



Peak Energy Demand Prediction Models

New River Light & Power

Alex Pettis | CIS 5860-101 Applied Analytics Project | August 3, 2024

Contents

Industry Background	3
Purpose	4
Business Problem	5
Project Goals	6
Literature Review.....	7
Introduction.....	7
Machine Learning Models for Peak Demand Prediction	7
Challenges and Future Directions	8
Literature Review Integration	8
Data Collection	9
Data Cleansing.....	11
Data Storage	12
Data Analysis	13
Results.....	15
Concord	15
Table 1 Concord Results.....	15
Concord Result Interpretation	16
Boone.....	17
Table 2 Boone Results	17
Boone Results Interpretation	18
Comparison of Concord and Boone Models	19
Challenges.....	20
Implications.....	21
Recommendations.....	22
References.....	24
Appendix.....	25
Data Details.....	25
Table A1. Hourly Energy Usage Data Description	25

Table A2. Weather Data	26
------------------------------	----

Industry Background

For over a century, Appalachian State University's New River Light and Power (NRLP) has been a cornerstone in providing reliable electricity to Western North Carolina's utility homes and businesses. Managed by the Division of Finance and Operations, NRLP serves nearly 9,000 residents and commercial clients in Boone and the surrounding areas. In January 2022, NRLP began sourcing electricity from Carolina Power Partners, a move that increased its capacity to incorporate more renewable energy. Partnering with App State's Office of Sustainability, Facilities Operations, and the Renewable Energy Initiative, NRLP has facilitated various energy efficiency projects, including the installation of solar panels and funding the Broyhill Wind Turbine. NRLP's commitment to dependable service has been recognized by the American Public Power Association's Reliable Public Power Provider (RP3) program. Annually, it contributes around \$650,000 to App State's general scholarship fund. NRLP's mission is to provide efficient and reliable electrical service to Appalachian State University, Boone, and the surrounding communities, while also supporting the university's financial needs. In doing so, NRLP strengthens the bond between Appalachian State University, Boone, and the local community, fostering economic development and positive public relations.

Purpose

The project aims to create a predictive model for Appalachian State University's New River Light and Power (NRLP) to forecast peak energy demand accurately. This model is designed to pinpoint times of maximum power usage, aiding NRLP in optimizing energy distribution and improve operational efficiency. By informing consumers about peak demand periods, the model helps them manage their energy consumption more efficiently, ultimately reducing costs.

Moreover, the project will support NRLP in enhancing operational efficiency and sustainability by lowering cost during peak demand and facilitating the integration of renewable energy sources. The initiative promotes energy conservation, encouraging consumers to adopt efficient energy practices, and contributes to reducing the overall carbon footprint, in line with NRLP's sustainable objectives.

By providing consumers with valuable insights and tools to manage their energy usage, the project aims to boost customer satisfaction. It also supports NRLP's mission of delivering reliable electrical service, strengthening the relationship between Appalachian State University, the town of Boone, and the surrounding communities, and fostering economic development through improved energy management. Overall, this project aligns with the objectives of enhancing energy distribution management, operational efficiency, and sustainability efforts.

Business Problem

The primary goal is to create an advanced predictive model that accurately forecasts peak energy demand by utilizing historical consumption data, weather conditions, and other pertinent factors. This model will be a crucial tool for NRLP in optimizing energy distribution and enhancing overall efficiency. By alerting consumers about anticipated peak demand times, the project encourages them to modify their energy usage patterns, leading to increased energy efficiency. As a result, consumers can reduce their energy bills by minimizing usage during peak hours, which are typically more expensive. Furthermore, by decreasing peak demand, NRLP can lower its costs related to energy procurement and distribution.

