**ỦY BAN NHÂN DÂN THÀNH PHỐ HỒ CHÍ MINH TRƯỜNG ĐẠI HỌC SÀI GÒN**

**Logo

Description automatically generated----------🙡🕮🙣----------**

**ĐỀ CƯƠNG NGHIÊN CỨU KHOA HỌC**

**PHƯƠNG PHÁP VÀ NGHIÊN CỨU KHOA HỌC TRONG CNTT**

**ĐỀ TÀI:**

**DỰ BÁO TIÊU THỤ ĐIỆN NĂNG BẰNG MACHINE LEARNING**

**MÃ HỌC PHẦN: 841324**

**NHÓM MÔN HỌC: 02**

**HỌC KỲ:** **02**

**NHÓM BÁO CÁO: Test**

**GVHD**: **ĐỖ NHƯ TÀI**

**THÀNH VIÊN**:

**3120410101 Đỗ Xuân Quý Dương**

**3121410515 Đặng Ngọc Đoan Trang**

**3123410306 Đình Đào Thái Sơn**

**3123410294 Nguyễn Anh Quân (Nhóm trưởng)**

**TP HỒ CHÍ MINH, THÁNG 5 NĂM 2025**

**ỦY BAN NHÂN DÂN THÀNH PHỐ HỒ CHÍ MINH**

**TRƯỜNG ĐẠI HỌC SÀI GÒN**

**Logo

Description automatically generated----------🙡🕮🙣----------**

**ĐỀ CƯƠNG NGHIÊN CỨU KHOA HỌC**

**PHƯƠNG PHÁP VÀ NGHIÊN CỨU KHOA HỌC TRONG CNTT**

**ĐỀ TÀI:**

**DỰ BÁO TIÊU THỤ ĐIỆN NĂNG BẰNG MACHINE LEARNING**

**MÃ HỌC PHẦN: 841324**

**NHÓM MÔN HỌC: 02**

**HỌC KỲ:** **02**

**NHÓM BÁO CÁO: Test**

**GVHD**: **ĐỖ NHƯ TÀI**

**THÀNH VIÊN**:

**3120410101 Đỗ Xuân Quý Dương**

**3121410515 Đặng Ngọc Đoan Trang**

**3123410306 Đình Đào Thái Sơn**

**3123410294 Nguyễn Anh Quân (Nhóm trưởng)**

**TP HỒ CHÍ MINH, THÁNG 5 NĂM 2025**

# Phân công và đánh giá

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Thứ tự** | **Họ tên** | **Mã số sinh viên** | **Nội dung công việc** | **Mức độ đóng góp** | **Giảng viên đánh giá** |
| 1 | Đỗ Xuân Quý Dương | 3120410101 | Đọc file kế hoạch, khảo sát các tài liệu liên quan, chuẩn bị đề cương nghiên cứu | 100% |  |
| 2 | Đặng Ngọc Đoan Trang | 3121410515 | Đọc file kế hoạch, khảo sát các tài liệu liên quan, viết bài khảo sát đề tài và chuẩn bị slide báo cáo | 100% |  |
| 3 | Đinh Đào Thái Sơn | 3123410306 | Đọc file kế hoạch, khảo sát các tài liệu liên quan, thực hiện survey dataset, kiểm nghiệm độ đo, input, output,…. | 100% |  |
| 4 | Nguyễn Anh Quân | 3123410294 | Lên kế hoạch hàng tuần, giám sát tiến độ và giao việc cho các thành viên nhóm, khảo sát các tài liệu liên quan | 100% |  |

# MỤC LỤC

[I. Lý do chọn đề tài 2](#_Toc197119669)

[1. Bối cảnh và tầm quan trọng của dự báo điện năng 2](#_Toc197119670)

[2. Thực trạng và những thách thức trong dự báo nhu cầu điện 2](#_Toc197119671)

[a. Các phương pháp dự báo truyền thống và hạn chế 2](#_Toc197119672)

[b. Xu hướng áp dụng Machine Learning trong dự báo điện năng 3](#_Toc197119673)

[II. Mục đích và phạm vi nghiên cứu 4](#_Toc197119674)

[1. Mục tiêu 4](#_Toc197119675)

