**TRƯỜNG ĐẠI HỌC TÀI NGUYÊN VÀ MÔI TRƯỜNG TP.HCM**

**KHOA: HỆ THỐNG THÔNG TIN VÀ VIỄN THÁM**

-----🙞🙜🕮🙞🙜-----



**ĐỒ ÁN TỐT NGHIỆP**

**CHUYÊN NGÀNH: TIN HỌC MÔI TRƯỜNG**

**XÂY DỰNG ỨNG DỤNG PHÂN TÍCH VÀ TRỰC QUAN HÓA DỮ LIỆU VỀ VIỆC SỬ DỤNG NĂNG LƯỢNG TOÀN CẦU**

Giảng viên hướng dẫn : **TS. Dương Thị Thúy Nga**

Sinh viên thực hiện : **Nguyễn Đỗ Nhất Vũ**

Mã số sinh viên : **0850080055**

Lớp : **08\_ĐH\_THMT**

Khóa : **2019-2023**

**TP. Hồ Chí Minh, tháng 5 năm 2024**

**TRƯỜNG ĐẠI HỌC TÀI NGUYÊN VÀ MÔI TRƯỜNG TP.HCM**

**KHOA: HỆ THỐNG THÔNG TIN VÀ VIỄN THÁM**



**ĐỒ ÁN TỐT NGHIỆP**

**CHUYÊN NGÀNH: TIN HỌC MÔI TRƯỜNG**

**XÂY DỰNG ỨNG DỤNG PHÂN TÍCH VÀ TRỰC QUAN HÓA DỮ LIỆU VỀ VIỆC SỬ DỤNG NĂNG LƯỢNG TOÀN CẦU**

Giảng viên hướng dẫn : **TS. Dương Thị Thúy Nga**

Sinh viên thực hiện : **Nguyễn Đỗ Nhất Vũ**

Mã số sinh viên : **0850080055**

Lớp : **08\_ĐH\_THMT**

Khóa : **2019-2023**

**TP. Hồ Chí Minh, tháng 5 năm 2024**

**MỞ ĐẦU**

Trong bối cảnh thế giới đang đối mặt với những thách thức cấp bách về biến đổi khí hậu và khủng hoảng năng lượng, việc đảm bảo an ninh năng lượng và phát triển bền vững trở thành ưu tiên hàng đầu của mọi quốc gia. Nhu cầu năng lượng ngày càng tăng cao trong khi nguồn cung năng lượng hóa thạch đang dần cạn kiệt, đồng thời gây ra những tác động tiêu cực đến môi trường. Do đó, việc phân tích, dự báo và tối ưu hóa việc sử dụng năng lượng là vô cùng cần thiết.

Nhận thức được tầm quan trọng của vấn đề này, đồ án “ Xây dựng ứng dụng phân tích và trực quan hóa dữ liệu về năng lượng toàn cầu” được thực hiện nhằm cung cấp một công cụ hữu ích cho việc đánh giá tình hình và dự báo xu hướng sử dụng năng lượng toàn cầu. Ứng dụng tập trung vào việc phân tích dữ liệu lịch sử và dự báo mức tiêu thụ năng lượng trong tương lai cho các quốc gia khác nhau, từ đó hỗ trợ các nhà hoạch định chính sách, nhà đầu tư, doanh nghiệp và người tiêu dùng trong việc đưa ra quyết định sáng suốt liên quan đến năng lượng.

Cụ thể, đồ án sử dụng bộ dữ liệu từ Cơ quan Năng lượng Quốc tế (IEA), một nguồn dữ liệu uy tín và toàn diện về năng lượng toàn cầu, bao gồm các chỉ số quan trọng về sản xuất, tiêu thụ, nhập khẩu, xuất khẩu năng lượng theo từng quốc gia và lĩnh vực. Bằng cách áp dụng các kỹ thuật phân tích dữ liệu và mô hình học máy, đặc biệt là mô hình hồi quy tuyến tính, đồ án dự báo xu hướng sử dụng năng lượng trong 5 năm tới cho các lĩnh vực chính: dân dụng (Residential), dịch vụ (Services), công nghiệp (Industry) và giao thông vận tải (Transport).

Kết quả của đồ án không chỉ cung cấp bức tranh toàn cảnh về tình hình năng lượng hiện tại mà còn đưa ra những dự báo về nhu cầu năng lượng trong tương lai, từ đó tạo cơ sở cho việc xây dựng các chính sách năng lượng hiệu quả, tối ưu hóa việc sử dụng năng lượng, giảm thiểu sự phụ thuộc vào năng lượng hóa thạch, và hướng đến một tương lai bền vững.

Và đây là bài báo đồ án tốt nghiệp của em. Rất mong được Qúy thầy cô đánh giá, góp ý để em có thể rút kinh nghiệm và hoàn thiện hơn. Em xin chân thành cảm ơn!

**LỜI CẢM ƠN**

Để có thể hoàn thành đồ án tốt nghiệp “Xây dựng ứng dụng phân tích và trực quan hóa dữ liệu về việc sử dụng năng lượng toàn cầu” thành công, em xin gửi lời cảm ơn sâu sắc đến Ban Giám hiệu Trường Đại học Tài nguyên và Môi trường TP.HCM, cùng Quý thầy cô khoa Hệ thống Thông tin và Viễn thám đã tạo điều kiện thuận lợi, cung cấp môi trường học tập và nghiên cứu chuyên nghiệp, trang bị cho em những kiến thức và kỹ năng cần thiết để hoàn thành đồ án này.

Em đặc biệt biết ơn và xin gửi lời tri ân sâu sắc đến cô TS. Dương Thị Thúy Nga, giảng viên hướng dẫn, người đã tận tâm dìu dắt, truyền đạt những kiến thức chuyên môn, kinh nghiệm thực tiễn quý báu, giúp em định hướng và hoàn thiện đồ án một cách tốt nhất. Sự nhiệt tình, tận tâm và sự hỗ trợ hết mình của thầy trong suốt quá trình nghiên cứu và xây dựng hệ thống đã giúp em phát triển khả năng tư duy, nâng cao trình độ chuyên môn và thêm yêu thích lĩnh vực nghiên cứu này.

Em cũng xin gửi lời cảm ơn đến gia đình, bạn bè đã luôn động viên và ủng hộ em trong suốt quá trình học tập và thực hiện đồ án.

Trong quá trình thực hiện đồ án, chắc hẳn em không tránh khỏi những thiếu sót. Em rất mong nhận được sự cảm thông và những góp ý quý báu từ thầy cô và các bạn để em có thể hoàn thiện bản thân hơn nữa.

Em xin trân trọng cảm ơn!

**LỜI CAM ĐOAN**

Em xin cam đoan rằng đồ án tốt nghiệp với đề tài “Xây dựng ứng dụng phân tích và trực quan hóa dữ liệu về việc sử dụng năng lượng toàn cầu " là một công trình nghiên cứu độc lập, với sự hướng dẫn tận tình của giảng viên hướng dẫn là TS. Dương Thị Thúy Nga. Tất cả nội dung và kết quả nghiên cứu trong đồ án này là thành quả của công việc của riêng em, và em cam đoan rằng chúng không bị ảnh hưởng bởi bất kỳ nguồn tài liệu nào khác. Em đã tuân thủ các nguyên tắc đạo đức và pháp luật liên quan đến việc sử dụng tài liệu tham khảo. Tất cả tài liệu tham khảo được sử dụng trong đồ án đã được trích dẫn và nêu rõ nguồn gốc trong mục "Tài liệu tham khảo" của báo cáo. Em cam đoan rằng việc sử dụng tài liệu tham khảo này là hợp pháp và tuân thủ quy định về bản quyền. Những kết quả nghiên cứu được trình bày trong đồ án này là trung thực và không chứa yếu tố sao chép, đạo nhái, hoặc vi phạm quyền sở hữu trí tuệ. Em đã thực hiện quy trình nghiên cứu và thu thập dữ liệu một cách trung thực và khách quan.

Em hiểu rằng nếu những lời cam đoan trên của em không chính xác, em sẽ chịu hoàn toàn trách nhiệm và sẵn sàng chấp nhận mọi hình thức kỷ luật từ khoa và nhà trường.

Sinh viên thực hiện

**NGUYỄN ĐỖ NHẤT VŨ**

**NHẬN XÉT**

**(Giảng viên hướng dẫn)**

………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………

**Kết luận: (Đồng ý hoặc Không đồng ý cho sinh viên nộp báo cáo)**

……………………………………………………………………………………………………………………………………………………………………………………

TP. Hồ Chí Minh, ngày….tháng….năm……

**GIẢNG VIÊN HƯỚNG DẪN**

*(ký tên)*

**TS. Dương Thị Thúy Nga**

**NHẬN XÉT**

**(Giảng viên phản biện)**

………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………

**Kết luận: (Đồng ý hoặc Không đồng ý cho sinh viên nộp báo cáo)**

……………………………………………………………………………………………………………………………………………………………………………………

TP. Hồ Chí Minh, ngày….tháng….năm……

**GIẢNG VIÊN PHẢN BIỆN**

*(ký tên)*

**LỊCH LÀM VIỆC**

|  |  |  |
| --- | --- | --- |
| **Thời gian** | **Nội dung** | **Địa điểm thực hiện** |
| **Tuần 1** | * Lên kế hoạch thực hiện ĐATN. * Tìm kiếm và thu thập dữ liệu. | Trường ĐH Tài Nguyên Và Môi Trường TP Hồ Chí Minh. |
| **Tuần 2** | * Phân tích dữ liệu. * Làm sạch dữ liệu. * Chuẩn hóa dữ liệu. | Trường ĐH Tài Nguyên Và Môi Trường TP Hồ Chí Minh. |
| **Tuần 3** | * Xác định đối tượng và phạm vi. * Tìm hiểu sơ bộ về vấn đề và ngành năng lượng. * Thực hiện đề cương ĐATN. | Trường ĐH Tài Nguyên Và Môi Trường TP Hồ Chí Minh. |
| **Tuần 4** | * Hoàn thiện đề cương ĐATN. * Chọn dữ liệu bao gồm các cột hoặc các biến số có giá trị đối với đề tài hoặc quá trình thực hiện đề tài. * Phân loại dữ liệu. | Trường ĐH Tài Nguyên Và Môi Trường TP Hồ Chí Minh. |
| **Tuần 5** | * Lên ý tưởng và thiết kế giao diện. * Xây dựng giao diện. * Xử lý và chuyển đổi dữ liệu. * Bắt đầu tiến hành trực quan hóa các dữ liệu có ý nghĩa và hiển thị chúng lên giao diện. * Chọn các loại biểu đồ phù hợp. * Sắp xếp thứ tự biểu đồ đã trực quan có tính thứ tự và khoa học. * Thêm các chú thích và diễn giải rõ ràng. | Trường ĐH Tài Nguyên Và Môi Trường TP Hồ Chí Minh. |
| **Tuần 6** | * Tìm hiểu sâu hơn về ngành năng lượng. * Tiếp tục thực hiện các yêu cầu từ tuần 5. | Trường ĐH Tài Nguyên Và Môi Trường TP Hồ Chí Minh. |
| **Tuần 7** | * Hoàn thiện các yêu cầu từ tuần 5.. * Thử nghiệm mô hình học máy ARIMA. * Huấn luyện mô hình cho mỗi quốc gia, mỗi lĩnh vực và mỗi dòng dữ liệu. | Trường ĐH Tài Nguyên Và Môi Trường TP Hồ Chí Minh. |
| **Tuần 8** | * Thử nghiệm mô hình học máy Prophet. * Huấn luyện mô hình cho mỗi quốc gia, mỗi lĩnh vực và mỗi dòng dữ liệu. * Kết quả RMSE vào khoảng 18 - 25% mức độ dao động của dữ liệu. | Trường ĐH Tài Nguyên Và Môi Trường TP Hồ Chí Minh. |
| **Tuần 9** | * Thử nghiệm mô hình học máy Moving Average. * Huấn luyện mô hình cho mỗi quốc gia, mỗi lĩnh vực và mỗi dòng dữ liệu. * Kết quả RMSE vào khoảng 15 – 20% mức độ dao động của dữ liệu. | Trường ĐH Tài Nguyên Và Môi Trường TP Hồ Chí Minh. |
| **Tuần 10** | * Thử nghiệm mô hình học máy Exponential Smoothing. * Tìm hiểu mô hình Linear Regression | Trường ĐH Tài Nguyên Và Môi Trường TP Hồ Chí Minh. |
| **Tuần 11** | * Thử nghiệm mô hình học máy Linear Regression. * Huấn luyện mô hình cho mỗi quốc gia, mỗi lĩnh vực và mỗi dòng dữ liệu. * Kết quả RMSE vào khoảng <10%mức độ dao động của dữ liệu và khoảng 10-15% đối với mức độ dao động dữ liệu cao. | Trường ĐH Tài Nguyên Và Môi Trường TP Hồ Chí Minh. |
| **Tuần 12** | * Lựa chọn mô hình Linear Regession. * Lưu trữ dữ liệu dữ báo ở dataframe mới. * Tiến hành trực quan hóa dữ liệu dự báo lên giao diện. | Trường ĐH Tài Nguyên Và Môi Trường TP Hồ Chí Minh. |
| **Tuần 13** | * Hoàn thiện các chức năng. * Kiểm tra và xử lý các lỗi. * Hoàn thiện giao diện | Trường ĐH Tài Nguyên Và Môi Trường TP Hồ Chí Minh. |
| **Tuần 14** | * Phân tích đánh giá về ngành năng lượng thông qua kiến thức đã tìm hiểu và học hỏi. * Đưa ra các đề xuất và biện pháp. * Hoàn thiện và kiểm tra toàn bộ dự án. | Trường ĐH Tài Nguyên Và Môi Trường TP Hồ Chí Minh. |
| **Tuần 15** | Hoàn thiện và nộp luận văn (word và ppt). | Trường ĐH Tài Nguyên Và Môi Trường TP Hồ Chí Minh. |

**MỤC LỤC**

[**CHƯƠNG I. TỔNG QUAN.** 1](#_Toc167693439)

[**1.1.** **Giới thiệu đề tài.** 1](#_Toc167693440)

[**1.2.** **Mục tiêu đề tài.** 2](#_Toc167693441)

[**1.3.** **Phạm vi đề tài.** 2](#_Toc167693442)

[**1.4.** **Đối tượng sử dụng.** 3](#_Toc167693443)

[**1.5.** **Phương pháp nghiên cứu.** 4](#_Toc167693444)

[**1.6.** **Tổng quan về ngành tài nguyên năng lượng.** 4](#_Toc167693445)

[**1.7.** **Tổng quan về ngành khoa học dữ liệu – Data Science.** 6](#_Toc167693446)

[**CHƯƠNG II. CƠ SỞ LÝ THUYẾT.** 8](#_Toc167693447)

[**2.1.Lý thuyết.** 8](#_Toc167693448)

[**2.2. Ngôn ngữ lập trình Python** 9](#_Toc167693449)

[**2.2.1. Ngôn ngữ lập trình Python là gì?** 9](#_Toc167693450)

[**2.2.2. Những lợi ích của ngôn ngữ lập trình Python.** 9](#_Toc167693451)

[**2.2.3. Đặc điểm của ngôn ngữ lập trình Python.** 10](#_Toc167693452)

[**2.3. Các thư viện sử dụng và mục đích sử dụng trong đề tài.** 13](#_Toc167693453)

[**2.3.1. Numpy.** 13](#_Toc167693454)

[**2.3.2. Pandas.** 15](#_Toc167693455)

[**2.3.3. Scikit-learn.** 17](#_Toc167693456)

[**2.3.4.Tkinter.** 19](#_Toc167693457)

[**2.3.5. Pillow.** 21](#_Toc167693458)

[**2.3.6. Matplotlib.** 22](#_Toc167693459)

[**2.3.7. Mplcursors.** 24](#_Toc167693460)

[**2.4. Pycharm.** 25](#_Toc167693461)

[**2.4.1. IDE Pycharm.** 25](#_Toc167693462)

[**2.4.2. Các tính năng nổi bật của Pycharm.** 25](#_Toc167693463)

[**2.4.3. Lựa chọn Pycharm cho đề tài.** 26](#_Toc167693464)

[**2.5.Các mô hình máy học.** 26](#_Toc167693465)

[2.5.1. ARMIMA (Autoregressive Integrated Moving Average). 26](#_Toc167693466)

[2.5.2. Prophet. 30](#_Toc167693467)

[2.5.3. Moving Average. 33](#_Toc167693468)

[2.5.4. Linear Regression. 35](#_Toc167693469)

[**CHƯƠNG III. CÀI ĐẶT THỰC NGHIỆM.** 39](#_Toc167693470)

[**3.1. Mô tả bài toán.** 39](#_Toc167693471)

[**3.2. Các vấn đề và khó khăn gặp phải.** 40](#_Toc167693472)

[**3.2.1. Về mô hình học máy:** 40](#_Toc167693473)

[**3.2.2. Về trực quan hóa dữ liệu:** 41](#_Toc167693474)

[**3.2.3. Về giao diện người dùng:** 41](#_Toc167693475)

[**3.2.4. Về kiến thức chuyên môn:** 41](#_Toc167693476)

[**3.3. Phương pháp thực hiện.** 42](#_Toc167693477)

[**3.3.1. Thu thập dữ liệu liệu.** 42](#_Toc167693478)

[**3.3.3. Phân tích dữ liệu lịch sử.** 47](#_Toc167693479)

[**3.3.5. Trực quan hóa dữ liệu dự báo.** 48](#_Toc167693480)

[**3.3.6. Xây dựng giao diện người dùng:** 48](#_Toc167693481)

[**3.2.7. Đề xuất giải pháp:** 49](#_Toc167693482)

[**3.4. Thiết kế - mô tả giao diện người dùng.** 49](#_Toc167693483)

[3.4.1. Cấu trúc giao diện: 49](#_Toc167693484)

[3.4.2. Mô tả chi tiết các thành phần: 50](#_Toc167693485)

[3.4.3. Tính năng: 50](#_Toc167693486)

[3.4.4. Ưu điểm: 51](#_Toc167693487)

[**3.5. Chức năng của ứng dụng.** 51](#_Toc167693488)

[**3.6. Áp dụng mô hình Linear Regression.** 52](#_Toc167693489)

[**3.7. Kết quả đạt được.** 53](#_Toc167693490)

[**3.7.1. Cấu trúc tổ chức file của ứng dụng.** 53](#_Toc167693491)

[**3.7.2. Giao diện tổng quan.** 55](#_Toc167693492)

[**3.7.3. Giao diện mức năng lượng sử dụng – Energy Use.** 55](#_Toc167693493)

[**3.7.4. Giao diện dự báo mức năng lượng sử dụng – Forecast Energy Use.** 57](#_Toc167693494)

[**3.7.5. Giao diện kết hợp dự báo và quá khứ.** 59](#_Toc167693495)

[**3.8. Các đề xuất và biện pháp cho các vấn đề về tài nguyên năng lượng.** 61](#_Toc167693496)

[**CHƯƠNG IV. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN.** 64](#_Toc167693497)

[**4.1. Kết luận.** 64](#_Toc167693498)

[**4.2. Hướng phát triển.** 65](#_Toc167693499)

[**DANH MỤC TÀI LIỆU THAM KHẢO.** 67](#_Toc167693500)

**DANH MỤC HÌNH ẢNH**

[Hình 2. 1.Ngôn ngữ lập trình Python 9](#_Toc167693501)

[Hình 2. 2. Thư viện Numpy. 13](#_Toc167693502)

[Hình 2. 3. Thư viện Pandas. 15](#_Toc167693503)

[Hình 2. 4. Thư viện Scikit-learn. 17](#_Toc167693504)

[Hình 2. 5. Thư viện Tkinter. 19](#_Toc167693505)

[Hình 2. 6. Thư viện Pillow. 21](#_Toc167693506)

[Hình 2. 7. Thư viện Matpolotlib. 22](#_Toc167693507)

[Hình 2. 8. IDE Pycharm. 25](#_Toc167693508)

[Hình 2. 9. Mô hình Prophet. 30](#_Toc167693509)

[Hình 3. 1. Thiết kế giao diện người dùng. 49](#_Toc167693510)

[Hình 3. 2. Cấu trúc tổ chức file. 53](#_Toc167693511)

[Hình 3. 3. Giao diện tổng quan. 55](#_Toc167693512)

[Hình 3. 4. Giao diện lĩnh vực dân dụng - Residential. 55](#_Toc167693513)

[Hình 3. 5. Giao diện lĩnh vực dịch vụ - Services. 56](#_Toc167693514)

[Hình 3. 6. Giao diện lĩnh vực công nghiệp - Industry. 56](#_Toc167693515)

[Hình 3. 7. Giao diện lĩnh vực Giao thông vận tải - Transport. 57](#_Toc167693516)

[Hình 3. 8. Giao diện dự báo lĩnh vực dân dụng - Residential. 57](#_Toc167693517)

[Hình 3. 9. Giao diện dự báo lĩnh vực dịch vụ - Services. 58](#_Toc167693518)

[Hình 3. 10. Giao diện dự báo lĩnh vực công nghiệp - Industry. 58](#_Toc167693519)

[Hình 3. 11. Giao diện dự báo lĩnh vực giao thông vận tải - Transport. 59](#_Toc167693520)

[Hình 3. 12. Giao diện kết hợp dữ liệu lĩnh vực dân dụng Residential. 59](#_Toc167693521)

[Hình 3. 13. Giao diện kết hợp dữ liệu lĩnh vực dịch vụ - Services. 60](#_Toc167693522)

[Hình 3. 14. Giao diện kết hợp dữ liệu lĩnh vực công nghiệp - Industry. 60](#_Toc167693523)

[Hình 3. 15. Giao diện kết hợp dữ liệu lĩnh vực giao thông vận tải - Transport. 61](#_Toc167693524)

**DANH MỤC VIẾT TẮT, CHÚ THÍCH**

|  |  |
| --- | --- |
| RMSE | Sai số trung bình phương gốc |
| GUI | Giao diện người dùng đồ họa |
| IEA | Cơ quan năng lượng quốc tế. |
| ARIMA | Mô hình hồi quy tự hồi quy tích hợp trung bình động. |
| IoT | Internet vạn vật, vạn vật kết nối Internet và cùng nhau thu thập trao đổi dữ liệu. |
| OOP | Lập trình hướng đối tượng. |
| API | Giao diện lập trình ứng dụng, nơi cho phép các phần mềm ứng dụng giao tiếp với nhau. |
| I/O | Input, Output đầu vào và đầu ra |
| URL | Đường dẫn đến 1 địa chỉ trên Internet |
| JPEG, PNG, GIF, BMP, TIFF, JPG, SVG | Định dạng file hình ảnh, thường được kí hiệu ở phía sau dấu chấm ở cuối mỗi hình ảnh. |
| PDF | Định dạng file tài liệu di động. |
| IDE | Môi trường phát triển tích hợp (Integrated Development Enviroment) |
| GPD | Tổng sản phẩm quốc nội (Gross Domestic Product) |
| MA | Mô hình máy học Moving Average |
| OLS | Phương pháp bình phương bé nhất (Ordinary Least Squares) |

**TÓM TẮT**

Đồ án “Xây dựng ứng dụng phân tích và trực quan hóa dữ liệu về việc sử dụng năng lượng toàn cầu” là một hệ thống GUI được xây dựng nhằm cung cấp cái nhìn trực quan và chi tiết về tình hình sử dụng năng lượng toàn cầu, đồng thời dự báo xu hướng tiêu thụ năng lượng trong 5 năm tới. Ứng dụng sử dụng ngôn ngữ lập trình Python cùng với các thư viện Tkinter, Matplotlib và sklearn để xử lý dữ liệu, xây dựng mô hình dự báo và trực quan hóa kết quả.

Nguồn dữ liệu chính được sử dụng là bộ dữ liệu của Cơ quan Năng lượng Quốc tế (IEA), bao gồm các chỉ số quan trọng về sản xuất, tiêu thụ, nhập khẩu, xuất khẩu năng lượng cho các quốc gia khác nhau và được phân loại theo 4 lĩnh vực chính: dân dụng, dịch vụ, công nghiệp và giao thông vận tải.

Ứng dụng cho phép người dùng lựa chọn quốc gia và lĩnh vực mong muốn để phân tích dữ liệu lịch sử và dự báo xu hướng tiêu thụ năng lượng. Mô hình hồi quy tuyến tính được sử dụng để dự báo mức sử dụng năng lượng, với độ chính xác được đánh giá thông qua chỉ số RMSE. Kết quả dự báo được trực quan hóa bằng các biểu đồ cột trực quan, giúp người dùng dễ dàng so sánh và nhận định xu hướng.

Thông qua đồ án này, người dùng có thể nắm bắt được tình hình sử dụng năng lượng hiện tại, dự đoán nhu cầu năng lượng trong tương lai, từ đó hỗ trợ việc hoạch định chính sách, đầu tư, sản xuất và tiêu dùng năng lượng hiệu quả hơn. Đồng thời, dựa trên kết quả phân tích và dự báo, đồ án sẽ đề xuất các phương án và giải pháp nhằm giải quyết những vấn đề về tài nguyên năng lượng toàn cầu, hướng đến một tương lai bền vững.

**CHƯƠNG I. TỔNG QUAN.**

* 1. **Giới thiệu đề tài.**

Năng lượng là nền tảng thiết yếu cho sự phát triển kinh tế, xã hội và nâng cao chất lượng cuộc sống. Tuy nhiên, việc sử dụng năng lượng cũng đặt ra nhiều thách thức, từ việc đảm bảo an ninh năng lượng, giảm thiểu tác động đến môi trường đến việc tối ưu hóa hiệu quả sử dụng và nguồn năng lượng hóa thạch đang dần cạn kiệt và gây ra những tác động tiêu cực đến môi trường. Bên cạnh đó, nhu cầu năng lượng ngày càng tăng cao do dân số thế giới tăng nhanh và sự phát triển của các ngành công nghiệp, công nghệ. Trong bối cảnh biến đổi khí hậu và khủng hoảng năng lượng toàn cầu, việc phân tích, dự báo và quản lý việc sử dụng năng lượng trở nên cấp bách hơn bao giờ hết. Điều này đặt ra thách thức lớn đối với an ninh năng lượng toàn cầu và sự phát triển bền vững.

Để ứng phó với thách thức này, việc phân tích, dự báo và tối ưu hóa việc sử dụng năng lượng trở nên cấp thiết hơn bao giờ hết. Việc nắm bắt được tình hình sử dụng năng lượng hiện tại, dự đoán xu hướng tiêu thụ trong tương lai sẽ cung cấp những thông tin hữu ích cho việc hoạch định chính sách năng lượng, đầu tư vào các nguồn năng lượng tái tạo, nâng cao hiệu quả sử dụng năng lượng và giảm thiểu tác động đến môi trường.

Đồ án được thực hiện với mục tiêu xây dựng một công cụ hỗ trợ phân tích và dự báo mức tiêu thụ năng lượng toàn cầu, dựa trên dữ liệu từ Cơ quan Năng lượng Quốc tế (IEA) - một nguồn dữ liệu uy tín về năng lượng.

Ứng dụng sẽ tập trung vào 4 lĩnh vực chính: Dân dụng (Residential): Phân tích mức tiêu thụ năng lượng trong các hộ gia đình cho các mục đích như sưởi ấm, làm mát, chiếu sáng, nấu nướng, sử dụng thiết bị gia dụng,... Dịch vụ (Services): Phân tích mức tiêu thụ năng lượng trong các ngành dịch vụ như thương mại, du lịch, y tế, giáo dục,... Công nghiệp (Industry): Phân tích mức tiêu thụ năng lượng trong các ngành công nghiệp sản xuất như chế tạo, khai khoáng, xây dựng,... Giao thông vận tải (Transport): Phân tích mức tiêu thụ năng lượng trong các hoạt động vận tải hành khách và hàng hóa bằng đường bộ, đường thủy, đường hàng không.

