

Electricity Consumption Analysis and Forecasting Report

UNIVERSITY OF MARYLAND, BALTIMORE COUNTY (UMBC)

DATA 699: Independent Study in Data Science

Prof. Ajinkya Borle

May 15,2024

By. NEHA C. NOOKA

Introduction

Electricity consumption is a critical aspect of modern society, impacting various sectors such as residential, commercial, and industrial. Understanding historical consumption patterns and forecasting future demand is essential for effective energy management, infrastructure planning, and policy formulation. In this report, I analyze historical electricity consumption data for the United Kingdom (UK) spanning from 2009 to 2024. Our objective is to gain insights into consumption trends, explore seasonality, identify influencing factors, and apply time series forecasting methods to predict electricity demand for the year 2024.

Dataset Background

The dataset used for analysis comprises historic electricity consumption data for the UK sourced from the National Grid Electricity System Operator (ESO). It includes information such as national demand, transmission system demand, embedded wind and solar generation, interconnector flows, and more. The data is time stamped at half-hourly intervals, providing a granular view of consumption patterns. Such detailed data is ideal for time series analysis and forecasting.

Exploratory Data Analysis (EDA) and Feature Engineering

Our analysis begins with exploratory data analysis (EDA) to understand the structure, distribution, and characteristics of the dataset. We perform the following steps:

1. Data Understanding: Initial exploration of the dataset to gain insights into its format.
2. Data Preparation: Handling null values, dropping irrelevant features, and removing outliers to ensure data quality.
3. Feature Understanding: Investigating trends, seasonality, and patterns within the data, including hourly, daily, weekly, monthly, and yearly variations.

Bank Holidays and Feature Creation

Bank holidays significantly influence electricity consumption patterns due to changes in economic activity and human behavior. Additionally, we create new features such as day of the week, month, year, and lagged variables to capture temporal dependencies and enhance predictive modeling.

Key Findings:

- **Peak Hours:** Peak electricity consumption occurs during daytime hours (8am to 6pm), indicating higher demand for electricity during this period.
- **Seasonal Variations:** Consumption levels are lowest during summer, potentially influenced by reduced heating requirements and increased daylight hours.
- **Bank Holidays:** On average, electricity consumption is lower on bank holidays compared to regular weekdays, with higher consumption observed on Saturdays.
- **Yearly Trends:** There is a consistent decreasing trend in electricity consumption over the years, indicating improvements in energy efficiency or shifts in consumption patterns.
- **Renewable Energy:** Solar and wind generation have shown significant growth over the years, highlighting the increasing adoption of renewable energy sources.

Time-Series Models: XGBoost

I employed the XGBoost algorithm for time series forecasting due to its capability to handle complex relationships and incorporate multiple features. Our modeling process involves the following steps:

1. **XGBoost Model:** Started with a baseline XGBoost model using default parameters and a basic feature set. This model serves as a reference point for further refinement.
2. **Cross Validation and Grid Search:** To improve model performance and prevent overfitting, we implement cross-validation with grid search to identify optimal hyperparameters. This step ensures better generalization and robustness of the model.

Model Evaluation and Prediction

We evaluate model performance using metrics such as Mean Absolute Percentage Error (MAPE) and Root Mean Squared Error (RMSE) on both test and hold-out datasets. The best-performing model is then utilized to predict electricity demand for future periods.

Conclusion

In conclusion, our analysis provides valuable insights into electricity consumption trends in the UK and demonstrates the effectiveness of XGBoost for time series forecasting. By leveraging historical data and advanced modeling techniques, stakeholders can make informed decisions regarding energy management, resource allocation, and infrastructure planning. Continuous monitoring and refinement of forecasting models will further enhance their accuracy and reliability, contributing to sustainable energy practices and economic development.