[a. Thu thập và xử lý dữ liệu 4](#_Toc197119676)

[b. Quy trình thực hiện 4](#_Toc197119677)

[c. Đánh giá và tối ưu mô hình 4](#_Toc197119678)

[d. Triển khai và ứng dụng 4](#_Toc197119679)

[2. Phạm vi nghiên cứu 5](#_Toc197119680)

[a. Phạm vi dữ liệu 5](#_Toc197119681)

[b. Phạm vi kỹ thuật 5](#_Toc197119682)

[c. Phạm vi ứng dụng 5](#_Toc197119683)

[III. Các giả thuyết và cách tiếp cận 6](#_Toc197119684)

[1. Giả thuyết nghiên cứu 6](#_Toc197119685)

[2. Cách tiếp cận nghiên cứu 6](#_Toc197119686)

[IV. Tài liệu kham khảo 7](#_Toc197119687)

# Lý do chọn đề tài

## Bối cảnh và tầm quan trọng của dự báo điện năng

Điện năng là một trong những nguồn năng lượng quan trọng nhất, đóng vai trò cốt lõi trong mọi lĩnh vực của đời sống xã hội, từ sinh hoạt hàng ngày đến hoạt động sản xuất công nghiệp, thương mại và dịch vụ. Trong thời đại công nghiệp hóa, hiện đại hóa và chuyển đổi số, nhu cầu sử dụng điện ngày càng gia tăng mạnh mẽ, kéo theo những thách thức lớn trong việc cân bằng cung - cầu, tối ưu hóa sản xuất và vận hành lưới điện.

Trong bối cảnh này, việc **dự báo chính xác nhu cầu tiêu thụ điện** không chỉ giúp các công ty điện lực điều tiết nguồn cung một cách hợp lý mà còn có vai trò quan trọng trong việc:

* **Đảm bảo vận hành ổn định của hệ thống điện**: Giúp tránh tình trạng quá tải hoặc thiếu hụt điện năng, gây ảnh hưởng đến đời sống và hoạt động sản xuất.
* **Tối ưu hóa chi phí sản xuất và truyền tải điện**: Hạn chế việc huy động quá mức các nhà máy nhiệt điện đắt đỏ hoặc phải mua điện từ nguồn khác với giá cao.
* **Tăng cường hiệu quả sử dụng năng lượng**: Giúp người tiêu dùng và các đơn vị cung cấp điện có kế hoạch sử dụng điện hợp lý hơn, giảm lãng phí điện năng.

## 2. Thực trạng và những thách thức trong dự báo nhu cầu điện

* + 1. **Các phương pháp dự báo truyền thống và hạn chế**

Trước đây, dự báo điện năng chủ yếu dựa trên các phương pháp **thống kê truyền thống**, bao gồm:

* **Hồi quy tuyến tính (Linear Regression)**: Xây dựng mô hình dựa trên mối quan hệ giữa lượng điện tiêu thụ và một số yếu tố ảnh hưởng như thời tiết, thời gian.
* **Trung bình trượt (Moving Average)**: Dự báo dựa trên trung bình của các giá trị trước đó.
* **Phân tích chuỗi thời gian (ARIMA, SARIMA)**: Mô hình dự báo dựa trên dữ liệu lịch sử theo thời gian.

Tuy nhiên, những phương pháp này có **nhiều hạn chế**, đặc biệt trong bối cảnh dữ liệu ngày càng phức tạp:

* **Không mô hình hóa được các quan hệ phi tuyến** giữa nhiều yếu tố như thời tiết, thời gian trong ngày, và xu hướng tiêu thụ điện.
* **Độ chính xác bị giới hạn**, đặc biệt khi có sự thay đổi đột ngột (ví dụ: do sự kiện lớn, lễ hội, thay đổi chính sách).
* **Không tận dụng được lợi thế của dữ liệu lớn (Big Data)** trong kỷ nguyên số

### b. Xu hướng áp dụng Machine Learning trong dự báo điện năng

Nhờ sự phát triển mạnh mẽ của trí tuệ nhân tạo (AI) và dữ liệu lớn, **Machine Learning (ML) đã trở thành một công cụ quan trọng trong việc dự báo điện năng tiêu thụ**. So với các phương pháp truyền thống, ML mang lại nhiều lợi ích:

* **Có khả năng học và mô hình hóa các quan hệ phi tuyến**, giúp dự báo chính xác hơn.
* **Tự động tối ưu hóa mô hình**, giảm thiểu sự phụ thuộc vào giả định thống kê.
* **Khai thác dữ liệu đa dạng**, bao gồm dữ liệu thời tiết, lịch sử tiêu thụ, dữ liệu kinh tế - xã hội.

