Energy Usage Prediction and Savings Estimation Using AI

Overview

The goal is to use historical weather data and energy consumption patterns to:

- Predict energy consumption before deploying AltoTech AI.
- Estimate potential energy savings after AI deployment.
- Visualize insights and predictions in an interactive Plotly Dash web application.

Medtrodology

Data Understanding & Processing

Datasets Used:

- **Before AI Plant Data (`before_ai_plant_data.csv`)**
  ➢ Energy usage from 2023-01-01 to 2024-07-28

- **After AI Plant Data (`after_ai_plant_data.csv`)**
  ➢ Energy usage from 2024-09-02 to 2025-03-03

- **Weather Data (`weather_data_Thailand.csv`)**
  ➢ Historical weather data for Bangkok from 2023-01-01 to 2025-03-03, including temperature, humidity, windspeed, and UV index.
  ➢ Source: Visual Crossing
  ➢ Link: [Visual Crossing Weather Query Builder]

Processing Steps:

- Renamed energy columns for clarity (`plant_energy_before`, `plant_energy_after`)
- Checked for null or missing values (none found)
- Converted date columns to `datetime` format

Exploratory Data Analysis (EDA)

Descriptive Statistics: Energy Consumption Statistics

Metric	Before AI (575 days)	After AI (183 days)	Change
Mean Usage	54,182.70 kWh	44,969.70 kWh	▼ 9,213.00 kWh
Median Usage	56,070.00 kWh	47,632.00 kWh	▼ 8,438.00 kWh
Minimum Usage	27,481.00 kWh	25,528.00 kWh	▼ 1,953.00 kWh
Maximum Usage	71,375.00 kWh	56,882.00 kWh	▼ 14,493.00 kWh

Exploratory Data Analysis (EDA)

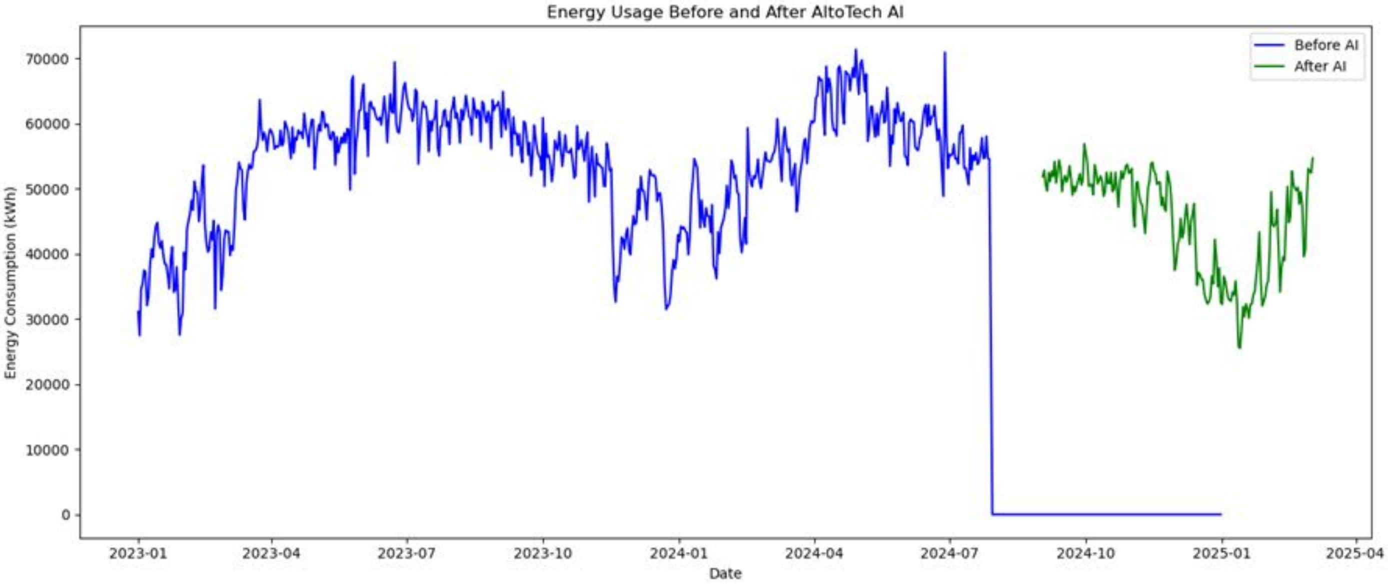
Descriptive Statistics: Weather Data Summary

Feature	Mean	Median	Min	Max	Std Dev
Temp (°C)	29.54	29.60	22.90	34.50	1.78
Temp Max (°C)	33.84	33.80	26.30	40.00	2.03
Temp Min (°C)	26.15	26.30	18.10	31.00	2.12
Feels Like (°C)	33.52	33.90	22.90	42.60	3.77
Humidity (%)	68.70	69.50	46.60	87.80	8.71
Wind Speed (km/h)	16.26	16.10	5.90	43.50	5.56
UV Index	6.61	7.00	2.00	10.00	1.57

Exploratory Data Analysis (EDA)

