[TheRealOnes x Udata]
Energy consumption prediction with machine learning
Members: Kim JongOh; Dien Ngo Bao; Quang Tran Ming; Nghia Pham Quang; Trang Doan Phan Thuy
Academic supervisor: Dr Son Dao Vu Truong, Dr Khuong Nguuyen Vinh
Industrial supervisor: Mr Nguyen Toan [Udata]

Background and Motivation

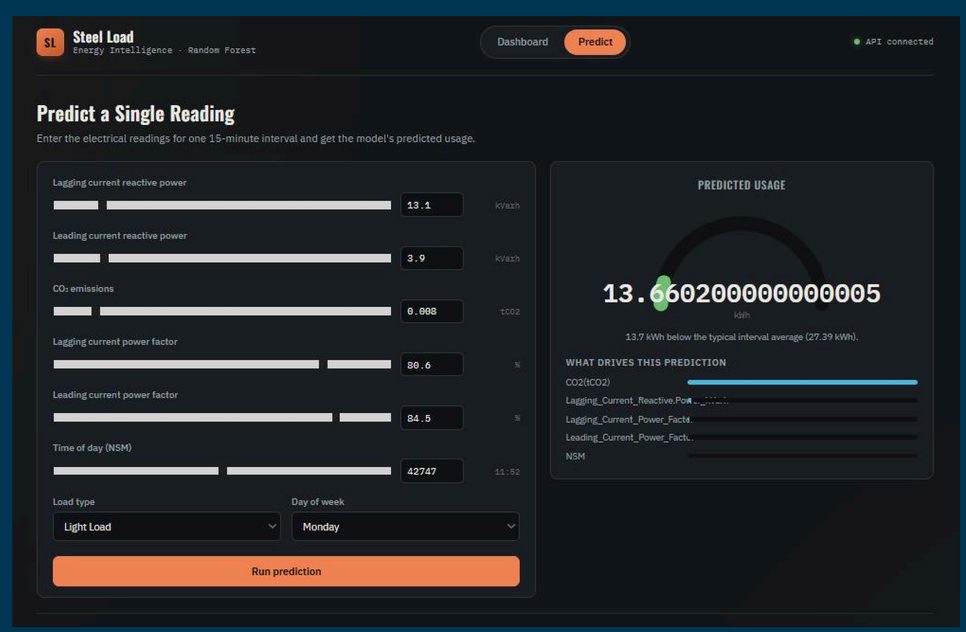
This capstone project addresses the critical gap in industrial manufacturing where traditional energy monitoring remains reactive and disconnected from operational planning. To transition facilities from reactive observation to proactive optimization, this project develops Uboard, a "schedule-aware" smart energy management platform designed for the manufacturing sector. While the project was initially framed around continuous time-series architectures (such as SARIMA), initial data analysis revealed that the manufacturing datasets are fundamentally job-based rather than purely time-based.

Consequently, the modeling strategy has pivoted to prioritize regression-based ensemble methods (Linear Regression, Random Forest, XGBoost, and LightGBM), with preliminary experiments yielding a peak R^2 score of 0.869 using the Random Forest Regressor. Upon progression of the project, the team has finalized the choice of model as a random forest model, as XGBoost and Linear regression proved to be lacking in terms of accuracy and had a tendency to show overfitting, due to the quantitative lack of provided resources.

Tested against real-world manufacturing datasets, this dual-module pipeline delivers both technical reliability and actionable business metrics, establishing a scalable, reproducible framework aligned with Vietnam's national priority of AI-driven energy optimization and Net Zero sustainability goals.

Objectives

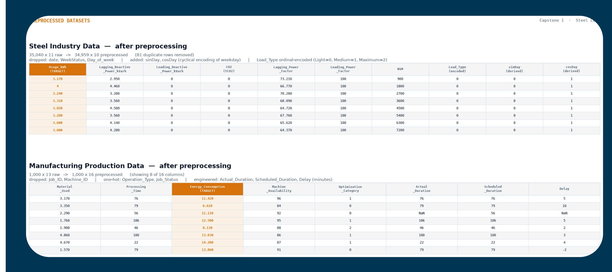
This project aims to provide a visualization of energy consumption and automate the process with the integration of machine learning model, assisting in making industrial energy saving a step more efficient and find out the patterns that cost more energy consumption, such as time, work type, work load.

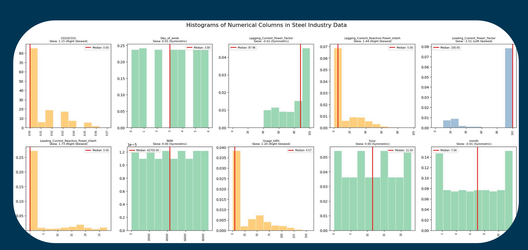


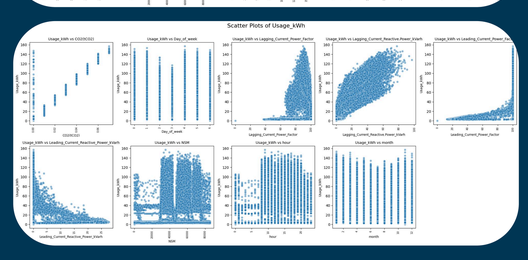
Methodology

Data acquisition

Along with the datasets provided by Udata, the team scoured for public datasets with meaningful datasets with nuance to energy consumption, work load and chronological data







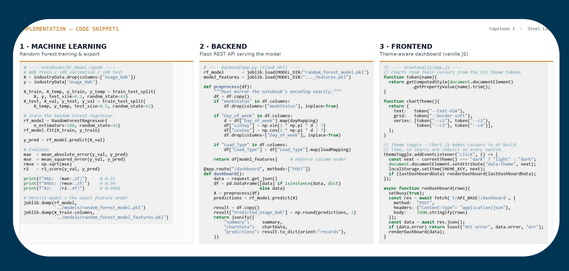
Model building

The datasets were pre-processed. These steps include:

1. **deleting unnecessary features**
2. **replacing empty entries**
3. **Creating new features with existing fields when necessary**
4. **Hot-encoding features for smoother ML learning**

Testing & development

The model was tested with 90% of the dataset to be learned, and the 10% for testing. Should the result show signs of overfitting or large error in accuracy, the model would be scrapped, the datasets revisited for further refining, and different models would be used for different attempts.



Dashboard visualization

In order to visualize the results, the team has assembled a simple dashboard with frontend-backend integration to show the prediction result in a more concise, clear and graphic display

Conclusion & Future developments

The team has assembled a learning model to learn the data of energy consumption patterns and predict the upcoming energy usage and have successfully visualized it onto a dashboard.

The suggestions for development comes in a few ways:

1. **Larger datasets and longer learning experiences enhance accuracies further.**
2. **For energy consumption reduction, a more keen monitoring of machinery power status is suggested.**

Kim JongOh - s3726103@rmit.edu.vn
Dien Ngo Bao - s4058810@rmit.edu.vn
Quang Tran Minh - s4098857@rmit.edu.vn
Nghia Pham Quang - s3927459@rmit.edu.vn
Trang Doan Phan Thuy - s4027648@rmit.edu.vn