**Báo cáo Nhóm 6**

**Đề tài chatbot thông tin bóng đá**

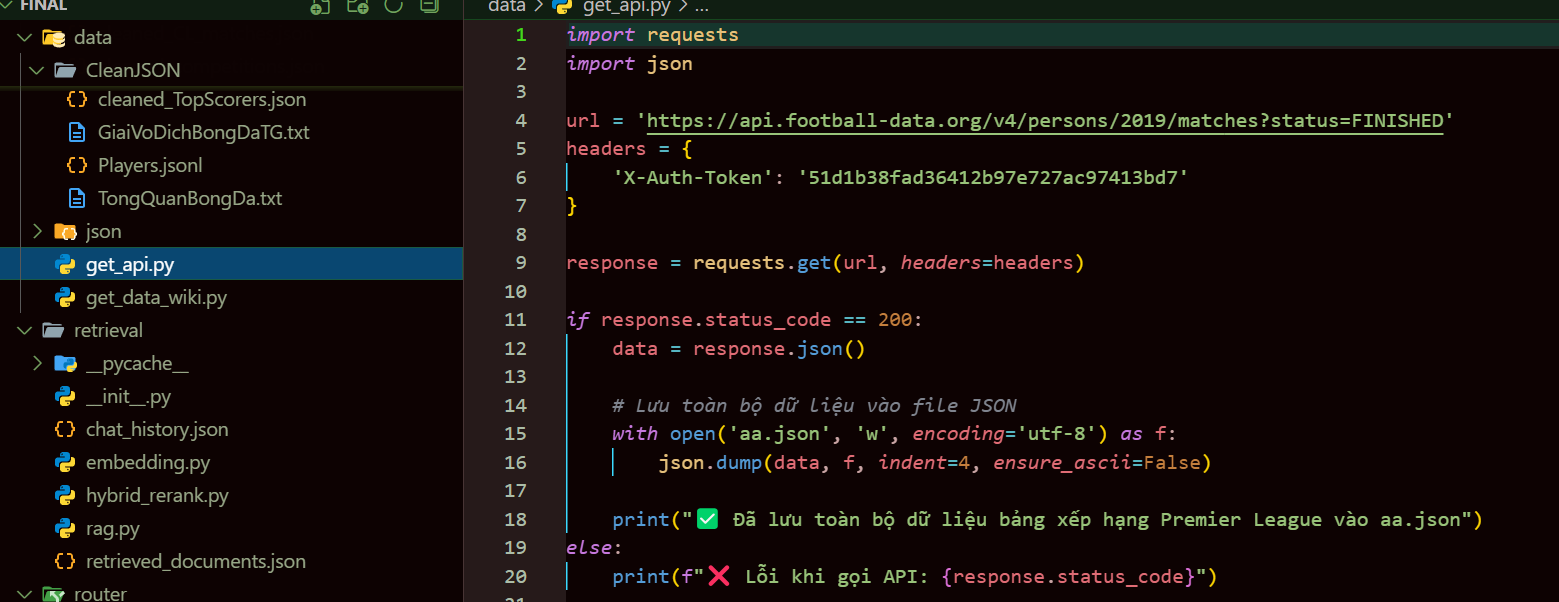
|  |  |  |  |
| --- | --- | --- | --- |
| **MSSV** | **Họ và Tên** | **Nhiệm Vụ** | **Đánh Giá** |
| 21139231 | Võ Hoàng Nhã Quyên | Thu thập và chuẩn bị dữ liệu, build kho vector | 120% |
| 21120491 | Nguyễn Võ Tú Uyên | Tìm kiếm nâng cao triển khai Hydrid search + reranking, build kho vector | 120% |
| 21117241 | Vũ Phạm Anh Thư | Backend sử dụng RAG | 100% |
| 21128631 | Phan Nguyễn Văn Phúc | Frontend có giao diện, hội thoại liên tục, có router điều hướng hội thoại | 100% |
| 21116261 | Lê Đức Nguyên | Build docker | 60% |

