05-11-2024

DATA ENGINEERING (CET3005B) MINIPROJECT REPORT

RENEWABLE ENERGY DATA ANALYSIS

GROUP MEMBERS:
Taksh Dhabalia
Panel A - 1032220493
Dharani Barigeda
Panel A - 1032220652
Prabhjot Bhatia
Panel A - 1032220689

TABLE OF CONTENTS

ABSTRACT

This project, titled Renewable Energy Data Analysis, investigates renewable energy trends in relation to demand and supply gaps, focusing specifically on solar power generation. Given the increasing reliance on renewable energy sources, understanding the daily and seasonal patterns of energy generation, as well as identifying any potential deficits, is crucial for efficient energy management. This analysis examines power generation from two solar plants, detects outliers in energy requirements, and explores correlations among key variables. Findings from this study reveal patterns that could guide future energy optimization and offer insights into daily and seasonal fluctuations, with implications for improving energy forecasting accuracy and supporting grid stability during peak demand periods.

INTRODUCTION

The transition to renewable energy has become a global priority as concerns about climate change and energy security continue to grow. Solar energy, in particular, is increasingly adopted due to its low environmental impact and long-term sustainability. However, the variable nature of solar power generation, influenced by factors like weather and seasonal changes, creates challenges for meeting continuous energy demand. In this context, accurate data analysis and trend identification are critical to balance energy availability with demand and mitigate deficits.

This project addresses the need for structured analysis of renewable energy generation and demand. By analyzing data from two solar plants, we aim to identify patterns in energy generation, assess the reliability of current supply levels, and detect areas of significant deficit.

The objectives of this project are threefold: (1) to analyze trends in energy generation and requirement, (2) to evaluate generation consistency across daily and seasonal timelines, and (3) to visualize areas where demand outpaces supply, suggesting room for improved forecasting.
The scope of this analysis is limited to data provided by two solar power plants, focusing on daily and seasonal patterns within a specific timeframe. While this scope allows for focused insights, limitations include the absence of real-time predictive modeling and the restriction to solar power data, excluding contributions from other renewable sources. This report seeks to provide actionable insights that can aid in more balanced and effective energy management practices.

METHODOLOGY

- Data Sources: The project utilized several datasets covering solar power generation, sensor data for ambient conditions, and energy demand and availability metrics. Each dataset was loaded from CSV files and included key variables such as DC and AC power generation, ambient and module temperatures, and irradiation levels.

- Preprocessing: Data preprocessing was crucial to ensure reliability in analysis. Missing values were assessed using heatmaps, while outliers in the energy requirement and availability data were detected using Z-score analysis. Outliers, defined as values with a Z-score exceeding ±3, were removed to avoid skewed results. Additionally, non-essential columns in the electricity generation dataset, including those unrelated to power generation, were dropped to focus the analysis on relevant factors.

- Aggregation and Analysis: Power generation data was aggregated on a daily basis to compute daily DC and AC power yields. Grouping by date and time allowed for the analysis of time-based generation trends and offered insight into daily power output consistency. This aggregation facilitated the analysis of patterns across different seasons, providing a view of solar power availability trends throughout the year.

- Visualization and Correlation: Visualizations, including line plots, bar charts, and heatmaps, played an integral role in trend identification. Line plots depicted daily power generation trends, while bar charts illustrated yearly energy deficits, highlighting demand-supply gaps. Correlation heatmaps for each plant were generated to evaluate relationships among variables such as DC power, AC power, and ambient conditions. These visualizations enabled a comprehensive understanding of the interplay between power output and influencing factors.

- Technologies Used: Python was the primary language for this project, with extensive use of libraries like Pandas, Matplotlib, Seaborn, and Scipy. These libraries facilitated data cleaning, aggregation, outlier detection, and correlation analysis, allowing for an in-depth exploration of renewable energy trends.

RESULTS AND DISCUSSION

The project's findings underscore both the potential and limitations of solar power in meeting continuous energy demand. Analysis revealed clear daily and seasonal trends in power generation, with generation peaking during sunny periods and decreasing during cloudy or low-irradiation days. This variability necessitates accurate forecasting to prevent shortages during low generation periods.

The outlier analysis showed significant variations in the energy requirement data, especially during peak demand seasons, when requirements often outpaced supply. By removing extreme outliers, we obtained a cleaner view of general trends, allowing for more reliable energy deficit analysis. The correlation heatmaps showed strong correlations between ambient conditions and power output, confirming the influence of environmental factors on solar power generation.

The visualization of yearly energy deficits highlighted the periods where supply gaps are most pronounced, suggesting the need for enhanced energy storage solutions or alternative renewable sources during these times. This analysis demonstrates the importance of monitoring seasonal patterns and incorporating predictive techniques for better resource allocation. Overall, while the project met its objectives in identifying trends and deficits, further advancements in forecasting methods could improve energy management outcomes.

CONCLUSION

This study provides valuable insights into the trends and challenges associated with renewable energy generation. By focusing on solar power data from two plants, the project identifies key trends in energy availability and the potential for generation shortages during certain periods. The results demonstrate that solar power generation is highly influenced by environmental factors, underlining the need for effective energy storage and diversification to ensure consistent supply.

The findings also suggest that predictive modeling, incorporating both historical data and real-time inputs, could further enhance the balance between demand and availability. Future research could explore machine learning techniques or integrate data from additional renewable sources, such as wind power, to improve forecasting accuracy and resource distribution.

REFERENCES

- Kanani, A. Solar Power Generation Data. Kaggle. Retrieved from https://www.kaggle.com/datasets/anikanani/solar-power-generation-data
- Government of India. Renewable Energy Data. Data.gov.in. Retrieved from https://data.gov.in/dataset-group-name/Renewable%20Energy
- Breiman, L. (2001). Random Forests. Machine Learning, 45(1), 5-32. doi:10.1023/A:1010933404324 (For Random Forest methodology)
- Kuhn, M., & Johnson, K. (2013). Applied Predictive Modeling. Springer. Retrieved from https://link.springer.com/book/10.1007/978-1-4614-6849-3 (For model evaluation techniques like MSE and R^2 Score)
- International Energy Agency. (2023). Renewable Energy Market Update. IEA. Retrieved from https://www.iea.org/reports/renewable-energy-market-update-2023 (Background on global renewable energy trends)