Peak Energy Demand Prediction Models

New River Light & Power

Alex Pettis | CIS 5860-101 Applied Analytics Project | July 14, 2024

Contents

Industry Background

For over a century, Appalachian State University's New River Light and Power (NRLP), a nonprofit electric utility overseen by the Division of Finance and Operations, has reliably powered the homes and businesses of Western North Carolina. Serving nearly 9,000 residents and commercial clients in and around Boone, NRLP began purchasing electricity from Carolina Power Partners in January 2022, enhancing its capacity to integrate more renewable energy sources. Collaborating with App State's Office of Sustainability, Facilities Operations, and the Renewable Energy Initiative, NRLP has supported numerous energy efficiency projects on campus, including the installation of solar panels and funding the Broyhill Wind Turbine. Recognized by the American Public Power Association's Reliable Public Power Provider (RP3) program, NRLP is commended for its dependable and safe electric service. Annually, it contributes around $650,000 of its budget to App State's general scholarship fund. NRLP's mission is to deliver efficient and reliable electrical service to Appalachian State University, Boone, and the surrounding communities, while also supporting the university's financial needs. In fulfilling this mission, NRLP fosters a positive connection between Appalachian State University, Boone, and the surrounding areas, promoting economic development and positive public relations.

Purpose

The purpose of the project is to develop a predictive model that accurately forecasts peak energy demand for Appalachian State University's New River Light and Power (NRLP). This model aims to identify periods of maximum power consumption, enabling NRLP to better manage energy distribution and operational efficiency. By providing consumers with insights into peak demand periods, it allows them to optimize their energy usage and reduce costs.

Additionally, the model will support NRLP's efforts in operational efficiency and sustainability, including cost reduction during high peak demand periods and the integration of renewable energy sources. The project promotes energy efficiency, encouraging more efficient energy use among consumers, lowering the overall carbon footprint, and supporting NRLP's sustainability goals.

By equipping consumers with actionable insights and tools to manage their energy consumption effectively, the project seeks to enhance customer satisfaction. Furthermore, it will support NRLP's mission to provide reliable electrical service, strengthen the connection between Appalachian State University, the town of Boone, and surrounding communities, and promote economic development through improved energy management. Overall, the project aligns with the goals of improving energy distribution management, enhancing operational and cost efficiencies, and contributing to sustainability efforts.

Literature Review

Introduction

The efficient management of peak electricity demand is crucial for enhancing grid stability and economic efficiency (Fu et al., 2023). Battery Energy Storage Systems (BESS) offer promising solutions by enabling peak demand reduction through instantaneous response capabilities (Fu et al., 2023). However, accurately predicting peak demand days and hours remains a significant challenge, impacting the optimal deployment and operation of BESS (Fu et al., 2023).

Machine Learning Models for Peak Demand Prediction

Recent advancements in machine learning (ML) have shown potential in improving the accuracy of peak day and hour predictions. Traditional methods like Autoregressive Integrated Moving Average (ARIMA) are being supplemented by nonlinear ML approaches that leverage multivariate predictors such as weather and economic factors (Fu et al., 2023). Studies have explored various ML models including K-nearest neighbors (KNN), Support Vector Machines (SVM), Gradient Boosting Machines (GBM), Random Forest (RF), and Artificial Neural Networks (ANN), either individually or in hybrid forms, to forecast peak electricity demand (Fu et al., 2023).

Fu et al. (2023) developed a supervised ML approach combining RF, GBM, and Logistic Regression (LR) to predict both the probability of the next operation day containing the peak hour of the month and the probability of an hour being the peak hour of the day. Their study, applied to the Duke Energy Progress system, achieved significant success in identifying peak days and hours with high accuracy under varying conditions (Fu et al., 2023). This approach not only enhances the operational efficiency of BESS but also addresses uncertainties associated with peak demand forecasts, crucial for effective dispatch decision-making (Fu et al., 2023).

Challenges and Future Directions

Despite the progress made, challenges such as data inadequacy and the need for robust uncertainty quantification persist (Fu et al., 2023). Future research should focus on further refining ML models to incorporate all relevant temporal and physical factors, improving the reliability and interpretability of peak demand predictions (Fu et al., 2023).