Các mô hình ML phổ biến trong dự báo điện năng bao gồm:

* **Random Forest, XGBoost:** Mô hình cây quyết định nâng cao, phù hợp với dữ liệu có nhiều đặc trưng.
* **Mạng nơ-ron nhân tạo (ANN):** Tự động học mẫu từ dữ liệu lớn.
* **Mô hình học sâu LSTM (Long Short-Term Memory):** Tối ưu cho dữ liệu chuỗi thời gian, giúp dự báo chính xác xu hướng tiêu thụ điện.

Với những lợi ích vượt trội, nghiên cứu này tập trung vào việc **ứng dụng Machine Learning để xây dựng mô hình dự báo điện năng tiêu thụ**, giúp nâng cao độ chính xác và tối ưu hóa việc quản lý hệ thống điện.

1. **Mục đích và phạm vi nghiên cứu**

## Mục tiêu

Dự án này nhằm xây dựng một mô hình Machine Learning để dự báo nhu cầu điện năng dựa trên các yếu tố như thời gian, thời tiết và các biến động đặc biệt (ví dụ: ngày lễ, kỳ nghỉ). Cụ thể, nghiên cứu này đặt ra các mục tiêu sau:

* + 1. **Thu thập và xử lý dữ liệu**
* Khảo sát và xử lý dữ liệu về về tiêu thụ điện năng
* Tiền xử lý các biến thời gian, các yếu tố liên quan
  + 1. **Quy trình thực hiện**
* Áp dụng các phương pháp Machine Learning để xây dựng mô hình dự báo điện năng.
* So sánh độ chính xác của các mô hình như hồi quy tuyến tính (Linear Regression), Random Forest, và XGBoost để tìm mô hình tối ưu
  + 1. **Đánh giá và tối ưu mô hình**
* Đánh giá mô hình bằng các thước đo như RMSE, MAE, MAPE.
* Tìm hiểu thêm cách cải thiện độ chính xác bằng cách tinh chỉnh tham số và chọn đặc trưng phù hợp
  + 1. **Triển khai và ứng dụng**
* Xây dựng API để cung cấp dự báo dưới dạng dịch vụ.
* Nghiên cứu tác động của biến đổi khí hậu đến như cầu điện trong tương lai
* Thu thập thêm dữ liệu để sử dụng cho các nghiên cứu sau này
  1. **Phạm vi nghiên cứu**
     1. **Phạm vi dữ liệu**
* **Nguồn dữ liệu:** Sử dụng tập dữ liệu “Electricity Load Forecasting” [[1](https://www.kaggle.com/datasets/saurabhshahane/electricity-load-forecasting)] từ Kaggle
* **Khoảng thời gian:** Tập dữ liệu bao gồm dữ liệu tiêu thụ điện trong khaongr thời gian x năm (tùy theo dữ liệu thực tế).
* **Biến số sử dụng:**Các đặc trưng như thời gian (datetime), tải điện (nat\_demand), các yếu tố thời tiết (T2M, QV2M, TQL, W2M), và biến liên quan đến ngày lễ (holiday, school).
  + 1. **Phạm vi kỹ thuật**
* **Phương pháp Machine Learning**: Chỉ tập trung vào **học có giám sát (Supervised Learning)**, không bao gồm học không giám sát hoặc reinforcement learning.
* **Thuật toán sử dụng**: Bao gồm **hồi quy tuyến tính, Random Forest, và XGBoost**. Không đi sâu vào các mô hình quá phức tạp như Transformer-based models.
* **Công cụ và thư viện**: Sử dụng **Python**, với các thư viện **Pandas, Scikit-learn, TensorFlow/Keras, Matplotlib, Flask/FastAPI**.
  + 1. **Phạm vi ứng dụng**
* **Mô hình dự báo ngắn hạn (Short-term Forecasting)**: Dự đoán nhu cầu điện trong vài giờ hoặc vài ngày tới.
* **Không nghiên cứu dự báo dài hạn (Long-term Forecasting)** như xu hướng tiêu thụ điện trong nhiều năm.
* **Không bao gồm yếu tố kinh tế, chính trị** như giá điện, chính sách năng lượng.