Project Goals

The primary goal of the project is to develop a predictive model that accurately forecasts peak energy demand for NRLP, identifying periods of maximum power consumption to enhance energy distribution management. By predicting peak demand, the model will improve NRLP's operational efficiency, optimizing resource management and reducing costs related to energy procurement and distribution. Additionally, the project supports NRLP's commitment to sustainability by promoting efficient energy use, lowering the carbon footprint, and facilitating the integration of renewable energy sources. It also aims to enhance consumer experience by providing timely information about peak demand periods, enabling users to adjust their energy consumption patterns and potentially save money by avoiding higher peak-hour rates. Ultimately, the project contributes to the broader sustainability goals of Appalachian State University and the surrounding community by fostering energy efficiency and supporting economic development.

Literature Review

Introduction

Effectively managing peak electricity demand is vital for maintaining grid stability and economic efficiency (Fu et al., 2023). Battery Energy Storage Systems (BESS) are emerging as valuable tools for reducing peak demand through their ability to respond instantly (Fu et al., 2023). However, predicting peak demand with precision remains a significant challenge, which affects the optimal use and placement of BESS (Fu et al., 2023).

Machine Learning Models for Peak Demand Prediction

Advances in machine learning (ML) are proving beneficial in refining peak demand predictions. While traditional forecasting methods like Autoregressive Integrated Moving Average (ARIMA) are still used, newer nonlinear ML techniques are enhancing predictions by incorporating various factors such as weather and economic indicators (Fu et al., 2023). Researchers have experimented with multiple ML models, including K-nearest neighbors (KNN), Support Vector Machines (SVM), Gradient Boosting Machines (GBM), Random Forest (RF), and Artificial Neural Networks (ANN). These models, either standalone or combined, aim to predict peak electricity demand more accurately (Fu et al., 2023). For instance, Fu et al. (2023) developed a hybrid supervised ML approach that combines RF, GBM, and Logistic Regression (LR) to forecast peak hours and days. Their method, applied to the Duke Energy Progress system, demonstrated high accuracy in identifying peak times, thereby enhancing BESS operation and addressing uncertainties in peak demand forecasts (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 these advancements, challenges such as insufficient data and the need for better uncertainty quantification remain (Fu et al., 2023). Future research should focus on refining ML models to better integrate temporal and physical factors, thereby improving the reliability and interpretability of peak demand predictions (Fu et al., 2023).

Literature Review Integration

The development of the predictive model for New River Light and Power (NRLP) was heavily influenced by the recent advancements in peak demand forecasting described in the literature. Fu et al. (2023) highlighted the significance of incorporating multivariate predictors, such as weather and economic factors, to improve prediction accuracy. Building on these insights, the project applied various machine learning techniques, including Multiple Regression, Logistic Regression, Support Vector Machines (SVM), and Fuzzy Regression, to explore the impact of weather variables on energy consumption patterns.

Data Collection

In my data collection process, I leveraged two primary sources: energy usage data and weather data. The energy usage data, generously provided by the project sponsor, offered detailed hourly records for five diverse locations, including Boone. This granular data, delivered in CSV format, was crucial for capturing the nuanced energy consumption patterns throughout each day. Complementing this, I sourced daily weather data from the NOAA Integrated Surface Database (ISD), following the project sponsor's insightful recommendation during our initial meeting. This database, accessible and reliable, provided essential daily weather metrics such as temperature, precipitation, and wind speed for most of the study locations. The convergence of these rich datasets forms a robust foundation for the analysis, ensuring a comprehensive understanding of the interplay between energy usage and weather conditions. By meticulously integrating these data sources, I have not only set the stage for a thorough and nuanced analysis but also paved the way for future research and practical applications in optimizing energy consumption based on weather patterns.

