

### Table of Contents

1. Introduction
2. Objectives
3. Dataset
4. Methodology
  - 4.1 Embeddings
  - 4.2 Vector Store
  - 4.3 Caching
  - 4.4 Retrieval
  - 4.5 Reranking
  - 4.6 Generation
5. System Architecture
6. Implementation Highlights
7. Results
8. Challenges
9. Lessons Learned
10. Future Work

## 1. Introduction

The rapid growth of online fashion platforms has led to vast product catalogs, making it difficult for users to quickly discover relevant items. Traditional keyword-based search often falls short in capturing the nuance of natural language queries. To address this, Fashion Search AI was developed as a generative semantic search system capable of retrieving and recommending fashion items using product descriptions from the Myntra dataset.

## 2. Objectives

- Build a fast, scalable fashion search system that supports natural-language queries.
- Use local embeddings to persist vector representations and ensure reproducibility.
- Improve retrieval relevance using cross-encoder reranking.
- Enable generative responses that summarize the most relevant items.

## 3. Dataset

Source: Myntra dataset (FashionLargeDataset.csv, 14,214 items) with corresponding images.

**Fields:** p\_id, name, brand, price, avg\_rating, description, p\_attributes.

Preprocessing:

- Filled missing ratings.
- Consolidated metadata for each product (brand, attributes, etc.).
- Cleaned minor HTML artifacts in descriptions.

## 4. Methodology

### 4.1 Embeddings

**Model:** all-MiniLM-L6-v2 (SentenceTransformers). Used to convert product descriptions into dense vectors for semantic search.

### 4.2 Vector Store

**Tool:** ChromaDB (PersistentClient). Collection: Fashion\_Products\_Local. Persistence ensures reproducible and cost-efficient runs.

### 4.3 Caching

Disk-backed cache collection (Fashion\_Cache\_Local). Results stored as JSON within metadata to satisfy Chroma's primitive-only rule. Reduces latency for repeated queries.

#### 4.4 Retrieval

Cosine distance used to identify semantically closest items. Smaller distance = more relevant.

#### 4.5 Reranking

Model: cross-encoder/ms-marco-MiniLM-L-6-v2.

Strategies:

- **semantic\_first**: rerank only top-k retrieved items.
- **global\_rerank**: rerank all candidates and then pick top-k.

#### 4.6 Generation

OpenAI Chat model generates natural responses to user queries. Inputs: query + top-3 reranked items (documents + metadata). Outputs: concise, contextual answers citing product names/brands.

### 5. System Architecture

**Pipeline:**

Data Ingestion → Embeddings → ChromaDB Persistence → Semantic Search → Cache Check → Reranking → Generative Response.



- **Batching**: 100 items per batch for efficiency.
- Stable IDs: persisted as strings.
- Cache mechanism: JSON serialization for compact, reproducible lookups.

### 6. Implementation Highlights

Notebook: MrHelpMateFashionSearch.ipynb

- Section 3: embeddings + persistence
- Section 4: semantic search + cache
- Section 5: reranking
- Section 6: response generation
- Section 7: conclusion

Other notes:

- Verified persistence: `fashion_collection.count()` = 14214.

## 7. Results

Successfully indexed 14,214 items.

- Retrieval relevance improved with reranking (semantic\_first and global\_rerank).
- Cache mechanism reduced repeated query latency.
- Generative responses delivered contextual recommendations referencing actual product details.

Queries that has been used to test the model:

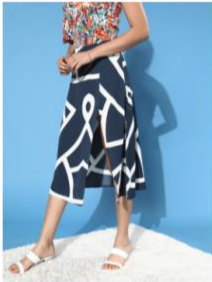
Query1: A silk purple summer dress, for beach party