***github****:* [link\_chatbot\_github](https://github.com/cherchezvous-uyen/soccer_chatbot.git)

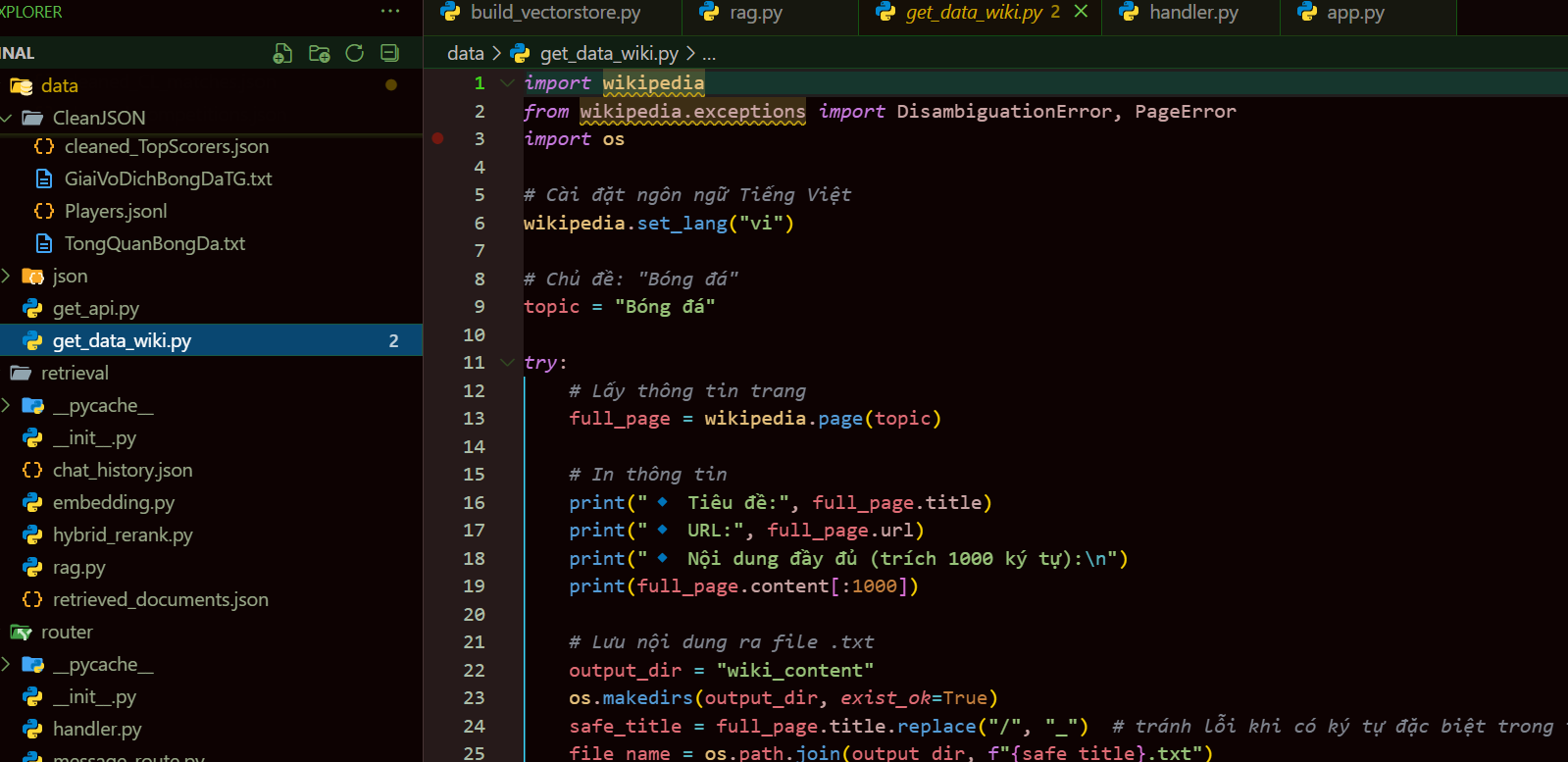
1. **Thu thập thông tin và xử lí dữ liệu (scrapy, api, txt, json, chunksize)**

