Explainability AI

Analysis and Optimization of Energy Consumption

Noémie MAZEPA

Lorrain MORLET

Elias NIERES-TAVERNIER

Aymeric MARTIN

Alexis VAROQUAUX

Antoine CLERVAL

Table of Contents

Introduction

Business and environmental context

With businesses and companies looking to be more energy-efficient and sustainable, managing energy consumption is now a huge priority. Offices, factories and retail spaces use a lot of energy and finding smarter ways to optimize that usage can cut costs while also reducing environmental impacts.

Most commercial buildings (offices, factories, retail spaces, warehouses, etc.) rely heavily on various energy sources, including electricity, natural gas, steam and renewables for lighting, heating, cooling, machinery, and IT systems. According to the International Energy Agency (IEA), commercial and residential buildings together use about 40-45% of the world's total energy. Commercial buildings are big energy users, especially in sectors like retail, healthcare, and IT, where there's a lot of demand for things like lights, computers, and other equipment. For example, in the U.S., commercial buildings account for around 19% of total energy use, and 70-80% of that energy is electricity, according to the US Department of Energy. So, energy consumption is a huge part of their operational costs.

The demand for energy in commercial buildings can peak during the summer due to air conditioning needs or during the winter due to heating. In the U.S., air conditioning accounts for approximately 17% of electricity consumption in buildings, according to the EIA. Energy costs are quite important in commercial buildings ($141 billion on energy in 2018 according to EIA) but it is also found that around 30% of used energy is wasted according to Envelo Solutions, and this leads to huge financial solutions.

However, a study by McKinsey found that companies could reduce their energy costs by 10-30% through energy optimization measures, like adopting more efficient lighting, HVAC systems, and better operational practices. This means there's a big chance to save money, making energy a top priority for improving efficiency.

Project Objectives

In this project, we want to build a predictive model and a dashboard to help businesses track their energy consumption, detect trends and make better decisions.

Our main goals are to:

- Analyze energy usage patterns to find peak times, anomalies and seasonal variations.
- Develop predictive models that forecast energy needs based on energy datasets as well as building and weather information.

- Suggest optimization strategies based on data insights to improve efficiency and reduce costs.

Besides saving money, these tools will help businesses hit their sustainability goals, decrease their carbon footprint, and stay on top of energy efficiency rules. With these insights thanks to data, companies can move toward smarter, greener energy management.

Energy Data and Tools Overview

Data sources

https://www.kaggle.com/c/ashrae-energy-prediction/data

train.csv

We selected the ASHRAE – Great Energy Prediction dataset from Kaggle. It's directly related to understanding and predicting energy consumption in buildings. The dataset has got a lot of detailed data about energy consumption in kilowatt-hours (kWh), the energy type and hourly timestamps. The measurements being taken at such close intervals will help us to understand trends within a day but also weekly and monthly. The whole dataset was constructed over a year and across several buildings, so we have a lot of data.

We can find the following features in **train.csv**:

- **building_id**: A unique identifier for each building, which links to the metadata file.
- **meter**: Identifies the type of meter used for energy recording. The values correspond to different energy types:
    - 0: Electricity
    - 1: Chilled water
    - 2: Steam
    - 3: Hot water

  Not every building has all meter types.

- **timestamp**: The precise time at which the energy consumption reading was recorded.
- **meter_reading**: The target variable, representing energy consumption in kWh (or its equivalent).

Correlation with Other Datasets

The ASHRAE dataset already comes with two other related datasets on building information and weather data. However, it is worth noting that energy consumption can vary based on several factors. Some of the main causes of fluctuations are:

- **Weather conditions (Winter/Summer):** Heating and air conditioning demands can change based on outdoor temperatures.
- **Work patterns (Weekdays/Weekends):** Different building usage habits affect energy consumption. Indeed, during the week, residential buildings are less occupied during the day so the consumption of energy could be less than during the weekends, etc.
- **Energy-intensive appliances:** The efficiency class of appliances can influence the overall energy consumption.
- **Occupancy levels:** More people in a building typically lead to higher energy usage.
- **Building insulation and age:** Older buildings or those with poor insulation may require more energy for heating and cooling.
- **Smart home systems:** Buildings with automated lighting and smart thermostats may optimize their energy use dynamically and this can lead to energy savings.

building_meta.csv and weather_train.csv

With the ASHRAE dataset, we also have other datasets on weather and building information. We think those datasets will be very useful to understand the correlations between energy consumption and other factors. We know that the seasons and temperature have a big impact on energy use. For example, buildings use more energy in the winter for heating and in the summer for air conditioning. We also know that details like building age, size and use matter. Older buildings and those with a bigger surface may need more energy to heat or cool.