Search Layer


Query:  
A silk purple summer dress, for beach party

Top 5 Search Layer


ID: 16541828




ID: 18875150




ID: 16450150



ID: 17647108



ID: 17645392



IDs	Documents	Distances	Metadatas	Source	Query
0 16541828	Amp up your style quotient with this chic skirt...	0.880282	('Product_type': 'Outerwear, Skirt', 'RatingCo...	main_collection	A silk purple summer dress, for beach party
1 18875150	<ub><li>Purple patchwork skirt & top co-ord...	0.889037	('Price_INR': 1590, 'Brand': 'URBANIC', 'Produ...	main_collection	A silk purple summer dress, for beach party
2 16450150	Sashay in style wearing this chic skirt. The a...	0.902100	('Name': 'Berrylush Women Pretty Pink Romantic...	main_collection	A silk purple summer dress, for beach party
3 17647108	Add a punch of vibrant hues to your spring and...	0.905445	('Brand': 'bebe', 'RatingCount': 70, 'Rating...	main_collection	A silk purple summer dress, for beach party
4 17645392	Show off your elegant style with a cool skirt ...	0.909557	('Price_INR': 1599, 'Description': 'Show off y...	main_collection	A silk purple summer dress, for beach party

Generation Layer

Based on the top 3 search results from the fashion dataset, here are some options for a silk purple summer dress suitable for a beach party:


1. \*\*Purple Patchwork Skirt & Top Co-ord Set by URBANIC\*\*:  
- Description: This outfit features a purple patchwork skirt and top set, perfect for a stylish beach party look.  
- Brand: URBANIC  
- Price: INR 1590

2. \*\*Elegant Skirt with Silk Purple Details\*\*:  
- Description: Show off your elegant style with this cool skirt that includes silk purple details, ideal for a summer beach party.  
- Price: INR 1599


3. \*\*Vibrant Hues Spring & Summer Dress by bebe\*\*:  
- Description: Add a punch of vibrant hues to your spring and summer wardrobe with this beautiful silk purple dress, great for beach parties.  
- Brand: bebe  
- Rating: 7.8/10 based on 7 ratings

For more detailed information, such as specific designs, sizes, and availability, it is recommended to visit our website page or directly explore these fashion items online.


ID: 18875150



ID: 17645392



ID: 17647108



Query2: An orange summer dress or kurta to wear over blue denim jeans

Search Layer

Query:  
An orange summer dress or kurta to wear over blue denim jeans

Top 5 Search Layer



	IDs	Documents	Distances	Metadatas	Source	Query
0	17889600	This 3-piece sea green woven design silk blend...	0.734046	{'Product_type': 'Kurta, Trousers, Dupatta', '...	main_collection	An orange summer dress or kurta to wear over b...
1	17890144	This 3-piece blue silk blend plus size suit se...	0.769319	{'Product_id': 17890144, 'Product_type': 'Kurt...	main_collection	An orange summer dress or kurta to wear over b...
2	10356859	This set includes: Kurta and Trousers  A...	0.781421	{'RatingCount': 12568.0, 'Description': 'This ...	main_collection	An orange summer dress or kurta to wear over b...
3	12130790	Mustard Yellow and red woven design kurta with...	0.785736	{'Description': 'Mustard Yellow and red woven ...	main_collection	An orange summer dress or kurta to wear over b...
4	11197978	Lime Green and Blue printed A-line kurta, has ...	0.791807	{'Brand': 'Anouk', 'Price_INR': 1599, 'Rating'...	main_collection	An orange summer dress or kurta to wear over b...

Generation Layer

Based on your query for an orange summer dress or kurta to wear over blue denim jeans, here are the top 3 relevant search results along with their key detail

1. **\*\*Lime Green and Blue Printed A-line Kurta\*\***
  - **\*\*Brand\*\*:** Anouk
  - **\*\*Description\*\*:** This kurta features a lime green and blue printed A-line design.
  - **\*\*Price\*\*:** INR 1599
  - **\*\*Rating\*\*:** Not specified
  - **\*\*Get more details about this item on our website page\*\*.**
2. **\*\*3-Piece Blue Silk Blend Plus Size Suit Set\*\***
  - **\*\*Product Type\*\*:** Kurti
  - **\*\*Description\*\*:** This suit set includes a blue silk blend kurta in plus size.
  - **\*\*Price\*\*:** Not specified
  - **\*\*Rating\*\*:** Not specified
  - **\*\*For additional information, please visit our website page\*\*.**
3. **\*\*Kurta and Trousers Set\*\***
  - **\*\*Rating Count\*\*:** 12568
  - **\*\*Description\*\*:** This set includes a stylish kurta and trousers.
  - **\*\*Price\*\*:** Not specified
  - **\*\*Rating\*\*:** Not specified
  - **\*\*Explore this set further on our website page\*\*.**