**Lấy dữ liệu bằng api và wiki**

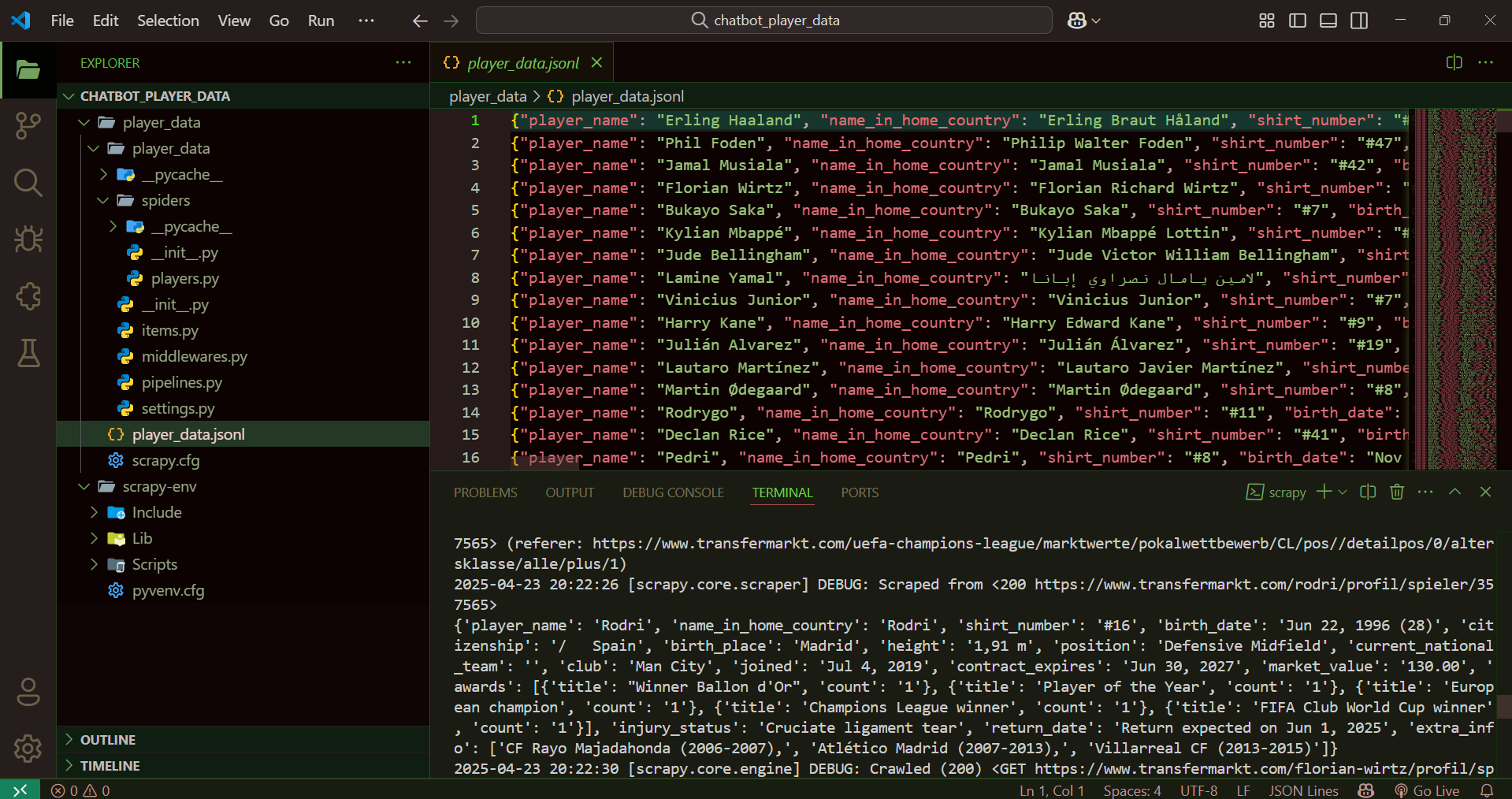
<https://www.football-data.org/>

****

https://vi.wikipedia.org/wiki/B%C3%B3ng\_%C4%91%C3%A1

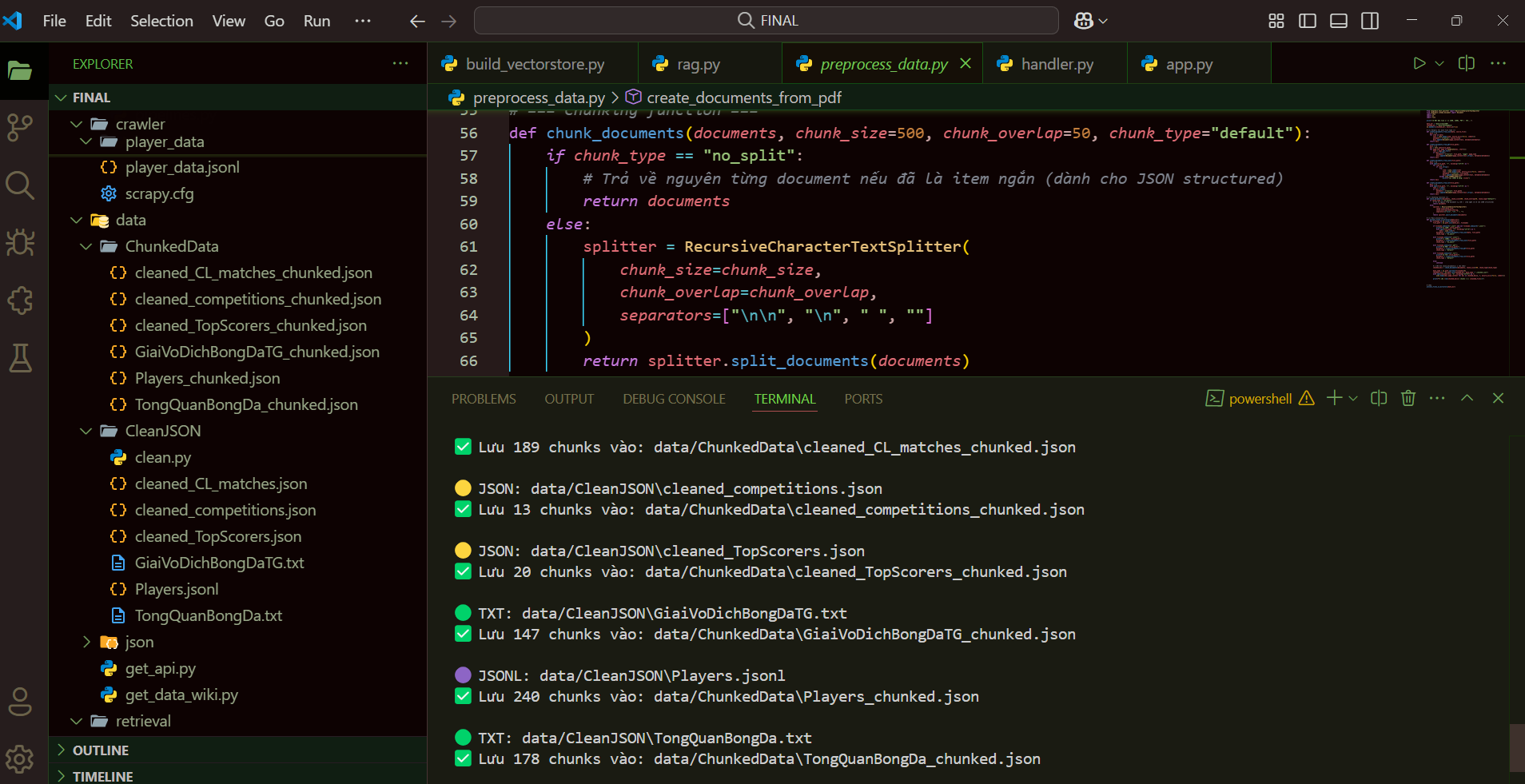
****

* **Crawl thông tin cầu thủ từ transfermrk.com**
* allowed\_domains = ["transfermarkt.com"]
* start\_urls = ["https://www.transfermarkt.com/uefa-champions-league/marktwerte/pokalwettbewerb/CL/pos//detailpos/0/altersklasse/alle/plus/1"]
* các thông tin gồm name, name in home town, shirt number, market value, birthday, info, citizenship, birthplace, club, awards, stats,...
* Lưu vào player\_data.jonl

