie-Ups', 'Type': 'Basic Jumpsuit', 'Wash Care': 'Machine Wash'}", 'Metadata': "{'name': 'SASSAFRAS
Women Black Solid Cropped Basic Tube Jumpsuit', 'color': 'Black', 'brand': 'SASSAFRAS'}", 'Text_Length': 27}),
 Document(page_content='17365646.0', metadata={'Unnamed: 0.1': 5445, 'Unnamed: 0': 495, 'name': 'I Saw It First Women Black &
Yellow Floral Print A-line Skirt', 'price': 1299.0, 'colour': 'Black', 'brand': 'I Saw It First', 'img': 'http://assets.mynta
ssets.com/assets/images/17365646/2022/4/11/0e04b8a5-aa86-4431-a448-410a56f715691649657055417DressesISawItFirstWomenDressesISa
wItFirstWomenDressesISawItF1.jpg', 'ratingCount': nan, 'avg_rating': nan, 'description': "Black and yellow floral printed A-l
ine above the knee length skirt, has elasticated waistband and slip-on closure.The model (height 5'8) is wearing a size 2895%
Polyester, 5% Elastane<br>Machine wash", 'p_attributes': "{'Add-Ons': 'NA', 'Body Shape ID': '333,424', 'Body or Garment Siz
e': 'To-Fit Denotes Body Measurements in', 'Character': 'NA', 'Closure': 'Slip-On', 'Fabric': 'Polyester', 'Fabric Type': 'N
A', 'Hemline': 'Flared', 'Knit or Woven': 'Knitted', 'Length': 'Above Knee', 'Lining': 'NA', 'Main Trend': 'Floral Print', 'N
umber of Pockets': 'NA', 'Occasion': 'Casual', 'Pattern': 'Printed', 'Print or Pattern Type': 'Floral', 'Slit Detail': 'No Sl
it', 'Sub Trend': 'New Basics', 'Surface Styling': 'NA', 'Sustainable': 'Regular', 'Transparency': 'Opaque', 'Type': 'A-Lin
e', 'Wash Care': 'Machine Wash', 'Wedding': 'NA'}", 'Metadata': "{'name': 'I Saw It First Women Black & Yellow Floral Print A
-line Skirt', 'color': 'Black', 'brand': 'I Saw It First'}", 'Text_Length': 29}),
```


2) **Embedding And Searching**: After loading data it is to be converted to embeddings and then semantic search is to be applied .

```
In [27]: # Store the split document in ChromaDB
         db = Chroma.from_documents(data, OpenAIEmbeddings())
```

Perform Similarity Search

```
In [29]: query = "find red kurta for me "
         docs = db.similarity_search(query)
         print(docs[0].metadata['description'])
```

```
Trendy and versatile, this playsuit is a worthy pick to add to your casualwear closet. This playsuit comes with an alluring eth
nic motifs print and V-neck. <ul><li>Stylish black shade</li><li>Beautiful ethnic motifs print</li><li>V-neck</li><li>Slee
veless</li><li>Polyester, machine wash</li></ul><b>Trend Alert</b><br>Playsuits flaunt sophisticated and charming details, like
knots, flared sleeves, plunge necklines, etc. in beautiful pastel palettes. It includes attractive prints and patterns, such as
checks, floral, nautical stripes, and Nu Boheme, which enhance your overall appeal.The model (height 5'8) is wearing a size SPo
lyester<br>Machine wash<br>Wash dark colour separately<br>Do not bleach<br>Warm iron if needed
```

LangChain also support all major vector stores and databases such as FAISS, ElasticSearch, LanceDB, Milvus, Pinecone etc. Refer to the API documentation for the complete list.

3)

4) Semantic Spotter Recommender example

Query Snapshots

```
In [49]:
llm = ChatOpenAI(model_name="gpt-3.5-turbo")
chain = create_stuff_documents_chain(llm, chat_prompt)

In [56]: query=input()

need a few red kurta

In [57]: chain.invoke({"context": data,"product_request":query})

Out[57]: 'Here are some red kurta options for you:\n\n1) Brand: 17365646.0\n   Description: Red Cotton Kurta\n   Price: 17365646.0\n\n2)
Brand: 17570446.0\n   Description: Red Silk Kurta with Golden Embroidery\n   Price: 17570446.0\n\n3) Brand: 17829400.0\n   Desc
ription: Red Linen Kurta with Floral Print\n   Price: 17829400.0\n\n4) Brand: 16012536.0\n   Description: Red Chanderi Kurta wi
th Zari Work\n   Price: 16012536.0\n\n5) Brand: 14579962.0\n   Description: Red Crepe Kurta with Sequin Detailing\n   Price: 14
579962.0\n\nFeel free to let me know if you need more options or details on any specific product!'