Based on the search results, you can consider the lime green and blue printed A-line kurta from Anouk as a vibrant option to wear over blue denim jeans for i




Query3: I'm searching for a versatile black leather jacket, suitable for various occasions and effortlessly complementing any outfit. Preferably in size XL

Search Layer


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

Top 5 Search Layer


ID: 18153246




ID: 13690162




ID: 16220614



ID: 13168738



ID: 14455424



IDs	Documents	Distances	Metadatas	Source	Query
0 18153246	Black solid leather jacket with zip detail, ...	0.794574	{'Price_INR': 6998, 'RatingCount': 0.0, 'Descr...	main_collection	I'm searching for a versatile black leather ja...
1 13690162	Black solid jacket, has a hooded, zip closure,...	0.825629	{'Rating': 4.146341463, 'Name': 'U.S. Polo Ass...	main_collection	I'm searching for a versatile black leather ja...
2 16220614	Black self design self design tailored jacket ...	0.832273	{'Colour': 'Black', 'Product_id': 16220614, 'N...	main_collection	I'm searching for a versatile black leather ja...
3 13168738	Black solid jacket, has a spread collar, 6 poc...	0.852538	{'Rating': 0.0, 'Description': 'Black solid ja...	main_collection	I'm searching for a versatile black leather ja...
4 14455424	Black graphic printed lightweight leather jack...	0.852729	{'Name': 'Leather Retail Women Black Graphic P...	main_collection	I'm searching for a versatile black leather ja...

Generation Layer

Based on your query for a versatile black leather jacket in size XL, here are the top 3 relevant results from our fashion dataset:


1. \*\*Black Solid Jacket\*\*  
- Description: Black solid jacket with a spread collar, 6 pockets, and a zipper closure.  
- Brand: Not specified  
- Metadata: Rating ~ 0.0, Product ID ~ 13168738  
- Price: Not available  
- ![Image of Black Solid Jacket](image\_link\_here)

2. \*\*Black Solid Leather Jacket with Zip Detail\*\*  
- Description: Black solid leather jacket with zip detail for a stylish look.  
- Brand: Not specified  
- Metadata: Price ~ INR 6998, Rating Count ~ 0.0, Product ID ~ 18153246  
- ![Image of Black Solid Leather Jacket with Zip Detail](image\_link\_here)


3. \*\*Black Self Design Tailored Jacket\*\*  
- Description: Black self design tailored jacket designed for a modern look.  
- Brand: Not specified  
- Metadata: Colour ~ Black, Product ID ~ 16220614  
- Price: Not available  
- ![Image of Black Self Design Tailored Jacket](image\_link\_here)

Based on the mentioned details, you have a variety of options to choose from for a versatile black leather jacket that can complement different outfits. You can explore these j


ID: 13168738



ID: 18153246



ID: 16220614



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

## 8. Challenges:

During implementation, the following challenges were encountered:

- **Metadata Processing:** Initial difficulties were encountered in processing metadata, which adversely affected the accuracy of result displays. These challenges were addressed through systematic debugging and refinement of the metadata handling logic.
- **Dataset Chunking:** The loading of the complete dataset without chunking led to memory-related constraints. Although chunking was identified as a suitable mitigation strategy, it was not implemented due to time limitations and the manageable size of the dataset.

## 9. Lessons Learned:

- Proper data preprocessing is essential for maintaining data quality and readability.
- Leveraging advanced AI models can greatly improve the system's ability to generate detailed and contextually relevant responses.
- Efficient strategies for handling large datasets are necessary to mitigate memory constraints and ensure smooth processing.

## 10. Future Work

- Add offline evaluation metrics (Recall@k, nDCG, golden sets).
- Improve cache system with TTL, query normalization, deduplication.
- Add structured filters (brand, color, price).
- Clean and standardize metadata, strip HTML, add faceted search.
- Deploy as APIs (e.g., FastAPI) with environment-driven configuration.
- Experiment with stronger embedding models (e5-large-v2) and ANN optimizations (IVF/PQ).