Bằng cách sử dụng các kỹ thuật phân tích dữ liệu và mô hình hồi quy tuyến tính, ứng dụng sẽ cung cấp cho người dùng: Dữ liệu lịch sử về mức tiêu thụ năng lượng theo từng quốc gia và lĩnh vực. Dự báo xu hướng tiêu thụ năng lượng trong 5 năm tới. Biểu đồ trực quan giúp dễ dàng so sánh và nhận định xu hướng.

Người dùng có thể dễ dàng lựa chọn quốc gia và lĩnh vực mong muốn để phân tích dữ liệu lịch sử và dự báo xu hướng tiêu thụ năng lượng trong 5 năm tới. Mô hình hồi quy tuyến tính được sử dụng để dự báo, với độ chính xác được đánh giá thông qua chỉ số RMSE. Kết quả được trực quan hóa bằng các biểu đồ cột trực quan, giúp người dùng dễ dàng so sánh và nhận định xu hướng.

Đề tài góp phần cung cấp cái nhìn toàn cảnh về tình hình sử dụng năng lượng hiện tại, dự đoán nhu cầu năng lượng trong tương lai, từ đó hỗ trợ việc hoạch định chính sách, đầu tư, sản xuất và tiêu dùng năng lượng hiệu quả hơn. Dựa trên kết quả phân tích và dự báo, đề tài cũng sẽ đề xuất các phương án và giải pháp nhằm giải quyết những vấn đề về tài nguyên năng lượng toàn cầu, hướng đến một tương lai bền vững.

* 1. **Mục tiêu đề tài.**

Đồ án cần đạt được mục tiêu xây dựng ứng dụng GUI trực quan., thân thiện và dễ sử dụng, trực quan hóa dữ liệu sử dụng năng lượng lịch sử và dự báo nhu cầu năng lượng trong 5 năm tới. Cùng với đó kết hợp kết quả lịch sử và dự báo để tiến hành so sánh và nắm bắt xu hướng năng lượng.

* 1. **Phạm vi đề tài.**

Đề tài được giới hạn trong phạm vi nghiên cứu sau:

* Loại dữ liệu sử dụng: Đề tài sử dụng dữ liệu năng lượng từ IEA (Cơ quan Năng lượng Quốc tế), một tổ chức liên chính phủ có uy tín trong lĩnh vực năng lượng. Dữ liệu IEA cung cấp thông tin toàn diện về sản xuất, tiêu thụ, nhập khẩu và xuất khẩu năng lượng cho các quốc gia trên thế giới.
* Lĩnh vực tập trung: Đề tài tập trung phân tích và dự báo mức sử dụng năng lượng trong bốn lĩnh vực chính: năng lượng dân dụng (Residential), dịch vụ (Services), công nghiệp (Industry), và giao thông vận tải (Transport).
* Quốc gia: Albania, Argentina, Armenia, Australia, Austria, Azerbaijan, Belarus, Belgium, Bosnia and Herzegovina, Brazil, Bulgaria, Canada, Chile, Chinese Taipei, Colombia, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Georgia, Germany, Greece, Hong Kong (China), Hungary, IEA Total, Ireland, Italy, Japan, Kazakhstan, Korea, Kosovo, Kyrgyzstan, Latvia, Lithuania, Luxembourg, Malta, Mexico, Morocco, Netherlands, New Zealand, Norway, Poland, Portugal, Republic of Moldova, Republic of North Macedonia, Republic of Türkiye, Romania, Serbia, Slovak Republic, Slovenia, South Africa, Spain, Sweden, Switzerland, Ukraine, United Kingdom, United States, Uruguay, Uzbekistan.
* Phương pháp dự báo: Đề tài sử dụng phương pháp hồi quy tuyến tính để dự báo mức sử dụng năng lượng trong 5 năm tiếp theo. Phương pháp này được lựa chọn do tính đơn giản, hiệu quả và khả năng dự báo tương đối chính xác trong ngắn hạn.

Phạm vi nghiên cứu được giới hạn như trên để đảm bảo tính khả thi và tập trung vào mục tiêu chính của đề tài là xây dựng ứng dụng phân tích và dự báo mức sử dụng năng lượng cho các quốc gia và lĩnh vực trọng yếu.

* 1. **Đối tượng sử dụng.**

Ứng dụng phân tích và dự báo mức sử dụng năng lượng được thiết kế nhằm phục vụ cho một loạt các đối tượng người dùng, mỗi người đều có những mối quan tâm và nhu cầu riêng biệt liên quan đến lĩnh vực năng lượng. Ứng dụng này đặc biệt hữu ích cho các nhà hoạch định chính sách, những người cần hiểu rõ xu hướng tiêu thụ năng lượng để xây dựng chiến lược phát triển năng lượng bền vững, hoạch định chính sách tiết kiệm năng lượng, phát triển nguồn năng lượng tái tạo và đảm bảo an ninh năng lượng quốc gia.

Đối với các nhà đầu tư, ứng dụng cung cấp thông tin giá trị để đánh giá tiềm năng của thị trường năng lượng, dự báo nhu cầu trong tương lai và đưa ra quyết định đầu tư hiệu quả. Việc phân tích dữ liệu theo từng quốc gia và lĩnh vực cụ thể giúp nhà đầu tư xác định được những thị trường tiềm năng và ngành nghề có nhu cầu năng lượng lớn.

Các doanh nghiệp hoạt động trong ngành năng lượng cũng được hưởng lợi từ ứng dụng này. Họ có thể sử dụng dữ liệu để theo dõi xu hướng tiêu thụ, dự báo nhu cầu, điều chỉnh kế hoạch sản xuất kinh doanh, và tối ưu hóa nguồn lực. Dữ liệu chi tiết theo từng loại năng lượng giúp doanh nghiệp đưa ra quyết định chính xác về đầu tư và phát triển sản phẩm phù hợp với nhu cầu thị trường.

Ứng dụng cũng là một công cụ hỗ trợ đắc lực cho các nhà nghiên cứu trong lĩnh vực năng lượng, biến đổi khí hậu và phát triển bền vững. Dữ liệu từ IEA cung cấp nền tảng vững chắc cho các nghiên cứu chuyên sâu, xây dựng và kiểm chứng mô hình dự báo, từ đó đưa ra các khuyến nghị về chính sách năng lượng hiệu quả.

Cuối cùng, với giao diện trực quan và dễ sử dụng, ứng dụng là công cụ học tập hữu ích cho sinh viên, học sinh các ngành liên quan đến năng lượng, kinh tế, môi trường. Ứng dụng giúp người dùng dễ dàng tiếp cận và tìm hiểu về dữ liệu năng lượng, từ đó nâng cao nhận thức về tầm quan trọng của việc sử dụng năng lượng hiệu quả.

* 1. **Phương pháp nghiên cứu.**

Phương pháp nghiên cứu được sử dụng trong dự án này là một sự kết hợp của các kỹ thuật khai phá dữ liệu, trực quan hóa dữ liệu và học máy, được áp dụng theo một quy trình tuần tự và có hệ thống.

Đầu tiên, quá trình bắt đầu bằng việc thu thập dữ liệu từ nguồn IEA uy tín, sau đó là các bước tiền xử lý dữ liệu quan trọng như làm sạch, chuẩn hóa và phân loại để đảm bảo chất lượng và tính nhất quán. Việc lựa chọn các biến số phù hợp với mục tiêu nghiên cứu là bước tiếp theo, nhằm đảm bảo tính hiệu quả và tập trung vào các khía cạnh quan trọng của lĩnh vực năng lượng.

Song song với việc xử lý dữ liệu, việc thiết kế và xây dựng giao diện người dùng được tiến hành dựa trên các nguyên tắc trực quan, thân thiện và dễ sử dụng. Dữ liệu được trực quan hóa bằng các biểu đồ phù hợp, được sắp xếp khoa học và có chú thích rõ ràng để người dùng dễ dàng nắm bắt thông tin.

Để dự báo mức sử dụng năng lượng trong tương lai, nhiều mô hình học máy đã được thử nghiệm và so sánh, bao gồm ARIMA, Prophet, Moving Average và Linear Regression. Mỗi mô hình được huấn luyện kỹ lưỡng trên tập dữ liệu của từng quốc gia, lĩnh vực và dòng dữ liệu để đảm bảo tính chính xác và phù hợp. Sau quá trình đánh giá toàn diện, mô hình Linear Regression được lựa chọn do hiệu suất vượt trội.

Kết quả dự báo sau đó được lưu trữ và tích hợp vào giao diện, cung cấp cho người dùng khả năng phân tích dữ liệu lịch sử và dự báo cho các quốc gia và lĩnh vực khác nhau. Cuối cùng, dự án được hoàn thiện với đầy đủ chức năng, giao diện thân thiện và đã được kiểm tra kỹ lưỡng để đảm bảo hoạt động ổn định.

Bên cạnh việc xây dựng ứng dụng, quá trình nghiên cứu còn bao gồm việc tìm hiểu sâu về ngành năng lượng, phân tích các xu hướng và thách thức, từ đó đưa ra các đề xuất và biện pháp nhằm nâng cao hiệu quả sử dụng năng lượng và hướng tới phát triển bền vững.

* 1. **Tổng quan về ngành tài nguyên năng lượng.**

Ngành năng lượng toàn cầu hiện nay đang trải qua một giai đoạn chuyển đổi sâu rộng và đầy biến động. Những yếu tố như biến đổi khí hậu, khủng hoảng địa chính trị, bứt phá công nghệ và nhu cầu năng lượng ngày càng tăng đang tạo nên bức tranh đa chiều với cả cơ hội và thách thức to lớn.

Biến đổi khí hậu là thách thức cấp bách nhất, đặt ra yêu cầu chuyển dịch mạnh mẽ sang các nguồn năng lượng sạch và tái tạo. Sự phụ thuộc vào nhiên liệu hóa thạch đang gây ra những hậu quả nghiêm trọng về môi trường, khiến việc giảm phát thải khí nhà kính trở thành ưu tiên hàng đầu. Tuy nhiên, thách thức này đồng thời cũng mở ra cơ hội to lớn cho sự phát triển của năng lượng tái tạo. Năng lượng mặt trời, gió, địa nhiệt và sinh khối đang ngày càng phổ biến nhờ chi phí giảm và hiệu suất tăng. Bên cạnh đó, các công nghệ năng lượng mới như pin nhiên liệu hydro, năng lượng hạt nhân thế hệ mới và công nghệ thu giữ carbon hứa hẹn tạo ra những đột phá trong tương lai.

Xu hướng số hóa và tự động hóa đang tái định hình ngành năng lượng, tạo ra hệ thống năng lượng thông minh hơn và hiệu quả hơn. Lưới điện thông minh, IoT và Big Data đang được ứng dụng để tối ưu hóa sản xuất, truyền tải và phân phối năng lượng. Sự kết hợp giữa công nghệ và năng lượng không chỉ mang lại hiệu quả kinh tế mà còn góp phần nâng cao độ tin cậy và an toàn cho hệ thống năng lượng.

Bối cảnh địa chính trị bất ổn cũng tác động lớn đến an ninh năng lượng toàn cầu. Các quốc gia đang nỗ lực đa dạng hóa nguồn cung năng lượng, tăng cường khả năng tự chủ và giảm thiểu rủi ro gián đoạn nguồn cung. Xu hướng này càng trở nên cấp bách trong bối cảnh xung đột và căng thẳng địa chính trị gia tăng, đòi hỏi các quốc gia phải chủ động và linh hoạt trong chiến lược năng lượng của mình.

Nhu cầu năng lượng toàn cầu dự kiến tiếp tục tăng trong những thập kỷ tới, chủ yếu do tăng trưởng kinh tế và dân số ở các nước đang phát triển. Đáp ứng nhu cầu này một cách bền vững và hiệu quả là một thách thức lớn. Để thực hiện được mục tiêu này, sự hợp tác chặt chẽ giữa các quốc gia, doanh nghiệp và tổ chức quốc tế là yếu tố then chốt. Các bên liên quan cần chung tay xây dựng khuôn khổ pháp lý, thúc đẩy đầu tư, chuyển giao công nghệ và chia sẻ kiến thức để hướng tới một tương lai năng lượng bền vững cho toàn cầu.

Ngành năng lượng hiện nay đối mặt với cả cơ hội và thách thức đều hiện hữu, đòi hỏi sự thay đổi mang tính chiến lược. Việc chuyển đổi sang một hệ thống năng lượng bền vững, an toàn và hiệu quả là nhiệm vụ cấp bách, yêu cầu sự đổi mới, đầu tư và hợp tác toàn cầu.

* 1. **Tổng quan về ngành khoa học dữ liệu – Data Science.**



Hình 1. 1 Khoa học dữ liệu (Data Science)

Khoa học dữ liệu (Data Science) đang trở thành một ngành nghề chủ chốt, tạo động lực cho sự đổi mới và chuyển đổi trong hầu hết các lĩnh vực, từ nghiên cứu khoa học đến ứng dụng thực tiễn. Khả năng khai thác thông tin giá trị từ khối lượng dữ liệu khổng lồ đã biến khoa học dữ liệu thành một công cụ thiết yếu cho việc giải quyết vấn đề phức tạp và đưa ra quyết định chiến lược hiệu quả.

Chúng ta đang sống trong kỷ nguyên bùng nổ dữ liệu, với lượng dữ liệu được tạo ra mỗi ngày vượt xa khả năng xử lý của con người. Khoa học dữ liệu đóng vai trò then chốt trong việc thu thập, xử lý, phân tích và diễn giải dữ liệu, biến dữ liệu thô thành thông tin hữu ích, kiến thức và hiểu biết sâu sắc.

Sự phát triển vượt bậc của các công nghệ như học máy (Machine Learning), học sâu (Deep Learning), trí tuệ nhân tạo (AI) và điện toán đám mây (Cloud Computing) đã tạo động lực mạnh mẽ cho khoa học dữ liệu. Các thuật toán ngày càng tinh vi, khả năng xử lý dữ liệu lớn và khả năng truy cập dễ dàng vào tài nguyên tính toán đã mở ra những khả năng mới cho việc phân tích, dự đoán và tự động hóa.

Khoa học dữ liệu được ứng dụng rộng rãi trong nhiều lĩnh vực, từ kinh doanh và tài chính đến y tế, giáo dục, sản xuất và nông nghiệp. Trong lĩnh vực y tế, khoa học dữ liệu giúp chẩn đoán bệnh sớm, cá nhân hóa phác đồ điều trị, và phát triển thuốc mới. Trong kinh doanh, khoa học dữ liệu hỗ trợ việc phân tích thị trường, tối ưu hóa chuỗi cung ứng, cá nhân hóa trải nghiệm khách hàng và phát hiện gian lận. Các ứng dụng trong sản xuất bao gồm tối ưu hóa quy trình sản xuất, bảo trì dự đoán và quản lý chuỗi cung ứng.

Sự gia tăng nhu cầu về các chuyên gia khoa học dữ liệu trên toàn cầu là minh chứng rõ ràng cho tầm quan trọng của ngành. Các kỹ năng phân tích dữ liệu, lập trình, hiểu biết về thuật toán học máy và khả năng diễn giải kết quả là những yếu tố quan trọng cho sự thành công trong lĩnh vực này. Nhu cầu về các chuyên gia có khả năng kết hợp kiến thức kỹ thuật với sự am hiểu về lĩnh vực ứng dụng ngày càng tăng cao.

Bên cạnh những cơ hội, khoa học dữ liệu cũng phải đối mặt với những thách thức như vấn đề về quyền riêng tư dữ liệu, sự thiên vị trong thuật toán, và việc giải thích kết quả phân tích một cách dễ hiểu cho người dùng. Việc đảm bảo tính minh bạch, công bằng và trách nhiệm giải trình trong việc sử dụng dữ liệu là yếu tố quan trọng để xây dựng niềm tin và sự chấp nhận của xã hội đối với khoa học dữ liệu.

Khoa học dữ liệu đang thay đổi cách chúng ta sống, làm việc và tương tác với thế giới. Ngành này tiếp tục phát triển với tốc độ chóng mặt, tạo ra những cơ hội mới cho sự đổi mới và giải quyết các vấn đề toàn cầu. Sự phát triển của khoa học dữ liệu hứa hẹn sẽ mang lại những bước tiến vượt bậc trong nhiều lĩnh vực, góp phần xây dựng một tương lai hiệu quả, thông minh và bền vững hơn.

**CHƯƠNG II. CƠ SỞ LÝ THUYẾT.**

**2.1.Lý thuyết.**

Trong bối cảnh biến đổi khí hậu và khủng hoảng năng lượng ngày càng trở nên cấp bách, việc thấu hiểu và dự đoán xu hướng sử dụng năng lượng là nền tảng cho việc xây dựng các giải pháp năng lượng bền vững. Đề tài này hướng đến việc phát triển một ứng dụng phân tích và dự báo mức sử dụng năng lượng, đóng góp vào việc nâng cao nhận thức và cung cấp công cụ hỗ trợ ra quyết định cho các bên liên quan.

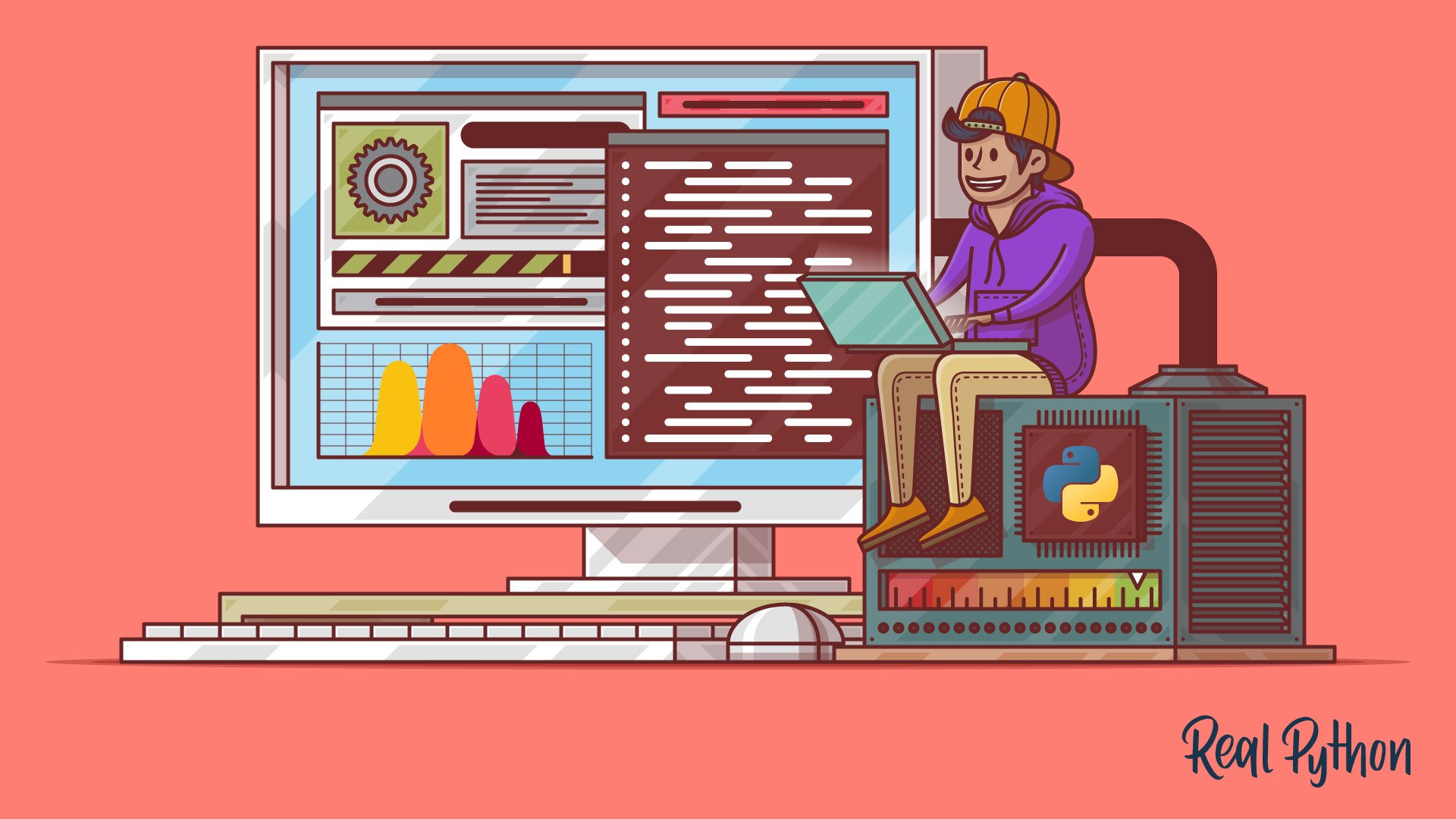
Mục đích cốt lõi của đề tài là xây dựng một hệ thống tích hợp khả năng phân tích dữ liệu lịch sử và dự báo xu hướng sử dụng năng lượng trong tương lai, tập trung vào các lĩnh vực tiêu thụ năng lượng chính như dân dụng, dịch vụ, công nghiệp và giao thông vận tải. Bằng cách khai thác nguồn dữ liệu năng lượng phong phú từ IEA, ứng dụng sẽ cung cấp cái nhìn toàn diện về tình hình sử dụng năng lượng theo quốc gia, lĩnh vực và loại hình năng lượng.

Hơn nữa, việc ứng dụng mô hình hồi quy tuyến tính, cho phép dự đoán chính xác và đáng tin cậy về mức sử dụng năng lượng trong tương lai. Điều này mang đến cho người dùng khả năng phân tích các kịch bản năng lượng, đánh giá tác động của các chính sách năng lượng và xây dựng chiến lược phát triển năng lượng bền vững.

Bên cạnh việc cung cấp thông tin và dự báo, đề tài cũng hướng đến việc thúc đẩy nhận thức về tầm quan trọng của việc sử dụng năng lượng hiệu quả và chuyển dịch sang các nguồn năng lượng tái tạo. Qua giao diện trực quan và dễ sử dụng, ứng dụng góp phần nâng cao nhận thức cộng đồng về các vấn đề năng lượng, từ đó khuyến khích sự tham gia của người dân vào việc xây dựng một tương lai năng lượng bền vững.

**2.2. Ngôn ngữ lập trình Python**

**2.2.1. Ngôn ngữ lập trình Python là gì?**



Hình 2. 1.Ngôn ngữ lập trình Python

Python là một ngôn ngữ lập trình bậc cao, thông dịch, hướng đối tượng, và có mục đích chung, được tạo ra bởi Guido van Rossum và phát hành lần đầu tiên vào năm 1991. Python được thiết kế với triết lý cốt lõi tập trung vào khả năng đọc code và cú pháp rõ ràng, cho phép lập trình viên diễn đạt ý tưởng trong số lượng code ít hơn so với một số ngôn ngữ khác. Khác với các ngôn ngữ yêu cầu biên dịch trước khi thực thi, Python sử dụng trình thông dịch để thực thi mã nguồn trực tiếp, từng dòng một, giúp cho việc phát triển và gỡ lỗi code trở nên nhanh chóng và thuận tiện hơn.

Python được biết đến với tính linh hoạt, cho phép nó được sử dụng trong nhiều lĩnh vực, bao gồm phát triển web, phân tích dữ liệu, khoa học máy tính, học máy, và nhiều hơn nữa. Python là một ngôn ngữ lập trình đa năng, có khả năng xây dựng các ứng dụng quy mô nhỏ đến các hệ thống phức tạp.

**2.2.2. Những lợi ích của ngôn ngữ lập trình Python.**

Sự phổ biến của Python đến từ một loạt lợi ích đáng kể, khiến nó trở thành lựa chọn lý tưởng cho cả lập trình viên mới bắt đầu lẫn chuyên nghiệp:

* Dễ học và dễ sử dụng: Cú pháp đơn giản, trực quan và dễ hiểu của Python giúp giảm thiểu rào cản gia nhập cho người mới bắt đầu, cho phép họ nhanh chóng nắm bắt được các khái niệm lập trình cơ bản và bắt đầu viết code.
* Năng suất cao: Python cho phép lập trình viên tập trung vào việc giải quyết vấn đề thay vì sa lầy vào các chi tiết cú pháp phức tạp. Cấu trúc rõ ràng, thư viện phong phú và nhiều công cụ hỗ trợ giúp rút ngắn thời gian phát triển, tăng năng suất làm việc.
* Thư viện phong phú: Python sở hữu một hệ sinh thái thư viện khổng lồ, được xây dựng và duy trì bởi cộng đồng, cung cấp các giải pháp cho hầu hết mọi nhu cầu lập trình. Các thư viện phổ biến như pandas, numpy, scipy cho khoa học dữ liệu, scikit-learn, TensorFlow cho học máy, matplotlib, seaborn cho trực quan hóa dữ liệu, Django, Flask cho phát triển web, và vô số thư viện khác đáp ứng mọi nhu cầu phát triển.
* Cộng đồng lớn mạnh và năng động: Cộng đồng Python là một trong những cộng đồng lớn mạnh và tích cực nhất trong thế giới lập trình. Cộng đồng này cung cấp một nguồn tài nguyên vô giá, từ các diễn đàn thảo luận, các trang web chia sẻ code, đến các hội thảo và sự kiện trực tuyến, giúp người dùng dễ dàng tìm kiếm thông tin, hỗ trợ và giải đáp thắc mắc.
* Tính linh hoạt: Python không bị giới hạn trong một lĩnh vực cụ thể. Nó có thể được sử dụng để xây dựng các ứng dụng web, phân tích dữ liệu, phát triển game, tự động hóa, tạo kịch bản, và nhiều hơn nữa. Sự linh hoạt này cho phép lập trình viên dễ dàng chuyển đổi giữa các dự án và lĩnh vực khác nhau.
* Miễn phí và mã nguồn mở: Python là ngôn ngữ lập trình miễn phí, có mã nguồn mở, được phân phối theo giấy phép cho phép sử dụng, sửa đổi và phân phối tự do. Điều này giúp giảm thiểu chi phí phát triển, khuyến khích sự sáng tạo và chia sẻ kiến thức trong cộng đồng.

**2.2.3. Đặc điểm của ngôn ngữ lập trình Python.**

* Thông dịch: Thay vì biên dịch code thành mã máy trước khi thực thi, Python sử dụng trình thông dịch để thực thi mã nguồn trực tiếp, từng dòng một. Điều này giúp cho việc phát triển và gỡ lỗi code trở nên nhanh chóng và thuận tiện hơn.
* Kiểu động: Python không yêu cầu lập trình viên phải khai báo kiểu dữ liệu cho biến. Kiểu dữ liệu được xác định động dựa trên giá trị được gán cho biến. Tính năng này giúp đơn giản hóa việc viết code, tuy nhiên cần chú ý đến các vấn đề tiềm ẩn về kiểu dữ liệu.
* Hướng đối tượng: Python hỗ trợ lập trình hướng đối tượng (OOP), cho phép người dùng định nghĩa các lớp, đối tượng, kế thừa, và đa hình. OOP giúp tổ chức code một cách hiệu quả, tăng khả năng tái sử dụng và bảo trì code.
* Cấp phát bộ nhớ tự động: Python sử dụng cơ chế cấp phát và giải phóng bộ nhớ tự động (garbage collection), giải phóng lập trình viên khỏi việc quản lý bộ nhớ thủ công. Điều này giúp giảm thiểu lỗi liên quan đến bộ nhớ, đơn giản hóa quá trình phát triển.