However, I encountered a significant challenge when collecting the weather data for Boone, as the dataset lacked crucial information on average wind speed. Determined to ensure a comprehensive dataset, I proactively sourced the missing wind data from another reliable website. This not only filled a critical gap but also enhanced the overall integrity of the dataset, ensuring that the analysis could account for all relevant weather variables across all locations. Python proved indispensable in this endeavor, providing the flexibility and power needed for efficient data management and integration. This meticulous and adaptive approach to data collection has laid a solid foundation for a thorough and accurate analysis, ultimately contributing to a deeper understanding of the interplay between energy usage and 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 Cleansing

In my data cleansing process, I undertook several steps to ensure the dataset was clean and ready for analysis. Using Python, along with the Pandas and NumPy packages, I first addressed the challenge of integrating energy usage data, which was recorded hourly, with daily weather data. To align these datasets, I started by checking for constant columns in the Weather Data (Table A2), identifying and removing those that did not vary and were thus unnecessary for the analysis. I also removed other columns that were deemed irrelevant from the outset. A significant challenge was reconciling the hourly energy usage data from the Energy Usage Data (Table A1) with the daily weather metrics. To address this, I aggregated the hourly energy consumption data by summing it up for each day, which ensured it matched the daily granularity of the weather data. I then incorporated these daily energy usage columns into the weather data. To account for temporal aspects, I split the DATE column into three separate columns—month, day, and year and recognized that different times of the year might require different models due to seasonal variations. Following this, I removed any additional unneeded columns and handled missing values by replacing them with either the mean or the previous value, thus maintaining data integrity. Lastly, I created a correlation matrix to identify and understand relationships between variables, which was instrumental in guiding model selection and feature engineering.

Data Storage

For data storage in my project, I utilized a combination of tools and technologies to ensure efficient management and integrity of the datasets. The primary data formats employed were CSV files, which facilitated ease of access and compatibility with analysis tools. The Energy Usage Data was consolidated into a single CSV file, containing hourly records for all locations. In contrast, the Weather Data was stored in separate CSV files, each dedicated to a specific location, allowing for organized management of daily weather metrics. This approach not only streamlined data handling but also enhanced the efficiency of data integration and analysis.

Data Analysis

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, specifically interacting PRCP and TMIN.

While most of the descriptive analytics had already been performed by a previous student, I added further depth to the analysis by creating a correlation matrix for each location. This matrix was instrumental in examining the relationships between different weather variables and energy demand. By analyzing the correlations, I was able to identify patterns and associations that might not be immediately obvious, such as how specific weather conditions influence energy consumption. This additional layer of descriptive analysis provided valuable insights into the variable interactions, helping to refine the understanding of how different weather factors impact energy demand. Ultimately, this comprehensive approach to descriptive analytics supported more informed decision-making in the development of predictive models.

It was previously determined that different models would be required for various times of the year due to seasonal variations. Therefore, I developed separate models for the summer months (June through September), the winter months (November through March), and distinct models for the transitional months of April and October. This approach ensures that the models can effectively capture seasonal patterns and fluctuations in energy demand. By tailoring the models to specific periods, I aimed to enhance their accuracy and relevance, addressing the

unique characteristics of each season. This segmentation also allows for more precise forecasting and better management of energy resources throughout the year.

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, which was later be applied to the Boone/New River usage data.

Results

Concord

Table 1 Concord 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)	0.4738
Logistic Regression with Interaction Term (PRCP, TMIN, Day, Year, PRCP_TMIN)	Winter (November-March)	0.6015
Logistic Regression with Interaction Term (PRCP, TMIN, Day, Year, PRCP_TMIN)	April	0.3838
Logistic Regression with Interaction Term (PRCP, TMIN, Day, Year, PRCP_TMIN)	October	0.5795
Decision Tree	All	0.7171
Support Vector Machine (SVM)	All	0.925
Fuzzy Regression	Summer (May-September)	0.848