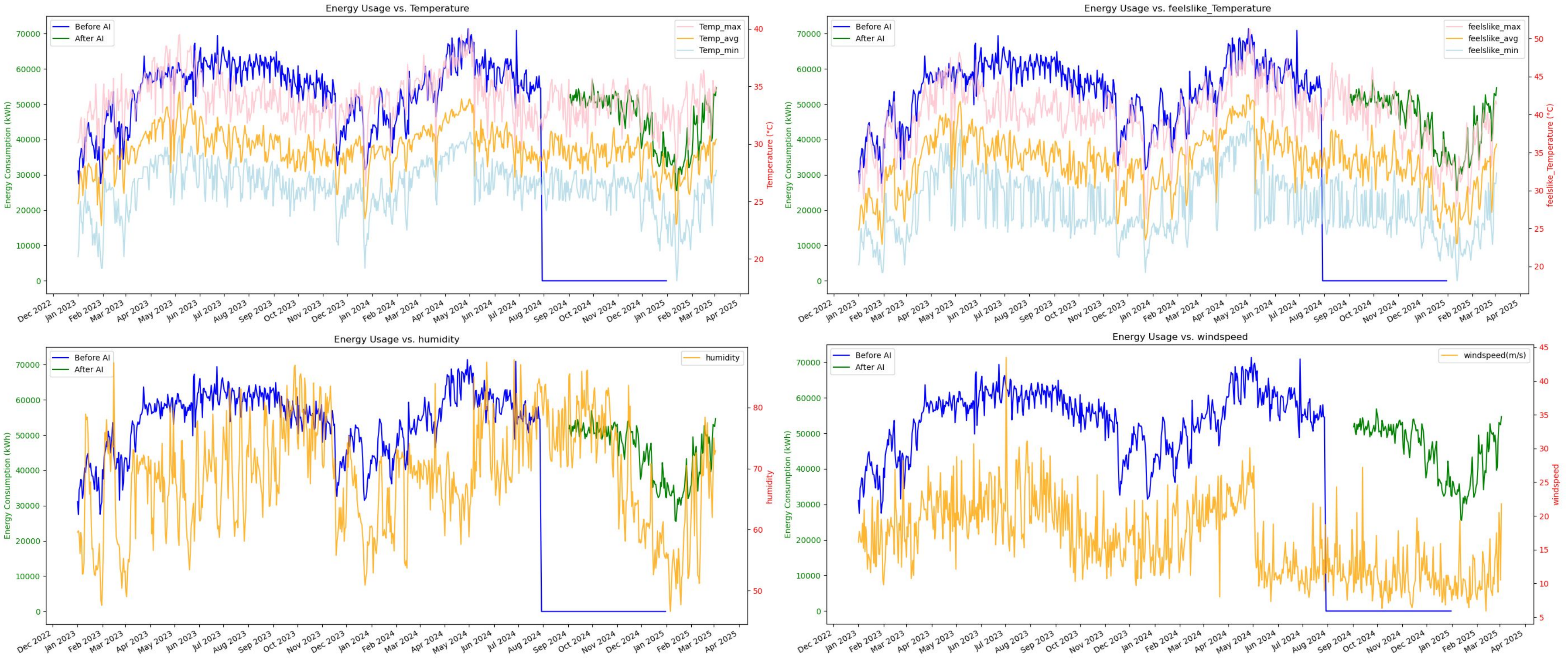
Visualization

Plot line graph to see characteristic graph of using energy before and after deploying AI.



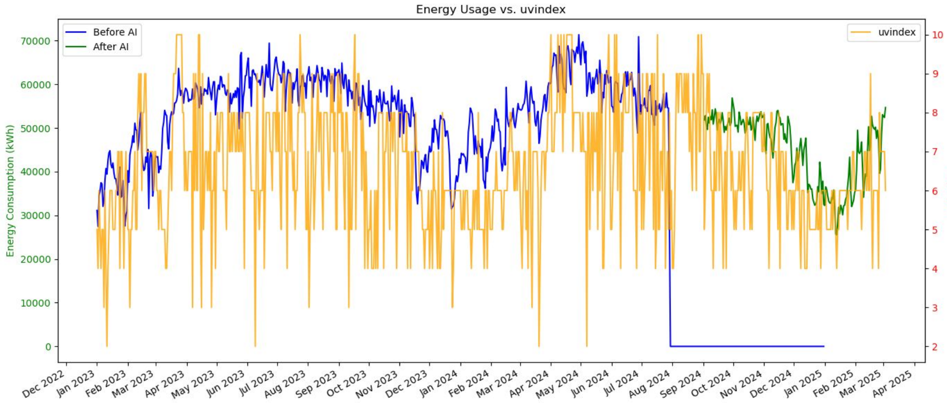
Exploratory Data Analysis (EDA)

