

SkinWise - Your New AI Skincare Agent

AML-GGYK

12/18/2025

Meet the AML-GGYK Team



Gabrielle Park



Gabriella Chu



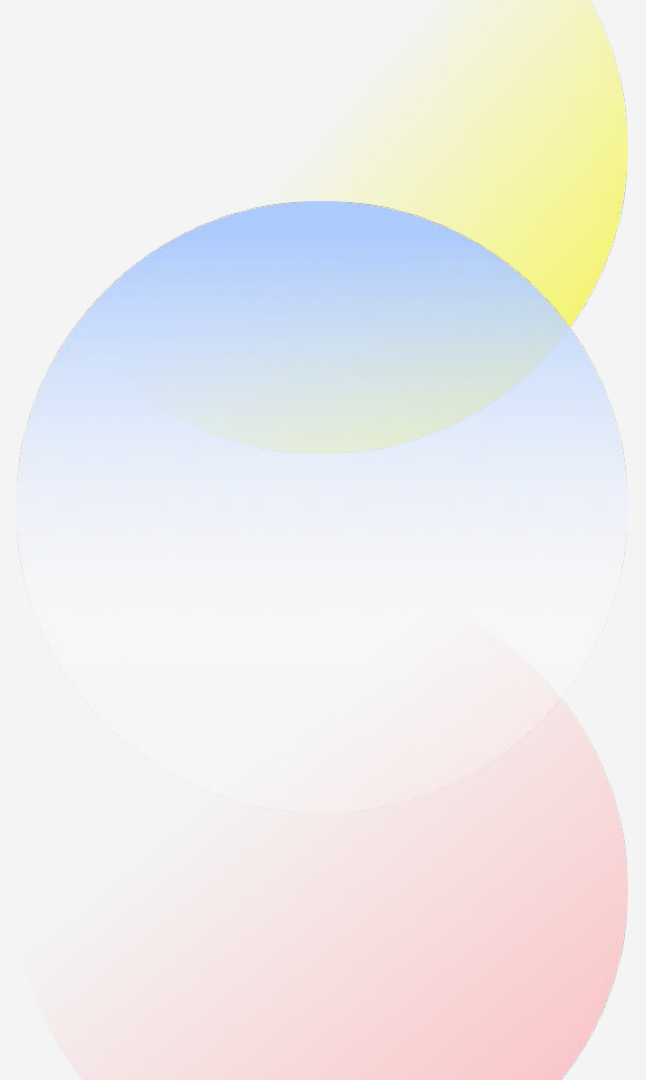
Yelene Cisse



Kshamaa Suresh

Introduction

Overview, Goals, and Motivation



Summary of content

1

Data collection & preparation

2

Sentiment Analysis

3

LLM Agent Architecture

4

Demo

Data collection and preparation. Medical Data and Sephora Products

Medical Data

- Hugging Face Dermatology Data
 - Dermatology skin condition info in Q&A format
 - 1,500+ rows, 2 columns
 - Feature engineering:
 - Answer categorization (Cause, Treatment, Symptom)
 - Condition categorization (e.g. Acne, Eczema)
 - Example:
 - “What are the treatment modalities for acne? Mild acne can often be managed with over-the-counter products containing ingredients like benzoyl peroxide or salicylic acid.”Renude Skin care product ingredients
- Kaggle dataset scraped from Renude’s “Ingredients in Skincare” database
 - 248 rows, 8 columns describing skincare ingredients and their purpose
 - Feature used: ingredient name, description, functions, target population

Sephora Data

- The data was sourced from [Kaggle](#) and was collected in March 2023.
 - 8K+ rows of products from the Sephora store, used about 2.5k
 - 27 columns
 - Product name, brand, price, highlights, etc.
 - Feature engineering: created 32 new binary features using highlights column
 - ex. “sensitive_skin”
- Embeddings
 - “product_text”: clean text description of each product
 - Hugging Face SentenceTransformer

Sentiment Analysis. VADER and TF-IDF+Logistic Regression Analyzers

Sentiment Analysis

Reviews data

Raw dataset contained 1M reviews, with rating, review text, and review title as fields of focus. After preprocessing and deduplication, the cleaned dataset comprised 581,785 unique reviews

Category	Rating	VADER
Positive	>= 4 Stars	>0.05
Neutral	3 Stars	[-0.05, 0.05]
Negative	<= 2 Stars	<-0.05

* **TF-IDF+LR** - Category assigned based on max probability

- Target: Sentiment based on 5 star rating field
- 3 types of models: reviews text, reviews title, combined text
- 2 Methods explored:
 - Valence Aware Dictionary and sEntiment Reasoner (VADER)
 - Term Frequency-Inverse Document Frequency (TF-IDF) + Logistic Regression (LR)
- TF-IDF + LR model outperformed VADER by 2%, reaching 85.5% accuracy, with major improvements on minority class (negative and neutral)
- Quality Review Score: 0.4 Length + 0.3 Confidence + 0.3 Accuracy to estimate quality for top review sampling
- Ulta Reviews scored with model, demonstrating adaptability to various data sources (Reddit, Blogs, Youtube reviews..)