**2.2.4. Ứng dụng của Python.**

* + - Python được sử dụng rộng rãi trong nhiều lĩnh vực, biến nó thành một trong những ngôn ngữ lập trình phổ biến nhất hiện nay:
    - Phát triển web: Python được dùng để xây dựng web server-side, tạo ứng dụng web và API. Các framework web phổ biến như Django và Flask giúp đơn giản hóa việc phát triển web, cho phép tạo ra các ứng dụng web mạnh mẽ và hiệu quả.
    - Phân tích và trực quan hóa dữ liệu: Python được sử dụng rộng rãi trong lĩnh vực khoa học dữ liệu, giúp phân tích, xử lý và trực quan hóa dữ liệu. Các thư viện như pandas, numpy, matplotlib và seaborn cung cấp các công cụ mạnh mẽ cho việc thao tác và trực quan hóa dữ liệu, cho phép trích xuất thông tin chi tiết từ dữ liệu thô.
    - Học máy và trí tuệ nhân tạo: Python là ngôn ngữ hàng đầu trong lĩnh vực học máy (machine learning) và trí tuệ nhân tạo (AI). Các thư viện như scikit-learn, TensorFlow, PyTorch cho phép xây dựng và triển khai các mô hình học máy phức tạp, được ứng dụng trong nhiều lĩnh vực như nhận diện hình ảnh, xử lý ngôn ngữ tự nhiên, dự báo, và nhiều hơn nữa.
    - Khoa học máy tính và kỹ thuật: Python được sử dụng rộng rãi trong các lĩnh vực khoa học máy tính, kỹ thuật, và nghiên cứu khoa học. Khả năng tính toán khoa học mạnh mẽ, khả năng trực quan hóa dữ liệu và tích hợp với các hệ thống khác khiến Python trở thành công cụ lý tưởng cho việc mô phỏng, phân tích, và giải quyết các bài toán khoa học phức tạp.
    - Phát triển phần mềm và ứng dụng: Python được sử dụng để xây dựng các ứng dụng desktop, ứng dụng di động, game, và nhiều loại phần mềm khác. Python cung cấp các công cụ và thư viện cần thiết để phát triển các ứng dụng đa dạng, từ ứng dụng dòng lệnh đơn giản đến các ứng dụng giao diện người dùng phức tạp.
    - Tự động hóa: Python là lựa chọn phổ biến cho việc tự động hóa các tác vụ lặp đi lặp lại, đơn giản hóa quy trình làm việc và tăng năng suất. Từ việc tự động hóa các tác vụ hệ thống đến việc thu thập dữ liệu web, Python cung cấp các công cụ linh hoạt cho việc tự động hóa mọi thứ.
    - Tự động hóa kiểm thử phần mềm và ứng dụng: Python được sử dụng rộng rãi trong việc tự động hóa kiểm thử phần mềm, giúp tăng hiệu quả và độ tin cậy của quá trình kiểm thử. Các framework kiểm thử như Selenium và pytest cho phép viết kịch bản kiểm thử tự động, kiểm tra chức năng của ứng dụng, và phát hiện lỗi sớm trong quá trình phát triển.

Trong đồ án này, Python được sử dụng để:

* + - Xử lý và làm sạch dữ liệu: Sử dụng thư viện pandas để đọc dữ liệu từ file Excel, xử lý dữ liệu bị thiếu, chuyển đổi kiểu dữ liệu, và sắp xếp dữ liệu.
    - Xây dựng mô hình dự báo: Sử dụng thư viện scikit-learn để xây dựng mô hình Linear Regression, huấn luyện mô hình trên dữ liệu lịch sử, và dự báo mức sử dụng năng lượng trong tương lai.
    - Trực quan hóa dữ liệu: Sử dụng thư viện matplotlib để tạo các biểu đồ trực quan hóa dữ liệu lịch sử và dự báo, giúp người dùng dễ dàng phân tích và so sánh các xu hướng.
    - Xây dựng giao diện người dùng: Sử dụng thư viện Tkinter để xây dựng giao diện đồ họa cho phép người dùng tương tác với ứng dụng, lựa chọn dữ liệu, và hiển thị kết quả.

**2.3. Các thư viện sử dụng và mục đích sử dụng trong đề tài.**

Dự án phân tích và dự báo mức sử dụng năng lượng tận dụng sức mạnh của nhiều thư viện Python để xử lý dữ liệu, xây dựng mô hình dự báo, trực quan hóa và xây dựng giao diện người dùng. Dưới đây là giới thiệu chi tiết về các thư viện đóng vai trò then chốt trong dự án:

**2.3.1. Numpy.**



Hình 2. 2. Thư viện Numpy.

Trong kỷ nguyên bùng nổ dữ liệu như hiện nay, việc xử lý hiệu quả các phép toán trên khối lượng dữ liệu khổng lồ là một bài toán then chốt trong nhiều lĩnh vực, từ khoa học máy tính đến phân tích dữ liệu, học máy và nhiều hơn thế nữa. NumPy (Numerical Python) ra đời như một giải pháp tối ưu cho bài toán này, trở thành thư viện nền tảng, không thể thiếu trong thế giới Python.

NumPy là một thư viện mã nguồn mở, cung cấp hỗ trợ mạnh mẽ cho mảng đa chiều, ma trận lớn và một bộ sưu tập đồ sộ các hàm toán học bậc cao được tối ưu hóa cho việc thao tác trên các cấu trúc dữ liệu này. Nói cách khác, NumPy là một phần mở rộng của Python, cho phép thực hiện các phép toán số học và logic trên mảng hiệu quả hơn rất nhiều so với việc sử dụng các cấu trúc dữ liệu có sẵn trong Python thuần túy.

Mục đích chính của NumPy là cung cấp một đối tượng mảng N chiều mạnh mẽ và hiệu quả. So với danh sách (list) trong Python, mảng NumPy có ưu thế vượt trội về hiệu suất, khả năng tiết kiệm bộ nhớ và tính tiện lợi. Hãy thử tưởng tượng bạn cần xử lý một danh sách chứa hàng triệu số nguyên trong Python. Việc thực hiện các phép toán như cộng, nhân, tìm kiếm trên danh sách này sẽ mất rất nhiều thời gian và tài nguyên. Trong khi đó, mảng NumPy cho phép bạn thực hiện các phép toán tương tự một cách nhanh chóng và hiệu quả hơn rất nhiều.

Lợi ích mà NumPy mang lại không chỉ dừng lại ở tốc độ xử lý. Do mảng NumPy được lưu trữ trong bộ nhớ liền kề, việc truy cập và thao tác dữ liệu trở nên nhanh chóng hơn, đặc biệt là với các mảng có kích thước lớn. Bên cạnh đó, việc lưu trữ dữ liệu cùng kiểu dữ liệu trong mảng NumPy giúp tối ưu hóa việc sử dụng bộ nhớ, giảm thiểu lãng phí tài nguyên. Ví dụ, một mảng NumPy chứa 1 triệu số nguyên 32-bit sẽ chỉ sử dụng khoảng 4MB bộ nhớ, trong khi đó một danh sách Python chứa 1 triệu số nguyên tương đương có thể sử dụng tới 10MB bộ nhớ hoặc hơn. Hơn nữa, NumPy cung cấp sẵn một bộ sưu tập lớn các hàm toán học và logic đã được tối ưu hóa, giúp việc thao tác và tính toán trên mảng trở nên dễ dàng và hiệu quả hơn bao giờ hết. Các hàm này bao gồm từ các phép toán đơn giản như cộng, trừ, nhân, chia đến các hàm phức tạp hơn như các phép toán lượng giác, hàm mũ, logarit và đại số tuyến tính.

Sự phổ biến của NumPy được minh chứng qua việc nó là một trong những thư viện Python được tải xuống nhiều nhất từ kho lưu trữ Python Package Index (PyPI), với hàng triệu lượt tải xuống mỗi tháng. Không chỉ vậy, NumPy còn đóng vai trò nền tảng cho rất nhiều thư viện Python khác trong lĩnh vực khoa học dữ liệu và học máy, chẳng hạn như Pandas, SciPy, Matplotlib, và Scikit-learn. Rất nhiều dự án, ứng dụng nổi tiếng trong lĩnh vực phân tích dữ liệu, học máy, thị giác máy tính và nhiều lĩnh vực khác đều dựa trên nền tảng NumPy.

Cụ thể hơn, trong lĩnh vực khoa học dữ liệu, NumPy được sử dụng rộng rãi cho việc thao tác dữ liệu, bao gồm làm sạch, biến đổi, sắp xếp, lọc và xử lý dữ liệu thô. Ví dụ, bạn có thể sử dụng NumPy để loại bỏ các giá trị khuyết, điền giá trị khuyết, chuẩn hóa dữ liệu, hoặc chuyển đổi dữ liệu giữa các kiểu dữ liệu khác nhau. NumPy cũng hỗ trợ tính toán các đại lượng thống kê mô tả như trung bình, độ lệch chuẩn, phương sai, hiệp phương sai,... và đóng vai trò quan trọng trong việc xây dựng mô hình học máy. Bên cạnh đó, NumPy còn được ứng dụng trong nhiều lĩnh vực khác như xử lý ảnh, trong đó nó được sử dụng để biểu diễn hình ảnh dưới dạng mảng và cho phép thực hiện các phép toán xử lý ảnh như lọc, phát hiện cạnh, phân đoạn,... ; mô phỏng hệ thống vật lý, chẳng hạn như mô phỏng chuyển động của các vật thể, mô phỏng dòng chảy của chất lỏng, v.v. ; và tính toán tài chính, ví dụ như tính toán các chỉ số tài chính, phân tích rủi ro, tối ưu hóa danh mục đầu tư,...

Tóm lại, với cộng đồng người dùng và nhà phát triển đông đảo, năng động, NumPy là một thư viện mạnh mẽ, linh hoạt, cung cấp nền tảng vững chắc cho tính toán khoa học và phân tích dữ liệu trong Python. Việc nắm vững NumPy là bước đệm thiết yếu cho bất kỳ ai muốn theo đuổi lĩnh vực khoa học dữ liệu, học máy, hay bất kỳ lĩnh vực nào liên quan đến xử lý dữ liệu số.

NumPy đóng vai trò quan trọng trong đề tài này bằng việc:

- Tạo và thao tác mảng: Dữ liệu năng lượng được lưu trữ trong các mảng NumPy, cho phép dễ dàng thực hiện các phép toán thống kê và biến đổi dữ liệu.

- Xây dựng mô hình dự báo: Các thuật toán học máy trong scikit-learn hoạt động trên dữ liệu dạng mảng NumPy, đảm bảo hiệu suất tính toán cao.

**2.3.2. Pandas.**



Hình 2. 3. Thư viện Pandas.

Pandas cung cấp cho người dùng những cấu trúc dữ liệu trực quan, dễ sử dụng như Series (dữ liệu một chiều) và DataFrame (dữ liệu hai chiều dạng bảng), cho phép thao tác dữ liệu một cách linh hoạt và hiệu quả. Giống như một chiếc bảng tính Excel thông minh, DataFrame trong Pandas cho phép người dùng dễ dàng truy cập, lọc, sắp xếp, nhóm, biến đổi và kết hợp dữ liệu từ nhiều nguồn khác nhau, bao gồm file CSV, Excel, SQL database, JSON, HTML và nhiều định dạng dữ liệu khác.

Sức mạnh của Pandas nằm ở khả năng xử lý dữ liệu một cách linh hoạt và hiệu quả. Hãy tưởng tượng bạn có một bảng dữ liệu lớn chứa thông tin về khách hàng, sản phẩm, doanh số. Với Pandas, bạn có thể dễ dàng thực hiện các tác vụ như:

- Làm sạch dữ liệu: Loại bỏ các giá trị khuyết, xử lý dữ liệu trùng lặp, chuyển đổi kiểu dữ liệu,...

- Khám phá dữ liệu: Tính toán các thống kê mô tả, trực quan hóa dữ liệu bằng biểu đồ, tìm kiếm các mẫu dữ liệu,...

- Biến đổi dữ liệu: Tạo cột mới, nhóm dữ liệu theo các tiêu chí, pivot bảng dữ liệu, kết hợp dữ liệu từ nhiều bảng,...

Pandas được xây dựng dựa trên nền tảng NumPy, tận dụng hiệu quả tính toán của mảng NumPy để tăng tốc độ xử lý dữ liệu. Nhờ đó, Pandas có thể xử lý hiệu quả các tập dữ liệu lớn, với hàng triệu dòng dữ liệu, một cách nhanh chóng và mượt mà.

Sự phổ biến của Pandas trong cộng đồng khoa học dữ liệu Python là không thể phủ nhận. Với hàng triệu lượt tải xuống mỗi tháng từ kho lưu trữ PyPI, Pandas đã trở thành công cụ đắc lực cho các nhà khoa học dữ liệu, nhà phân tích dữ liệu, nhà nghiên cứu, và bất kỳ ai muốn thao tác và phân tích dữ liệu có cấu trúc. Pandas được tích hợp với nhiều thư viện Python khác như NumPy, Scikit-learn, Matplotlib, Seaborn, tạo nên một hệ sinh thái khoa học dữ liệu Python mạnh mẽ, linh hoạt và toàn diện.

Pandas được ứng dụng rộng rãi trong nhiều lĩnh vực, từ phân tích tài chính đến nghiên cứu khoa học, từ tiếp thị đến chăm sóc sức khỏe. Trong phân tích tài chính, Pandas được sử dụng để xử lý dữ liệu thị trường chứng khoán, phân tích rủi ro đầu tư, xây dựng mô hình dự báo tài chính. Trong nghiên cứu khoa học, Pandas giúp các nhà khoa học xử lý dữ liệu thí nghiệm, phân tích kết quả nghiên cứu, trực quan hóa dữ liệu.

Trong quá trình thực hiện, Pandas là công cụ chính được sử dụng để:

- Đọc dữ liệu từ file Excel: Pandas giúp dễ dàng tải dữ liệu năng lượng từ file Excel vào DataFrame, một cấu trúc dữ liệu bảng mạnh mẽ.

- Làm sạch và chuẩn bị dữ liệu: Pandas cho phép xử lý dữ liệu bị thiếu, chuẩn hóa tên cột, chuyển đổi kiểu dữ liệu, và sắp xếp dữ liệu để chuẩn bị cho việc phân tích và dự báo.

- Thao tác và biến đổi dữ liệu: Pandas cung cấp các phương thức linh hoạt để lựa chọn dữ liệu, tạo cột mới, gộp dữ liệu, và biến đổi dữ liệu theo nhu cầu phân tích.

Pandas là một thư viện rất phổ biến trong cộng đồng Python, với hàng triệu lượt tải xuống mỗi tháng. Pandas được đánh giá cao về tính linh hoạt, hiệu quả, và khả năng xử lý dữ liệu lớn.

**2.3.3. Scikit-learn.**



Hình 2. 4. Thư viện Scikit-learn.

Scikit-learn, một thư viện Python mã nguồn mở, đóng vai trò như một công cụ cung cấp một bộ sưu tập phong phú các thuật toán học máy, giúp việc xây dựng và triển khai các mô hình học máy trở nên dễ dàng và hiệu quả hơn bao giờ hết.

Scikit-learn được xây dựng dựa trên nền tảng NumPy, SciPy và Matplotlib, mang đến cho người dùng một giao diện nhất quán, dễ sử dụng, cho phép họ dễ dàng thử nghiệm và so sánh hiệu suất của các thuật toán khác nhau. Thư viện này cung cấp một loạt các công cụ cho các tác vụ học máy phổ biến, bao gồm:

- Phân loại: Xác định xem một đối tượng thuộc về nhóm nào (ví dụ: phân loại email spam/không spam, phân loại hình ảnh).

- Hồi quy: Dự đoán một giá trị liên tục (ví dụ: dự đoán giá nhà, dự đoán doanh thu).

- Phân cụm: Gom nhóm các đối tượng có tính chất tương đồng (ví dụ: phân khúc khách hàng, phân nhóm gen).

- Giảm chiều dữ liệu: Giảm số lượng biến trong dữ liệu mà vẫn giữ được thông tin quan trọng (ví dụ: nén hình ảnh, trích xuất đặc trưng).

Scikit-learn không chỉ cung cấp các thuật toán học máy mà còn cung cấp một loạt các công cụ hữu ích khác cho việc tiền xử lý dữ liệu, lựa chọn đặc trưng, đánh giá mô hình và điều chỉnh siêu tham số, giúp người dùng xây dựng và tối ưu hóa mô hình một cách hiệu quả.

Với giao diện đơn giản, dễ sử dụng, hiệu suất cao, tài liệu hướng dẫn chi tiết và cộng đồng người dùng đông đảo, Scikit-learn đã trở thành thư viện học máy phổ biến nhất trong Python, được sử dụng rộng rãi trong cộng đồng khoa học dữ liệu, học máy và trí tuệ nhân tạo.

Các ứng dụng của Scikit-learn trong thực tế vô cùng đa dạng, trải dài trên nhiều lĩnh vực:

- Y tế: Chẩn đoán bệnh, phát hiện bất thường, cá nhân hóa điều trị.

- Tài chính: Phát hiện gian lận, dự đoán rủi ro tín dụng, phân tích thị trường chứng khoán.

- Tiếp thị: Phân khúc khách hàng, khuyến nghị sản phẩm, tối ưu hóa chiến dịch quảng cáo.

- Thương mại điện tử: Hệ thống đề xuất sản phẩm, dự đoán nhu cầu khách hàng.

Scikit-learn là một thư viện Python mạnh mẽ, dễ sử dụng và hiệu quả, là công cụ đắc lực cho bất kỳ ai muốn khám phá thế giới học máy và ứng dụng nó vào việc giải quyết các bài toán thực tế. Với Scikit-learn, việc xây dựng và triển khai các mô hình học máy trở nên dễ dàng và tiếp cận hơn bao giờ hết.

**2.3.4.Tkinter.**



Hình 2. 5. Thư viện Tkinter.

Trong việc lập trình, giao diện người dùng đồ họa (GUI) đóng vai trò như một cầu nối quan trọng, cho phép người dùng tương tác trực quan và hiệu quả với các ứng dụng. Tkinter, thư viện GUI chuẩn được tích hợp sẵn trong Python, mang đến cho người dùng một công cụ đơn giản, dễ tiếp cận để tạo ra các ứng dụng desktop với giao diện đẹp mắt và thân thiện.

Tkinter cung cấp một bộ sưu tập phong phú các widget, những khối xây dựng cơ bản cho giao diện người dùng, như nút bấm, nhãn, hộp văn bản, khung, menu, thanh cuộn, canvas, v.v., cho phép người dùng dễ dàng sắp xếp và bố trí các thành phần giao diện theo ý muốn. Việc sử dụng Tkinter rất đơn giản, người dùng chỉ cần khởi tạo các widget, thiết lập thuộc tính, bố trí chúng trên cửa sổ ứng dụng và xử lý các sự kiện tương tác từ người dùng.

Tkinter hoạt động dựa trên mô hình lập trình hướng sự kiện (event-driven programming), nơi các sự kiện như click chuột, gõ phím, di chuyển chuột sẽ kích hoạt các hàm xử lý tương ứng. Mô hình này giúp tạo ra các ứng dụng GUI phản hồi nhanh nhạy và tương tác mượt mà với người dùng.

Mặc dù là một thư viện GUI đơn giản, Tkinter vẫn có thể được sử dụng để tạo ra các ứng dụng desktop đa dạng, từ các ứng dụng đơn giản như máy tính, trình soạn thảo văn bản đến các ứng dụng phức tạp hơn như trình quản lý cơ sở dữ liệu, trò chơi đơn giản.

Ưu điểm nổi bật của Tkinter là tính đơn giản, dễ học, dễ sử dụng và được tích hợp sẵn trong Python, không yêu cầu cài đặt thêm. Điều này khiến Tkinter trở thành lựa chọn lý tưởng cho người mới bắt đầu học lập trình GUI hoặc cho các dự án nhỏ, yêu cầu giao diện đơn giản, nhanh chóng.

Tuy nhiên, Tkinter cũng có một số hạn chế. Do là một thư viện GUI "cổ điển", giao diện của các ứng dụng Tkinter thường mang phong cách truyền thống, ít tùy biến và khó sánh bằng với các framework GUI hiện đại như PyQt hay Kivy. Ngoài ra, Tkinter cũng thiếu các công cụ hỗ trợ cho việc xây dựng các ứng dụng GUI phức tạp, đòi hỏi nhiều tùy biến và hiệu ứng đồ họa.

Dù vậy, với tính đơn giản, dễ sử dụng và khả năng đáp ứng nhu cầu tạo giao diện cho các ứng dụng desktop cơ bản, Tkinter vẫn là một thư viện GUI phổ biến trong Python. Nó là lựa chọn phù hợp cho các dự án nhỏ, yêu cầu giao diện đơn giản, nhanh chóng và dễ phát triển.Việc thiết kế giao diện trong đề tài rất thích hợp với Tkinter vì vậy Tkinter được sử dụng để:

- Xây dựng giao diện chính của ứng dụng: Tkinter được dùng để tạo ra cửa sổ chính của ứng dụng, chứa các widget cho phép người dùng lựa chọn quốc gia, lĩnh vực, và hiển thị các biểu đồ.

- Tạo các widget tương tác: Tkinter cung cấp các widget như nút, hộp danh sách, và menu, cho phép người dùng tương tác với ứng dụng, lựa chọn dữ liệu, và thực hiện các hành động khác.

Tkinter là thư viện phổ biến, thường được sử dụng cho các ứng dụng GUI đơn giản trong Python. Mặc dù không mạnh mẽ như các framework GUI khác, Tkinter vẫn là lựa chọn hiệu quả cho các ứng dụng nhỏ và dễ phát triển.

**2.3.5. Pillow.**



Hình 2. 6. Thư viện Pillow.

Pillow, còn được biết đến là "PIL fork," là một thư viện Python mạnh mẽ và phổ biến được sử dụng để xử lý hình ảnh. Được xây dựng dựa trên thư viện PIL (Python Imaging Library) ban đầu, Pillow cung cấp các tính năng mở rộng và hỗ trợ nhiều định dạng hình ảnh hơn, bao gồm JPEG, PNG, GIF, BMP, TIFF, WebP, và nhiều định dạng khác.

Pillow cung cấp một loạt các công cụ cho phép thao tác hình ảnh ở nhiều cấp độ khác nhau. Người dùng có thể sử dụng Pillow để:

- Mở và hiển thị hình ảnh: Pillow cho phép mở hình ảnh từ nhiều nguồn khác nhau, bao gồm file, URL, và thậm chí từ bộ nhớ. Nó cũng cung cấp các hàm để hiển thị hình ảnh trong các ứng dụng GUI.

- Thay đổi kích thước và cắt hình ảnh: Pillow cung cấp các hàm để thay đổi kích thước, cắt, xoay, và lật hình ảnh, cho phép người dùng điều chỉnh hình ảnh theo ý muốn.

- Chỉnh sửa màu sắc và độ tương phản: Pillow cho phép thay đổi độ sáng, độ tương phản, độ bão hòa, và các thuộc tính màu sắc khác của hình ảnh, cho phép người dùng tinh chỉnh hình ảnh theo nhu cầu.

- Áp dụng bộ lọc và hiệu ứng: Pillow hỗ trợ nhiều bộ lọc và hiệu ứng hình ảnh, chẳng hạn như làm mờ, làm sắc nét, dập nổi, và phát hiện cạnh.

- Vẽ hình học và văn bản: Pillow cho phép vẽ các hình dạng hình học như đường thẳng, hình chữ nhật, hình tròn, và văn bản lên hình ảnh, cho phép người dùng tạo ra các hình ảnh mới hoặc chú thích hình ảnh hiện có.

- Xử lý pixel: Pillow cho phép truy cập và thao tác từng pixel của hình ảnh, cho phép thực hiện các tác vụ xử lý hình ảnh phức tạp.

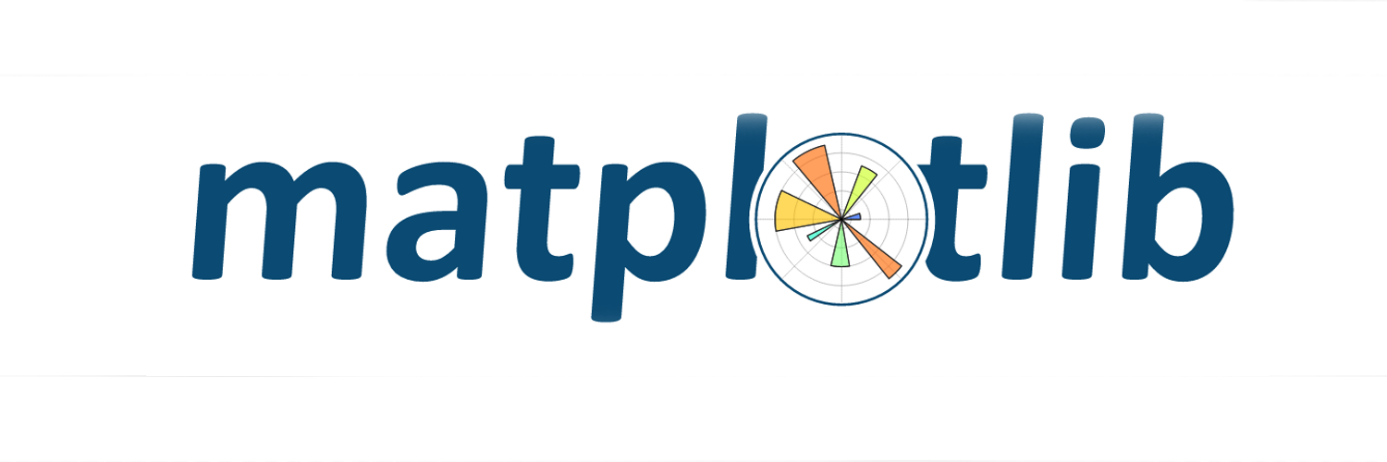
Trong đề tài này, Pillow được sử dụng để:

- Xử lý biểu tượng trong giao diện: Pillow được sử dụng để tải và hiển thị các biểu tượng (icon) trong giao diện người dùng, góp phần tạo nên một giao diện trực quan và hấp dẫn hơn cho người dùng.

- Tạo và chỉnh sửa biểu đồ: Mặc dù Matplotlib là công cụ chính để tạo biểu đồ, Pillow có thể được sử dụng để chỉnh sửa và tinh chỉnh các biểu đồ được tạo ra bởi Matplotlib, chẳng hạn như thêm logo, watermark, hoặc điều chỉnh kích thước và định dạng hình ảnh trước khi lưu.

Pillow là một thư viện xử lý hình ảnh vô cùng phổ biến trong Python, được đánh giá cao về tính linh hoạt, hiệu suất, và khả năng hỗ trợ nhiều định dạng hình ảnh. Nó là công cụ lý tưởng cho mọi tác vụ xử lý hình ảnh trong Python, từ đơn giản đến phức tạp.

**2.3.6. Matplotlib.**



Hình 2. 7. Thư viện Matpolotlib.

Matplotlib là thư viện Python toàn diện cho việc tạo hình ảnh và trực quan hóa dữ liệu. Nó cung cấp một hệ thống cực kỳ linh hoạt cho việc tạo ra nhiều loại biểu đồ chất lượng cao, bao gồm biểu đồ đường, biểu đồ cột, biểu đồ phân tán, biểu đồ histogram, biểu đồ hộp, biểu đồ bánh, biểu đồ heatmap, biểu đồ contour, và nhiều loại biểu đồ khác.

Matplotlib cho phép người dùng tùy chỉnh mọi khía cạnh của biểu đồ, từ màu sắc, kiểu dáng, nhãn, font chữ, đến vị trí, kích thước, legend, và chú thích. Nó cung cấp khả năng kiểm soát chi tiết về mọi yếu tố của biểu đồ, cho phép người dùng tạo ra những hình ảnh trực quan hóa dữ liệu chính xác và đẹp mắt.

Ngoài việc tạo ra các biểu đồ tĩnh, Matplotlib cũng hỗ trợ tạo ra các biểu đồ động và tương tác, cho phép người dùng tương tác với dữ liệu thông qua các thao tác như zoom, pan, và hover.

Trong lập trình Python, Matplotlib được sử dụng rộng rãi trong các lĩnh vực khoa học dữ liệu, phân tích dữ liệu, học máy, và bất kỳ lĩnh vực nào yêu cầu trực quan hóa dữ liệu. Nó là công cụ lý tưởng cho việc:

- Người dùng dễ dàng hình dung dữ liệu, khám phá các xu hướng, và phát hiện các mẫu ẩn trong dữ liệu.