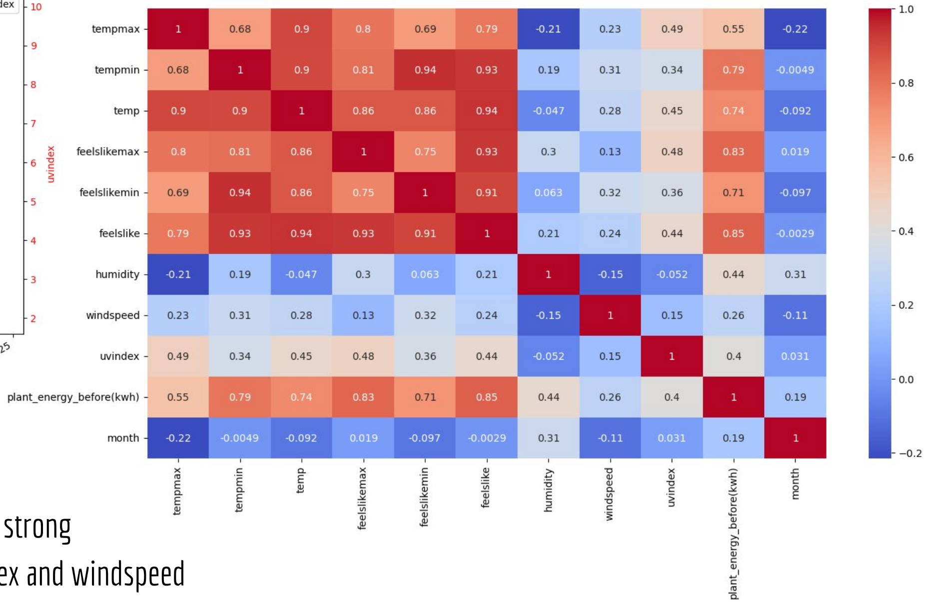
Plot line graph to compare historical weather data with energy consumption and heat map to see correlations.



Exploratory Data Analysis (EDA)

Plot line graph to compare historical weather data with energy consumption and heat map to see correlations.





Visual Insights

- All weather data have relationships with energy consumption
- Tempmax, tempmin, temp, feelslikemax, feelslikemin and feelslike had strong relationship with energy consumption and follow with humidity, uvindex and windspeed
- Energy usage decreases in cooler months (January, February, December).

Feature Selection

Selected Features:
['tempmin', 'humidity', 'feelslike', 'windspeed', 'uvindex', 'month']

Justifications:

The selection of features was guided by both **correlation heatmaps** and **visual exploration graphs**, which revealed both **linear and non-linear relationships** with energy consumption.

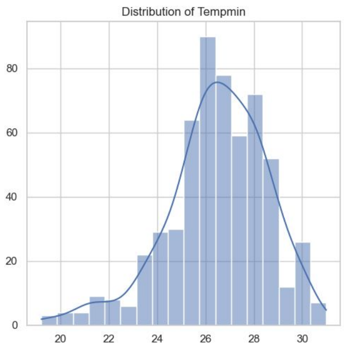
Linear Relationships:

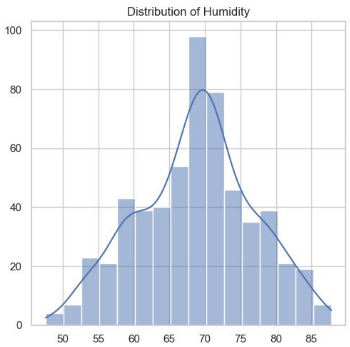
- **Temperature Variables**: Multiple temperature-related variables (tempmin, tempmax, temp) showed strong linear relationships with energy demand. Among them, **tempmin** had the **highest correlation** and was chosen to avoid redundancy. This choice simplifies the model while preserving predictive power.
- **Humidity**: Clearly showed a positive linear trend with energy usage in scatter plots — higher humidity often leads to higher air conditioning load.
- **Feelslike**: This feature, which reflects perceived temperature by combining **temperature**, **humidity**, and **wind**, also showed a **strong linear correlation**. Only the general feelslike value was kept, while **feelslikemax** and **feelslikemin** were dropped to avoid duplication.
- **Windspeed & UV Index**: Both had a visible trend in scatter plots indicating some linear relationship with energy demand. For instance, higher windspeed may reduce cooling loads in some settings, while higher UV index might increase indoor cooling needs due to solar gain.

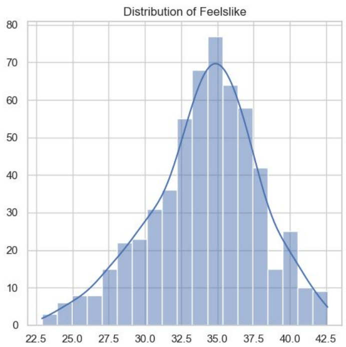
Non-linear Relationship:

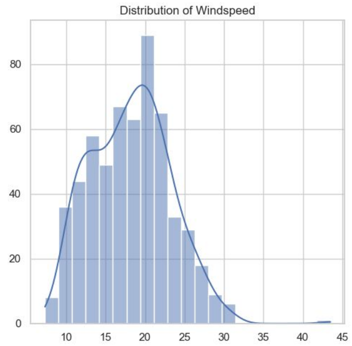
- **Month** demonstrated a weaker linear correlation but revealed a **clear seasonal pattern** in visual plots. Energy usage changes across the months, likely due to weather variations such as hotter or cooler seasons affecting heating or cooling needs. Because of this, we kept "month" as a feature to help the model capture these seasonal patterns.