Concord Result Interpretation

The analysis of summer energy demand reveals that the fuzzy regression model and the updated OLS model identify temperature at observation (TOBS) and maximum temperature (TMAX) as significant predictors. The fuzzy regression model demonstrates a strong fit with an R-squared of 0.85, indicating its effectiveness in modeling summer energy usage. Despite these strengths, challenges such as multicollinearity and autocorrelation persist, suggesting that further refinement is necessary. Additionally, some predictors, like precipitation (PRCP), are not significant, emphasizing the predominant role of temperature-related factors in influencing peak demand during the summer months. In the winter period, models highlight snow depth (SNWD) and minimum temperature (TMIN) as key predictors of peak energy demand. The OLS model explains approximately 15.3% of the variance, with SNWD showing a positive relationship with peak demand. However, issues such as high multicollinearity and autocorrelation impact the reliability of these models, indicating the need for adjustments to improve their performance and accuracy. For October, the logistic regression model indicates that minimum temperature (TMIN) is a significant predictor of peak energy demand, accounting for 57.95% of the variance. Other variables, such as precipitation (PRCP), are not significant in this model. The presence of multicollinearity suggests that the model's effectiveness could be compromised by high correlation among predictors, necessitating potential adjustments to enhance its predictive power. The logistic regression model for April shows TMIN as a significant predictor of peak energy demand, although it only explains 38.38% of the variance. Precipitation does not significantly impact peak days in this period. The model also faces challenges with multicollinearity, which

affects its effectiveness. Given the lower explanatory power compared to other months, additional predictors or model adjustments may be needed to improve its accuracy.

Boone

Table 2 Boone Results

Model	Season	R-Squared/Accuracy
Multiple Regression (All Variables)	Summer (May-September)	0.511
Multiple Regression (DAPR, MDP, SNOW, SNWD, TMAX, TBOS, Month, Day, Year)	Summer (May-September)	0.508
Multiple Regression (All Variables)	Winter (November-March)	0.847
Multiple Regression (All Variables)	April	0.333
Multiple Regression (All Variables)	October	0.047
Logistic Regression with Interaction Term (PRCP, TMIN, Day, Year, PRCP_TMIN)	Summer (May-September)	0.082
Logistic Regression with Interaction Term (PRCP, TMIN, Day, Year, PRCP_TMIN)	Winter (November-March)	0.487
Logistic Regression with Interaction Term (PRCP, TMIN, Day, Year, PRCP_TMIN)	April	0.263
Logistic Regression with Interaction Term (PRCP, TMIN, Day, Year, PRCP_TMIN)	October	0.091
Decision Tree	All	0.209
Support Vector Machine (SVM)	All	0.895
Fuzzy Regression	Summer (May-September)	0.511

Boone Results Interpretation

The analyses provide a nuanced understanding of predicting peak energy usage for Boone, highlighting the impact of various factors across seasons, including transitional months like April and October. Temperature-related variables, particularly minimum temperature (TMIN), and weather patterns such as snow depth (SNWD) and precipitation (PRCP), emerge as crucial in forecasting peak energy demand. During summer, elevated levels of DAPR and MDPR strongly correlate with increased energy usage, whereas SNOW and SNWD tend to reduce demand. In contrast, winter models underscore the significance of minimum temperature and wind speed in predicting peak demand, with precipitation playing a less critical role. For transitional months, April shows a mix of significant predictors, including DAPR, MDPR, and SNWD, which exhibit both positive and negative relationships with energy demand. October, however, presents limited predictive power, with significant factors such as DAPR, MDPR, SNOW, and SNWD contributing minimally to the overall model. The findings suggest that while seasonal models provide robust predictions, addressing the unique patterns of transitional months can enhance the accuracy and effectiveness of peak energy demand forecasting and energy management strategies for Boone.

Comparison of Concord and Boone Models

The comparison between the Concord and Boone models highlights how insights from one location can improve predictions for the other. Concord's strong performance in summer with the fuzzy regression model ($R\text{-squared} = 0.848$) underscores the value of temperature-related variables, suggesting that similar methods could be beneficial for Boone, where summer models are less accurate. In winter, both Concord and Boone identify SNWD and TMIN as crucial predictors, but Concord's issues with multicollinearity and autocorrelation are reflected in Boone's model performance, indicating that addressing these issues could enhance Boone's accuracy. For the transitional months of April and October, Concord's models show that TMIN is a significant predictor, which offers a potential improvement for Boone's models, particularly where October shows weak performance. The success of advanced techniques like SVM and fuzzy regression in Concord suggests that adopting these approaches could enhance Boone's predictive accuracy. Overall, the challenges and successes observed in Concord provide valuable guidance for refining Boone's models, especially in managing multicollinearity and improving the predictive power of temperature-related variables.