Conclusion

In conclusion, the integration of ensemble ML techniques represents a significant advancement in forecasting peak electricity demand, offering potential benefits for grid stability and economic efficiency through optimized BESS deployment (Fu et al., 2023). Addressing remaining challenges through enhanced data augmentation and model selection techniques will be crucial for advancing the field (Fu et al., 2023).

Methodology

Literature Review Integration

The methodology for developing the predictive model for New River Light and Power (NRLP) drew extensively from recent advancements in peak energy demand forecasting, as highlighted in the literature. Fu et al. (2023) emphasized the importance of integrating multivariate predictors such as weather and economic factors to enhance the accuracy of peak demand predictions. Building upon this foundation, the project utilized machine learning techniques including Multiple Regression, Logistic Regression, Support Vector Machines (SVM), and Fuzzy Regression to analyze the relationship between weather variables and energy consumption patterns.

Data Collection

In my data collection process, I utilized two primary sources: the energy usage data and weather data. The energy usage data was provided to me by the project sponsor, encompassing detailed records of energy consumption for each location involved in the study. To complement this, I collected weather data from the NOAA Integrated Surface Database (ISD). This database offered comprehensive and reliable weather metrics, including temperature, precipitation, wind speed, and other relevant variables for each location in the study. This dual data collection approach ensured that I had a robust dataset, integrating both energy consumption patterns and the corresponding weather conditions. Detailed information regarding the specific data variables and their sources can be found in Table A1 (Energy Usage Data) and Table A2 (Weather Data).

Data Preparation

In my data preparation process, I undertook several steps to ensure the dataset was clean and ready for analysis. First, I checked for constant columns in my Weather Data (Table A2), identifying and removing those that did not vary and were thus unnecessary for the analysis. Alongside these, I also removed other columns that were deemed irrelevant from the outset. I then added a column named "Concord" to the weather data, which represented the sum of the daily energy usage for the Concord location which came from the Energy Usage Data (Table A1). To account for the temporal aspects of energy usage, I split the DATE column into three separate columns: month, day, and year, recognizing that different times of the year might require different models due to seasonal variations. I proceeded by removing any additional unneeded columns, followed by handling missing values, replacing them with the mean to maintain data integrity. Lastly, I created a correlation matrix to identify which columns were correlated, aiding in understanding the relationships between variables and informing model selection and feature engineering.

Analysis Techniques

I have conducted an extensive analysis to predict peak energy demand days, employing a variety of advanced statistical and machine learning models. Initially, I utilized multiple regression to understand the linear relationship between weather variables and energy demand. Then, I applied logistic regression to classify days as peak or non-peak demand based on the same variables. To capture potential interaction effects between variables, I also tested logistic regression with an interaction term. The terms I interacted were PRCP and TMIN. Furthermore, I explored Support Vector Machines (SVM) to leverage their powerful classification capabilities. Lastly, I experimented with fuzzy regression to handle any uncertainty and imprecision in the data. Each of these models provided unique insights and contributed to a comprehensive understanding of the factors influencing peak energy demand days. For these models, I have used Concord energy usage data as testing data that will be then applied to the Boone/New River usage data.

Challenges

Throughout the project, several challenges surfaced that required strategic solutions to enhance model performance and ensure data consistency. Initially, the models showed low R-squared values, indicating inadequate predictive power. To address this, I reintroduced the date column and segmented it into three distinct variables (Month, Day, Year). This adjustment allowed the models to capture temporal nuances more effectively, thereby improving their accuracy in forecasting peak energy demand. Another significant challenge stemmed from the disparity in data granularity: energy usage data was recorded hourly, whereas weather data was available only on a daily basis. This inconsistency posed a hurdle during model execution, necessitating the aggregation of energy usage data into daily averages. By aligning the temporal granularity of both datasets, I ensured compatibility and coherence in the analysis, facilitating more accurate insights into the relationship between weather conditions and energy consumption patterns.