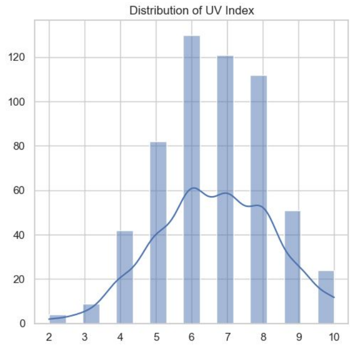
Feature Engineering











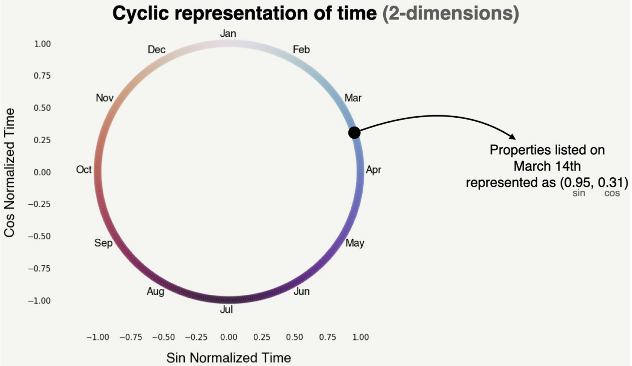
Distribution Analysis:
Most features ≈ normal → suitable for StandardScaler

Feature Scaling:
Applied StandardScaler for consistency.

This is appropriate given the mostly normal distributions of the input features.

Encoding of Month:

➤ Used sinusoidal encoding:
month_sin = sin(2π * month / 12)
month_cos = cos(2π * month / 12)



Model Selection

Models Compared (via Cross-Validation):

1. Linear Regression
2. Ridge Regression
3. Random Forest
4. Gradient Boosting

Model	R^2	RMSE	CVRMSE	NMBE
Linear Regression	0.8752	2,983.03	0.0551	0.0001
Ridge Regression	0.8752	2,982.93	0.0551	0.0001
Random Forest	0.8839	2,888.11	0.0534	-0.0019
Gradient Boosting	0.8831	2,893.94	0.0534	-0.0001

Model Insights:

- Tree-based models (Random Forest & Gradient Boosting) performed better than linear models.
- Both had higher R^2 and lower RMSE, CVRMSE, and near-zero NMBE.
- Captured non-linear patterns more effectively.

Model Selection

Model Tuning (via GridSearchCV):

➡ **Random Forest (Best Params)**

- `n_estimators=200`, `max_depth=20`, optimal split settings
- **R^2**: 0.884 | **RMSE**: ~2,774

➡ **Gradient Boosting (Best Params)**

- `n_estimators=100`, `learning_rate=0.05`, `max_depth=5`, `subsample=0.8`
- **R^2**: 0.887 | **RMSE**: ~2,743
- **Best Performing Model**

Metric Interpretations

The **R^2 score** is 0.8866, meaning the model explains about 89% of the variation in the target data. This indicates a very good fit and shows that the model captures patterns in the data effectively.

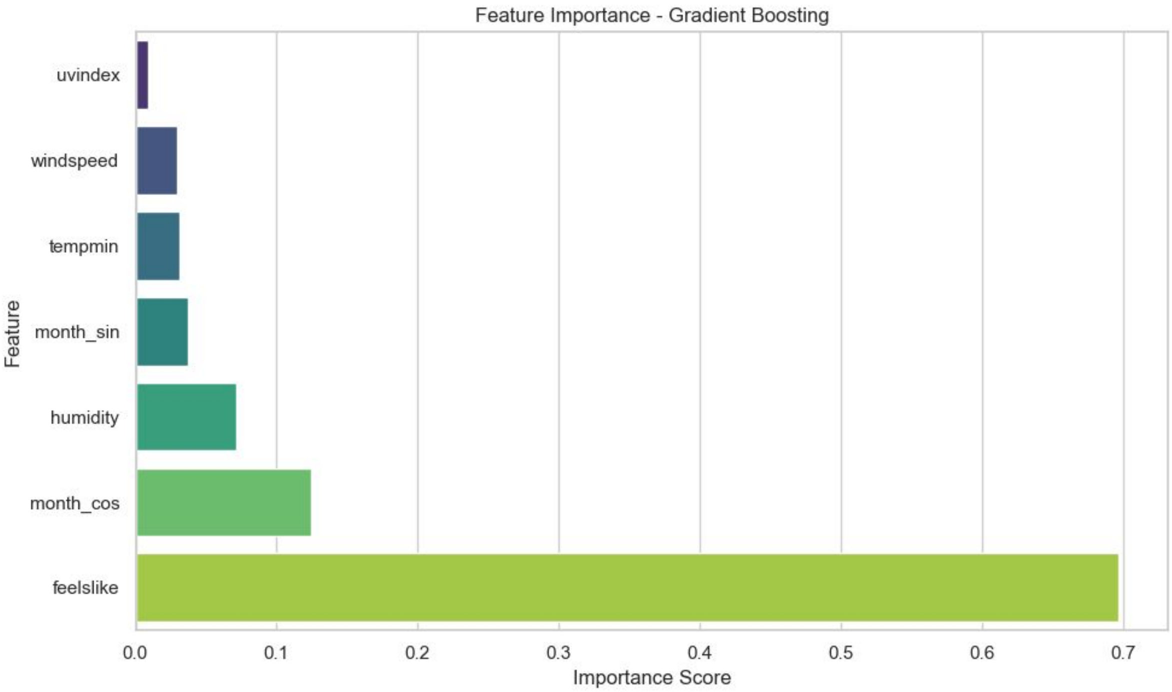
The **RMSE (Root Mean Squared Error)** is 2742.68, which shows the average difference between predicted and actual values. A lower RMSE means better prediction accuracy, and this value suggests the model performs well.