- Cho phép người dùng tạo ra các biểu đồ trực quan và dễ hiểu để truyền đạt thông tin một cách hiệu quả.

- Xuất bản các biểu đồ ở nhiều định dạng khác nhau, bao gồm PNG, JPG, PDF, và SVG, cho phép người dùng dễ dàng chia sẻ và trình bày kết quả phân tích.

Trong ứng dụng, Matplotlib đóng vai trò quan trọng trong việc:

- Sử dụng để tạo ra các biểu đồ trực quan hóa dữ liệu năng lượng theo thời gian, cho phép người dùng dễ dàng quan sát, phân tích và so sánh các xu hướng sử dụng năng lượng trong quá khứ và dự báo cho tương lai.

- Tùy chỉnh mọi khía cạnh của biểu đồ, chẳng hạn như màu sắc, nhãn, tiêu đề, chú thích, và nhiều hơn nữa, để tạo ra các biểu đồ trực quan, dễ hiểu và phù hợp với mục đích trình bày.

Matplotlib là thư viện trực quan hóa dữ liệu phổ biến nhất trong Python, được đánh giá cao về tính linh hoạt, hiệu suất, và khả năng tùy biến cao. Nó là công cụ không thể thiếu cho bất kỳ nhà khoa học dữ liệu, nhà phân tích dữ liệu, hoặc nhà nghiên cứu nào sử dụng Python.

**2.3.7. Mplcursors.**

Matplotlib, thư viện trực quan hóa dữ liệu chủ lực trong Python, cho phép tạo ra các biểu đồ chất lượng cao với khả năng tùy biến đáng kinh ngạc. Tuy nhiên, biểu đồ Matplotlib mặc định thường là tĩnh, thiếu khả năng tương tác trực tiếp với dữ liệu. mplcursors ra đời như một giải pháp cho vấn đề này, mang đến một lớp phủ tương tác thông minh, giúp người dùng dễ dàng khám phá và hiểu rõ hơn về dữ liệu được hiển thị trên biểu đồ.

mplcursors là một thư viện Python mã nguồn mở, hoạt động như một tiện ích mở rộng cho Matplotlib. Nó cho phép người dùng thêm các chú thích (annotation) động vào biểu đồ, hiển thị thông tin chi tiết về các điểm dữ liệu khi di chuyển chuột qua. Chỉ với vài dòng code đơn giản, mplcursors có thể biến những biểu đồ Matplotlib tĩnh thành những công cụ tương tác trực quan, cho phép người dùng "zoom" vào từng điểm dữ liệu và khám phá thông tin ẩn giấu bên trong.

Sức mạnh của mplcursors nằm ở sự đơn giản và hiệu quả. Người dùng không cần phải viết những đoạn code phức tạp để xử lý sự kiện chuột hay cập nhật chú thích. mplcursors tự động phát hiện vị trí con trỏ chuột trên biểu đồ và hiển thị chú thích tương ứng với điểm dữ liệu gần nhất. Người dùng có thể tùy chỉnh nội dung, vị trí và định dạng của chú thích, cũng như cách thức hiển thị và ẩn chú thích.

mplcursors mang đến nhiều lợi ích cho việc phân tích và khám phá dữ liệu.

* + - Tương tác trực quan: Cho phép người dùng dễ dàng "drill down" vào từng điểm dữ liệu, xem thông tin chi tiết mà không cần phải tìm kiếm thủ công.
    - Khám phá dữ liệu: Giúp người dùng phát hiện các mẫu dữ liệu, xu hướng, giá trị ngoại lệ, và các đặc điểm thú vị khác mà có thể bị bỏ qua trong biểu đồ tĩnh.
    - Trình bày dữ liệu: Biểu đồ tương tác với mplcursors trở nên sinh động và hấp dẫn hơn, thu hút sự chú ý của người xem và giúp truyền tải thông tin hiệu quả hơn.

Mặc dù không được tích hợp sẵn trong Matplotlib, mplcursors rất dễ cài đặt và sử dụng. Thư viện này tương thích với hầu hết các loại biểu đồ Matplotlib, bao gồm biểu đồ đường, biểu đồ cột, biểu đồ phân tán, biểu đồ heatmap, v.v.

**2.4. Pycharm.**

**2.4.1. IDE Pycharm.**



Hình 2. 8. IDE Pycharm.

PyCharm là một Môi trường Phát triển Tích hợp (IDE) mạnh mẽ được thiết kế dành riêng cho ngôn ngữ lập trình Python. Được phát triển bởi JetBrains, PyCharm cung cấp một bộ công cụ toàn diện giúp đơn giản hóa và nâng cao hiệu quả quá trình phát triển ứng dụng Python.

**2.4.2. Các tính năng nổi bật của Pycharm.**

PyCharm sở hữu nhiều tính năng hữu ích hỗ trợ lập trình viên Python:

- Hoàn thiện mã thông minh: PyCharm phân tích mã nguồn và đề xuất các lựa chọn hoàn thiện mã, giúp người dùng viết code nhanh chóng và chính xác hơn.

- Kiểm tra lỗi và gợi ý sửa lỗi: IDE này tự động kiểm tra lỗi cú pháp, lỗi logic và đưa ra gợi ý để sửa lỗi hiệu quả.

- Điều hướng mã nguồn: PyCharm cung cấp các công cụ điều hướng mạnh mẽ, giúp người dùng dễ dàng di chuyển giữa các tệp, lớp, hàm và biến trong dự án.

- Tích hợp gỡ lỗi: PyCharm tích hợp trình gỡ lỗi cho phép người dùng thực thi mã từng bước, kiểm tra giá trị biến và phát hiện lỗi trong thời gian thực.

- Hỗ trợ kiểm thử: IDE này tích hợp với các framework kiểm thử phổ biến như unittest và pytest, giúp người dùng viết và chạy test dễ dàng.

- Quản lý phiên bản: PyCharm tích hợp với các hệ thống quản lý phiên bản như Git, cho phép người dùng theo dõi thay đổi mã nguồn, cộng tác với người khác và quản lý các phiên bản của dự án.

- Hỗ trợ các framework và thư viện Python: PyCharm hỗ trợ nhiều framework và thư viện Python phổ biến như Django, Flask, NumPy, Pandas, Matplotlib, và nhiều hơn nữa.

- Giao diện tùy biến: Người dùng có thể tùy chỉnh giao diện, phím tắt và các thiết lập khác của PyCharm để phù hợp với nhu cầu và thói quen làm việc của mình.

**2.4.3. Lựa chọn Pycharm cho đề tài.**

PyCharm được lựa chọn làm môi trường phát triển cho dự án này vì các lý do sau:

- Tăng cường năng suất: Các tính năng như hoàn thiện mã thông minh, kiểm tra lỗi, điều hướng mã nguồn, gỡ lỗi giúp người dùng viết code nhanh chóng và hiệu quả hơn.

- Giảm thiểu lỗi: Khả năng kiểm tra lỗi và gợi ý sửa lỗi của PyCharm giúp phát hiện và khắc phục lỗi sớm trong quá trình phát triển, nâng cao chất lượng mã nguồn.

- Hỗ trợ dự án Python: PyCharm cung cấp hỗ trợ chuyên sâu cho ngôn ngữ Python, tích hợp với các framework và thư viện phổ biến, giúp quá trình phát triển ứng dụng Python trở nên dễ dàng và thuận tiện hơn.

Với những ưu điểm này, PyCharm là một IDE lý tưởng để phát triển dự án phân tích và dự báo mức sử dụng năng lượng, đảm bảo quá trình phát triển diễn ra suôn sẻ và hiệu quả.

**2.5.Các mô hình máy học.**

### 2.5.1. ARMIMA (Autoregressive Integrated Moving Average).

#### **2.5.1.1. Giới thiệu về ARIMA.**

Mô hình ARIMA, viết tắt của Autoregressive Integrated Moving Average, là một phương pháp thống kê mạnh mẽ và phổ biến được sử dụng để phân tích và dự báo dữ liệu chuỗi thời gian. Nó dựa trên ý tưởng rằng các giá trị trong tương lai của một chuỗi thời gian có thể được dự đoán dựa trên các giá trị trong quá khứ của chính chuỗi thời gian đó và các giá trị sai số ngẫu nhiên trong quá khứ.

#### **2.5.1.2. Định nghĩa mô hình ARIMA.**

Mô hình ARIMA là một mô hình tuyến tính tổng quát hóa của mô hình Autoregressive Moving Average (ARMA). Nó kết hợp ba thành phần chính:

* + - Autoregressive (AR): Thành phần này sử dụng mối tương quan giữa các giá trị hiện tại và các giá trị trễ trong quá khứ của chuỗi thời gian.
    - Integrated (I): Thành phần này liên quan đến việc sử dụng sai phân của chuỗi thời gian để làm cho chuỗi thời gian trở nên ổn định.
    - Moving Average (MA): Thành phần này sử dụng mối tương quan giữa các giá trị hiện tại và các giá trị sai số ngẫu nhiên trong quá khứ của chuỗi thời gian.

#### **2.5.1.3. Lý do sử dụng ARIMA trong dự báo chuỗi thời gian.**

ARIMA được sử dụng rộng rãi trong dự báo chuỗi thời gian bởi vì:

* Khả năng xử lý chuỗi thời gian phức tạp: ARIMA có thể mô hình hóa các chuỗi thời gian có xu hướng, chu kỳ và tính mùa vụ.
* Dễ dàng kiểm tra và điều chỉnh: Các thành phần AR, I và MA có thể được kiểm tra và điều chỉnh độc lập để tối ưu hóa mô hình.
* Cung cấp dự báo chính xác: ARIMA có thể cung cấp dự báo chính xác cho nhiều loại chuỗi thời gian.

#### **2.5.1.4. Cấu trúc mô hình ARIMA**

Mô hình ARIMA được biểu diễn bằng ký hiệu ARIMA(p, d, q), trong đó:

* p: Bậc của thành phần Autoregressive (AR).
* d: Bậc của thành phần Integrated (I).
* q: Bậc của thành phần Moving Average (MA).

#### **2.5.1.5. Giải thích các thành phần**

Autoregressive (AR): Thành phần này giả định rằng giá trị hiện tại của chuỗi thời gian phụ thuộc tuyến tính vào các giá trị trễ trong quá khứ của nó.

Integrated (I): Thành phần này liên quan đến việc sử dụng sai phân của chuỗi thời gian. Sai phân bậc d của một chuỗi thời gian là chuỗi thời gian được tạo bằng cách lấy hiệu số giữa các giá trị hiện tại và các giá trị trễ d giai đoạn. Mục đích của sai phân là làm cho chuỗi thời gian trở nên ổn định, tức là loại bỏ xu hướng và tính mùa vụ.

Moving Average (MA): Thành phần này giả định rằng giá trị hiện tại của chuỗi thời gian phụ thuộc tuyến tính vào các giá trị sai số ngẫu nhiên trong quá khứ của nó.

#### **2.5.1.6. Công thức toán học của ARIMA**

Công thức toán học của mô hình ARIMA(p, d, q) có dạng:

y\_t = c + φ\_1 \* y\_{t-1} + ... + φ\_p \* y\_{t-p} + θ\_1 \* ε\_{t-1} + ... + θ\_q \* ε\_{t-q} + ε\_t

Trong đó:

* y\_t là giá trị của chuỗi thời gian tại thời điểm t
* c là hằng số
* φ\_1, ..., φ\_p là các tham số của thành phần AR
* θ\_1, ..., θ\_q là các tham số của thành phần MA
* ε\_t là giá trị sai số ngẫu nhiên tại thời điểm t

#### **2.5.1.7. Quy trình xây dựng mô hình ARIMA.**

Bước 1: Kiểm tra tính ổn định của chuỗi thời gian: Sử dụng các biểu đồ chuỗi thời gian và các kiểm định thống kê (ví dụ: kiểm định Dickey-Fuller) để đánh giá tính ổn định của chuỗi thời gian.

Bước 2: Sử dụng phép biến đổi để làm chuỗi thời gian ổn định (differencing): Nếu chuỗi thời gian không ổn định, áp dụng phép sai phân để loại bỏ xu hướng và tính mùa vụ.

Bước 3: Xác định các giá trị p, d, q cho mô hình: Sử dụng các hàm tự tương quan (ACF) và hàm tự tương quan riêng phần (PACF) để xác định bậc của các thành phần AR, I và MA.

Bước 4: Ước lượng các tham số của mô hình: Sử dụng các phương pháp ước lượng thống kê (ví dụ: phương pháp khả năng cực đại) để ước lượng các tham số của mô hình.

Bước 5: Kiểm tra tính phù hợp của mô hình: Sử dụng các kiểm định thống kê và phân tích phần dư để đánh giá tính phù hợp của mô hình.

Bước 6: Sử dụng mô hình để dự báo: Sau khi mô hình được xây dựng và kiểm tra, sử dụng mô hình để dự báo các giá trị trong tương lai của chuỗi thời gian.

#### **2.5.1.8. Ưu điểm và nhược điểm**

Ưu điểm:

* + - Khả năng xử lý chuỗi thời gian phức tạp.
    - Dễ dàng kiểm tra và điều chỉnh các thành phần.

Nhược điểm:

* + - Yêu cầu kiến thức chuyên sâu về chuỗi thời gian.
    - Cần nhiều bước kiểm tra và điều chỉnh.

#### **2.5.1.9. Ứng dụng của ARIMA.**

Trong thực tế:

* Tài chính: Dự báo giá cổ phiếu, lãi suất.
* Kinh tế: Dự báo GDP, lạm phát.
* Dự báo sản lượng, nhu cầu năng lượng, lượng mưa,...

Trong đề tài này:

ARIMA được thử nghiệm như một phương pháp tiềm năng để dự báo mức sử dụng năng lượng. Tuy nhiên, sau khi thử nghiệm độ sai lệch và chỉ số RMSE quá lớn nên ARIMA không được lựa chọn là mô hình cuối cùng cho dự án.

### 2.5.2. Prophet.



Hình 2. 9. Mô hình Prophet.

#### **2.5.2.1. Giới thiệu về Prophet.**

Prophet là một mô hình dự báo chuỗi thời gian được phát triển bởi Facebook, được thiết kế để xử lý dữ liệu chuỗi thời gian có tính mùa vụ cao và chịu ảnh hưởng bởi các ngày lễ. Mô hình này được xây dựng dựa trên ý tưởng phân tách chuỗi thời gian thành các thành phần riêng biệt, bao gồm xu hướng, tính mùa vụ và các tác động của ngày lễ, sau đó kết hợp các thành phần này để tạo ra dự báo.

#### **2.5.2.2. Định nghĩa mô hình Prophet.**

Prophet là một mô hình hồi quy cộng, trong đó chuỗi thời gian được biểu diễn bằng tổng của ba thành phần chính:

- Xu hướng (trend): Biểu diễn sự thay đổi theo thời gian của chuỗi thời gian, thường được mô hình hóa bằng hàm tuyến tính hoặc hàm logistic.

- Mùa vụ (seasonality): Biểu diễn các chu kỳ lặp đi lặp lại trong chuỗi thời gian, ví dụ như chu kỳ ngày, tuần, tháng hoặc năm. Prophet sử dụng chuỗi Fourier để mô hình hóa tính mùa vụ.

- Ngày lễ (holidays): Biểu diễn tác động của các ngày lễ lên chuỗi thời gian. Prophet cho phép người dùng xác định danh sách các ngày lễ và mô hình hóa tác động của chúng.

#### **2.5.2.3. Mục đích và lý do sử dụng.**

Prophet được phát triển bởi Facebook để giải quyết các vấn đề dự báo trong các ứng dụng thực tế của họ. Mô hình này được sử dụng để dự báo các chỉ số như lượt truy cập website, doanh thu bán hàng, lượt tương tác trên mạng xã hội, v.v.

Lý do Prophet được sử dụng rộng rãi trong dự báo chuỗi thời gian:

- Dễ sử dụng: Prophet được thiết kế với API đơn giản, dễ sử dụng, không yêu cầu người dùng có kiến thức chuyên sâu về chuỗi thời gian.

- Xử lý tốt tính mùa vụ và ngày lễ: Prophet có khả năng mô hình hóa hiệu quả các chuỗi thời gian có yếu tố mùa vụ phức tạp và chịu ảnh hưởng bởi các ngày lễ.

- Khả năng tùy chỉnh cao: Prophet cho phép người dùng điều chỉnh nhiều tham số để tối ưu hóa mô hình cho từng trường hợp cụ thể.

#### **2.5.2.4. Cấu trúc mô hình.**

Cấu trúc mô hình Prophet có dạng:

y(t) = g(t) + s(t) + h(t) + epsilon(t)

Trong đó:

- y(t): Giá trị của chuỗi thời gian tại thời điểm t.

- g(t): Thành phần xu hướng tại thời điểm t.

- s(t): Thành phần mùa vụ tại thời điểm t.

- h(t): Thành phần ngày lễ tại thời điểm t.

- epsilon(t): Sai số ngẫu nhiên tại thời điểm t.

#### **2.5.2.5. Quy trình xây dựng mô hình.**

- Bước 1: Chuẩn bị dữ liệu đầu vào: Dữ liệu đầu vào của Prophet cần có hai cột: "ds" (thời gian) và "y" (giá trị chuỗi thời gian).

- Bước 2: Xác định các tham số của mô hình: Người dùng có thể điều chỉnh nhiều tham số của mô hình, bao gồm loại hàm xu hướng, chu kỳ mùa vụ, danh sách ngày lễ, v.v.

- Bước 3: Huấn luyện mô hình với dữ liệu lịch sử: Prophet sử dụng phương pháp ước lượng khả năng cực đại để tìm các tham số tối ưu cho mô hình.

- Bước 4: Tối ưu hóa và điều chỉnh mô hình: Người dùng có thể đánh giá hiệu suất của mô hình và điều chỉnh các tham số để cải thiện độ chính xác.

- Bước 5: Sử dụng mô hình để dự báo: Sau khi mô hình được xây dựng và tối ưu hóa, Prophet có thể được sử dụng để dự báo các giá trị trong tương lai của chuỗi thời gian.

#### **2.5.2.6. Ưu điểm và nhược điểm**

Ưu điểm:

- Dễ sử dụng, không yêu cầu kiến thức chuyên sâu về chuỗi thời gian.

- Xử lý tốt các chuỗi thời gian có yếu tố mùa vụ và các ngày lễ.

- Khả năng tùy chỉnh cao.

Nhược điểm:

- Không phù hợp với tất cả các loại chuỗi thời gian.

- Có thể không chính xác nếu dữ liệu có nhiều biến động không theo quy luật.

#### **2.5.2.7. Ứng dụng của Prophet**

- Marketing: Dự báo doanh số bán hàng, lượt truy cập website, hiệu quả chiến dịch quảng cáo.

- Logistics: Dự báo nhu cầu sản phẩm, tối ưu hóa chuỗi cung ứng.

- Tài chính: Dự báo giá cổ phiếu, lãi suất.

Trong đề tài này: Prophet được thử nghiệm như một phương pháp hiệu quả để dự báo mức sử dụng năng lượng. Sau khi thử nghiệm kết quả dự báo và chỉ số RMSE rơi vào khoảng 18 – 25% giá tri dao động của dữ liệu, đây là mức có thể chấp nhận được. Mô hình này được lưu trữ và tiếp tục tiến hành so sánh với các mô hình khác. Tuy nhiên, khi so sánh với các mô hình khác, Prophet không được lựa chọn là mô hình cuối cùng cho dự án.

### 2.5.3. Moving Average.

#### **2.5.3.1. Định nghĩa và nguyên lý hoạt động.**

- Mô hình Moving Average (MA), hay còn gọi là mô hình Trung bình trượt, là một phương pháp dự báo chuỗi thời gian dựa trên ý tưởng sử dụng trung bình của một số giá trị quá khứ nhất định để dự đoán giá trị tương lai.

- Mô hình MA hoạt động dựa trên nguyên tắc rằng giá trị hiện tại của chuỗi thời gian bị ảnh hưởng bởi các nhiễu ngẫu nhiên trong quá khứ. Bằng cách tính trung bình của các nhiễu này, mô hình MA có thể làm mượt các biến động ngẫu nhiên và đưa ra dự báo cho giá trị tương lai.

#### **2.5.3.2. Cấu trúc mô hình.**

Thành phần chính của mô hình MA là bậc của mô hình (q), đại diện cho số lượng giá trị quá khứ được sử dụng để tính trung bình. Mô hình MA bậc q được ký hiệu là MA(q).

Công thức toán học của mô hình MA(q):

*yt*​=*c*+*θ*1​∗*εt*−1​+*θ*2​∗*εt*−2​+...+*θq*​∗*εt*−*q*​+*εt*

Trong đó:

* *yt*​: Giá trị dự báo của chuỗi thời gian tại thời điểm t
* 𝑐*c*: Hằng số
* 𝜃1,𝜃2,...,𝜃𝑞*θ*1​,*θ*2​,...,*θq*​: Các tham số của mô hình MA(q)
* 𝜀𝑡−1,𝜀𝑡−2,...,𝜀𝑡−𝑞*εt*−1​,*εt*−2​,...,*εt*−*q*​: Sai số ngẫu nhiên tại các thời điểm t-1, t-2, ..., t-q
* *εt*​: Sai số ngẫu nhiên tại thời điểm t

#### **2.5.3.3. Quy trình xây dựng mô hình.**

- Bước 1: Chọn và đánh giá chuỗi thời gian cần dự báo: Xác định chuỗi thời gian cần dự báo và đánh giá các đặc điểm của nó, bao gồm xu hướng, tính mùa vụ và tính ngẫu nhiên.

- Bước 2: Xác định và chọn lựa các tham số cho mô hình MA: Xác định bậc của mô hình MA(q) dựa trên hàm tự tương quan (ACF) và hàm tự tương quan riêng phần (PACF).

- Bước 3: Ước lượng và tính toán dự báo với mô hình MA: Sử dụng các phương pháp ước lượng thống kê để ước lượng các tham số của mô hình MA và tính toán giá trị dự báo.

- Bước 4: Kiểm tra và đánh giá hiệu suất của mô hình MA: Sử dụng các chỉ số đánh giá hiệu suất để đánh giá độ chính xác của mô hình MA.

#### **2.5.3.4. Ưu điểm.**

- Dự báo tốt cho chuỗi thời gian ổn định: Mô hình MA hoạt động tốt với các chuỗi thời gian có xu hướng biến động đều và không có tính mùa vụ rõ rệt.

- Đơn giản và dễ hiểu: Cấu trúc và nguyên lý hoạt động của mô hình MA tương đối đơn giản, dễ hiểu và triển khai, ít yêu cầu kiến thức chuyên sâu về chuỗi thời gian.

#### **2.5.3.5. Nhược điểm.**

- Không phù hợp với chuỗi thời gian phức tạp: Mô hình MA không thể xử lý hiệu quả các chuỗi thời gian có xu hướng biến động không đều, có tính mùa vụ hoặc chứa các giá trị ngoại lệ.

- Dễ bị ảnh hưởng bởi ngoại lệ: Các giá trị ngoại lệ trong dữ liệu có thể làm sai lệch kết quả dự báo của mô hình MA.

#### **2.5.3.6. Ứng dụng.**

- Dự báo doanh số bán hàng: Dự đoán doanh số bán hàng trong tương lai dựa trên doanh số của các kỳ trước đó.

- Dự báo sản lượng: Dự đoán sản lượng của một nhà máy hoặc một quy trình sản xuất dựa trên sản lượng của các kỳ trước đó.

- Dự báo giá cổ phiếu: Dự đoán giá cổ phiếu trong tương lai dựa trên giá cổ phiếu của các ngày giao dịch trước đó.

- Trong đề tài này: MA được thử nghiệm như một mô hình để dự báo mức sử dụng năng lượng. Sau khi thử nghiệm kết quả dự báo và chỉ số RMSE rơi vào khoảng 15 – 20% giá tri dao động của dữ liệu, đây là mức trung bình và có thể được sự dụng để áp dụng cho đề tài và trực quan hóa dữ liệu sau này.Mô hình này được lưu trữ và tiếp tục tiến hành so sánh với các mô hình khác. Tuy nhiên, khi so sánh với các mô hình khác, MA không được lựa chọn là mô hình cuối cùng cho dự án.

### 2.5.4. Linear Regression.

#### **2.5.4.1. Giới thiệu.**

Trong bối cảnh thế giới đang đối mặt với những thách thức về biến đổi khí hậu và khủng hoảng năng lượng, việc phân tích và dự báo mức sử dụng năng lượng trở nên vô cùng quan trọng. Dự báo chính xác về nhu cầu năng lượng trong tương lai giúp các chính phủ, doanh nghiệp, và nhà nghiên cứu hoạch định chính sách năng lượng hiệu quả, tối ưu hóa sản xuất và tiêu thụ năng lượng, và hướng đến một tương lai bền vững hơn.

Hồi quy tuyến tính là một mô hình dự đoán đơn giản nhưng mạnh mẽ, được sử dụng rộng rãi trong nhiều lĩnh vực, bao gồm dự báo mức sử dụng năng lượng. Mô hình này giả định một mối quan hệ tuyến tính giữa biến phụ thuộc (biến cần dự đoán) và một hoặc nhiều biến độc lập (biến đặc trưng). Trong dự án phân tích và dự báo mức sử dụng năng lượng, chúng ta sử dụng hồi quy tuyến tính để dự đoán mức tiêu thụ năng lượng trong 5 năm tiếp theo (2022-2027) cho các quốc gia khác nhau, dựa trên dữ liệu lịch sử về mức tiêu thụ năng lượng từ năm 2000 đến 2021.

#### **2.5.4.2. Công thức và Giải thích.**

Công thức chung của mô hình hồi quy tuyến tính đơn biến:

y = β0 + β1x + ε

Trong đó:

* y: Biến phụ thuộc, đại diện cho mức tiêu thụ năng lượng cho một loại năng lượng cụ thể (ví dụ: Nhiên liệu sinh học và chất thải, Than đá và sản phẩm than, Điện,...).
* x: Biến độc lập, đại diện cho năm.
* β0: Hệ số chặn (intercept), đại diện cho giá trị của y khi x bằng 0. Trong trường hợp này, β0 thể hiện mức tiêu thụ năng lượng của loại năng lượng đó vào năm 2000 (năm gốc).
* β1: Hệ số góc (slope), đại diện cho mức độ thay đổi của y khi x thay đổi 1 đơn vị. Nói cách khác, β1 thể hiện mức tăng trưởng trung bình hàng năm của mức tiêu thụ năng lượng cho loại năng lượng đó.
* ε: Sai số ngẫu nhiên, đại diện cho phần dư của mô hình, tức là phần không thể giải thích được bởi mô hình. Sai số này có thể xuất phát từ nhiều yếu tố, ví dụ như sai số đo lường, sự kiện bất thường, hay các yếu tố khác chưa được xem xét trong mô hình.

Công thức chung của mô hình hồi quy tuyến tính đa biến:

y = β0 + β1x1 + β2x2 + ... + βnxn + ε

Công thức đa biến cho phép chúng ta xem xét nhiều yếu tố ảnh hưởng đến mức tiêu thụ năng lượng, ví dụ như dân số, GDP, giá năng lượng,...

#### **2.5.4.3. Phương pháp ước lượng.**

Để tìm ra đường thẳng (hoặc siêu phẳng) phù hợp nhất với dữ liệu, chúng ta sử dụng phương pháp bình phương bé nhất (OLS - Ordinary Least Squares). OLS tìm cách tối thiểu hóa tổng bình phương sai số giữa giá trị dự đoán của mô hình và giá trị thực tế của dữ liệu.

Công thức OLS:

β = (XTX)-1XTy

Trong đó:

* β là vector chứa các hệ số hồi quy (β0, β1,...).
* X là ma trận chứa các giá trị của biến độc lập.
* y là vector chứa các giá trị của biến phụ thuộc.