Results

Table 1 Results

Model	Season	R-Squared/Accuracy
Multiple Regression (All Variables)	Summer (May-September)	0.636
Multiple Regression (AWND, SWND, TMAX, TMIN, TOBS, Month)	Summer (May-September)	0.635
Multiple Regression (All Variables)	Winter (November-March)	0.153
Multiple Regression (All Variables)	April	0.499
Multiple Regression (All Variables)	October	0.493
Logistic Regression (All Variables)	Summer (May-September)	0.717
Logistic Regression with Interaction Term (PRCP, TMIN, Day, Year, PRCP_TMIN)	Summer (May-September)	-0.427
Support Vector Machine (SVM)	All	0.925
Fuzzy Regression	Summer (May-September)	0.848

Result Interpretation

Multiple Regression

The analysis involved applying various models to predict peak energy demand, revealing significant insights. For the summer months (May to September), multiple regression using all variables achieved an R-squared value of 0.636, explaining approximately 63.6% of the variance in peak energy demand. Even when reducing the number of variables to only include AWND, SWND, TMAX, TMIN, TOBS, and Month, the model's performance remained almost unchanged, with an R-squared of 0.635. In contrast, the winter model (November to March) had a much lower R-squared value of 0.153, indicating that only 15.3% of the variance in peak demand was explained, suggesting that other factors might be more influential in winter. For the transitional months, April and October, the R-squared values were 0.499 and 0.493, respectively, demonstrating moderate explanatory power. From this, we can see that multiple regression works better in predicting peak energy demand during the summer months when more variables are included.

Logistic Regression, Support Vector Machine, and Fuzzy Regression

Logistic regression for the summer months showed an accuracy of 71.7%, correctly predicting peak and non-peak days most of the time. However, adding an interaction term (PRCP_TMIN) resulted in a negative R-squared value of -0.427, indicating a poor model fit worse than a simple mean prediction. The support vector machine (SVM) model, applied to all seasons, performed exceptionally well, achieving a high accuracy of 92.5%, making it highly effective in predicting peak and non-peak days. However, it's important to note that SVM is particularly good at predicting non-peak days, or true negatives, which limits its usefulness for identifying peak days. Fuzzy regression for the summer months yielded an R-squared value of 0.848, indicating a high level of explanatory power, capturing the nuances in the data for this particular season. From the results of these models, we can see that the fuzzy regression model is the best at accurately predicting peak energy demand.

Conclusions

Based on a comprehensive analysis of the data, several key conclusions emerge regarding the predictive modeling of peak energy demand for New River Light and Power (NRLP). During the summer months, models incorporating weather related variables demonstrate robust performance in forecasting peak energy demand. This underscores the significant impact of weather conditions, such as temperature and precipitation, on energy consumption patterns during warmer periods. These findings suggest that strategies focused on weather sensitive forecasting can effectively guide energy distribution and operational planning during peak demand seasons.

Conversely, predictive models exhibit limited effectiveness during winter months, indicating that factors beyond weather variables, such as socioeconomic conditions and heating demands, play a pivotal role in shaping energy consumption trends during colder periods. This highlights the need for broader considerations in energy demand forecasting, encompassing both environmental and socioeconomic factors to enhance predictive accuracy across seasonal variations.

Multiple regression models offer reasonable explanatory power in certain seasons but show variability in effectiveness across different times of the year. Further refinement of these models could improve consistency and reliability in predicting energy demand patterns throughout the year. Logistic regression, while moderately accurate in distinguishing peak and non-peak energy demand periods, requires careful management of model complexity, especially when incorporating interaction terms that can impact performance.

Machine learning models, such as Support Vector Machines (SVM), demonstrate high accuracy in predicting energy demand patterns but exhibit sensitivity to the specific characteristics of the dataset and seasonal variations. These models prove valuable in capturing complex relationships between variables but necessitate ongoing adaptation and refinement to maintain predictive robustness across different seasons.

Variables related to weather such as temperature, precipitation, and wind speed—consistently emerge as critical predictors of energy demand. Incorporating these variables into predictive models is essential for enhancing accuracy, particularly in seasonal forecasting efforts. However, the observed seasonal variability in model performance underscores the importance of continuous research and refinement. Future efforts should explore additional variables beyond weather, including economic indicators and local events, to capture nuanced aspects of energy demand variation throughout the year.