The **CVRMSE (Coefficient of Variation of RMSE)** is 0.0503, a low value that indicates consistent and reliable predictions, with small errors relative to the average values.

The **NMBE (Normalized Mean Bias Error)** is -0.0028, which is very close to zero. This means the model has very little bias, and does not consistently overpredict or underpredict.

Analysis

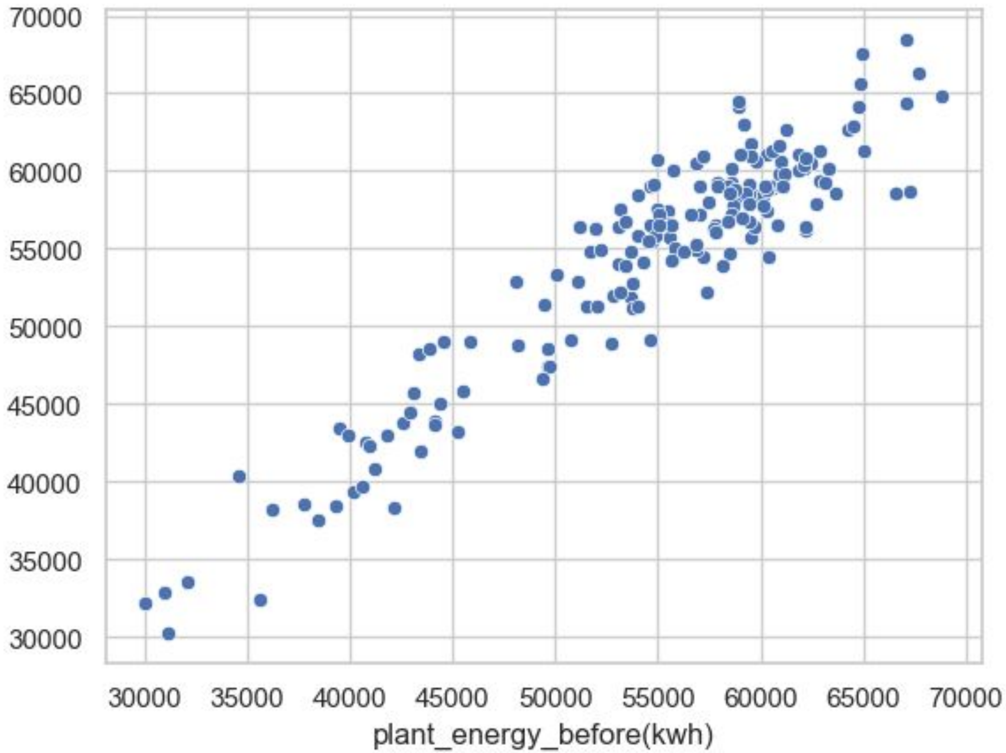
Feature Importance



Key Insights:

- **`feelslike` dominates** the feature space, accounting for ~70% of the total importance. This suggests that perceived temperature has the strongest influence on energy usage, likely because it most directly affects heating and cooling demand.
- **Temporal effects** (`month_cos` and `month_sin`) together account for ~16% of the importance. This captures seasonal patterns in energy usage, where cosine and sine encoding help reflect the cyclical nature of months.
- **`humidity`** is moderately important (~7.1%), which aligns with expectations, as it can influence how indoor temperatures feel and thus impact HVAC usage.
- **`tempmin` and `windspeed`** have relatively small contributions (~3% each), suggesting that while they do affect energy demand, their impact is less than `feelslike` or seasonal patterns.
- **`uvindex` has minimal influence**, likely because its effect on daily energy consumption is indirect or negligible in this context.