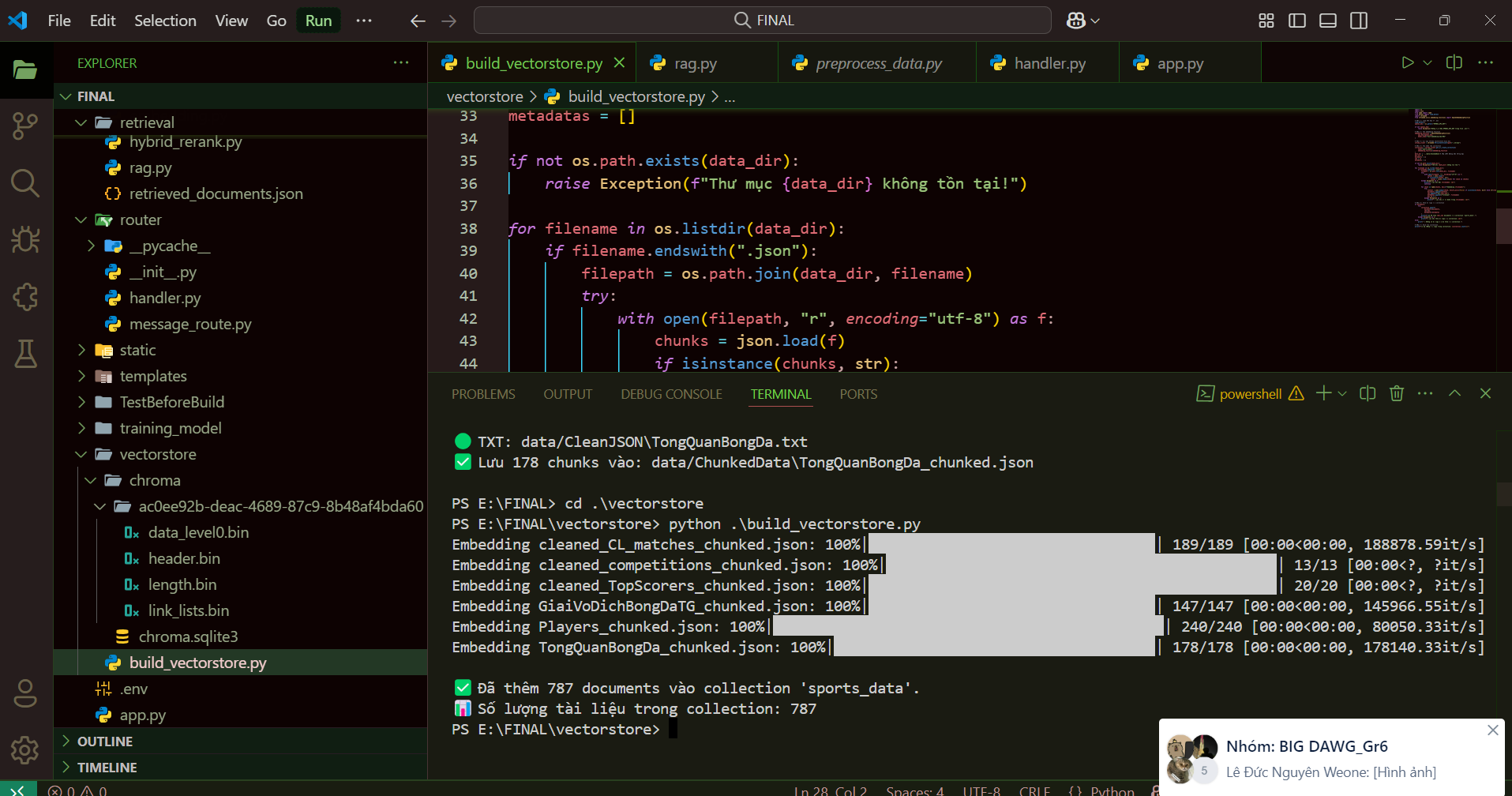


**Xử lí chunksize dữ liệu**

Dữ liệu từ cleanJSON đổ vào ChunkedDataa

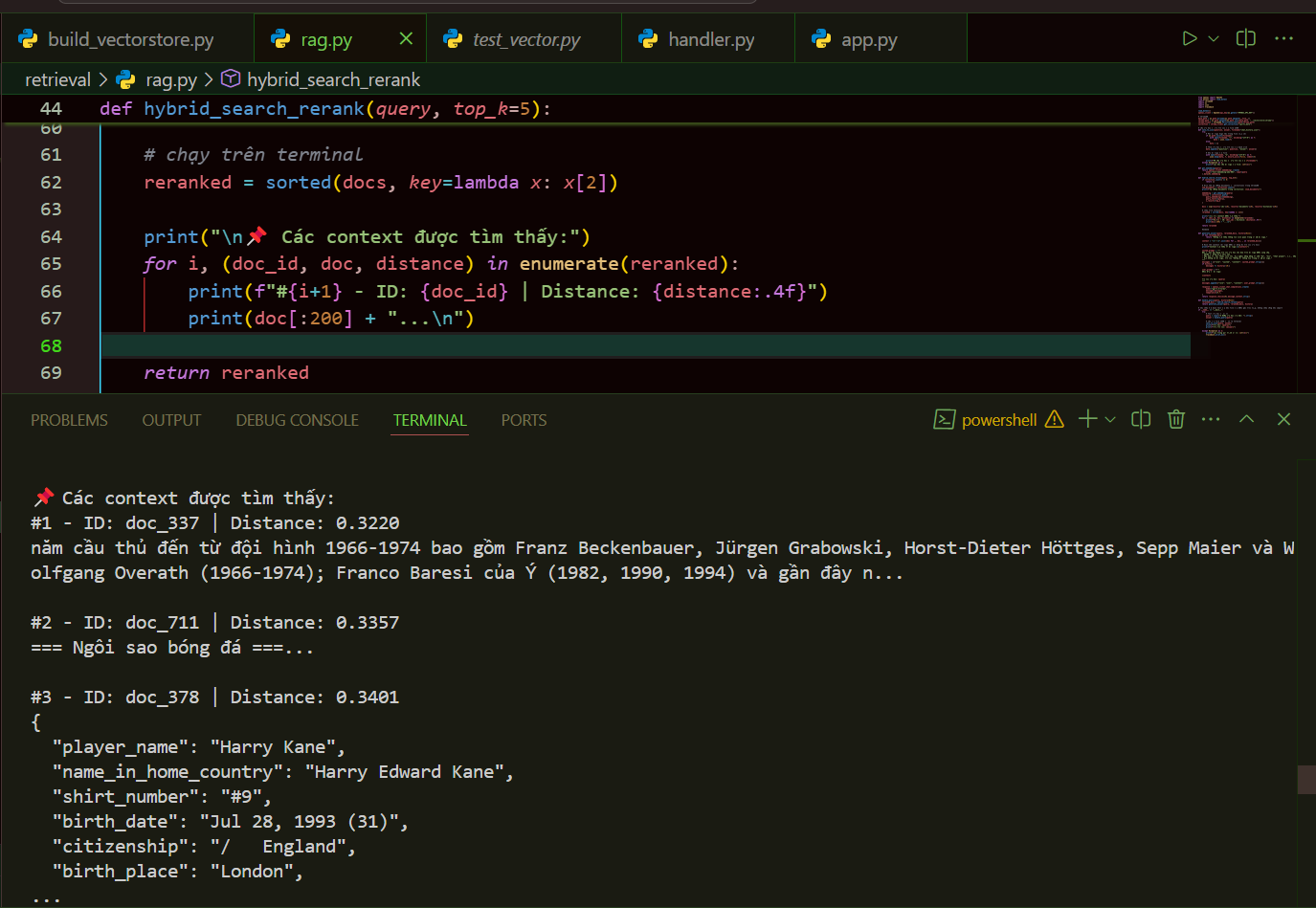
****

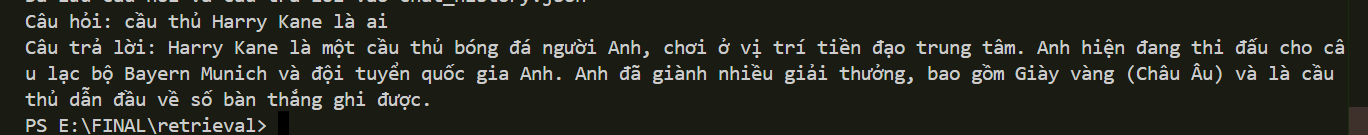
1. **Xây dựng kho vector**

****

1. **Triển khai Hybrid search và Reranking**

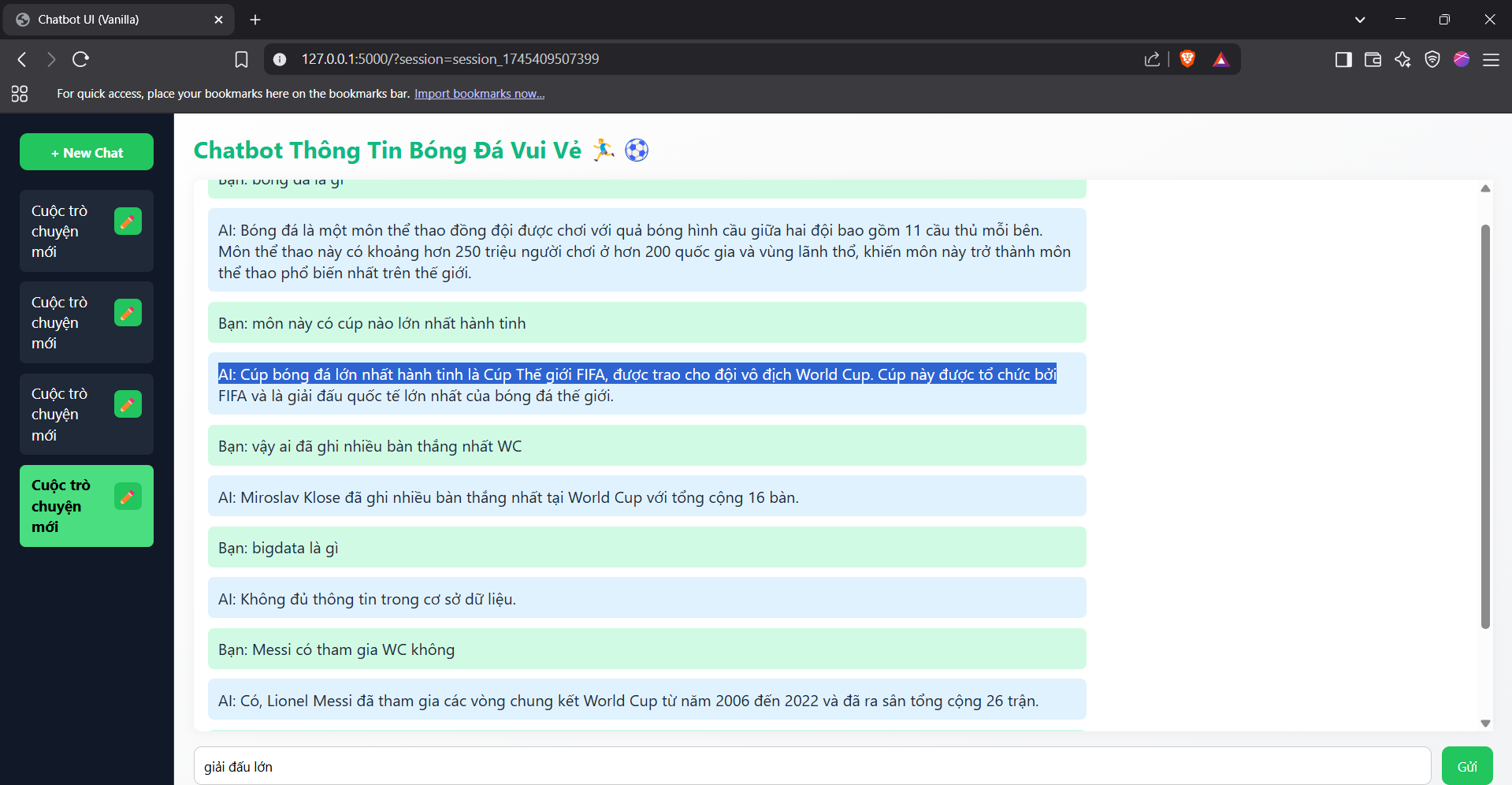
Khi hỏi câu hỏi bất kì thì hàm hybrid search rerank sẽ tìm và so khoảng cách giữa các doc từ đó, docs nào có kc gần 1 nhất nó sẽ lấy ra và trả lời cho mình

****

****

1. **Frontend:**

* **Giao diện hội thoại liên tục và kết quả hội thoại**
* Nếu có dữ liệu trong chroma thì mới được trả lời
* Không có thông tin thì không được lấy nền đã học từ mô hình mà trả ra câu trả lời là không có thông tin



