

# HƯỚNG DẪN SỬ DỤNG HỆ THỐNG DỰ ĐOÁN KHÍ THẢI TÀU BIỂN

Tài liệu Kỹ thuật

Ngày 30 tháng 12 năm 2025

## Mục lục

1	Tổng quan hệ thống	2
2	Yêu cầu Môi trường (Prerequisites)	2
3	Mô tả chi tiết các tệp tin	2
4	Quy trình vận hành (Step-by-Step Guide)	3
4.1	Bước 1: Huấn luyện Mô hình và đánh giá SHAP . . . . .	3
4.2	Bước 2: Dánh giá Hiệu suất . . . . .	3
4.3	Bước 3: Sử dụng Ứng dụng Web . . . . .	3
4.4	Model và Data . . . . .	3
5	Xử lý sự cố thường gặp	3

# 1 Tổng quan hệ thống

Hệ thống dự đoán khí thải tàu biển được xây dựng dựa trên các mô hình học máy (LightGBM) để ước lượng lượng khí thải ( $NO_x, SO_x, CO_2, \dots$ ) dựa trên các thông số kỹ thuật của tàu. Bộ mã nguồn bao gồm các quy trình từ huấn luyện mô hình, đánh giá hiệu suất đến triển khai ứng dụng web.

## 2 Yêu cầu Môi trường (Prerequisites)

Để chạy được các tệp tin trong hệ thống, môi trường Python cần được cài đặt các thư viện sau:

- **Xử lý dữ liệu:** pandas, numpy
- **Mô hình hóa:** scikit-learn, lightgbm
- **Trực quan hóa:** matplotlib, seaborn
- **Giao diện Web:** gradio

Cài đặt nhanh

```
pip install pandas numpy scikit-learn lightgbm matplotlib seaborn gradio
```

## 3 Mô tả chi tiết các tệp tin

Dưới đây là chức năng của từng notebook trong thư mục dự án:

### 1. ship-exhaust.ipynb (Core Training):

- *Chức năng:* Đây là tệp chính dùng để xử lý dữ liệu thô, huấn luyện mô hình MultiOutputRegressor với LightGBM.
- *Đầu ra:* Mô hình đã huấn luyện và các tệp dự đoán sơ bộ.

### 2. ship-emissions-base-evaluate.ipynb (Baseline Evaluation):

- *Chức năng:* Dánh giá hiệu suất của mô hình cơ sở (baseline) chưa qua tinh chỉnh.
- *Mục đích:* Thiết lập mức chuẩn để so sánh hiệu quả sau khi tuning.

### 3. ship-emissions-evaluate.ipynb (Final Evaluation):

- *Chức năng:* Dánh giá chi tiết mô hình sau khi đã tối ưu hóa. Tệp này vẽ các biểu đồ so sánh  $R^2$  và MAPE cho từng loại khí thải (E1 đến E5).
- *Đầu ra:* Biểu đồ trực quan hóa sai số.

### 4. ship-emissions.ipynb (Deployment):

- *Chức năng:* Chứa mã nguồn cho giao diện người dùng (GUI) sử dụng thư viện Gradio.
- *Cách dùng:* Cho phép người dùng tải lên file CSV thông số tàu và nhận về file kết quả dự đoán khí thải.

## 4 Quy trình vận hành (Step-by-Step Guide)

### 4.1 Bước 1: Huấn luyện Mô hình và đánh giá SHAP

Khởi chạy tệp `ship-exhaust.ipynb`.

- Đảm bảo đường dẫn tới bộ dữ liệu huấn luyện là chính xác. (Đã làm)
- Chạy tất cả các cell (Run All). (Đã làm)
- Quá trình này sẽ thực hiện tiền xử lý (Encoder), training và lưu lại các kết quả trung gian. (Đã làm)
- Xem SHAP ở các cell đầu tiên.

### 4.2 Bước 2: Đánh giá Hiệu suất

Sau khi có kết quả từ bước 1, sử dụng hai tệp đánh giá để kiểm tra chất lượng mô hình.

- Mở `ship-emissions-evaluate.ipynb`.
- Mở `ship-emissions-base-evaluate.ipynb`.
- Chạy cell vẽ biểu đồ để quan sát chỉ số  $R^2$
- Tiêu chí đạt:  $R^2 > 0.9$  và MAPE thấp (dưới 10%).

### 4.3 Bước 3: Sử dụng Ứng dụng Web

Để sử dụng mô hình trong thực tế hoặc demo:

1. Mở tệp `ship-emissions.ipynb`.
2. Chạy toàn bộ cell chứa lệnh.
3. Một đường link sẽ hiện ra. Truy cập link này trên trình duyệt.
4. **Thao tác trên Web:**

- Nhấn nút **Upload File** để tải lên file CSV chứa thông số tàu mới.
- Nhấn **Predict** để hệ thống xử lý.
- Tải về file kết quả dự đoán tại khung đầu ra.

### 4.4 Model và Data

1. model: model
2. data: data

## 5 Xử lý sự cố thường gặp

Liên hệ