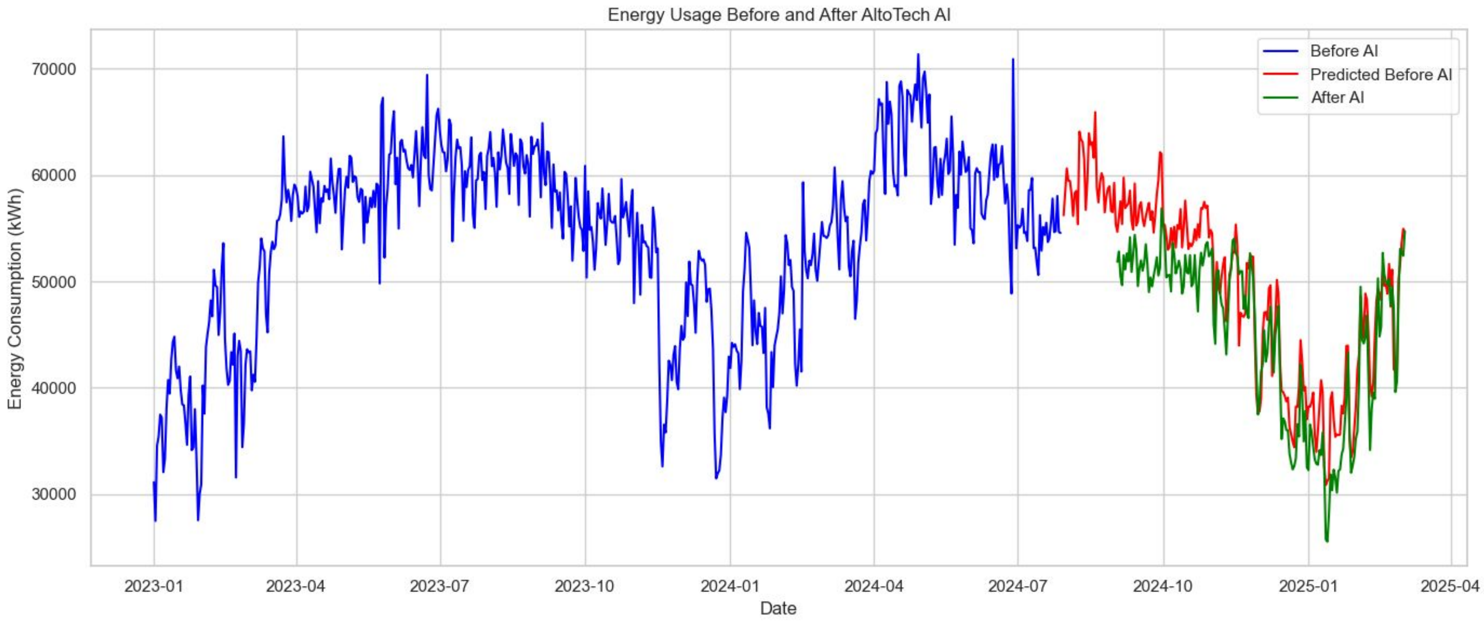
Analysis


70000
65000
60000
55000
50000
45000
40000
35000
30000
30000
35000
40000
45000
50000
55000
60000
65000
70000
plant_energy_before(kwh)

Predicting with the best model

Energy Prediction

Predict energy consumption before deploying AltoTech AI.
From 29-07-2024 to 03-03-2025



Predicted VS. Actual Energy Use

Compare predicted energy with actual energy

- **Total predicted (baseline):** 8740213.54 kWh
- **Total actual usage:** 8229454.25 kWh

Savings Calculation:

- **Savings (kWh)** = Predicted Baseline (before AI) - Actual Energy Consumption (after AI)
- **Total Savings**: 510,759.29 kWh
- **Savings Percentage**: 5.86% of the predicted energy consumption.

Cost Savings Estimate:

- Assuming a cost of 3.72 THB per kWh, the **estimated cost savings** over the 7-month period was **1,900,024.58 THB**.
- **Monthly Cost Savings**: 271,432.08 THB.

Dash Web Application

Introduce Web Application

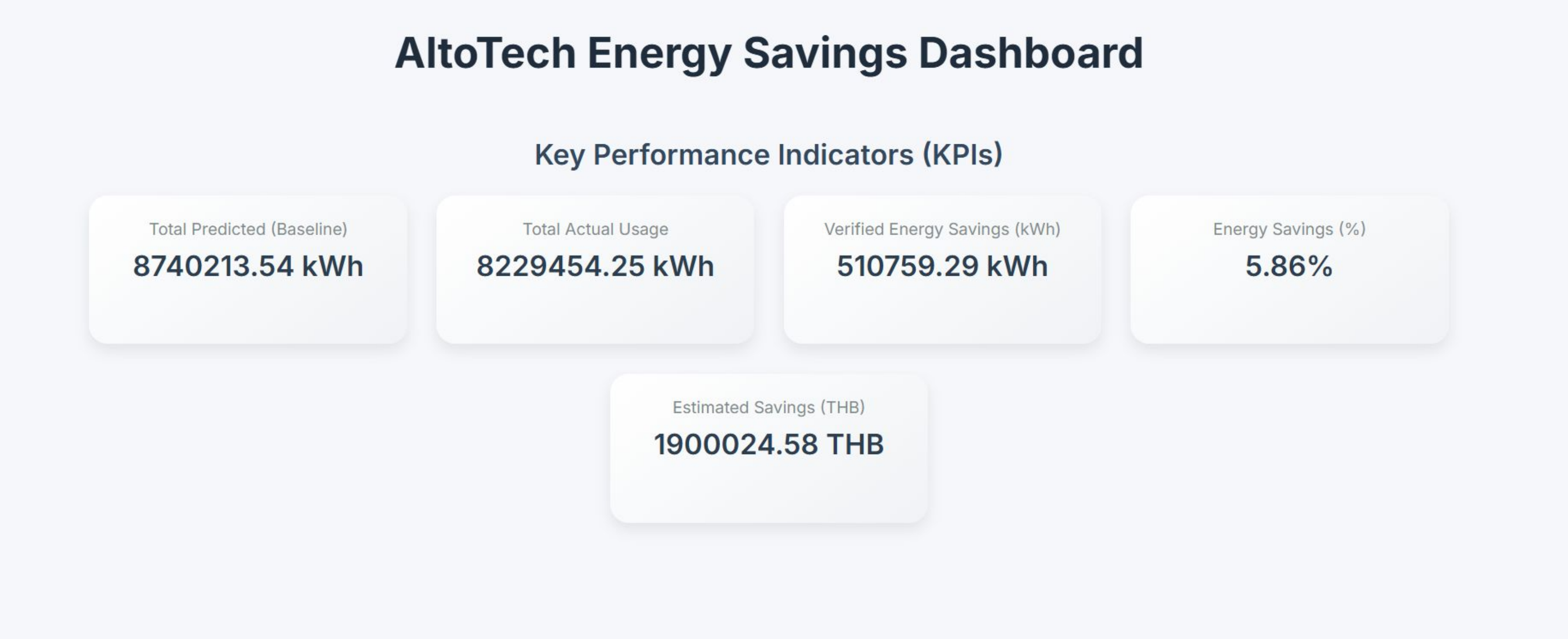
A functional Plotly Dash web application was developed to showcase the analysis to explore the energy savings.

