Energy Consumption Prediction and Optimization

Introduction

In the age of smart infrastructure and sustainable development, predicting and optimizing energy consumption has emerged as a priority for urban planners, energy providers, and environmentalists. With escalating energy demands, it is imperative to develop intelligent systems that can forecast energy usage and optimize consumption patterns for improved efficiency. This case study presents an end-to-end approach to energy consumption forecasting and optimization using historical electricity usage data and meteorological variables, employing data preprocessing techniques, machine learning models, and performance evaluation metrics.

Problem Statement

The aim of this project is to create a predictive model capable of forecasting energy usage in commercial buildings, followed by optimization of energy consumption patterns. Specifically, this study focuses on forecasting hourly energy consumption for a single office building using time, temperature, and calendar-based features, and applying K-Nearest Neighbors (KNN) regression for prediction.

Dataset Overview

Two key datasets were used:

1. **Electricity Consumption Dataset**: Contains hourly energy consumption data from various buildings across two years (2016–2017). The data consists of 555 building columns and over 17,000 hourly entries.
2. **Weather Dataset**: Includes site-specific hourly weather information such as air temperature, cloud coverage, dew temperature, precipitation, sea level pressure, wind direction, and wind speed.

For this case study, one building named **Panther_office_Hannah** was selected for analysis, along with weather data specific to the "Panther" site.

Data Preprocessing

Preprocessing steps included:

- **Truncating time**: Focused only on the year 2017 for modeling.
- **Forward-filling missing values**: Ensured time series continuity.
- **Outlier filtering**: Removed unrealistic weather values such as temperatures below -40°C.
- **Resampling**: Weather data was resampled to hourly intervals.
- **Feature construction**: Time-based features like hour and day of the week were extracted. Temperature was included as a primary external factor.

Feature Engineering

Features were constructed using:

- **One-hot encoding** for hours (0–23)
- **One-hot encoding** for days of the week (0–6)
- **Temperature values** during corresponding hours

These features were combined into a design matrix suitable for feeding into machine learning algorithms.

Model Selection: K-Nearest Neighbors (KNN)

A KNN regression model was chosen due to its simplicity and interpretability. It performs well on small datasets and is non-parametric, making no assumptions about the underlying data distribution.

Training and Testing:

- **Training Months**: April, May, June (Q2)
- **Testing Month**: July (Q3)

Data was split accordingly:

- **Training Set**: 2,184 hours
- **Testing Set**: 744 hours

Model Training:

The KNN model was trained using:

model = KNeighborsRegressor().fit(train_features, trainingdata.values)

Prediction and Evaluation:

The model predicted energy usage for July 2017 and the results were compared against actual usage.

Results

Sample output:

Timestamp	Actual	Predicted
2017-07-01 00:00:00	5.3370	5.4946
2017-07-01 01:00:00	3.8547	5.0342
2017-07-01 02:00:00	5.5751	4.1846

Visualization of predictions vs. actual values showed that the model followed the trend but with noticeable variation.

Performance Metric:

- **Mean Absolute Percentage Error (MAPE)**: 33.59%

This indicates moderate prediction accuracy, suitable for trend forecasting but not exact value matching.

Optimization Insights

While the primary focus was on prediction, insights from the model can inform energy optimization strategies such as:

- **Smart scheduling** of devices during non-peak predicted hours
- **Temperature-controlled HVAC management**
- **Proactive maintenance alerts** during abnormal energy surges
- **Energy efficiency audits** based on historical peaks and consumption behavior

Challenges Faced

- Handling missing data across both electricity and weather datasets
- Incorporating only a limited set of weather variables (mainly temperature)
- Model underperformance during extreme weather conditions
- KNN limitations in scalability and real-time deployment

Future Enhancements

- Integrate deep learning models (e.g., LSTM) for sequential data forecasting
- Incorporate additional weather and occupancy-based features
- Use hybrid models combining clustering and prediction
- Build a live dashboard for dynamic energy optimization

Conclusion

This case study demonstrates that energy consumption prediction using KNN regression, when combined with engineered time and weather features, can provide valuable foresight into energy patterns. Although the prediction accuracy has room for improvement, the model serves as a solid baseline for deploying more complex predictive and optimization systems in commercial infrastructure.

By integrating predictive modeling with smart control systems, institutions can pave the way for cost-effective and environmentally sustainable energy management solutions.