The key features in **building_meta.csv** are:

- **site_id**: Links to the weather dataset for location-based climate data.
- **building_id**: Connects to train.csv for energy consumption tracking.
- **primary_use**: Describes the main function of the building (for example education, retail, healthcare, etc.) based on EnergyStar property classifications.
- **square_feet**: Total floor area of the building.
- **year_built**: The year the building was constructed.
- **floor_count**: The total number of floors in the building.

The key features in **weather_train.csv** are:

- **site_id**: Identifier matching weather data to the corresponding building site.
- **air_temperature**: Temperature recorded in degrees Celsius.
- **cloud_coverage**: Measures how much of the sky is covered by clouds (in oktas).
- **dew_temperature**: The temperature at which air moisture condenses, also in degrees Celsius.
- **precip_depth_1_hr**: Rainfall depth recorded over one hour, in millimeters.
- **sea_level_pressure**: Air pressure at sea level, measured in millibars or hectopascals.
- **wind_direction**: The direction from which the wind is blowing, measured in degrees (0-360).
- **wind_speed**: The wind speed in meters per second.

We believe that these factors can really help us understand the causes of energy consumption peaks and predict when energy peaks happen, which is key for optimization and cost reduction.

Technologies and Tools

- **Data processing and modeling**: we are using Python for programming and data processing, with Pandas and NumPy to manipulate our dataset. Then, we used machine learning libraries such as Scikit-learn, XGBoost and LightGBM to develop and train different models to predict the energy consumption of the different buildings.

- **Visualization**: we used Matplotlib and Seaborn to explore the data and visualize trends. Then, we used PowerBI to develop our dashboard.

- **Development environment**: we have worked in VS Code and Google Colab mostly to ensure flexibility in coding. Google colab was also great for some of us because it provides a GPU which allowed the training of some models to be more efficient.

- **Version control**: this was managed using Git and GitHub so we could easily track the changes made to the Notebook and create branches so that collaboration and teamwork was efficient.

Methodology

Phase 1: Data collection and Preparation

Data collection

For this project, we selected the ASHRAE – Great Energy Prediction dataset from Kaggle, which provides detailed records of energy consumption in buildings. The dataset contains hourly energy consumption data. This dataset was built over a year and includes several building types, making it great for analyzing trends in energy consumption for different types of commercial buildings.

Additionally, we utilized two other datasets:

- **Building metadata (building_meta.csv)**: Contains information such as building type, size, and year built. This helps correlate energy consumption patterns with building characteristics.
- **Weather data (weather_train.csv)**: Includes temperature, humidity, and other weather-related features that can influence energy consumption patterns.

Data merging and preparation

To ensure that our data is well-organized for analysis, we first merged the train.csv with the building_meta.csv based on building_id to integrate energy consumption data with building-specific details.

Then, we merged this dataset with the weather data (weather_train.csv) using the site_id and timestamp as matching keys. The timestamp was converted into the datetime format to ensure the 2 datasets would align well during the merge.

Data cleaning and preprocessing

We performed several data cleaning steps:

1. **Handling Missing Values**: Missing values were identified and replaced. Most missing values were filled with the mean value for each respective column.
2. **Outlier Detection**: We used the Interquartile Range (IQR) method to detect and handle extreme outliers in the data. These values were replaced by the nearest acceptable limit to avoid impacting the analysis too badly.
3. **Feature Engineering**: We created several new features to better understand the temporal and seasonal influences on energy consumption:

- **Time-based features**: Extracted hour, day of the week, and month from the timestamp to understand daily and weekly consumption patterns.
- **Cyclical Features**: We applied time cyclic encoding using sine and cosine transformations on the hour, day of the week, and month to capture cyclical relationships in these time-based features.
- **Seasonal Features**: We created a column to categorize months into seasons (for example winter, spring) based on typical energy consumption patterns.

Phase 2: Analysis and Modeling

Exploratory Data Analysis – EDA

We began by exploring the dataset to understand key trends and consumption patterns. We mostly used visulizations to do so:

1. Visualizations energy consumption trends using time series plots to detect seasonal, daily, and weekly patterns.
2. Analyzing correlations between energy consumption and weather features (for example temperature, humidity) to identify factors contributing to energy peaks.
3. Exploring the relationship between building features (for example size, age) and energy consumption.