Key features of the dashboard include:

1. **KPIs Display**: Showcasing verified savings in kWh, percentage, and potential cost savings.
2. **Visual Comparison**: Interactive plots to compare predicted baseline energy use with actual energy consumption over time.
3. **Weather Insights**: Weather trends (e.g., temperature, humidity) alongside energy consumption data.
4. **Model Transparency**: Displaying key model metrics like R^2, CVRMSE, and NMBE.
5. **Interactivity**: Users can zoom in on specific time periods and filter data to explore different trends.

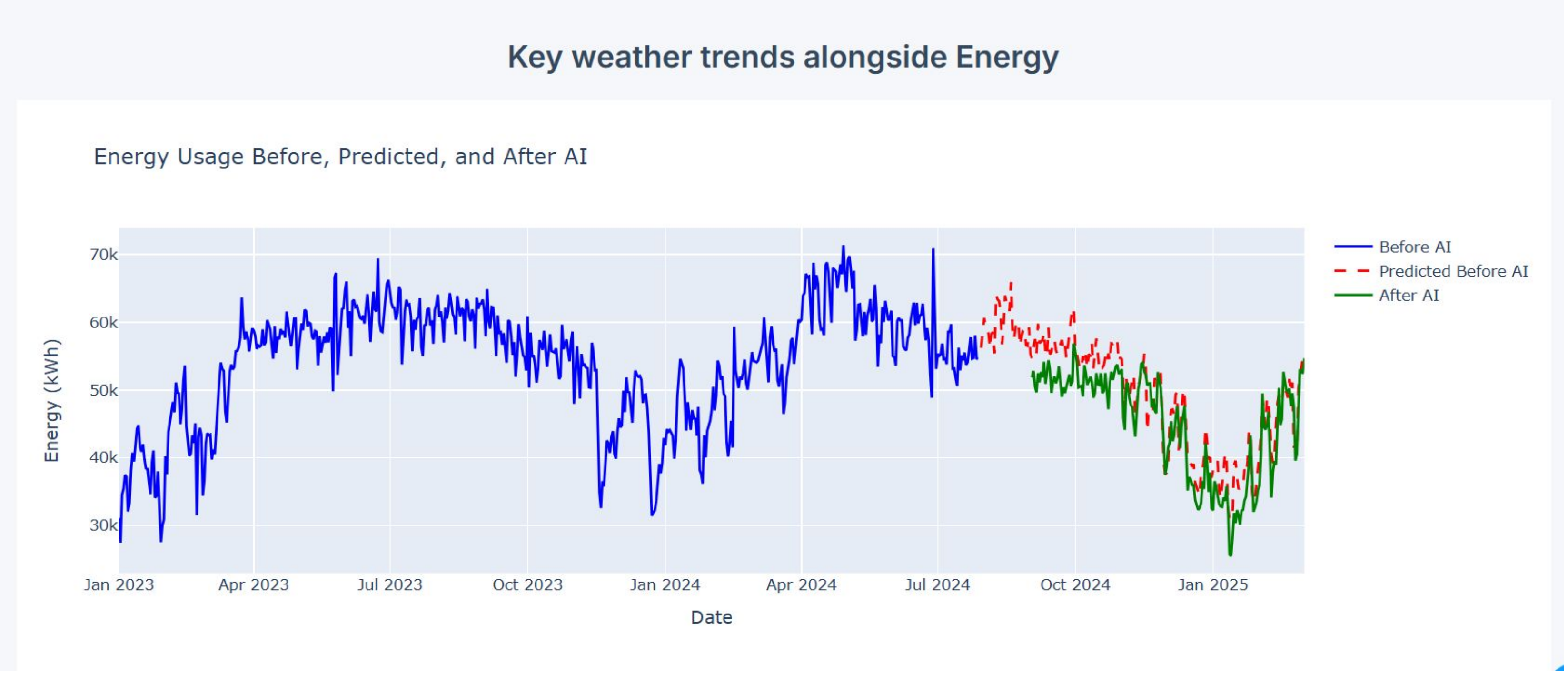
KPIs Display

Showcasing verified savings in kWh, percentage, and potential cost savings.