In [58]: query=input()
chain.invoke({"context": data,"product_request":query})

jeans please

Out[58]: 'Here are some affordable and stylish jeans for you:\n\n1) Brand: 17365646.0\n   Dress Description: Skinny High Ankle Jeans\n
Price: 17365646.0\n\n2) Brand: 17570446.0\n   Dress Description: Slim Fit Mid-Rise Jeans\n   Price: 17570446.0\n\n3) Brand: 178
29400.0\n   Dress Description: Bootcut Jeans\n   Price: 17829400.0\n\n4) Brand: 16012536.0\n   Dress Description: Distressed Bo
yfriend Jeans\n   Price: 16012536.0\n\n5) Brand: 14579962.0\n   Dress Description: Flare Leg Jeans\n   Price: 14579962.0\n\n6)
Brand: 18534256.0\n   Dress Description: High Waisted Skinny Jeans\n   Price: 18534256.0\n\n7) Brand: 19147756.0\n   Dress Desc
ription: Cropped Straight Leg Jeans\n   Price: 19147756.0\n\n8) Brand: 16196158.0\n   Dress Description: Wide Leg Jeans\n   Pri
ce: 16196158.0\n\n9) Brand: 18780564.0\n   Dress Description: Mom Jeans\n   Price: 18780564.0\n\n10) Brand: 12101936.0\n   Dre
ss Description: Distressed Skinny Jeans\n   Price: 12101936.0\n\nFeel free to let me know if you need more options or details
on any specific product!'

In [59]:
query=input()
chain.invoke({"context": data,"product_request":query})

I am looking for white tops for women
```


```
In [58]: query=input()
chain.invoke({"context": data,"product_request":query})

jeans please

Out[58]: 'Here are some affordable and stylish jeans for you:\n\n1) Brand: 17365646.0\n   Dress Description: Skinny High Ankle Jeans\n
Price: 17365646.0\n\n2) Brand: 17570446.0\n   Dress Description: Slim Fit Mid-Rise Jeans\n   Price: 17570446.0\n\n3) Brand: 178
29400.0\n   Dress Description: Bootcut Jeans\n   Price: 17829400.0\n\n4) Brand: 16012536.0\n   Dress Description: Distressed Bo
yfriend Jeans\n   Price: 16012536.0\n\n5) Brand: 14579962.0\n   Dress Description: Flare Leg Jeans\n   Price: 14579962.0\n\n6)
Brand: 18534256.0\n   Dress Description: High Waisted Skinny Jeans\n   Price: 18534256.0\n\n7) Brand: 19147756.0\n   Dress Desc
ription: Cropped Straight Leg Jeans\n   Price: 19147756.0\n\n8) Brand: 16196158.0\n   Dress Description: Wide Leg Jeans\n   Pri
ce: 16196158.0\n\n9) Brand: 18780564.0\n   Dress Description: Mom Jeans\n   Price: 18780564.0\n\n10) Brand: 12101936.0\n   Dre
ss Description: Distressed Skinny Jeans\n   Price: 12101936.0\n\nFeel free to let me know if you need more options or details
on any specific product!'

In [59]:
query=input()
chain.invoke({"context": data,"product_request":query})

I am looking for white tops for women
```


```
Out[59]: 'Here are some white tops for women that you might like:\n\n1) Brand: 17365646.0\n   Dress Description: White Button-Up Blouse
\n   Price: $49.99\n\n2) Brand: 18534256.0\n   Dress Description: White Off-Shoulder Top\n   Price: $29.99\n\n3) Brand: 1914775
6.0\n   Dress Description: White Ruffle Sleeve Blouse\n   Price: $39.99\n\n4) Brand: 17158526.0\n   Dress Description: White La
ce Trim Tank Top\n   Price: $19.99\n\n5) Brand: 18004088.0\n   Dress Description: White Peplum Hem Blouse\n   Price: $34.99\n\n
6) Brand: 18003804.0\n   Dress Description: White Crochet Detail Top\n   Price: $27.99\n\n7) Brand: 17931206.0\n   Dress Descri
ption: White V-Neck Tunic\n   Price: $22.99\n\n8) Brand: 17765672.0\n   Dress Description: White Cold Shoulder Top\n   Price:
$31.99\n\n9) Brand: 18068668.0\n   Dress Description: White Tie-Front Blouse\n   Price: $25.99\n\n10) Brand: 18813862.0\n   Dr
ess Description: White Embroidered Peasant Top\n   Price: $37.99\n\nFeel free to ask if you need more information on any speci
fic product!'
```

Challenges and Future Scope

Challenges

1) Identifying minimal subset that would yield proper data for semantic search documents in the dataset .
2) The support code was written on the older openai version, the common modules using openai apis had to be rewritten to handle response structure and api call changes.
3) The score generation and query response is limited by the dataset size and user requirement. Since the sample dataset is small, the query suggestions are not much satisfactory however working well with respect to the kind of content that is present.
4) Frequent data updates in the chroma db database for very large datasets is taxing.

Future Scope

1) Store user shopping patterns and when the user logs in provide users with similar suggestions even if the user has not yet searched anything. This approach can be used to also refine user likes and dislikes.
2) We can also include price in search response in future so that we can provide user with more refined searches.However since price keeps changing , we should not make it part of embeddings but of metadata.

4. Conclusion

This project showcases the utility of semantic search combined with the power langchain stuff chaining to improve and enhance user experience to more humane and also reduce online shopping time expenditure of the user .