# Các giả thuyết và cách tiếp cận

* 1. **Giả thuyết nghiên cứu**
* **Giả thuyết 1:** Có mối quan hệ giữa các yếu tố thời tiết (nhiệt độ, độ ẩm, tốc độ gió, lượng mưa) và nhu cầu tiêu thụ điện năng.
* **Giả thuyết 2:** Các mô hình học máy hiện đại (Random Forest, XGBoost) có thể dự báo tải điện chính xác hơn so với các phương pháp thống kê truyền thống (ARIMA, hồi quy tuyến tính).
* **Giả thuyết 3:** Các yếu tố như ngày lễ và lịch học có ảnh hưởng đáng kể đến nhu cầu tiêu thụ điện.

1. **Cách tiếp cận nghiên cứu**

**Thu thập và tiền xử lý dữ liệu**

* + Kiểm tra và làm sạch dữ liệu (loại bỏ giá trị thiếu, xử lý dữ liệu bị nhiễu).
  + Chuẩn hóa dữ liệu để phù hợp với mô hình.

**Khám phá dữ liệu (EDA - Exploratory Data Analysis)**

* + Phân tích mối quan hệ giữa các biến đầu vào và tải điện.
  + Xây dựng biểu đồ để trực quan hóa xu hướng tiêu thụ điện.

**Xây dựng mô hình dự báo**

* + So sánh các phương pháp truyền thống (ARIMA, hồi quy tuyến tính).
  + Áp dụng các mô hình học máy và học sâu (Random Forest, XGBoost).

**Đánh giá mô hình**

* + Sử dụng các chỉ số như RMSE, MAPE, R² để đo lường độ chính xác.
  + So sánh hiệu suất của các mô hình để chọn ra mô hình tối ưu nhất.

**Triển khai và thử nghiệm**

* + Nếu có API, có thể thử triển khai mô hình để nhận dữ liệu đầu vào theo thời gian thực.
  + Đánh giá hiệu suất trong thực tế.

# Tài liệu kham khảo

[[1](https://www.kaggle.com/datasets/saurabhshahane/electricity-load-forecasting)] Shahane, S. (n.d.). “Electricity Load Forcasting [Data set]. Kaggle. <https://www.kaggle.com/datasets/saurabhshahane/electricity-load-forecasting>

[[2](https://www.mdpi.com/1996-1073/15/10/3566)] Aisyah, S., & Simaremare, A. (2022). Exploratory weather data analysis for electricity load forecasting using SVM and GRNN, case study in Bali, Indonesia. \*Energies\*, 15(20), 7648. <https://www.mdpi.com/1996-1073/15/10/3566>

[[3](https://orca.cardiff.ac.uk/id/eprint/103793/1/Manuscript.pdf)] Kuster, C., Rezgui, Y., & Mourshed, M. (2022). Electrical load forecasting models: A critical systematic review. \*Sustainable Cities and Society\*, 85, 104047. <https://orca.cardiff.ac.uk/id/eprint/103793/1/Manuscript.pdf>

[[4](https://ieeexplore.ieee.org/document/9604919/metrics#metrics)] IEEE Transactions on Power Systems. (2021). Anomaly detection in power consumption. \*IEEE Transactions on Power Systems\*, 36(4), 3456-3465. <https://ieeexplore.ieee.org/document/9604919/metrics#metrics>

[[5](https://www.researchgate.net/publication/348690914_Short-Term_Electricity_Load_Forecasting_with_Machine_Learning)] IEEE Transactions on Power System. (2022). Short-term load forecasting using LSTM. \**IEEE Transactions on Power Systems*\*, 35(3), 2345-1354. <https://www.researchgate.net/publication/348690914_Short-Term_Electricity_Load_Forecasting_with_Machine_Learning>