1. **Xây dựng Docker**

**Tạo Dockerfile**

# Sử dụng Python 3.11 làm base image

FROM python:3.11

# Thiết lập thư mục làm việc trong container

WORKDIR /app

# Sao chép tệp requirements.txt vào container

COPY requirements.txt /app/

# Cài đặt các thư viện từ requirements.txt

RUN pip install --no-cache-dir -r requirements.txt

# Cài đặt Gunicorn

RUN pip install gunicorn

# Sao chép mã nguồn của bạn vào container

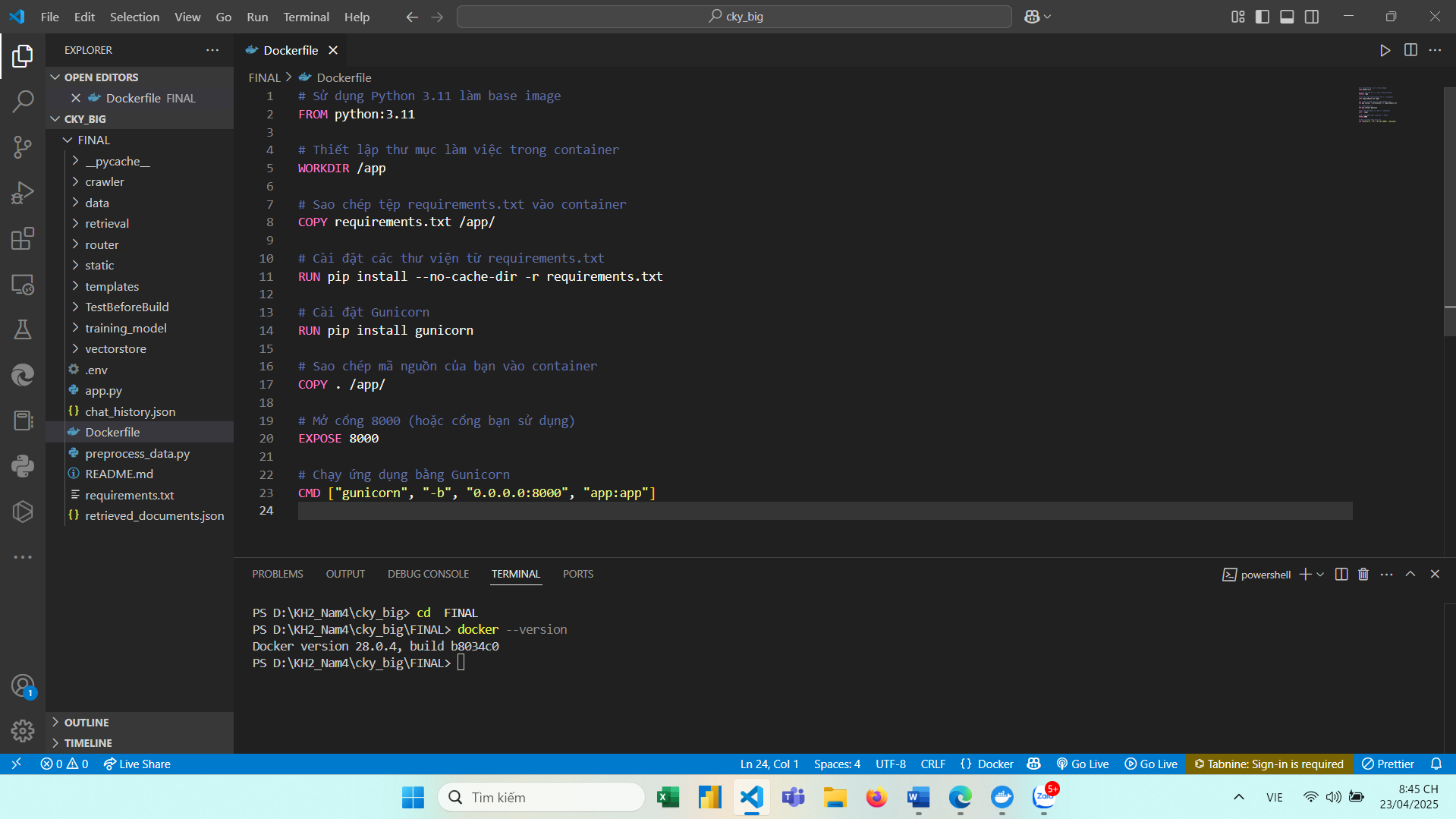
COPY . /app/

# Mở cổng 8000 (hoặc cổng bạn sử dụng)

EXPOSE 8000

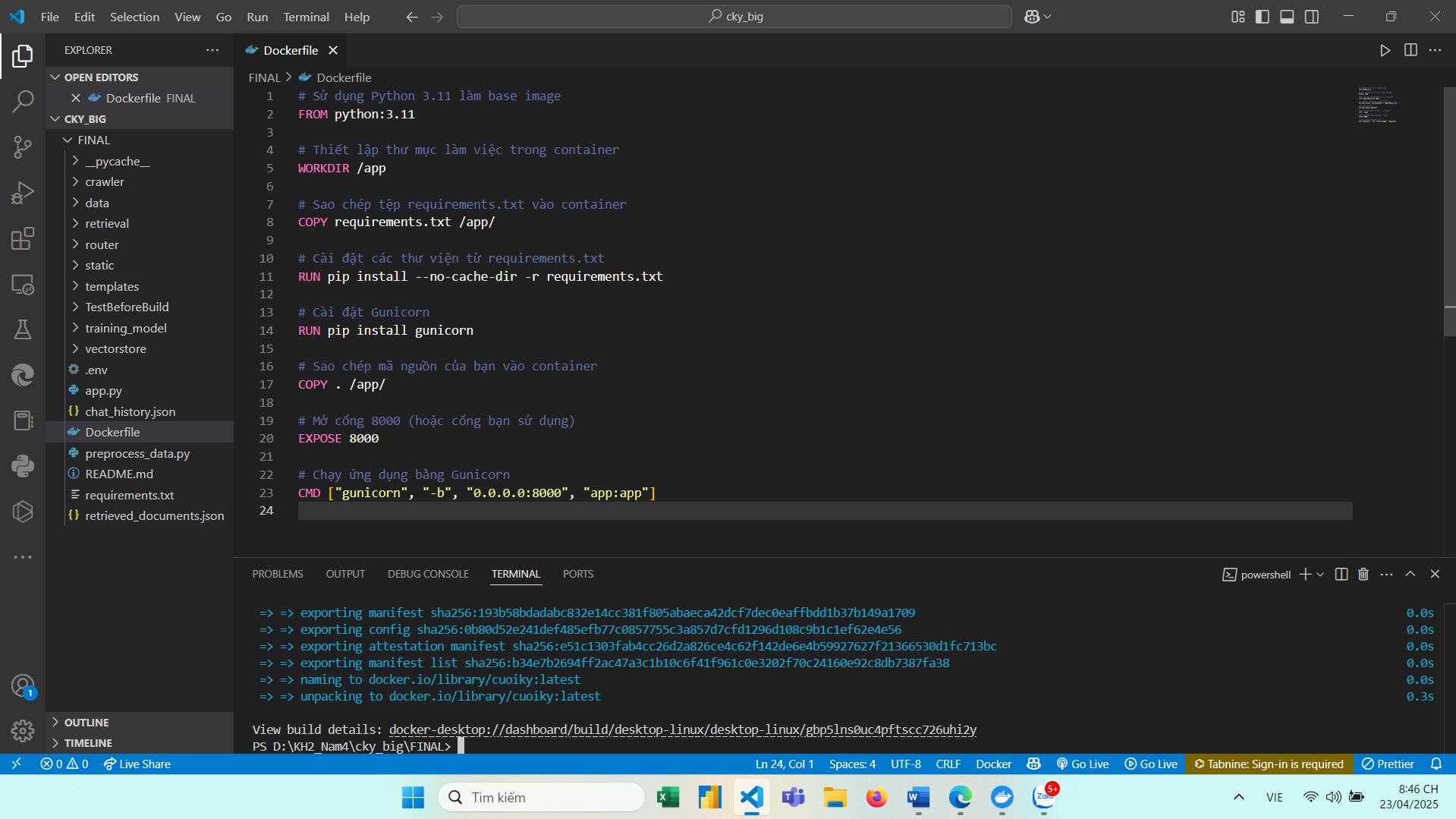
# Chạy ứng dụng bằng Gunicorn

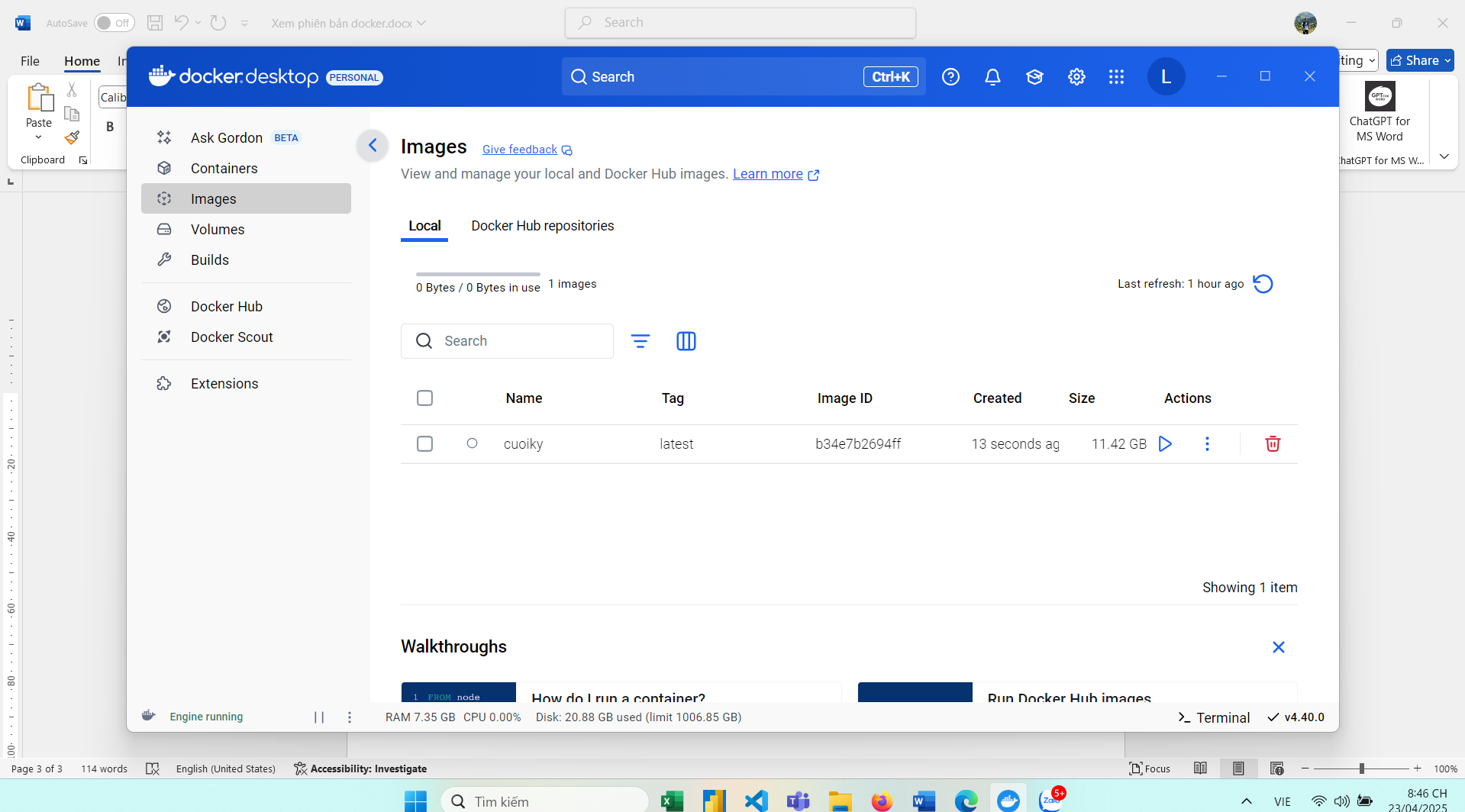
CMD ["gunicorn", "-b", "0.0.0.0:8000", "app:app"]



**Xây dựng Docker Image**

docker build -t cuoiky .





**Chạy Docker Container**

docker run -p 8000:8000 cuoiky