#### **2.5.4.4. Áp dụng Hồi quy Tuyến tính trong Dự án.**

Trong dự án phân tích và dự báo mức sử dụng năng lượng, chúng ta áp dụng mô hình hồi quy tuyến tính đơn biến cho mỗi loại năng lượng và cho mỗi quốc gia. Dữ liệu lịch sử về mức tiêu thụ năng lượng được sử dụng để huấn luyện mô hình. Sau khi huấn luyện, mô hình được sử dụng để dự đoán mức tiêu thụ năng lượng cho 5 năm tiếp theo (2022-2027).

Trong đề tài Linear Regression được thử nghiệm với mong muốn đạt được mức đột phá về độ chính xác đối với bộ dữ liệu của đề trong việc dự báo mức sử dụng năng lượng. Sau khi thử nghiệm kết quả dự báo và chỉ số RMSE rơi vào khoảng dưới 10% giá tri dao động của dữ liệu và khoảng 15% với mức độ giá trị dữ liệu có dao động cao đây là mức vô cùng tốt và tốt nhất so với những mô hình đã được thử nghiệm. Vì vậy, mô hình này đã được áp dụng vào đề tài để tiến hành tiếp tục dự báo và trực quan hóa kết quả dự báo sau khi hoàn thành.

#### **2.5.4.5. Đánh giá Mô hình.**

Để đánh giá hiệu suất của mô hình hồi quy tuyến tính, chúng ta sử dụng các chỉ số như:

Hệ số xác định (R-squared): R-squared đo lường mức độ phù hợp của mô hình với dữ liệu, tức là tỷ lệ biến động của biến phụ thuộc được giải thích bởi mô hình. Giá trị R-squared nằm trong khoảng từ 0 đến 1, giá trị càng cao thì mô hình càng phù hợp với dữ liệu.

Sai số bình phương trung bình gốc (RMSE): RMSE đo lường trung bình sai số dự đoán của mô hình. RMSE càng thấp thì mô hình càng dự đoán chính xác.

Ngoài ra, chúng ta cũng thực hiện kiểm định giả thuyết để kiểm tra ý nghĩa thống kê của các hệ số hồi quy.

#### **2.5.4.6. Ưu nhược điểm.**

Ưu điểm:

* Dễ hiểu và dễ cài đặt: Hồi quy tuyến tính là một mô hình tương đối đơn giản, dễ hiểu và dễ cài đặt.
* Hiệu quả tính toán cao: Việc huấn luyện và sử dụng mô hình hồi quy tuyến tính không yêu cầu nhiều tài nguyên tính toán.
* Có thể diễn giải được: Các hệ số hồi quy có ý nghĩa thực tế rõ ràng, giúp chúng ta hiểu được mối quan hệ giữa biến phụ thuộc và biến độc lập.

Nhược điểm:

* Giả định tuyến tính: Mô hình hồi quy tuyến tính giả định một mối quan hệ tuyến tính giữa biến phụ thuộc và biến độc lập. Trong thực tế, mối quan hệ này có thể phức tạp hơn, không phải lúc nào cũng tuyến tính.
* Nhạy cảm với dữ liệu nhiễu: Dữ liệu năng lượng thường chứa nhiễu do nhiều yếu tố, ví dụ như sai số đo lường, sự kiện bất thường,... Điều này có thể ảnh hưởng đến độ chính xác của mô hình.
* Overfitting: Mô hình hồi quy tuyến tính có thể bị overfitting, tức là mô hình quá khớp với dữ liệu huấn luyện, dẫn đến dự đoán kém chính xác cho dữ liệu mới.

#### **2.5.4.7. Kết luận.**

Hồi quy tuyến tính là một mô hình đơn giản nhưng hiệu quả, được sử dụng rộng rãi trong dự báo mức sử dụng năng lượng. Tuy nhiên, để đạt được kết quả dự báo chính xác và đáng tin cậy, cần phải xem xét kỹ lưỡng các giả định và hạn chế của mô hình, cũng như lựa chọn dữ liệu huấn luyện phù hợp và đánh giá hiệu quả của mô hình một cách cẩn thận.

Ngoài ra, để nâng cao độ chính xác của mô hình, có thể xem xét sử dụng các phương pháp hồi quy phi tuyến, kết hợp với các kỹ thuật xử lý dữ liệu nâng cao để loại bỏ nhiễu và xử lý dữ liệu bị thiếu.

**CHƯƠNG III. CÀI ĐẶT THỰC NGHIỆM.**

**3.1. Mô tả bài toán.**

Trong bối cảnh biến đổi khí hậu và khủng hoảng năng lượng ngày càng trở nên cấp bách, việc dự báo và phân tích tình hình sử dụng năng lượng toàn cầu đóng vai trò then chốt trong việc định hình chiến lược năng lượng bền vững. Bài toán đặt ra cho đồ án này là xây dựng một ứng dụng phân tích và dự báo mức sử dụng năng lượng, cung cấp cái nhìn chi tiết và trực quan về tình hình sử dụng năng lượng ở các quốc gia trên thế giới, đồng thời đưa ra dự báo xu hướng tiêu thụ năng lượng trong 5 năm tới.

Dữ liệu đầu vào cho bài toán là bộ dữ liệu từ Cơ quan Năng lượng Quốc tế (IEA), bao gồm các chỉ số về sản xuất, tiêu thụ, nhập khẩu, xuất khẩu năng lượng từ năm 2000 đến 2021 cho 63 quốc gia và vùng lãnh thổ. Dữ liệu được phân loại theo 4 lĩnh vực chính:

- Dân dụng (Residential): Tiêu thụ năng lượng trong các hộ gia đình cho các mục đích như sưởi ấm, làm mát, chiếu sáng, nấu nướng, sử dụng thiết bị gia dụng,...

- Dịch vụ (Services): Tiêu thụ năng lượng trong các ngành dịch vụ như thương mại, du lịch, y tế, giáo dục,...

- Công nghiệp (Industry): Tiêu thụ năng lượng trong các ngành công nghiệp sản xuất như chế tạo, khai khoáng, xây dựng,...

- Giao thông vận tải (Transport): Tiêu thụ năng lượng trong các hoạt động vận tải hành khách và hàng hóa bằng đường bộ, đường thủy, đường hàng không.

Bài toán yêu cầu:

* Xử lý và làm sạch dữ liệu: Loại bỏ dữ liệu nhiễu, xử lý dữ liệu bị thiếu, chuẩn hóa dữ liệu đầu vào để đảm bảo tính chính xác và nhất quán.
* Phân tích dữ liệu lịch sử: Phân tích dữ liệu tiêu thụ năng lượng theo từng quốc gia, lĩnh vực và loại hình năng lượng, từ đó đưa ra đánh giá về tình hình sử dụng năng lượng hiện tại.
* Xây dựng mô hình dự báo: Xây dựng mô hình hồi quy tuyến tính để dự báo xu hướng tiêu thụ năng lượng cho từng quốc gia, lĩnh vực và loại hình năng lượng trong 5 năm tiếp theo (2022-2027). Đánh giá độ chính xác của mô hình thông qua chỉ số RMSE.
* Trực quan hóa dữ liệu: Sử dụng biểu đồ để trực quan hóa dữ liệu lịch sử và dữ liệu dự báo, giúp người dùng dễ dàng quan sát, so sánh và nhận định xu hướng.
* Xây dựng giao diện người dùng: Xây dựng giao diện ứng dụng thân thiện, dễ sử dụng cho phép người dùng lựa chọn quốc gia, lĩnh vực và loại hình năng lượng mong muốn, hiển thị dữ liệu lịch sử, dữ liệu dự báo và biểu đồ trực quan.

Kết quả đầu ra của bài toán là một ứng dụng phân tích và dự báo mức sử dụng năng lượng, cung cấp thông tin hữu ích cho các nhà hoạch định chính sách, nhà đầu tư, doanh nghiệp, nhà nghiên cứu, và người tiêu dùng. Ứng dụng giúp người dùng nắm bắt được tình hình sử dụng năng lượng hiện tại, dự đoán nhu cầu năng lượng trong tương lai, từ đó hỗ trợ việc hoạch định chính sách, đầu tư, sản xuất và tiêu dùng năng lượng hiệu quả hơn.

Ngoài việc giải quyết bài toán chính, đồ án cũng đề xuất các giải pháp và biện pháp nhằm cải thiện hiệu quả sử dụng năng lượng và giảm thiểu tác động tiêu cực đến môi trường, dựa trên kết quả phân tích và dự báo của ứng dụng.

**3.2. Các vấn đề và khó khăn gặp phải.**

Trong quá trình thực hiện, em đã đối mặt với một loạt các khó khăn và thách thức, đòi hỏi sự kiên trì, sáng tạo và khả năng thích ứng để tìm ra giải pháp tối ưu. Những vấn đề này trải dài từ khâu xử lý dữ liệu, lựa chọn và tinh chỉnh mô hình học máy, đến việc trực quan hóa dữ liệu và thiết kế giao diện người dùng. Cụ thể, những khó khăn chính bao gồm:

**3.2.1. Về mô hình học máy:**

- Lựa chọn mô hình: có rất nhiều mô hình máy học khác nhau mỗi loại có ưu nhược điểm riêng. Việc lựa chọn mô hình phù hợp cho dự án là vô cùng khó khăn. Vì lẽ đó nên em phải cân nhắc kỹ lưỡng các yếu tố như:

- Đặc điểm dữ liệu: Dữ liệu năng lượng có tính chu kỳ, mùa vụ, hay chịu ảnh hưởng bởi các yếu tố bất thường như biến động kinh tế, chính trị?

- Mục tiêu dự báo: Dự báo ngắn hạn hay dài hạn? Mức độ chính xác mong muốn là bao nhiêu?

-Khả năng tính toán:Mô hình có yêu cầu tài nguyên tính toán quá lớn so với khả năng hiện có?

- Điều chỉnh siêu tham số: việc tinh chỉnh siêu tham số cho mô hình học máy là một quá trình thử nghiệm và điều chỉnh khó khăn và yêu cầu về mặt toán học khá cao

- Hiệu suất dự báo: Thực tế phức tạp hơn lý thuyết. Dữ liệu năng lượng thường chứa nhiễu do nhiều yếu tố ngoại lai, chẳng hạn như:Sai số đo lường,biến động kinh tế, chính trị, thay đổi hành vi người tiêu dùng.

**3.2.2. Về trực quan hóa dữ liệu:**

- Lựa chọn biểu đồ phù hợp: có nhiều loại biểu đồ khác nhau và em phải cân nhắc lựa chọn biểu đồ sao cho phù hợp với đề tài và người dùng dễ hiểu.

- Màu sắc: Lựa chọn bảng màu hài hòa, dễ phân biệt, đồng thời tuân theo quy tắc về khả năng tiếp cận cho người dùng mù màu.

- Nhãn và chú thích: Cung cấp đầy đủ thông tin cần thiết để người dùng hiểu rõ biểu đồ.

- Bố cục:Sắp xếp các biểu đồ một cách logic, khoa học, giúp người xem dễ dàng theo dõi và so sánh thông tin.

**3.2.3. Về giao diện người dùng:**

- Thiết kế giao diện thân thiện: Giao diện người dùng là phần tiếp xúc với người dùng của ứng dụng, ảnh hưởng trực tiếp đến trải nghiệm của người dùng. Em đã gặp khó khăn trong việc cân bằng giữa tính thẩm mỹ và tính tiện dụng, đảm bảo giao diện vừa đẹp mắt, vừa dễ sử dụng, cho phép người dùng dễ dàng thao tác và tương tác với ứng dụng.

- Tích hợp các thành phần: Việc kết hợp các thành phần như biểu đồ, bảng biểu, menu, và các widget khác vào giao diện người dùng sao cho hài hòa, đồng bộ và hiệu quả là một thử thách không nhỏ. Tôi phải đảm bảo các thành phần được sắp xếp hợp lý, không gây rối mắt, và tương tác với nhau một cách mượt mà.

**3.2.4. Về kiến thức chuyên môn:**

**-** Ngành năng lượng: Lĩnh vực năng lượng là một mảng kiến thức rộng lớn và phức tạp. Việc tìm hiểu, phân tích và đánh giá dữ liệu liên quan đòi hỏi tôi phải nghiên cứu, tổng hợp thông tin từ nhiều nguồn khác nhau, bao gồm báo cáo và phân tích của IEA.

- Học máy: Thế giới học máy luôn thay đổi và phát triển với tốc độ chóng mặt. Để xây dựng ứng dụng hiệu quả, tôi phải liên tục cập nhật kiến thức về các thuật toán mới, các kỹ thuật tối ưu hóa, và các thư viện Python. Việc tự học, tham gia các khóa học trực tuyến, và tham khảo ý kiến chuyên gia là những hoạt động không thể thiếu.

Những khó khăn và thách thức gặp phải trong quá trình thực hiện dự án là những bài học quý báu, giúp tôi trưởng thành hơn về kỹ năng chuyên môn, khả năng giải quyết vấn đề, và tinh thần kiên trì. Tôi tin rằng những kinh nghiệm này sẽ là hành trang vững chắc cho con đường học tập và nghiên cứu của tôi trong tương lai.

**3.3. Phương pháp thực hiện.**

Để giải quyết bài toán đã nêu, đồ án áp dụng kết hợp các phương pháp phân tích dữ liệu, học máy và thiết kế giao diện người dùng, cụ thể như sau:

**3.3.1. Thu thập dữ liệu liệu.**

Nguồn dữ liệu:Dữ liệu được thu thập từ bộ dữ liệu của Cơ quan Năng lượng Quốc tế (IEA), một nguồn uy tín và đáng tin cậy về thông tin năng lượng toàn cầu. Bộ dữ liệu này cung cấp thông tin chi tiết về sản xuất, tiêu thụ, nhập khẩu, xuất khẩu năng lượng cho 63 quốc gia và vùng lãnh thổ từ năm 2000 đến 2021, được phân loại theo 4 lĩnh vực chính: dân dụng, dịch vụ, công nghiệp, và giao thông vận tải.

Về cấu trúc dữ liệu được chia làm 4 sheet đại diện cho 4 lĩnh vực quan trọng của ngành năng lượng.

**3.3.3.1. Sheet 1 Residential – Energy.**

Sheet "Residential - Energy" chứa dữ liệu về mức sử dụng năng lượng trong khu vực dân dụng (Residential) của 63 quốc gia và vùng lãnh thổ. Dữ liệu được thu thập từ Cơ quan Năng lượng Quốc tế (IEA) và bao gồm thông tin về mức sử dụng năng lượng cho từng quốc gia trong 10 năm từ 2000 đến 2021.

Cấu trúc dữ liệu:Mỗi dòng trong sheet đại diện cho một quốc gia và một loại sử dụng năng lượng cụ thể trong khu vực dân dụng. Cấu trúc dữ liệu bao gồm:

* Country: Tên quốc gia hoặc vùng lãnh thổ.
* End use: Loại sử dụng năng lượng trong khu vực dân dụng, ví dụ: "Residential appliances", "Residential cooking", "Residential lighting", "Residential space cooling", "Residential space heating", "Residential water heating".
* Product: Nguồn năng lượng sử dụng, ví dụ: "Biofuels and waste (PJ)", "Coal and coal products (PJ)", "Electricity (PJ)", "Gas (PJ)", "Heat (PJ)", "Oil and oil products (PJ)", "Other sources (PJ)".
* Năm: Từ năm 2000 đến 2021, mỗi cột đại diện cho mức sử dụng năng lượng của quốc gia tương ứng trong năm đó, tính theo PJ (Petajoule).

Đặc điểm dữ liệu:

Độ bao phủ: Dữ liệu bao phủ 63 quốc gia và vùng lãnh thổ có mức tiêu thụ năng lượng cao nhất trên toàn cầu, cung cấp một bức tranh toàn diện về tình hình sử dụng năng lượng trong khu vực dân dụng.

Phân loại theo quốc gia và loại sử dụng năng lượng: Dữ liệu được phân loại theo quốc gia và loại sử dụng năng lượng, cho phép phân tích chi tiết về mức tiêu thụ năng lượng theo từng quốc gia, từng loại năng lượng và theo từng năm.

Phân loại theo nguồn năng lượng: Dữ liệu được phân loại theo nguồn năng lượng, cho phép phân tích mức tiêu thụ từng loại năng lượng, từ đó có thể đánh giá mức độ phụ thuộc vào từng nguồn năng lượng cụ thể.

Sự thay đổi theo thời gian: Dữ liệu cung cấp thông tin về mức tiêu thụ năng lượng trong 10 năm, cho phép phân tích xu hướng sử dụng năng lượng theo thời gian.

Dạng số: Mức tiêu thụ năng lượng được biểu thị bằng số liệu, cho phép phân tích định lượng và so sánh giữa các quốc gia, loại sử dụng năng lượng và loại năng lượng.

Sheet "Residential - Energy" trong đề tài được sử dụng để phân tích mức sử dụng năng lượng trong khu vực dân dụng của các quốc gia trên toàn cầu, đánh giá xu hướng tiêu thụ năng lượng theo thời gian, xây dựng mô hình dự báo mức sử dụng năng lượng cho từng quốc gia và từng loại sử dụng năng lượng và đề xuất các giải pháp nhằm cải thiện hiệu quả sử dụng năng lượng trong khu vực dân dụng và hướng tới phát triển năng lượng bền vững.

**3.3.3.2. Sheet 2 Services – Energy.**

Sheet "Services - Energy" chứa dữ liệu về mức sử dụng năng lượng trong lĩnh vực dịch vụ (Services) của 63 quốc gia và vùng lãnh thổ. Dữ liệu được thu thập từ Cơ quan Năng lượng Quốc tế (IEA) và bao gồm thông tin về mức sử dụng năng lượng cho từng quốc gia trong 10 năm từ 2000 đến 2021.

Cấu trúc dữ liệu:

* Mỗi dòng trong sheet đại diện cho một quốc gia và một loại sử dụng năng lượng cụ thể trong lĩnh vực dịch vụ. Cấu trúc dữ liệu bao gồm:
* Country: Tên quốc gia hoặc vùng lãnh thổ.
* End use: Loại sử dụng năng lượng trong lĩnh vực dịch vụ, ví dụ: "Services lighting", "Services space cooling", "Services space heating".
* Product: Nguồn năng lượng sử dụng, ví dụ: "Biofuels and waste (PJ)", "Coal and coal products (PJ)", "Electricity (PJ)", "Gas (PJ)", "Heat (PJ)", "Oil and oil products (PJ)", "Other sources (PJ)".
* Năm: Từ năm 2000 đến 2021, mỗi cột đại diện cho mức sử dụng năng lượng của quốc gia tương ứng trong năm đó, tính theo PJ (Petajoule).

Đặc điểm dữ liệu:

Độ bao phủ: Dữ liệu bao phủ 63 quốc gia và vùng lãnh thổ có mức tiêu thụ năng lượng cao nhất trên toàn cầu, cung cấp một bức tranh toàn diện về tình hình sử dụng năng lượng trong lĩnh vực dịch vụ.

Phân loại theo quốc gia và loại sử dụng năng lượng: Dữ liệu được phân loại theo quốc gia và loại sử dụng năng lượng, cho phép phân tích chi tiết về mức tiêu thụ năng lượng theo từng quốc gia, từng loại năng lượng và theo từng năm.

Phân loại theo nguồn năng lượng: Dữ liệu được phân loại theo nguồn năng lượng, cho phép phân tích mức tiêu thụ từng loại năng lượng, từ đó có thể đánh giá mức độ phụ thuộc vào từng nguồn năng lượng cụ thể.

Sự thay đổi theo thời gian: Dữ liệu cung cấp thông tin về mức tiêu thụ năng lượng trong 10 năm, cho phép phân tích xu hướng sử dụng năng lượng theo thời gian.

Dạng số: Mức tiêu thụ năng lượng được biểu thị bằng số liệu, cho phép phân tích định lượng và so sánh giữa các quốc gia, loại sử dụng năng lượng và loại năng lượng.

Mục đích sử dụng sheet này cũng giống như sheet 1.

**3.3.3.3. Sheet 3 Industry – Energy.**

Sheet "Industry - Energy" chứa dữ liệu về mức sử dụng năng lượng trong lĩnh vực công nghiệp (Industry) của 63 quốc gia và vùng lãnh thổ. Dữ liệu được thu thập từ Cơ quan Năng lượng Quốc tế (IEA) và bao gồm thông tin về mức sử dụng năng lượng cho từng quốc gia trong 10 năm từ 2000 đến 2021.

Cấu trúc dữ liệu:

Mỗi dòng trong sheet đại diện cho một quốc gia, một ngành công nghiệp cụ thể và một loại sử dụng năng lượng. Cấu trúc dữ liệu bao gồm:

* Country: Tên quốc gia hoặc vùng lãnh thổ.
* Subsector: Ngành công nghiệp, ví dụ: "Agriculture forestry fishing [ISIC 01-03]", "Basic metals [ISIC 24]", "Chemicals and chemical Products [ISIC 20-21]", "Construction [ISIC 41-43]", "Mining [ISIC 05-09]", "Non-metallic minerals [ISIC 23]", "Paper pulp and printing [ISIC 17-18]".
* Product: Nguồn năng lượng sử dụng, ví dụ: "Biofuels and waste (PJ)", "Coal and coal products (PJ)", "Electricity (PJ)", "Gas (PJ)", "Heat (PJ)", "Oil and oil products (PJ)", "Other sources (PJ)".
* Năm: Từ năm 2000 đến 2021, mỗi cột đại diện cho mức sử dụng năng lượng của quốc gia tương ứng trong năm đó, tính theo PJ (Petajoule).

Đặc điểm dữ liệu:

Độ bao phủ: Dữ liệu bao phủ 63 quốc gia và vùng lãnh thổ có mức tiêu thụ năng lượng cao nhất trên toàn cầu, cung cấp một bức tranh toàn diện về tình hình sử dụng năng lượng trong lĩnh vực công nghiệp.

Phân loại theo quốc gia, ngành công nghiệp và loại sử dụng năng lượng: Dữ liệu được phân loại theo quốc gia, ngành công nghiệp và loại sử dụng năng lượng, cho phép phân tích chi tiết về mức tiêu thụ năng lượng theo từng quốc gia, từng ngành, từng loại năng lượng và theo từng năm.

Phân loại theo nguồn năng lượng: Dữ liệu được phân loại theo nguồn năng lượng, cho phép phân tích mức tiêu thụ từng loại năng lượng, từ đó có thể đánh giá mức độ phụ thuộc vào từng nguồn năng lượng cụ thể.

Sự thay đổi theo thời gian: Dữ liệu cung cấp thông tin về mức tiêu thụ năng lượng trong 10 năm, cho phép phân tích xu hướng sử dụng năng lượng theo thời gian.

Dạng số: Mức tiêu thụ năng lượng được biểu thị bằng số liệu, cho phép phân tích định lượng và so sánh giữa các quốc gia, ngành công nghiệp, loại sử dụng năng lượng và loại năng lượng.

Mục đích:

Sheet "Industry - Energy" được sử dụng với mục đích giống sheet 1 và sheet 2.

**3.3.3.4. Sheet 4 Transport – Energy.**

Sheet "Transport - Energy" chứa dữ liệu về mức sử dụng năng lượng trong lĩnh vực giao thông vận tải (Transport) của 63 quốc gia và vùng lãnh thổ. Dữ liệu được thu thập từ Cơ quan Năng lượng Quốc tế (IEA) và bao gồm thông tin về mức sử dụng năng lượng cho từng quốc gia trong 10 năm từ 2000 đến 2021.

Cấu trúc dữ liệu:

Mỗi dòng trong sheet đại diện cho một quốc gia, một loại phương tiện/hình thức vận tải cụ thể và một loại năng lượng sử dụng. Cấu trúc dữ liệu bao gồm:

* Country: Tên quốc gia hoặc vùng lãnh thổ.
* Mode/vehicle type: Loại phương tiện/hình thức vận tải, ví dụ: "Buses", "Cars/light trucks", "Freight trucks", "Motorcycles", "Total airplanes", "Total freight transport", "Total passenger transport", "Total road", "Total ships", "Total trains".
* Product: Nguồn năng lượng sử dụng, ví dụ: "Coal and coal products (PJ)", "Diesel and light fuel oil (PJ)", "Electricity (PJ)", "Gas (PJ)", "Heavy fuel oil (PJ)", "Jet fuel and aviation gasoline (PJ)", "LPG (PJ)", "Motor gasoline (PJ)", "Other sources (PJ)".
* Năm: Từ năm 2000 đến 2021, mỗi cột đại diện cho mức sử dụng năng lượng của quốc gia tương ứng trong năm đó, tính theo PJ (Petajoule).

Đặc điểm dữ liệu:

Độ bao phủ: Dữ liệu bao phủ 63 quốc gia và vùng lãnh thổ có mức tiêu thụ năng lượng cao nhất trên toàn cầu, cung cấp một bức tranh toàn diện về tình hình sử dụng năng lượng trong lĩnh vực giao thông vận tải.

Phân loại theo quốc gia, loại phương tiện/hình thức vận tải và loại năng lượng: Dữ liệu được phân loại theo quốc gia, loại phương tiện/hình thức vận tải và loại năng lượng, cho phép phân tích chi tiết về mức tiêu thụ năng lượng theo từng quốc gia, từng loại phương tiện/hình thức vận tải, từng loại năng lượng và theo từng năm.

Phân loại theo nguồn năng lượng: Dữ liệu được phân loại theo nguồn năng lượng, cho phép phân tích mức tiêu thụ từng loại năng lượng, từ đó có thể đánh giá mức độ phụ thuộc vào từng nguồn năng lượng cụ thể.

Sự thay đổi theo thời gian: Dữ liệu cung cấp thông tin về mức tiêu thụ năng lượng trong 10 năm, cho phép phân tích xu hướng sử dụng năng lượng theo thời gian.

Dạng số: Mức tiêu thụ năng lượng được biểu thị bằng số liệu, cho phép phân tích định lượng và so sánh giữa các quốc gia, loại phương tiện/hình thức vận tải, loại sử dụng năng lượng và loại năng lượng.

**3.3.2.Tiền xử lý dữ liệu:**

Làm sạch dữ liệu: sử dụng thư viện Pandas trong Python, em đã thực hiện các bước làm sạch dữ liệu sau:

* Xử lý dữ liệu bị thiếu: Thay thế các giá trị thiếu bằng giá trị trung bình của các quốc gia khác trong cùng năm, đảm bảo tính nhất quán và giảm thiểu sai lệch trong dữ liệu.
* Chuẩn hóa tên cột: Đảm bảo tính thống nhất và dễ đọc của tên cột, thuận tiện cho việc truy cập và sử dụng dữ liệu.
* Chuyển đổi kiểu dữ liệu: Chuyển đổi các cột dữ liệu sang kiểu dữ liệu phù hợp (số nguyên, số thực,...) để thực hiện các phép toán và phân tích chính xác.
* Sắp xếp dữ liệu: Sắp xếp dữ liệu theo các tiêu chí phù hợp (quốc gia, năm, lĩnh vực, loại hình năng lượng) để dễ dàng truy cập và trực quan hóa.

Tổ chức dữ liệu: Để dễ dàng thao tác và phân tích, dữ liệu được tổ chức thành các DataFrame riêng biệt cho từng lĩnh vực và loại hình năng lượng, sử dụng các hàm pivot\_table và melt của Pandas.