Predictive Modeling

We developed several predictive models to forecast energy consumption:

1. **Random Forest Regressor**: Chosen for its ability to handle complex, non-linear relationships in the data. This model performed exceptionally well with an RMSE of *0.6201* and an R^2 of *0.9167*.
2. **LightGBM** and **XGBoost**: These tree-based algorithms were also tested. Although they didn't perform as well as Random Forest, they showed potential with RMSE scores of *1.3062* and *1.3179*, and R^2 scores of *0.63*.
3. **Linear Regression**: We also tried Linear Regression, which performed the worst with an RMSE of *1.9485* and an R^2 of *0.1771*.

Model Optimization

The best 3 models that we identified after our first testing were Random Forest Regressor, LightGBM, and XGBoost. To improve even more their performance, we applied hyperparameter tuning with optuna.

After tuning the hyperparameters, we had the following results:

- The **Random Forest Regressor** kept a great result with an RMSE of 0.6201 and R^2 of 0.92.
- The **LightGBM** model had a huge improvement in RMSE (from 1.31 to 0.70) and R^2 (from 0.63 to 0.89).
- The **XGBoost** model had also a great reduction in RMSE (from 1.32 to 0.60) and an increase in R^2 (from 0.62 to 0.92).

As we can see after tuning the hyperparameters, the best 2 models are Random Forest and XGBoost with both the lowest RMSE (0.62) and r2 score (0.92). However, we will prioritize the XGBoost model though, because it is much faster to train and test.

Clustering for Pattern and Factors Detection

To identify consumption patterns, we applied K-means clustering to get insights based on energy usage, weather conditions, and building characteristics. This clustering helped in detecting different energy usage profiles and provided insights into when certain consumption patterns are more likely to occur, like energy peaks during specific weather conditions or different times of day.

Phase 3: Visualization and Insights

We developed a dashboard with PowerBI to present the results of our analysis and predictions. The dashboard provides graphs to understand energy consumption trends over time and to detect energy peaks, correlated with weather data and building characteristics.

These tools will allow us to communicate the insights effectively to both technical and non-technical audiences.

Key Insigths

After looking at energy consumption trends, we've spotted a few key takeaways that show where the biggest challenges are and where improvements can be made. These insights reveal patterns in energy use across different buildings and the factors that impact their energy needs.

- **High energy consumption in large buildings, especially in educational facilities:** Schools, universities, and similar buildings often use a lot of energy due to their large size and the constant demand for heating, cooling, and lighting. This makes them major contributors to overall energy consumption.

- **Seasonal and time-of-day energy peaks:** There are clear spikes in energy use during specific times of the day and year. For example, heating tends to increase around midday when temperatures drop, and lighting demands peak in the evening. These patterns contribute significantly to overall energy costs.

- **Weather conditions impacting energy usage:** Energy consumption is closely tied to the weather. For instance, temperature, cloud coverage, and wind speed all affect how much heating or cooling a building needs, making energy use fluctuate. A colder day will naturally push up heating needs, while a sunny day might reduce the need for artificial lighting.

- **Differences in energy efficiency between building types:** Public service buildings and educational facilities don't always have the same energy efficiency. Some buildings, like schools, are older and have poorer insulation or less efficient systems, which makes them use more energy compared to newer, more optimized buildings.

Optimization Strategies

Smart Energy Management Systems

Recommendation: The best move here is to set up real-time monitoring and automation to keep track of energy usage and make adjustments on the fly.

Benefits:

- You can spot when energy usage is peaking and tweak things like heating, cooling, and lighting to avoid wasting energy.

- Using IoT sensors and AI will help fine-tune these systems automatically, making sure you're not using more energy than necessary.

- When the building is less occupied, energy waste can be reduced by adjusting usage to fit actual needs.

Example: Smart thermostats that adjust the temperature based on the real-time data about how many people are in the building.

Optimize HVAC and Lighting Systems

Recommendation: Upgrading to more energy-efficient HVAC systems and lighting models is key to cutting down on energy usage, especially in larger spaces like schools or offices.

Benefits:

- Installing LED lights can drastically reduce energy use, and this is particularly useful in schools and other big buildings that use a lot of lighting.

- Motion sensors and daylight sensors can help cut down on unnecessary lighting when rooms aren't in use or when there’s enough natural light.

- Upgrading to high-efficiency HVAC systems will improve heating and cooling management, keeping the building comfortable without wasting energy.

Example: Schools and office buildings using automated lighting and smart HVAC systems to reduce waste and optimize comfort.

Adjust Operations Based on Energy Peaks

Recommendation: The goal here is to shift energy-heavy activities to times when electricity prices are lower, avoiding those costly peak demand hours.

Benefits:

- By shifting operations outside peak times, you’ll lower your costs since electricity prices can spike during peak hours.

- The use of machines and equipment can be spread out throughout the day, balancing the load and reducing energy strain at certain times.