In conclusion, NRLP can optimize energy management strategies by selecting models that accommodate seasonal variations and integrate relevant variables influencing energy demand patterns. Utilizing predictive insights effectively can enhance operational efficiency, inform energy distribution strategies, and empower consumers with information to encourage energy-saving behaviors. Investing in advanced modeling techniques and expanding data sources will further enhance the accuracy and reliability of energy demand forecasting, supporting NRLP's sustainability goals and operational efficiency objectives throughout all seasons.

Future Directions

My future research direction involves prioritizing the refinement of predictive models to enhance their accuracy and robustness across all seasons. This includes exploring additional variables beyond weather, such as economic indicators and local events, to capture nuanced aspects of energy demand variation throughout the year. Integrating more comprehensive datasets, such as real-time data feeds or socioeconomic data, holds promise for significantly improving model performance by providing deeper insights into energy consumption patterns, especially during critical periods. Moving forward, I aim to integrate these advanced predictive models into NRLP's operational framework, establishing protocols for real-time monitoring and decision-making based on forecasted peak energy demand. This initiative is designed to optimize energy distribution and enhance operational efficiency. Engaging stakeholders, including consumers and regulatory bodies, will be essential for gathering feedback on the utility and usability of predictive insights, guiding iterative improvements, and ensuring alignment with NRLP's sustainability goals and customer satisfaction objectives. Continuous evaluation and refinement of predictive models will be ongoing, involving regular performance assessments against actual data to identify opportunities for enhancement and adaptation to evolving environmental and operational conditions.

References

Fu, Tao, Huifen Zhou, Xu Ma, Z. Jason Hou, and Di Wu. "*Predicting Peak Day and Peak Hour of Electricity Demand with Ensemble Machine Learning*." Frontiers in Energy Research 10 (November 8, 2022): 944804. https://doi.org/10.3389/fenrg.2022.944804.

Appendix

Data Details

Table A1. Hourly Energy Usage Data Description

Hourly Energy Usage Data Description		
Colum Name	Data Type	Description
UTC	Object	Timestamp in UTC timezone.
Eastern	Object	Timestamp in Eastern timezone
Year	int64	Year of the observation. Usage data ranges from 2015-2021.
OrdDay	int64	Ordinal day of the year (1 to 365).
OrdHr	int64	Ordinal hour of the year (1 to 8760).
Weekday	int64	Day of the week (1 to 7). Ex. Monday 1, Tuesday 2, etc.
Month	int64	Month of the year (1 to 12).
Day	int64	Day of the month (1 to 31).
HourEnd	int64	Ending hour of the observation period (typically in military time).
DST	int64	Binary indicator (0 or 1) for Daylight Saving Time.
Concord	int64	Energy usage for Concord for the hour.
Greenwood	int64	Energy usage for Greenwood for the hour.
NewRiver	int64	Energy usage for New River for the hour.
KingsMountain	int64	Energy usage for Kings Mountain for the hour.
Winterville	int64	Energy usage for Winterville for the hour.
Total	int64	Total energy usage for the hour.

Peak	int64	Binary Indicator (0 or 1) for peak time.

Table A2. Weather Data

Weather Data Description		
Colum Name	Data Type	Description
AWND	float64	Average daily windspeed in miles per hour (mph).
DAPR	float64	Number of days with measurable precipitation (at least 0.01 inches) in the last 24 hours.
MDPR	float64	Maximum daily precipitation (in inches) recorded in the last 24 hours.
PRCP	float64	Daily precipitation (in inches).
SNOW	float64	Snowfall (in inches) on the ground.
SNWD	float64	Snow depth (in inches) on the ground.
TMAX	float64	Maximum temperature (in degrees Fahrenheit) during the day.
TMIN	float64	Minimum temperature (in degrees Fahrenheit) during the day.
TBOS	float64	Temperature at the time of observation (in degrees Fahrenheit).
Month	int64	Month of the year.
Day	int64	Day of the month.
Year	int64	Year of the observation.
Concord	float64	Daily energy usage.