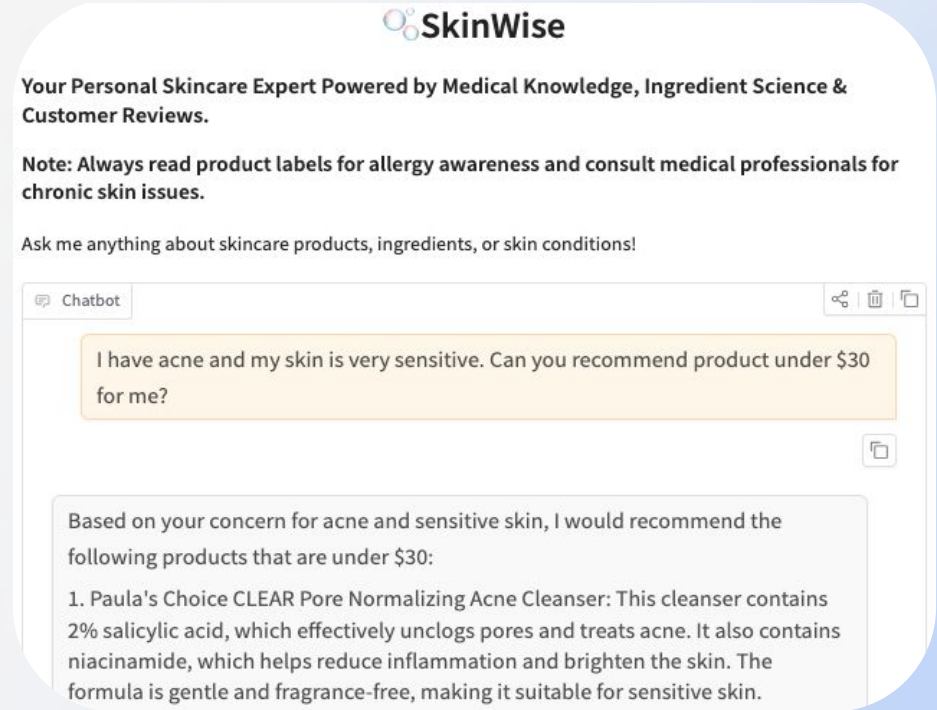
LLM Agent Architecture. Mistral 7B Powered Agent

LLM Agent Architecture

- Embeddings and Vectorstore creation
- LLM from Mistral
- Skincare Agent Class Definition
- Gradio UI implementation

Steps involved:

- Setting up the environment.
- Installing packages (langchain).
- Downloading the kaggle dataset.
- Loading the medical data, ingredients data and sentiment analysis.
- Process product data and do feature extraction.
- Creating knowledge documents.
- Create the vector store and load the model.
- Setup the LLM agent.



SkinWise

Your Personal Skincare Expert Powered by Medical Knowledge, Ingredient Science & Customer Reviews.

Note: Always read product labels for allergy awareness and consult medical professionals for chronic skin issues.

Ask me anything about skincare products, ingredients, or skin conditions!

Chatbot

I have acne and my skin is very sensitive. Can you recommend product under \$30 for me?

Based on your concern for acne and sensitive skin, I would recommend the following products that are under \$30:

1. Paula's Choice CLEAR Pore Normalizing Acne Cleanser: This cleanser contains 2% salicylic acid, which effectively unclogs pores and treats acne. It also contains niacinamide, which helps reduce inflammation and brighten the skin. The formula is gentle and fragrance-free, making it suitable for sensitive skin.

AML-GGYK

12/18/2025

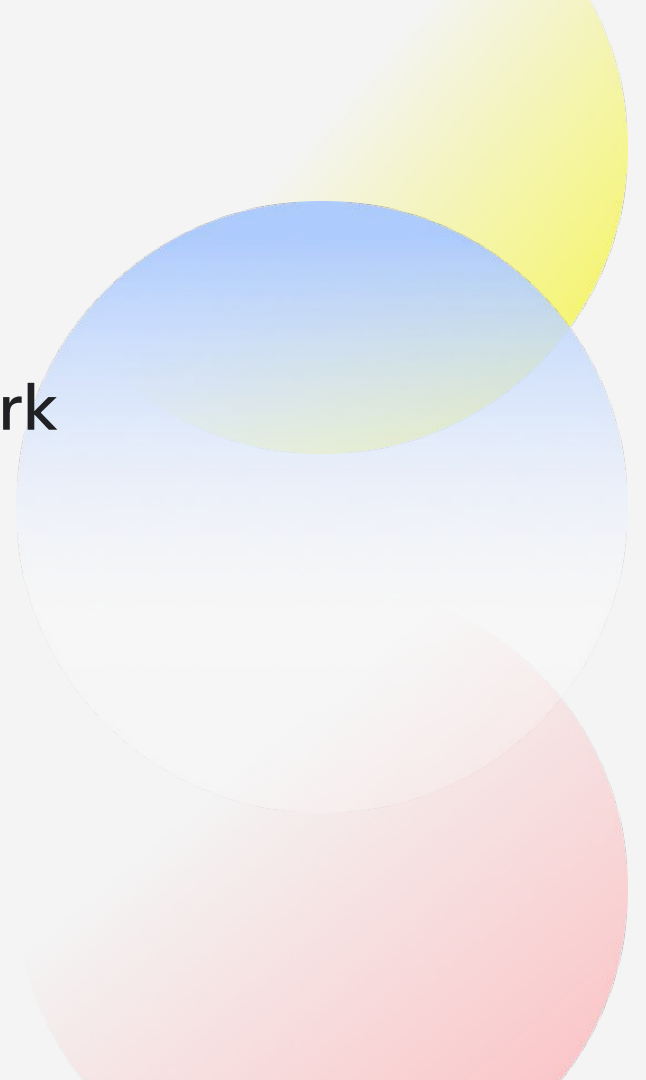
Demo

Hugging Face link for UI: <https://huggingface.co/spaces/kshamaasuresh/skincare-agent>

Github link: <https://github.com/AML-GGYK/DermaLLM>

Conclusion

Wrap-up, Limitations, and Future Work



Thank you