Challenges

Throughout the project, several challenges emerged that required strategic solutions to enhance model performance and ensure data consistency. Initially, low R-squared values indicated inadequate predictive power, prompting the reintroduction of the date column segmented into Month, Day, and Year to capture temporal nuances more effectively. Additionally, the disparity in data granularity, hourly energy usage data versus daily weather data, necessitated aggregating the energy data into daily averages to ensure compatibility and coherence in the analysis. Time constraints due to commitments in another course also limited the depth of exploratory analysis and model refinement. Despite these limitations, focusing on essential adjustments and strategic solutions helped address the challenges and improve overall model performance.

Implications

The results from this project offer valuable insights for improving the prediction of peak energy usage and guiding future efforts. The analysis reveals significant predictors of peak demand for each season, such as temperature-related factors like TMAX and TMIN for summer and SNWD and TMIN for winter. To enhance model performance, it is crucial to refine existing models by incorporating these significant variables more effectively. Additionally, addressing data granularity issues where energy usage is recorded hourly and weather data daily requires aggregating energy usage data into daily averages to align temporal scales and improve model reliability. Reintroducing and segmenting the date column into Month, Day, and Year has proven beneficial for capturing temporal nuances, and this approach should be maintained in future models to better account for seasonal variations. Furthermore, managing multicollinearity through careful variable selection and regularization techniques, such as Lasso or Ridge regression, will improve model reliability and interpretability. Continuous validation and refinement of models are essential, given time constraints and the need for deeper exploratory analysis, to ensure that models remain accurate and relevant over time. Implementing these insights will not only improve predictive accuracy but also support more informed decision-making. Finally, the insights gained from these models can inform strategic energy management practices, allowing organizations to implement targeted strategies for managing energy consumption, optimizing usage during peak periods, and ultimately reducing energy costs while enhancing efficiency. By proactively addressing these aspects, organizations can better align their energy management strategies with actual demand patterns, leading to more effective and cost-efficient energy solutions.

Recommendations

To enhance NRLP's predictive capabilities through machine learning, focus on several key recommendations. First, tackle multicollinearity issues in winter models by employing advanced feature selection and dimensionality reduction techniques such as Principal Component Analysis (PCA) or regularization methods like Lasso. These methods will help simplify the models and improve interpretability, leading to more accurate predictions. Additionally, refine the models by developing more granular seasonal models that account for specific weather patterns and incorporating interaction terms to better capture the complex relationships between predictors and peak demand.

Integrate additional weather variables, such as humidity and dew point, to provide a more nuanced understanding of energy demand patterns. These variables can offer deeper insights into how different weather conditions impact energy usage, potentially leading to more precise forecasts. For summer, focus on analyzing high-temperature periods and their impact on peak demand, utilizing advanced techniques like Long Short-Term Memory (LSTM) networks or recurrent neural networks (RNNs) to manage the complex time series data and variability in energy use during peak heat events.

Future research should explore these advanced machine learning techniques and conduct comparative studies to evaluate their effectiveness in improving prediction accuracy across different seasons. Investigate how deep learning models, including LSTM and RNN, can enhance forecasting by capturing intricate patterns and trends in the data. Establish a robust framework for continuous performance monitoring to assess model effectiveness regularly, adjusting based on performance metrics and user feedback. This ongoing optimization will

ensure that NRLP's energy demand forecasting remains accurate and effective throughout the year, ultimately improving operational efficiency and supporting sustainability goals.

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/Boone	float64	Daily energy usage for each respective location.