**3.3.3. Phân tích dữ liệu lịch sử.**

* Thống kê mô tả: Áp dụng các phép tính thống kê mô tả như trung bình, độ lệch chuẩn, giá trị lớn nhất, giá trị nhỏ nhất,... để khám phá đặc điểm phân bố và xu hướng của dữ liệu tiêu thụ năng lượng theo từng quốc gia, lĩnh vực và loại hình năng lượng. Kết quả thống kê mô tả cung cấp cái nhìn tổng quan về tình hình sử dụng năng lượng hiện tại.
* Trực quan hóa dữ liệu: Sử dụng thư viện Matplotlib trong Python, em đã tạo ra các biểu đồ trực quan hóa dữ liệu lịch sử, nhằm giúp người dùng dễ dàng quan sát, so sánh và nhận định xu hướng tiêu thụ năng lượng. Việc lựa chọn loại biểu đồ phù hợp được cân nhắc kỹ lưỡng dựa trên loại dữ liệu, mối quan hệ giữa các biến và mục tiêu truyền tải thông tin. Bên cạnh đó, tôi cũng chú trọng đến việc thiết kế biểu đồ sao cho trực quan, dễ hiểu và thu hút người xem, sử dụng màu sắc hài hòa, kiểu dáng phù hợp, nhãn và chú thích rõ ràng.

**3.3.4. Xây dựng mô hình dự báo.**

* Thử nghiệm và lựa chọn mô hình: Để tìm ra mô hình dự báo phù hợp nhất, tôi đã thử nghiệm và so sánh hiệu suất của nhiều mô hình học máy, bao gồm ARIMA, Prophet, Moving Average, và Linear Regression. Mỗi mô hình được huấn luyện và đánh giá trên dữ liệu lịch sử của từng quốc gia, lĩnh vực và loại hình năng lượng. Sau khi so sánh kết quả dự báo dựa trên chỉ số RMSE và các tiêu chí khác, mô hình Linear Regression được lựa chọn cho dự án.
* Lý do chọn Linear Regression: Mô hình này được ưa chuộng do tính đơn giản, hiệu quả và khả năng dự báo tương đối chính xác trong ngắn hạn, phù hợp với mục tiêu dự báo xu hướng tiêu thụ năng lượng trong 5 năm tới của đồ án.
* Huấn luyện và đánh giá: Mô hình Linear Regression được huấn luyện kỹ lưỡng trên dữ liệu lịch sử, sử dụng phương pháp bình phương bé nhất (OLS) để tìm ra đường thẳng (hoặc siêu phẳng) phù hợp nhất với dữ liệu. Sau khi huấn luyện, mô hình được đánh giá độ chính xác dựa trên chỉ số RMSE và so sánh kết quả dự báo với dữ liệu thực tế để xác định mức độ phù hợp.

**3.3.5. Trực quan hóa dữ liệu dự báo.**

- Biểu đồ dự báo: Sử dụng Matplotlib để tạo các biểu đồ trực quan hóa dữ liệu dự báo cho từng quốc gia, lĩnh vực và loại hình năng lượng.

- So sánh dữ liệu lịch sử và dự báo: Hiển thị cả dữ liệu lịch sử và dữ liệu dự báo trên cùng một biểu đồ, giúp người dùng dễ dàng so sánh và nhận định xu hướng.

**3.3.6. Xây dựng giao diện người dùng:**

- Thư viện Tkinter: Sử dụng thư viện Tkinter trong Python để xây dựng giao diện người dùng đồ họa (GUI) cho ứng dụng.

- Giao diện thân thiện: Giao diện được thiết kế đơn giản, trực quan và dễ sử dụng, cho phép người dùng thao tác và tương tác một cách dễ dàng.

-Tính năng lựa chọn: Giao diện cho phép người dùng lựa chọn quốc gia, lĩnh vực và loại hình năng lượng mong muốn để phân tích và dự báo.

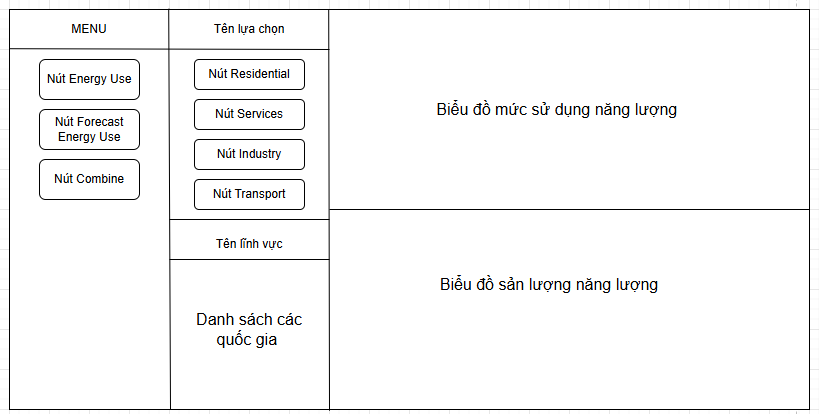
- Hiển thị kết quả: Giao diện hiển thị dữ liệu lịch sử, dữ liệu dự báo và biểu đồ trực quan theo lựa chọn của người dùng.

**3.2.7. Đề xuất giải pháp:**

**-** Dựa trên kết quả phân tích và dự báo: Phân tích kết quả dự báo và so sánh với tình hình sử dụng năng lượng hiện tại để xác định các vấn đề tiềm ẩn.

- Đề xuất giải pháp cụ thể:Đưa ra các đề xuất và biện pháp cụ thể, khả thi và có tính ứng dụng cao nhằm cải thiện hiệu quả sử dụng năng lượng, giảm thiểu tác động tiêu cực đến môi trường, và hướng đến phát triển năng lượng bền vững.

**3.4. Thiết kế - mô tả giao diện người dùng.**

****

Hình 3. 1. Thiết kế giao diện người dùng.

### 3.4.1. Cấu trúc giao diện:

Giao diện ứng dụng được chia thành 3 phần chính:

- Menu chính: Nằm ở phía bên trái màn hình, hiển thị các lựa chọn chính cho người dùng:

- ENERGY USE: Hiển thị dữ liệu lịch sử về mức sử dụng năng lượng.

- FORECAST ENERGY USE: Hiển thị dữ liệu dự báo mức sử dụng năng lượng trong 5 năm tới.

- COMBINED DATA: Hiển thị dữ liệu kết hợp cả lịch sử và dự báo, cho phép so sánh trực quan.

- Lựa chọn lĩnh vực và quốc gia:Nằm bên phải Menu chính, cho phép người dùng lựa chọn lĩnh vực (Residential, Services, Industry, Transport) và quốc gia mong muốn. Mỗi lĩnh vực được đại diện bởi một nút bấm kèm biểu tượng trực quan, giúp người dùng dễ dàng nhận biết. Danh sách quốc gia được hiển thị dưới dạng các nút bấm, sắp xếp theo thứ tự bảng chữ cái.

- Khu vực hiển thị biểu đồ: Chiếm phần lớn diện tích màn hình bên phải, hiển thị các biểu đồ trực quan hóa dữ liệu theo lựa chọn của người dùng. Biểu đồ được tạo động dựa trên lựa chọn lĩnh vực, quốc gia và loại dữ liệu (lịch sử, dự báo, kết hợp).

### 3.4.2. Mô tả chi tiết các thành phần:

-Menu chính: Các nút bấm trong menu chính được thiết kế nổi bật, rõ ràng, giúp người dùng dễ dàng nhận biết và lựa chọn chức năng mong muốn.

- Nút bấm lĩnh vực: Mỗi nút bấm đại diện cho một lĩnh vực, kèm theo biểu tượng trực quan giúp người dùng dễ dàng nhận biết.

- Nút bấm quốc gia: Các nút bấm quốc gia được sắp xếp theo thứ tự bảng chữ cái, giúp người dùng dễ dàng tìm kiếm quốc gia mong muốn.

Biểu đồ:

- Loại biểu đồ:Sử dụng biểu đồ cột (bar chart) để hiển thị mức sử dụng năng lượng theo thời gian. Biểu đồ cột cho phép so sánh trực quan giữa các năm và dễ dàng nhận biết xu hướng tăng/giảm của mức sử dụng năng lượng.

- Màu sắc: Sử dụng bảng màu hài hòa, dễ phân biệt, đồng thời tuân theo quy tắc về khả năng tiếp cận cho người dùng mù màu.

- Nhãn và chú thích: Biểu đồ được trang bị đầy đủ nhãn trục, tiêu đề, chú thích và đơn vị đo lường, giúp người dùng hiểu rõ thông tin được hiển thị.

### 3.4.3. Tính năng:

- Lựa chọn linh hoạt: Người dùng có thể dễ dàng lựa chọn lĩnh vực, quốc gia và loại dữ liệu mong muốn. Giao diện sẽ tự động cập nhật biểu đồ tương ứng với lựa chọn của người dùng.

- Trực quan hóa dữ liệu: Biểu đồ trực quan hóa dữ liệu giúp người dùng dễ dàng quan sát, so sánh và nhận định xu hướng sử dụng năng lượng.

- So sánh dữ liệu: Giao diện cho phép hiển thị đồng thời dữ liệu lịch sử và dự báo trên cùng một biểu đồ, giúp người dùng dễ dàng so sánh và nhận biết sự thay đổi theo thời gian.

### 3.4.4. Ưu điểm:

- Đơn giản, dễ sử dụng: Giao diện được thiết kế đơn giản, trực quan, dễ sử dụng cho cả những người dùng không quen thuộc với ứng dụng.

- Linh hoạt: Người dùng có thể dễ dàng lựa chọn dữ liệu mong muốn và giao diện sẽ tự động cập nhật biểu đồ tương ứng.

- Trực quan: Biểu đồ được thiết kế đẹp mắt, dễ hiểu, giúp người dùng dễ dàng nắm bắt thông tin.

**3.5. Chức năng của ứng dụng.**

* Hiển thị dữ liệu lịch sử: Ứng dụng cho phép người dùng xem dữ liệu lịch sử về mức sử dụng năng lượng theo từng quốc gia, lĩnh vực và loại hình năng lượng. Dữ liệu được hiển thị dưới dạng biểu đồ cột, cho phép so sánh trực quan mức tiêu thụ năng lượng theo thời gian.
* Dự báo mức sử dụng năng lượng: Ứng dụng sử dụng mô hình Linear Regression đã được huấn luyện trên dữ liệu lịch sử để dự báo mức tiêu thụ năng lượng trong 5 năm tiếp theo. Kết quả dự báo cũng được hiển thị dưới dạng biểu đồ cột, giúp người dùng dễ dàng nắm bắt xu hướng tiêu thụ năng lượng trong tương lai.
* So sánh dữ liệu: Ứng dụng cho phép hiển thị đồng thời dữ liệu lịch sử và dự báo trên cùng một biểu đồ, tạo điều kiện cho người dùng so sánh và nhận định sự thay đổi của mức tiêu thụ năng lượng theo thời gian.
* Lựa chọn linh hoạt: Người dùng có thể dễ dàng lựa chọn lĩnh vực, quốc gia và loại dữ liệu (lịch sử, dự báo, kết hợp) mong muốn. Giao diện sẽ tự động cập nhật biểu đồ tương ứng với lựa chọn của người dùng.

**3.6. Áp dụng mô hình Linear Regression.**

**-** Chuẩn bị dữ liệu: Dữ liệu lịch sử về mức tiêu thụ năng lượng từ năm 2000 đến 2021 cho từng quốc gia, lĩnh vực và loại năng lượng được sử dụng để huấn luyện mô hình.

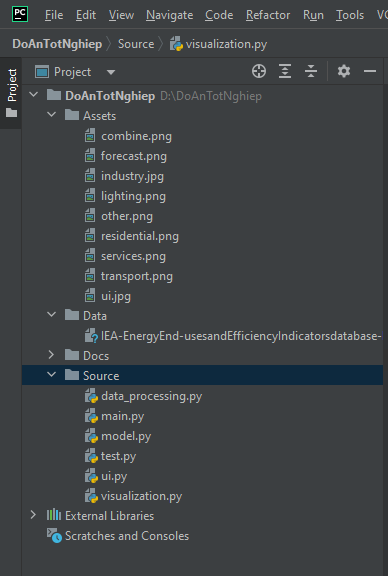
- Huấn luyện mô hình: Mô hình Linear Regression được huấn luyện bằng phương pháp bình phương bé nhất (OLS), tìm cách tối thiểu hóa tổng bình phương sai số giữa giá trị dự đoán của mô hình và giá trị thực tế của dữ liệu.

- Dự báo: Sau khi huấn luyện, mô hình được sử dụng để dự đoán mức tiêu thụ năng lượng cho 5 năm tiếp theo (2022-2027).

- Kết quả: Sau khi đã tiến hành huấn luyện và mô hình dự báo đã cho ra chỉ số RMSE rơi vào dưới 10% giá trị dao động của dữ liệu. Với chỉ số RMSE này có thể cho thấy việc dự báo đã thực hiện với độ chính xác cao.

**3.7. Kết quả đạt được.**

**3.7.1. Cấu trúc tổ chức file của ứng dụng.**

****

Hình 3. 2. Cấu trúc tổ chức file.

Dự án "Ứng dụng phân tích và dự báo mức sử dụng năng lượng toàn cầu" được tổ chức theo cấu trúc file rõ ràng, nhằm đảm bảo tính logic, dễ quản lý và bảo trì. Cấu trúc bao gồm các thư mục chính sau:

DoAnTotNghiep: Thư mục gốc của dự án, chứa toàn bộ mã nguồn và tài nguyên.

- Assets:Chứa các hình ảnh (icon) được sử dụng trong giao diện người dùng, giúp tăng tính trực quan và thẩm mỹ cho ứng dụng.

- Data: Chứa file dữ liệu Excel gốc từ IEA, là nguồn dữ liệu đầu vào cho ứng dụng.

- Docs: Chứa và lưu trữ các tài liệu hướng dẫn sử dụng ứng dụng, báo cáo phân tích,...

- Source: Thư mục quan trọng nhất, chứa toàn bộ mã nguồn Python của ứng dụng:

- data\_processing.py: Chứa các hàm Python để đọc, làm sạch, xử lý và tổ chức dữ liệu đầu vào từ file Excel. File này đảm bảo dữ liệu được chuẩn hóa và sẵn sàng cho phân tích và dự báo.

- main.py: File chính của ứng dụng, khởi chạy giao diện người dùng và kết nối các thành phần khác của ứng dụng.

- model.py: Chứa mã nguồn Python để xây dựng, huấn luyện và đánh giá mô hình Linear Regression, đồng thời lưu trữ dữ liệu dự báo.

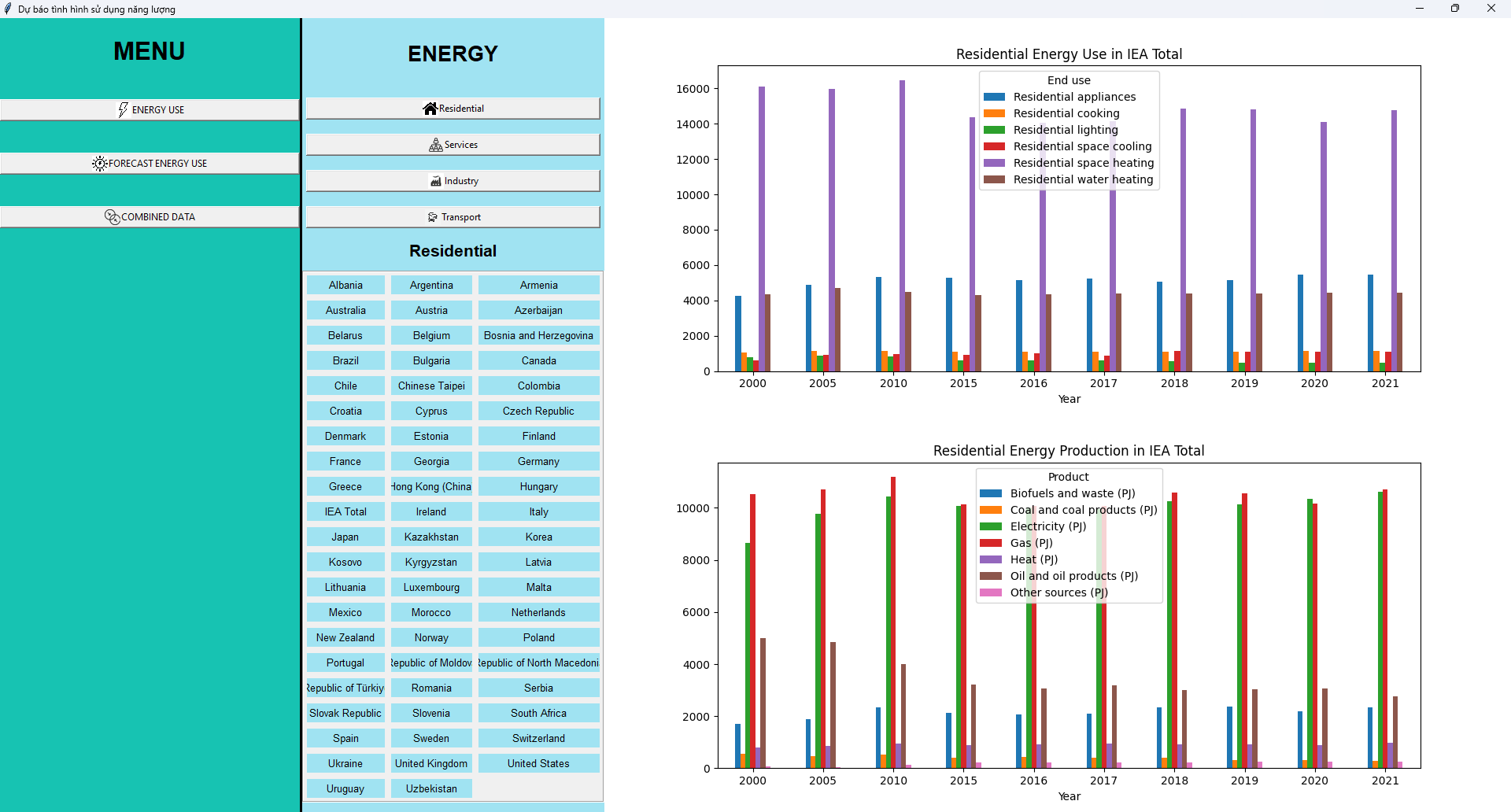
- test.py: (Có thể được sử dụng để viết các test case kiểm thử chức năng của ứng dụng.)

- ui.py: Chứa mã nguồn Python để thiết kế và xây dựng giao diện người dùng bằng thư viện Tkinter. File này quản lý các thành phần giao diện, xử lý sự kiện người dùng, và hiển thị dữ liệu.

- visualization.py: Chứa các hàm Python để tạo các biểu đồ trực quan hóa dữ liệu, sử dụng thư viện Matplotlib. File này đảm nhận việc hiển thị dữ liệu lịch sử, dữ liệu dự báo và dữ liệu kết hợp theo yêu cầu của người dùng.

Cấu trúc tổ chức file trên giúp cho việc quản lý mã nguồn trở nên dễ dàng, thuận tiện cho việc phát triển, bảo trì và mở rộng ứng dụng trong tương lai. Mỗi file Python có chức năng riêng biệt, rõ ràng, đảm bảo tính module hóa và khả năng tái sử dụng mã nguồn.

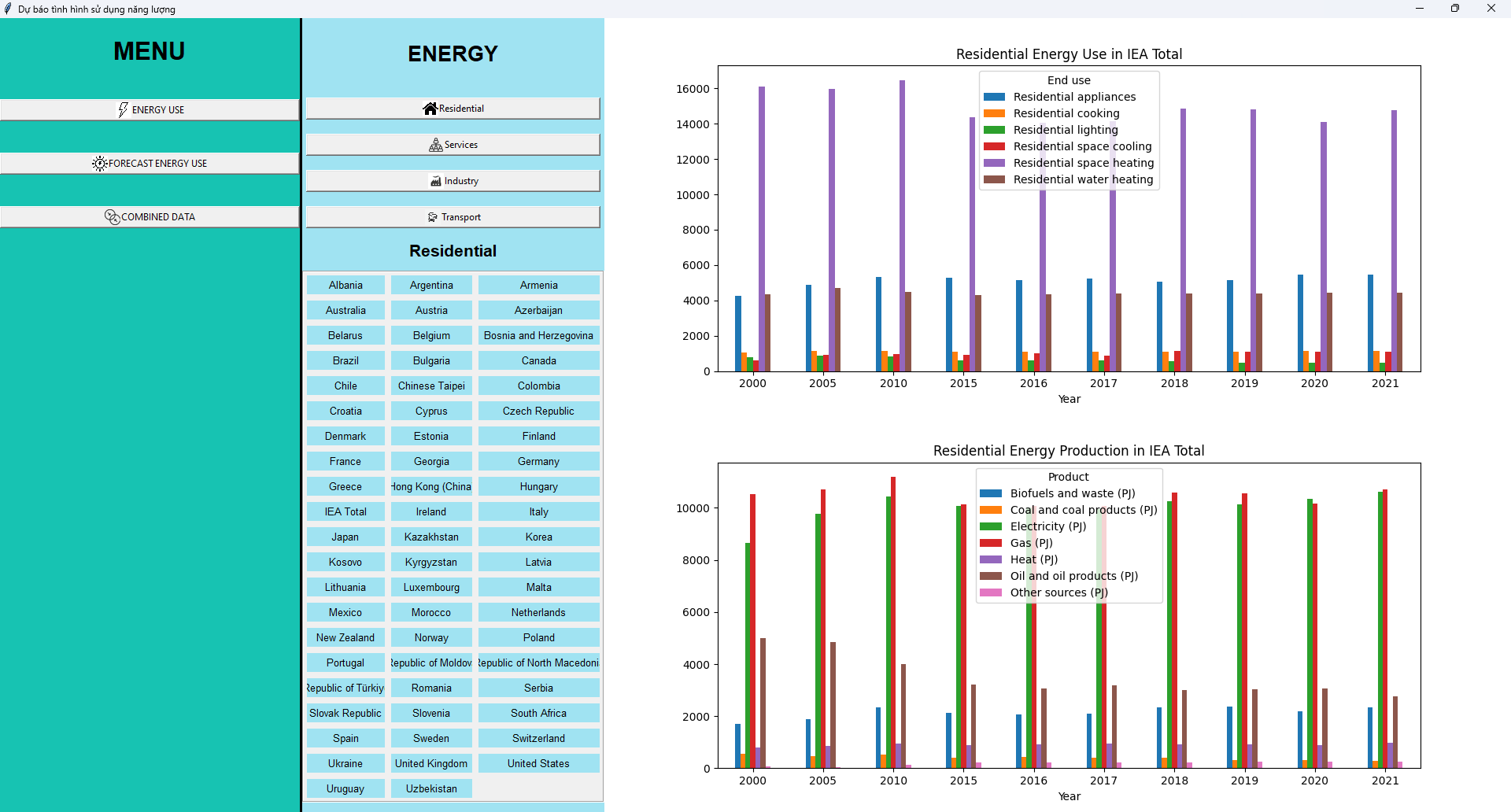
**3.7.2. Giao diện tổng quan.**



Hình 3. 3. Giao diện tổng quan.

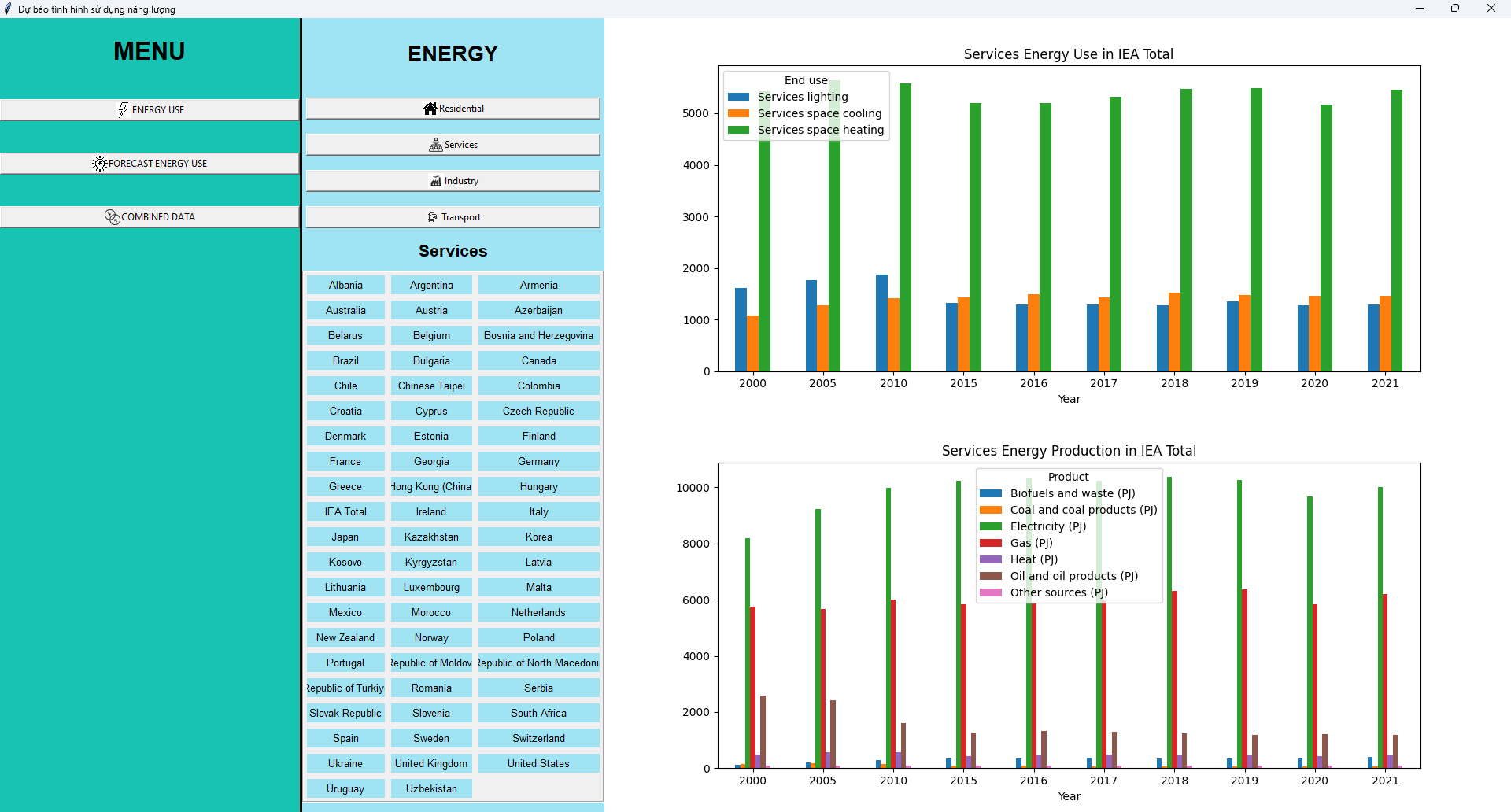
**3.7.3. Giao diện mức năng lượng sử dụng – Energy Use.**

**3.7.3.1. Lĩnh vực dân dụng – Residential.**



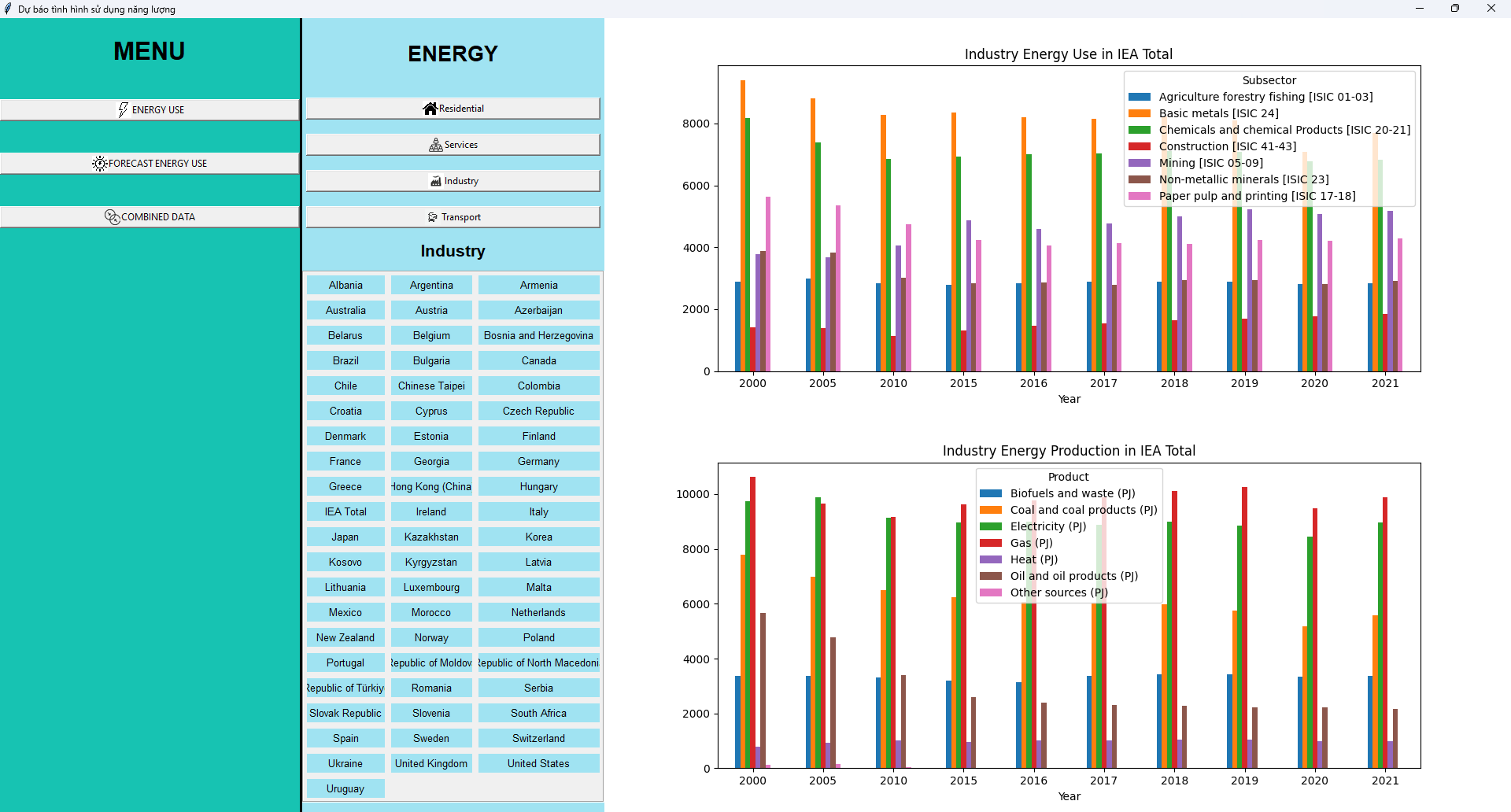
Hình 3. 4. Giao diện lĩnh vực dân dụng - Residential.