- You can also encourage more flexible work hours to reduce the demand for energy during those busy times.

Example: Offices and factories adjusting heating and cooling schedules to off-peak hours when energy prices are lower.

Building Retrofitting and Insulation Improvements

Recommendation: It's important to improve building insulation and upgrade windows and roofs to make the space more energy-efficient, especially if the building is older.

Benefits:

- Adding better insulation reduces heating and cooling losses, which is especially beneficial in older educational buildings that may not be well-insulated.
- With better insulation, you can lower energy bills because the building stays naturally warm in the winter and cool in the summer.
- Energy-efficient windows and reflective roofing materials can boost insulation, helping to keep the temperature steady without cranking up the HVAC systems.

Example: Retrofitting old educational buildings with better insulation and energy-efficient windows or roofs to cut down on heating and cooling costs.

Conclusion

To conclude, our project shows how using data and smart models can help businesses save energy and reduce costs. By understanding energy usage patterns and optimizing how energy is used, companies can make a big difference in both their bills and their environmental impact. We found that factors like the weather, building type, and time of year have a certain effect on energy consumption. With models like XGBoost and Random Forest, we were able to predict and fine-tune energy use. Our suggestions, like upgrading systems or using real-time monitoring, can help businesses make energy-saving decisions. Overall, this project proves that with the right tools, businesses can cut costs and become more sustainable at the same time.

Bibliography

TotalEnergies (n.d.) *Consommation anormale d'électricité : causes et solutions*. Available at: https://services.totalenergies.fr/particuliers/conseils/energies-habitat/electricite/consommation-anormale-electricite-causes-solutions (Accessed: 4 March 2025).

EDF (n.d.) *Pourquoi votre facture d'électricité augmente ?* Available at: https://particulier.edf.fr/fr/accueil/guide-energie/electricite/augmentation-electricite.html (Accessed: 4 March 2025).

Hellio (n.d.) *Facteurs influençant la consommation d'énergie*. Available at: https://faq.hellio.com/definition-facteurs-influence-energie (Accessed: 4 March 2025).

Vattenfall (n.d.) *Les paramètres qui influencent la consommation d'électricité*. Available at: https://www.vattenfall.fr/le-mag-energie/parametres-qui-influent-sur-la-consommation-d-electricite (Accessed: 4 March 2025).

GlobalABC (n.d.) *IEA tracking report: Buildings*. Available at: https://globalabc.org/index.php/resources/publications/iea-tracking-report-buildings (Accessed: 4 March 2025).

LED Lighting Supply (n.d.) *What types of commercial buildings consume the most energy?* Available at: https://www.ledlightingsupply.com/blog/what-types-of-commercial-buildings-consume-the-most-energy (Accessed: 4 March 2025).

U.S. Energy Information Administration (EIA) (n.d.) *Use of energy in commercial buildings*. Available at: https://www.eia.gov/energyexplained/use-of-energy/commercial-buildings.php (Accessed: 4 March 2025).

International Energy Agency (IEA) (n.d.) *Buildings*. Available at: https://www.iea.org/energy-system/buildings (Accessed: 4 March 2025).

 U.S. Department of Energy (DOE), Office of Energy Efficiency and Renewable Energy (EERE) (n.d.) *Commercial Buildings Initiative Fact Sheet*. Available at: https://www1.eere.energy.gov/buildings/publications/pdfs/commercial_initiative/cbi_fs.pdf (Accessed: 4 March 2025).

Budderfly (n.d.) *Building energy consumption*. Available at: https://blog.budderfly.com/building-energy-consumption (Accessed: 22 March 2025).

Energy Solutions Oxfordshire (n.d.) *Research shows that commercial buildings must reduce energy use by 80% to meet net zero targets*. Available at:

https://www.energysolutionsoxfordshire.org/research-shows-that-commercial-buildings-must-reduce-energy-use-by-80-to-meet-net-zero-targets/ (Accessed: 22 March 2025).

LED Lighting Supply (n.d.) *What types of commercial buildings consume the most energy?* Available at: https://www.ledlightingsupply.com/blog/what-types-of-commercial-buildings-consume-the-most-energy (Accessed: 22 March 2025).

Envelo Solutions (n.d.) *Energy efficiency for commercial buildings*. Available at: https://envelo.solutions/energy-efficiency-for-commercial-buildings/ (Accessed: 22 March 2025).

U.S. Energy Information Administration (EIA) (n.d.) *Commercial buildings energy consumption survey (CBECS)*. Available at: https://www.eia.gov/consumption/commercial/ (Accessed: 22 March 2025).