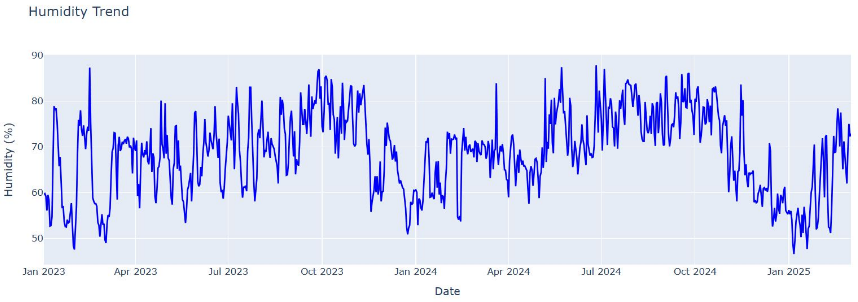
Visual Comparison

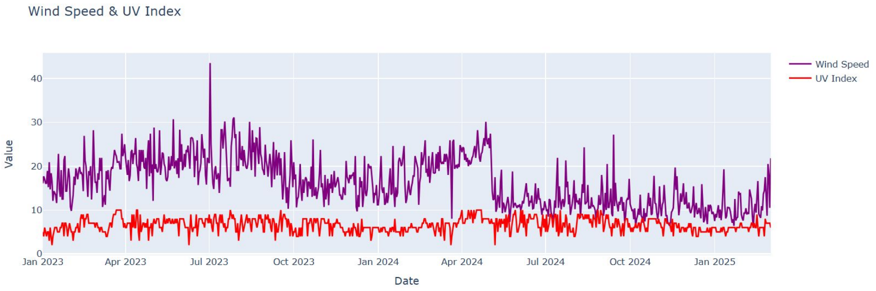
Interactive plots to compare predicted baseline energy use with actual energy consumption over time.



Weather Insights

Weather trends (e.g., temperature, humidity) alongside energy consumption data.





Model Transparency

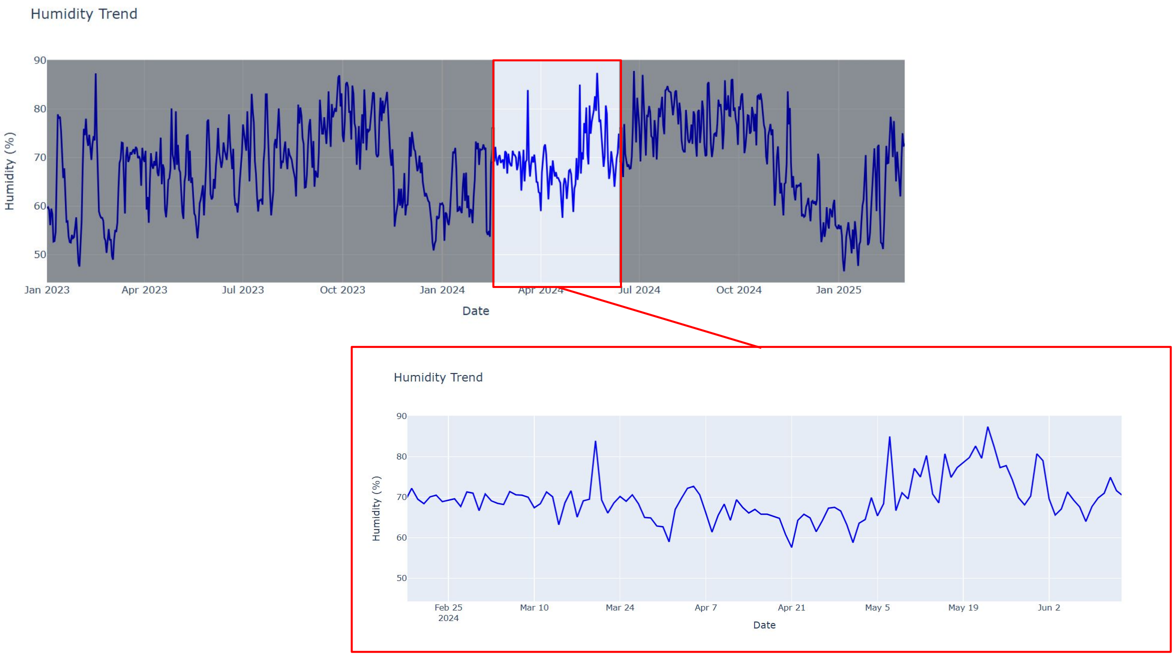
Displaying key model metrics like R^2, CVRMSE, and NMBE.

Model Performance Metrics

R^2: 0.8866	RMSE: 2742.68	CVRMSE: 0.0503	NMBE: -0.0028

Interactivity

Users can zoom in on specific time periods and filter data to explore different trends.



Thank You