Report: Austrian Energy Consumption Prediction and Optimization

By Saeed khan &

Muhammad Saad

Abstract

This report details an analysis of Austrian energy consumption data with the goal of predicting energy load and identifying potential optimization strategies. Utilizing historical load and weather data, a LightGBM model was developed to forecast energy demand. Exploratory data analysis revealed significant correlations between energy load and factors such as temperature, dew point, and time-based patterns (hourly, daily, and weekly seasonality). Based on these findings, two potential optimization techniques – temperature-based thermostat adjustment and weekday peak hour load reduction – were simulated, demonstrating potential energy savings.

1. Introduction

Predicting and optimizing energy consumption is crucial for efficient grid management, cost reduction, and environmental sustainability. This project focuses on analyzing Austrian energy consumption data to build a predictive model and explore potential strategies for optimizing energy usage. The dataset used in this analysis combines energy data from Open Power System Data (OPSD) with weather data from the Open-Meteo API. It provides hourly observations for Austria covering a 5-year period (Jan 2015 – Jan 2020) with 43,848 rows and 24 columns.

2. Data Loading and Cleaning

The dataset "energy_weather_data.csv" was loaded into a pandas DataFrame. Initial exploration revealed missing values in the **AT_price_day_ahead , AT_solar_generation_actual**, and **AT_wind_onshore_generation_actual** columns. Missing values in **AT_price_day_ahead** were imputed using linear interpolation to maintain data integrity for time series analysis.

3. Feature Engineering

To enhance the predictive capabilities of the model, several relevant features were engineered:

Time-based features: Hour of the day, day of the week, weekend indicator, holiday indicator (using a custom Austrian holiday calendar), and month were extracted from the datetime index.

Lagged load features: Energy load values from the preceding 24, 48, and 72 hours were created as features to capture temporal dependencies and seasonality in energy consumption.

4. Exploratory Data Analysis (EDA)

Comprehensive EDA was conducted to understand the characteristics of the data and identify key patterns influencing energy consumption.

- Time series plots of actual energy load revealed clear annual and weekly seasonality.
- Analysis of daily seasonality showed distinct load profiles for weekdays and weekends, with higher consumption and prominent morning and evening peaks during weekdays.
- Seasonal decomposition of the daily load data confirmed the presence of trend and strong weekly seasonality.
- Correlation analysis between weather and energy features highlighted the significant inverse relationship between temperature/dew point and energy load.
- -Scatter plots further illustrated the non-linear relationship between temperature/dew point and energy load.

5. Machine Learning Model for Energy Load Prediction

A LightGBM Regressor model was chosen for energy load prediction due to its efficiency and effectiveness with tabular data and time series.

- The dataset was split into training and testing sets chronologically to simulate real-world forecasting scenarios.
- Features were scaled using StandardScaler to ensure that all features contribute equally to the model training.

- The LightGBM model was trained on the scaled training data and evaluated on the test set using early stopping to prevent overfitting.

6. Model Evaluation

The performance of the LightGBM model was evaluated using standard regression metrics on the test set:

- Mean Squared Error (MSE): 74216.33
- Root Mean Squared Error (RMSE): 272.43
- R-squared (R2): 0.96

An R-squared value of **0.96** indicates that the model explains a high percentage of the variance in the energy load and provides a good fit to the data. Visualizations comparing the actual and predicted load on the test set further demonstrated the model's ability to capture the patterns in energy consumption.

7. **Energy Consumption Optimization**

Based on the insights gained from EDA and the predictive model, potential energy consumption optimization techniques were explored.

Key factors for optimization: Temperature, dew point, and weekday/weekend patterns were identified as significant drivers of energy load.

Suggested techniques:

- **Temperature-based optimization:** Adjusting thermostat setpoints based on temperature ranges or implementing smart thermostat schedules.
- **Weekday/weekend optimization:** Shifting non-essential consumption or reducing load during peak weekday hours.

Simulation of Optimization: Simplified simulations were conducted to estimate the potential impact of these techniques on historical load data.

- **Temperature optimization:** Simulated adjustments based on temperatures outside a predefined operational range. This simulation showed a potential increase in load due to pre-heating in cold temperatures.
- **Weekday peak hour optimization:** Simulated load reduction during identified peak weekday hours. This simulation showed a positive potential for energy savings.

8. Potential Energy Savings

The simulations provided estimates of potential energy savings:

- **Potential Total Savings from Temperature Optimization:** 2852922.64 MWh
- **Potential Total Savings from Weekday/Weekend Peak Hour Optimization:** 4827181.25 MWh

These simulations suggest that implementing optimization strategies based on temperature and weekday/weekend patterns could lead to considerable energy savings and potentially reduce energy costs.

9. Challenges and Future Work

Data collection: due to high preservation of the energy data its becoming very challenging to find energy and weather data .

Missing Data: The significant amount of missing data in the **AT_price_day_ahead** column posed a challenge. While interpolation was used, exploring more sophisticated imputation methods or investigating the reasons for missing data could improve accuracy.

Simplified Optimization Simulation The optimization simulations were based on simplified assumptions. Future work could involve more detailed and realistic simulations incorporating factors like building types, user behavior, and the economic feasibility of implementing these techniques.

Model Improvement: While the LightGBM model performed well, exploring other time series forecasting models (e.g., ARIMA, LSTM) and hyperparameter tuning could potentially improve prediction accuracy.

Impact of Other Features: Further analysis of the impact of other weather features and exploring interactions between features could provide additional insights for optimization.

10. Conclusion

This project successfully demonstrated the process of analyzing Austrian energy consumption data, building a predictive model using LightGBM, and simulating the potential impact of data-driven optimization techniques. The findings highlight the importance of considering weather conditions and time-based patterns for accurate energy load forecasting and effective demand-side management strategies. While challenges related to data quality and simulation complexity exist, this work provides a solid foundation for further research and implementation of energy optimization solutions.