**3.7.3.2. Lĩnh vực dịch vụ - Services.**



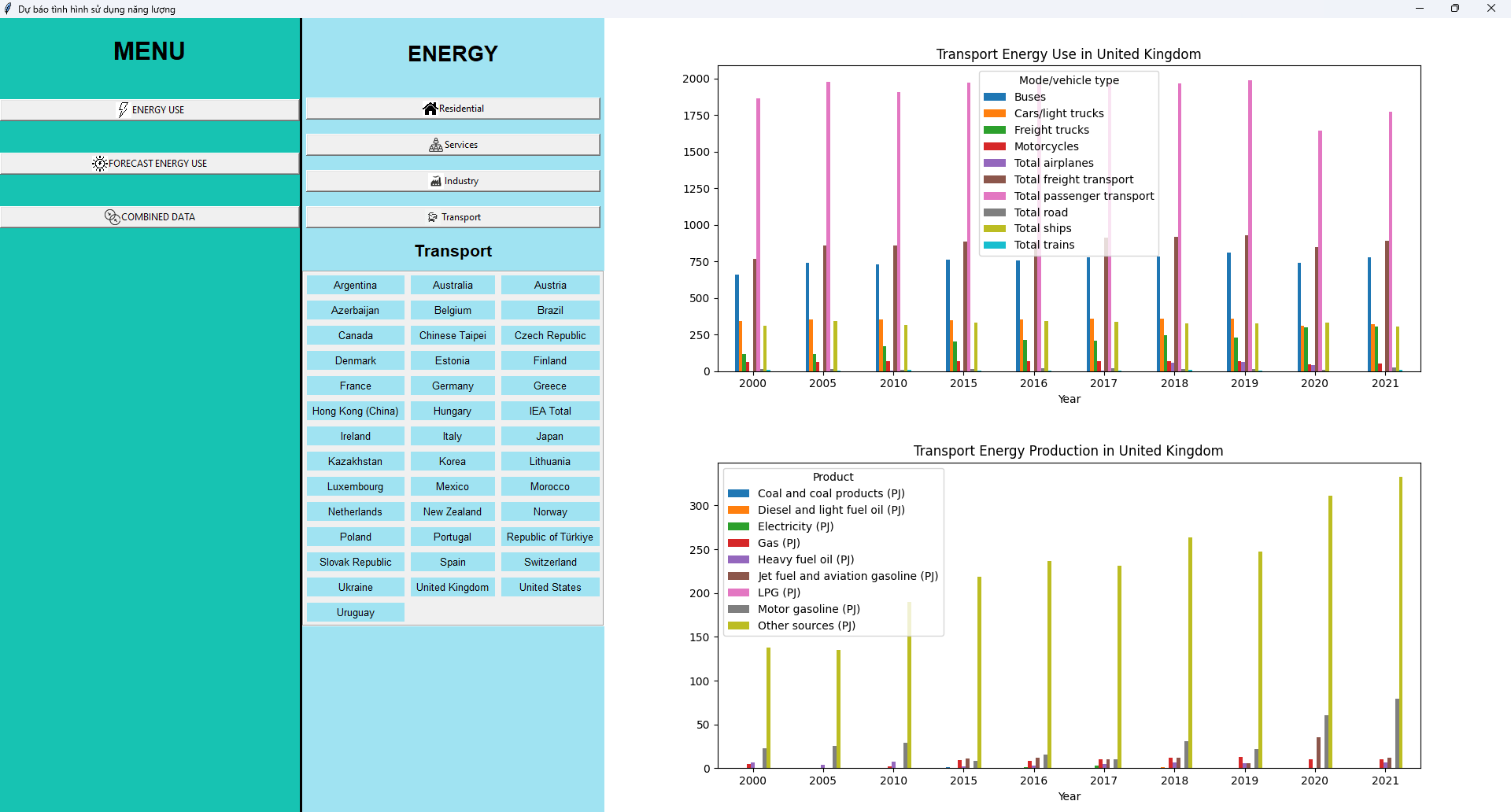
Hình 3. 5. Giao diện lĩnh vực dịch vụ - Services.

**3.7.4.3. Lĩnh vực công nghiệp – Industry.**



Hình 3. 6. Giao diện lĩnh vực công nghiệp - Industry.

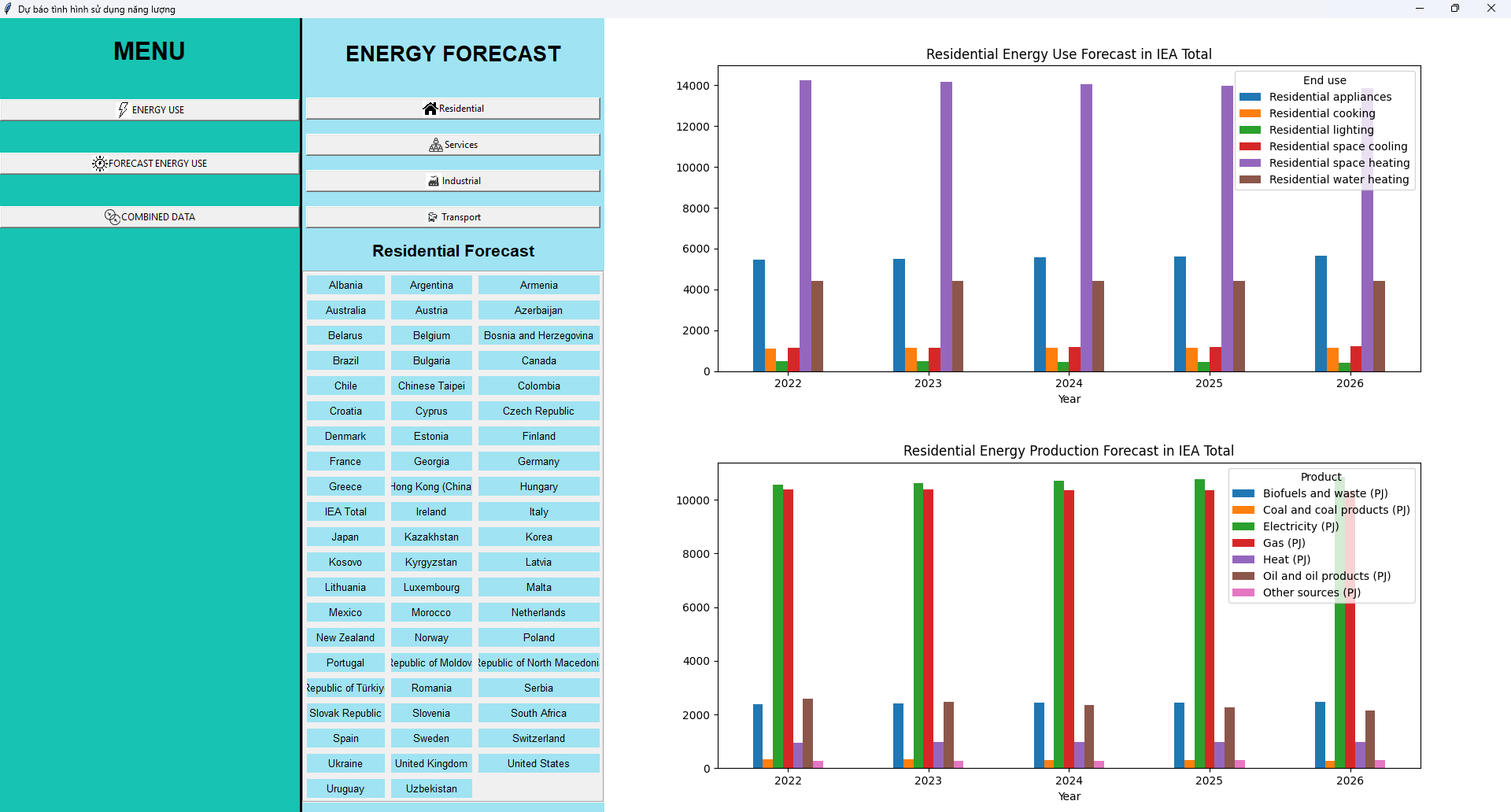
**3.7.5.4. Lĩnh vực giao thông vận tải – Transport.**



Hình 3. 7. Giao diện lĩnh vực Giao thông vận tải - Transport.

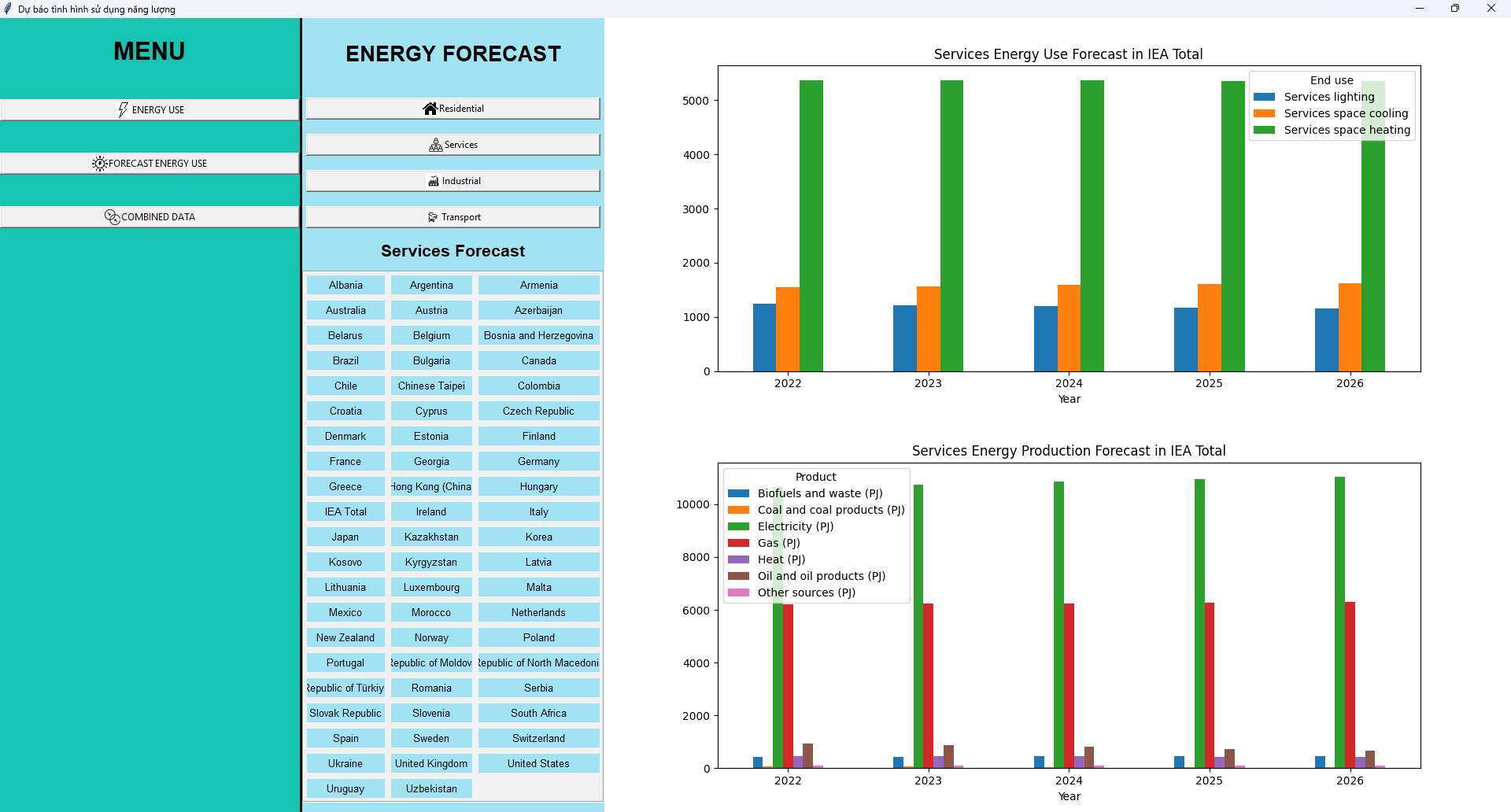
**3.7.4. Giao diện dự báo mức năng lượng sử dụng – Forecast Energy Use.**

**3.7.4.1. Lĩnh vực dân dụng – Residential.**



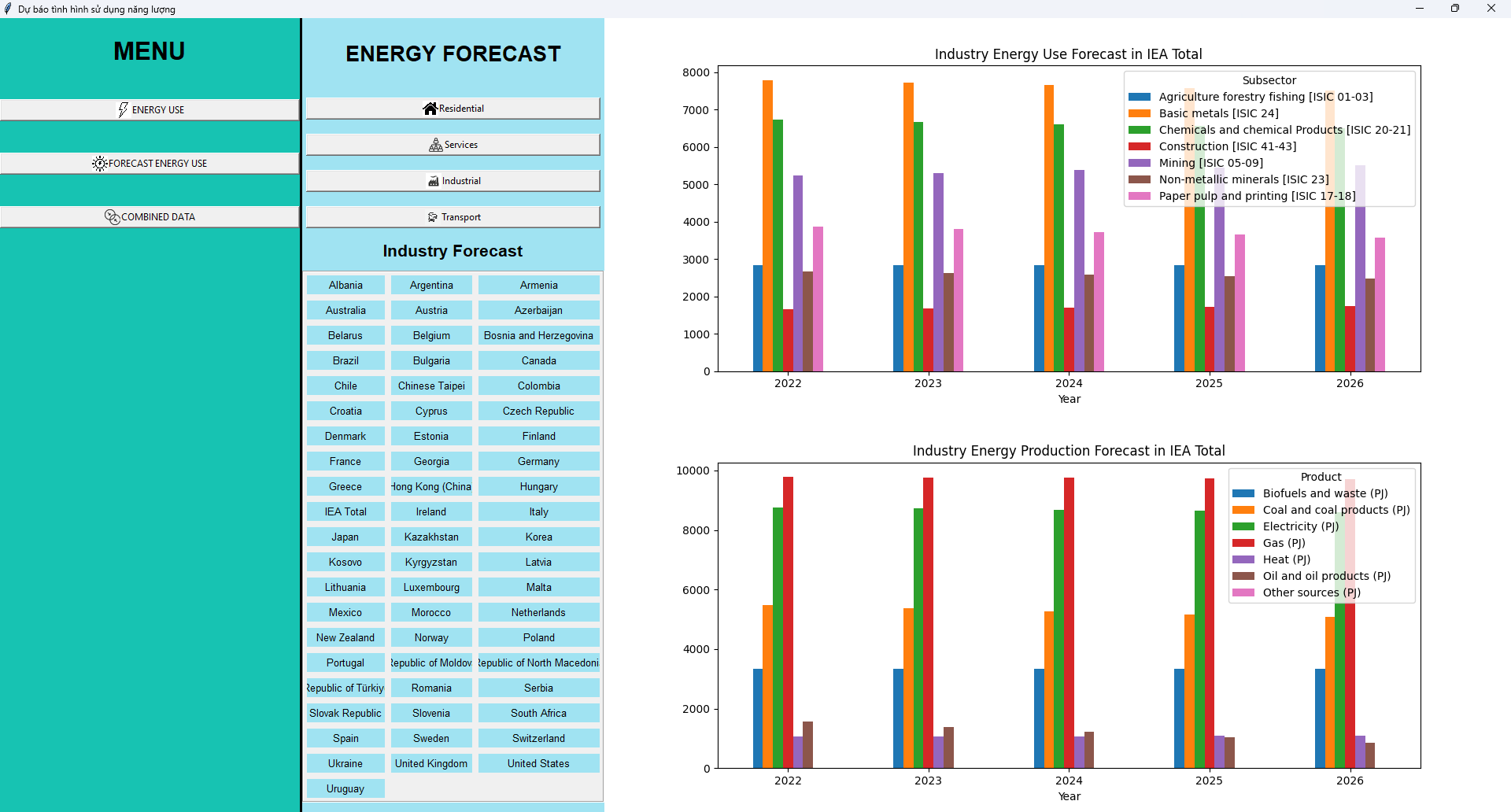
Hình 3. 8. Giao diện dự báo lĩnh vực dân dụng - Residential.

**3.7.4.2. Lĩnh vực dịch vụ - Services.**



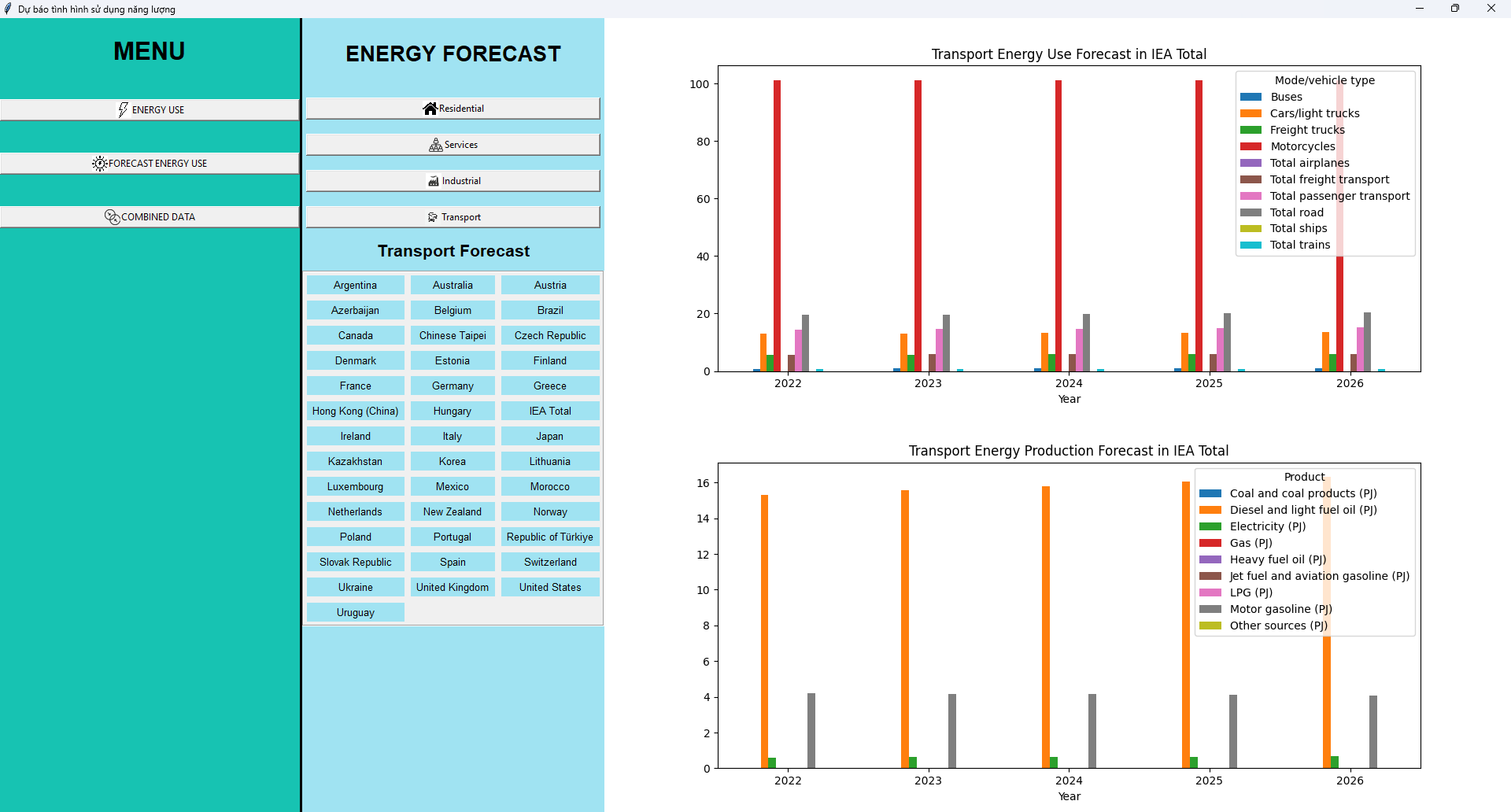
Hình 3. 9. Giao diện dự báo lĩnh vực dịch vụ - Services.

**3.7.4.3. Lĩnh vực công nghiệp – Industry.**



Hình 3. 10. Giao diện dự báo lĩnh vực công nghiệp - Industry.

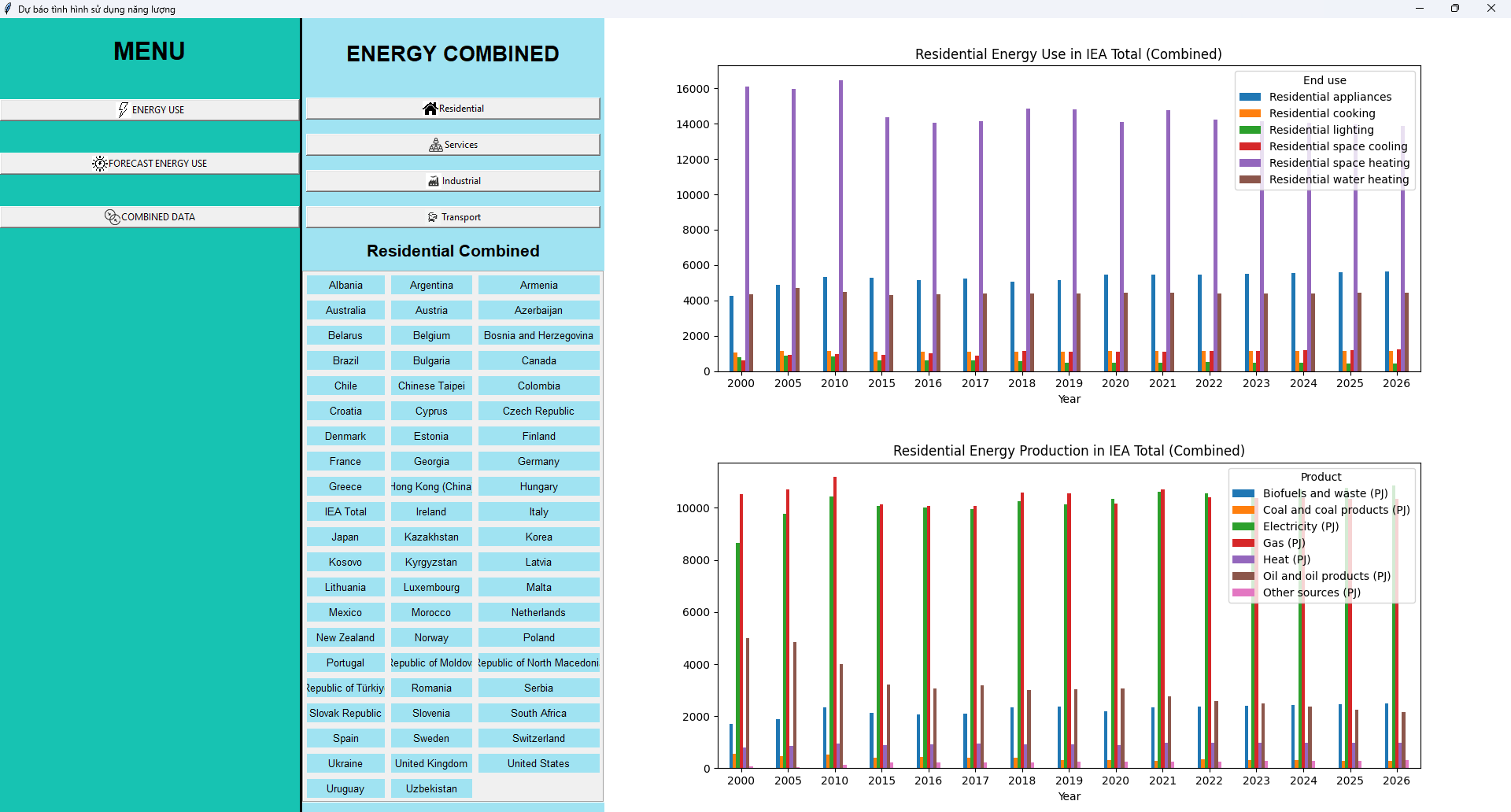
**3.7.4.4. Lĩnh vực giao thông vận tải – Transport.**



Hình 3. 11. Giao diện dự báo lĩnh vực giao thông vận tải - Transport.

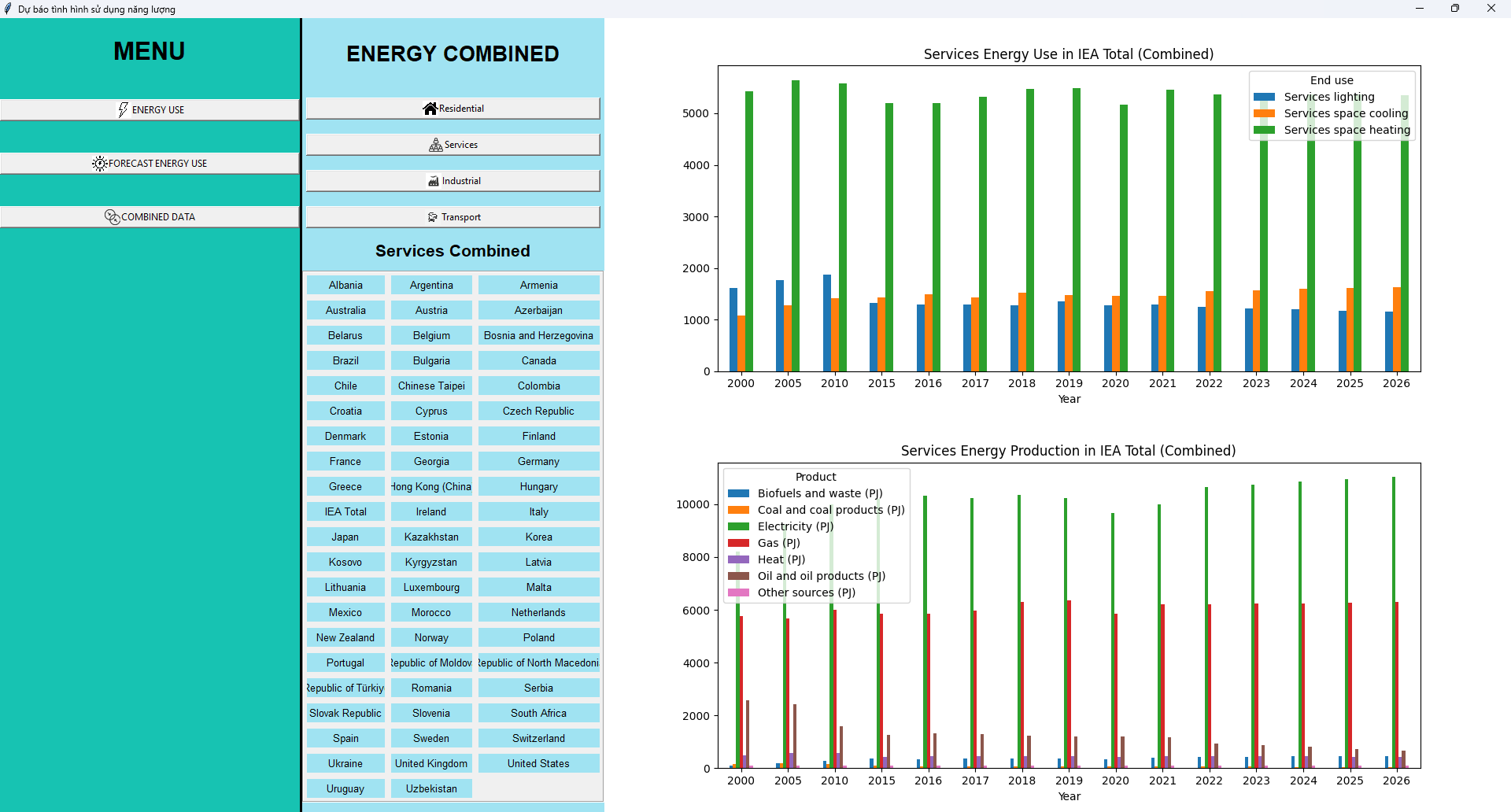
**3.7.5. Giao diện kết hợp dự báo và quá khứ.**

**3.7.5.1. Lĩnh vực dân dụng – Residential.**



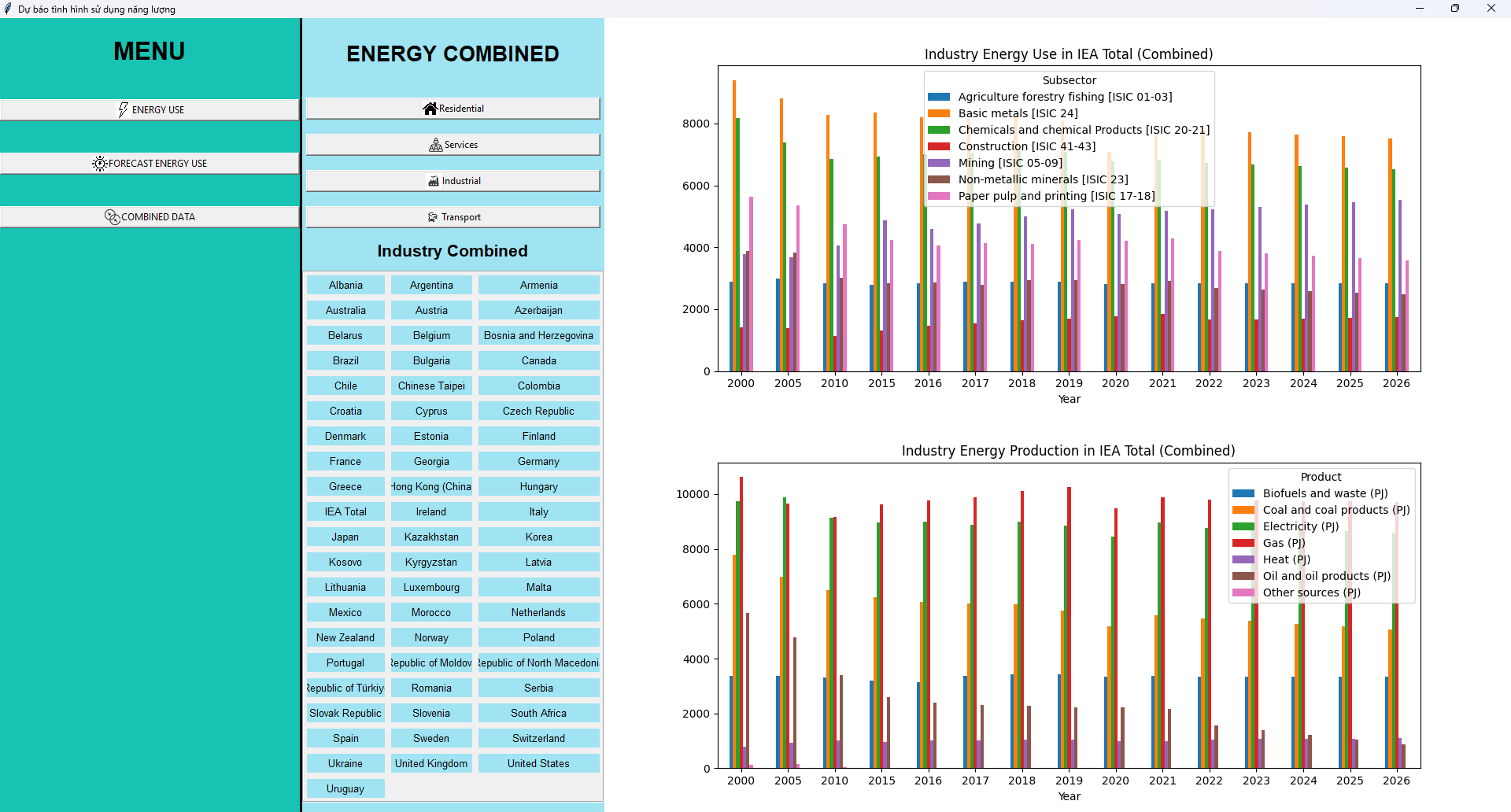
Hình 3. 12. Giao diện kết hợp dữ liệu lĩnh vực dân dụng Residential.

**3.7.5.2. Lĩnh vực dịch vụ - Services.**



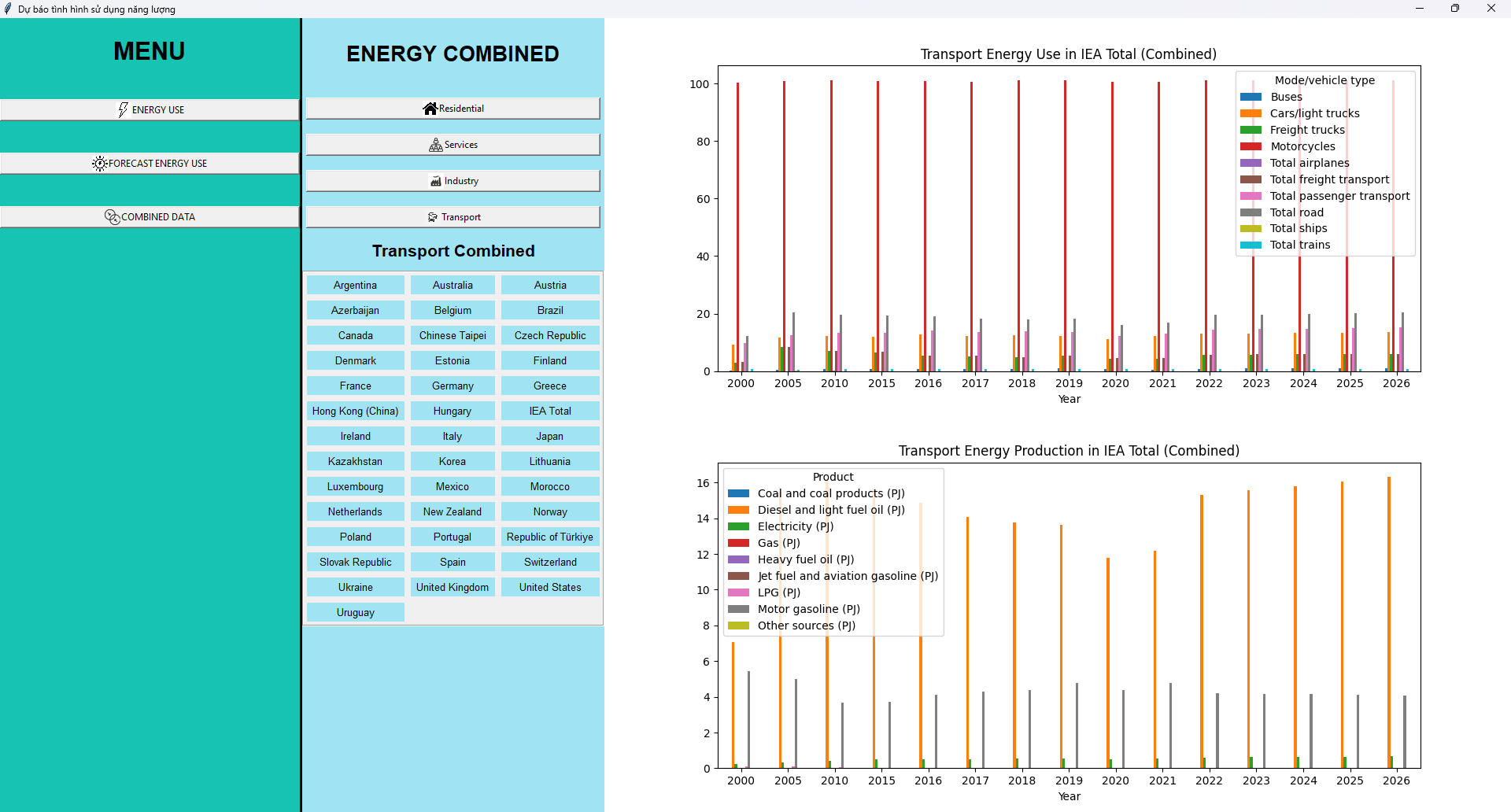
Hình 3. 13. Giao diện kết hợp dữ liệu lĩnh vực dịch vụ - Services.

**3.7.5.3. Lĩnh vực công nghiệp – Industry.**



Hình 3. 14. Giao diện kết hợp dữ liệu lĩnh vực công nghiệp - Industry.

**3.7.5.4. Lĩnh vực giao thông vận tải – Transport.**



Hình 3. 15. Giao diện kết hợp dữ liệu lĩnh vực giao thông vận tải - Transport.

**3.8. Các đề xuất và biện pháp cho các vấn đề về tài nguyên năng lượng.**

Ngành năng lượng toàn cầu đang trải qua giai đoạn chuyển đổi đầy thách thức, đòi hỏi sự đổi mới và hợp tác toàn diện để đảm bảo nguồn cung năng lượng ổn định, bền vững và thân thiện với môi trường. Dưới đây là một số đề xuất và biện pháp cụ thể, dựa trên phân tích dữ liệu, dự báo xu hướng và am hiểu về bối cảnh năng lượng toàn cầu:

- Nâng cao hiệu quả sử dụng năng lượng trong các ngành công nghiệp là chìa khóa để giảm thiểu tiêu thụ năng lượng, chi phí sản xuất và phát thải khí nhà kính. Cần thúc đẩy áp dụng các công nghệ sản xuất tiên tiến, tối ưu hóa quy trình sản xuất, sử dụng thiết bị tiết kiệm năng lượng, tái sử dụng năng lượng thải từ các quy trình công nghiệp, và đào tạo nâng cao nhận thức cho công nhân về sử dụng năng lượng hiệu quả.

- Các tòa nhà hay cao ốc là một trong những lĩnh vực tiêu thụ năng lượng lớn. Cần thiết kế và xây dựng các tòa nhà mới theo tiêu chuẩn xanh, sử dụng vật liệu cách nhiệt hiệu quả, hệ thống chiếu sáng LED tiết kiệm năng lượng, hệ thống thông gió tự nhiên và cơ khí hiệu quả. Cải tạo các tòa nhà hiện có để nâng cao hiệu suất năng lượng cũng là một giải pháp quan trọng. Bên cạnh đó, việc tích hợp năng lượng mặt trời vào các tòa nhà để sưởi ấm nước, chiếu sáng và làm mát cũng góp phần giảm thiểu tiêu thụ năng lượng từ lưới điện.

- Khuyến khích người tiêu dùng sử dụng các thiết bị gia dụng tiết kiệm năng lượng bằng cách đẩy mạnh chương trình dán nhãn năng lượng, cung cấp thông tin rõ ràng về hiệu suất năng lượng của các thiết bị, hỗ trợ tài chính cho người dân tiếp cận các thiết bị tiết kiệm năng lượng, và nâng cao nhận thức cộng đồng về việc sử dụng năng lượng hiệu quả trong gia đình.

- Cải thiện hiệu suất nhiên liệu cho phương tiện giao thông là một giải pháp quan trọng để giảm thiểu tiêu thụ năng lượng và phát thải khí nhà kính. Cần thúc đẩy sử dụng phương tiện giao thông công cộng, phát triển hệ thống giao thông công cộng hiệu quả, hiện đại và tiện lợi. Bên cạnh đó, việc khuyến khích sử dụng xe điện, xe hybrid và phát triển hạ tầng giao thông thông minh, kết nối, sử dụng năng lượng hiệu quả cũng góp phần giảm thiểu tiêu thụ năng lượng và phát thải.

- Đầu tư mạnh mẽ vào phát triển năng lượng tái tạo từ các nguồn như mặt trời, gió, địa nhiệt, thủy điện, và sinh khối. Đa dạng hóa danh mục năng lượng tái tạo giúp tận dụng tối đa tiềm năng của từng nguồn năng lượng theo điều kiện tự nhiên và địa lý của mỗi quốc gia.

- Xây dựng và phát triển lưới điện thông minh là yếu tố then chốt để tích hợp năng lượng tái tạo vào hệ thống điện quốc gia một cách hiệu quả. Lưới điện thông minh cho phép quản lý năng lượng linh hoạt, tối ưu hóa sản xuất và phân phối điện, tăng cường độ tin cậy và ổn định của hệ thống điện.

- Đầu tư vào nghiên cứu và phát triển các công nghệ năng lượng tái tạo thế hệ mới, năng lượng hydro, năng lượng hạt nhân thế hệ mới an toàn và hiệu quả hơn, và công nghệ thu giữ và lưu trữ carbon. Những công nghệ đột phá này hứa hẹn sẽ tạo ra những bước tiến lớn trong việc chuyển dịch sang nền kinh tế năng lượng sạch và bền vững.

- Cần từng bước giảm sự phụ thuộc vào năng lượng hóa thạch, hướng đến mục tiêu sử dụng năng lượng tái tạo và các nguồn năng lượng thay thế khác. Việc này đòi hỏi sự chuyển dịch mạnh mẽ trong cơ cấu năng lượng, đầu tư vào cơ sở hạ tầng năng lượng mới, và thay đổi thói quen tiêu dùng năng lượng.

- Tăng cường hợp tác quốc tế về năng lượng là chìa khóa để đảm bảo an ninh năng lượng toàn cầu. Các quốc gia cần chia sẻ kinh nghiệm, công nghệ và nguồn lực để cùng nhau phát triển năng lượng bền vững, giải quyết các thách thức chung, và ứng phó với biến đổi khí hậu.

- Ban hành các chính sách hỗ trợ phát triển năng lượng tái tạo và tiết kiệm năng lượng. Cung cấp ưu đãi về thuế, tài chính, đất đai, và thủ tục hành chính cho các dự án năng lượng tái tạo, khuyến khích doanh nghiệp và người dân đầu tư vào lĩnh vực này. Hỗ trợ các chương trình tiết kiệm năng lượng trong các ngành công nghiệp, tòa nhà, và giao thông vận tải.

- Thiết lập các tiêu chuẩn hiệu quả năng lượng bắt buộc cho các ngành công nghiệp, tòa nhà, và thiết bị gia dụng. Áp dụng các biện pháp giám sát, kiểm tra và xử phạt nghiêm minh đối với các trường hợp vi phạm.

- Thúc đẩy thị trường carbon để định giá phát thải khí nhà kính, khuyến khích các doanh nghiệp giảm phát thải và đầu tư vào công nghệ sạch.

- Tăng cường các hoạt động tuyên truyền, giáo dục và phổ biến kiến thức về năng lượng tái tạo, tiết kiệm năng lượng, và bảo vệ môi trường. Nâng cao nhận thức cộng đồng về vai trò quan trọng của năng lượng bền vững đối với sự phát triển kinh tế - xã hội và bảo vệ môi trường.

- Phát triển và ứng dụng các hệ thống quản lý năng lượng thông minh, sử dụng công nghệ IoT, Big Data và trí tuệ nhân tạo để giám sát, phân tích, dự báo và tối ưu hóa việc sử dụng năng lượng trong thời gian thực.

- Thúc đẩy quá trình số hóa trong ngành năng lượng, ứng dụng công nghệ số để nâng cao hiệu quả hoạt động, quản lý và điều hành hệ thống năng lượng, từ sản xuất, truyền tải, phân phối đến tiêu thụ năng lượng.

Các biện pháp và đề xuất này không chỉ giúp nâng cao hiệu quả sử dụng năng lượng và giảm thiểu tác động tiêu cực đến môi trường mà còn góp phần thúc đẩy sự phát triển bền vững và ổn định của ngành năng lượng toàn cầu trong tương lai.

**CHƯƠNG IV. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN.**

**4.1. Kết luận.**

Đồ án "Xây dựng ứng dụng phân tích và trực quan hóa dữ liệu về việc sử dụng năng lượng toàn cầu" đã đạt được một thành công đáng kể trong việc tạo ra một công cụ hữu ích, cung cấp cái nhìn trực quan và chi tiết về tình hình sử dụng năng lượng trên toàn thế giới. Ứng dụng này, được phát triển bằng ngôn ngữ lập trình Python và tích hợp các thư viện mạnh mẽ như Pandas, NumPy, Matplotlib, Scikit-learn, Tkinter và Pillow, cho phép xử lý dữ liệu hiệu quả, xây dựng mô hình dự báo, trực quan hóa kết quả và thiết kế giao diện người dùng thân thiện.

Ứng dụng cho phép người dùng dễ dàng truy cập, phân tích và trực quan hóa dữ liệu lịch sử về mức tiêu thụ năng lượng từ năm 2000 đến 2021 cho 63 quốc gia và vùng lãnh thổ, được phân loại theo 4 lĩnh vực chính: dân dụng, dịch vụ, công nghiệp và giao thông vận tải. Biểu đồ cột trực quan giúp người dùng dễ dàng so sánh mức tiêu thụ năng lượng theo từng loại năng lượng, theo thời gian và giữa các quốc gia.

Hơn nữa, ứng dụng sử dụng mô hình Linear Regression để dự báo xu hướng tiêu thụ năng lượng trong 5 năm tiếp theo (2022-2027) cho từng quốc gia, lĩnh vực và loại hình năng lượng, với độ chính xác cao được thể hiện qua chỉ số RMSE thấp. Giao diện người dùng được thiết kế đơn giản, trực quan và dễ sử dụng, cho phép người dùng linh hoạt lựa chọn quốc gia, lĩnh vực và loại dữ liệu mong muốn, ứng dụng sẽ tự động cập nhật biểu đồ tương ứng.

Đồ án này không chỉ cung cấp một công cụ hữu ích cho việc phân tích, đánh giá và dự báo tình hình sử dụng năng lượng toàn cầu, hỗ trợ việc ra quyết định liên quan đến năng lượng, mà còn góp phần nâng cao nhận thức về tầm quan trọng của việc sử dụng năng lượng hiệu quả và bền vững, thúc đẩy việc chuyển dịch sang các nguồn năng lượng sạch và tái tạo. Đây là một ví dụ điển hình cho việc ứng dụng khoa học dữ liệu và học máy để giải quyết vấn đề thực tế trong lĩnh vực năng lượng.

Mặc dù đạt được nhiều kết quả tích cực, đồ án vẫn còn một số hạn chế, chẳng hạn như việc sử dụng mô hình Linear Regression, tuy hiệu quả nhưng vẫn còn một số hạn chế nhất định, và việc sử dụng dữ liệu chỉ đến năm 2021. Trong tương lai, có thể nghiên cứu và áp dụng các mô hình dự báo phức tạp hơn để nâng cao độ chính xác và cập nhật dữ liệu mới nhất để đảm bảo tính thời sự cho ứng dụng.

Tóm lại, đồ án đã thành công trong việc tạo ra một công cụ hữu ích, đóng góp vào việc giải quyết vấn đề năng lượng toàn cầu. Ứng dụng có tiềm năng được phát triển và ứng dụng rộng rãi trong tương lai, hỗ trợ việc hoạch định chính sách, đầu tư, sản xuất và tiêu dùng năng lượng hiệu quả hơn.

**4.2. Hướng phát triển.**

Trong tương lai, giao diện người dùng có thể được nâng cấp bằng cách sử dụng các framework GUI mạnh mẽ hơn như PyQt hoặc Kivy, mang đến giao diện đẹp mắt, chuyên nghiệp và tùy biến cao hơn. Việc bổ sung thêm các tính năng tương tác cho biểu đồ, chẳng hạn như zoom, pan, hover, sẽ cho phép người dùng khám phá dữ liệu một cách chi tiết và trực quan hơn. Bên cạnh đó, tích hợp bản đồ để hiển thị dữ liệu năng lượng theo vùng địa lý sẽ giúp người dùng dễ dàng nhận biết sự phân bố và xu hướng tiêu thụ năng lượng trên toàn cầu. Việc thiết kế bảng điều khiển (Dashboard) tùy biến, cho phép người dùng lựa chọn và sắp xếp các biểu đồ, chỉ số quan trọng theo nhu cầu cũng là một hướng phát triển đáng cân nhắc.

Ngoài việc nâng cấp giao diện, chức năng của ứng dụng cũng có thể được mở rộng. Việc tích hợp trợ lý ảo sẽ cho phép người dùng tương tác với ứng dụng bằng giọng nói, đặt câu hỏi và nhận thông tin về dữ liệu năng lượng một cách tự nhiên và thuận tiện. Ứng dụng có thể được mở rộng để phân tích và dự báo lượng phát thải CO2 liên quan đến việc sử dụng năng lượng, giúp đánh giá tác động đến môi trường và hỗ trợ việc xây dựng các giải pháp giảm thiểu khí thải. Bổ sung chức năng đánh giá hiệu quả sử dụng năng lượng cho từng quốc gia, lĩnh vực thông qua các chỉ số và phân tích chuyên sâu sẽ giúp xác định những điểm yếu và đề xuất các giải pháp tiết kiệm năng lượng. Ứng dụng cũng có thể cho phép người dùng tạo ra các kịch bản năng lượng khác nhau, dự báo tác động của các chính sách năng lượng và đánh giá tính hiệu quả của các giải pháp năng lượng bền vững. Cuối cùng, việc cung cấp chức năng so sánh chi tiết giữa các quốc gia về mức tiêu thụ năng lượng, xu hướng, hiệu quả sử dụng năng lượng và phát thải CO2 sẽ hỗ trợ việc học hỏi kinh nghiệm và áp dụng các giải pháp hiệu quả.

Để nâng cao độ chính xác của mô hình dự báo, việc nghiên cứu và áp dụng các thuật toán học máy phức tạp hơn, chẳng hạn như mạng nơ-ron (neural networks) hoặc học tăng cường (reinforcement learning) là một hướng đi tiềm năng. Song song với đó, việc áp dụng các kỹ thuật xử lý dữ liệu nâng cao, chẳng hạn như phân tích chuỗi thời gian, xử lý nhiễu, và xử lý dữ liệu bị thiếu, sẽ cải thiện chất lượng dữ liệu đầu vào và nâng cao hiệu suất của mô hình. Việc sử dụng các kỹ thuật tối ưu hóa siêu tham số tự động, chẳng hạn như Bayesian Optimization hoặc Genetic Algorithms, sẽ giúp tìm kiếm tổ hợp siêu tham số tối ưu một cách hiệu quả hơn.

Việc cập nhật và mở rộng dữ liệu cũng là một yếu tố quan trọng. Tích hợp dữ liệu năng lượng thời gian thực từ các nguồn trực tuyến sẽ giúp ứng dụng luôn cập nhật và phản ánh tình hình sử dụng năng lượng một cách chính xác.

**DANH MỤC TÀI LIỆU THAM KHẢO.**

**Tiếng Việt:**

[1] Python cơ bản – Bùi Việt Hà, 2020, NXB Đại học Quốc gia Hà Nội.

[2] Đường vào lập trình Python – TS. Nguyễn Ngọc Giang, 2020, NXB Đại học Quốc gia Hà Nội.

[3] Năng lượng tái tạo và công nghệ khai thác – Nguyễn Đức Hạnh, 2020, NXB Xây dựng.

[4] Machine Learning cơ bản – Vũ Hữu Tiệp, NXB Khoa học và Kỹ Thuật.

[5] Lập trình với Python (Hành trang cho tương lai) – Học viện VIETSTEM, 2023, NXB Đại học Quốc gia Hà Nội

[6] Lập trình Python – Đoàn Minh Phụng, 2022, NXB Khoa học và Kỹ Thuật.

[7] Trí tuệ nhân tạo học máy và ứng dụng – Nguyễn Tất Bảo Thiên và Nguyễn Quốc Huy, 2022, NXB Thanh Niên.

[8] Data Strategy – Bernard Marr, 2019, NXB Tổng hợp TPHCM.

[9] Python lập trình thuật toán – Bùi Việt Hà, 2022, NXB Đại học Quốc Gia Hà Nội.

[10] Cơ sở năng lượng tái tạo – Vũ Ngọc Huyền, 2022, NXB Bách Khoa Hà Nội

**Tiếng Anh:**

[1] Python for Data Analysis, 3rd Edition - Wes McKinney, 2022, O'Reilly Media.

[2] Hands-On Machine Learning with Scikit-Learn, Keras & TensorFlow, 3rd Edition- Aurélien Géron, 2022, O'Reilly Media.

[3] Deep Learning with Python, 2nd Edition- François Chollet, 2021, Manning Publications.

[4] Energy for Future Presidents: The Science Behind the Headlines - Richard A. Muller, 2012, W. W. Norton & Company.

[5] Sustainable Energy - without the hot air - David J. C. MacKay, 2008, UIT Cambridge.

[6] Storytelling with Data: A Data Visualization Guide for Business Professionals - Cole Nussbaumer Knaflic, 2015, Wiley

[7] Data Visualization: A Practical Introduction - Kieran Healy, 2018, Princeton University Press.

[8] Python GUI Programming with Tkinter- Alan D. Moore, 2020, Packt Publishing

[9] Modern Tkinter for Busy Python Developers - Mark Roseman, 2020, Independently published.

[10] Regression: Book One, Series of Machine Learning with Scikit-Learn, Shu Liu, 2021.

[11] [Global Energy Review 2020 – Analysis - IEA](https://www.iea.org/reports/global-energy-review-2020) Báo cáo về dữ liệu năng lượng toàn cầu 2020 theo IEA.

[12] [Global Energy Review 2021 – Analysis - IEA](https://www.iea.org/reports/global-energy-review-2021) Báo cáo về dữ liệu năng lượng toàn cầu 2021 theo IEA.