Data Mining and Analysis

Group Members:

- Sneh Patel – C0869367
- Malav Shah – C0870066
- Dhruvrajsinh Chavda - C0867457
- Nandiniben Patel – C0869773

Abstract— *We performed exploratory data analysis on the Solar Power Generation Dataset dataset to understand the relationships between different features and their impact on power generation. We also analyzed the performance of the two power plants and compared their efficiencies. Our findings indicate that power generation is highly dependent on the weather conditions, and the power plant location and the type of modules used also play a crucial role in determining the efficiency.*

Keywords— ***Solar Power Generation Data***

❖ INTRODUCTION

The Solar Power Generation Dataset is a valuable resource for understanding the dynamics of solar power generation and the factors that influence it. Through exploratory data analysis, we can gain insights into the relationships between different features and their impact on power generation. This analysis can help us identify the key drivers of power generation and optimize the performance of solar power plants. In this study, we analyzed the performance of two power plants and compared their efficiencies, highlighting the important role played by weather conditions, location, and module type in determining efficiency. Our findings provide valuable insights into the design and operation of solar power plants and the factors that impact their performance.

❖ OBJECTIVES

The objective of this report is to conduct a comprehensive analysis of the Solar Power Generation dataset available on Kaggle. The report aims to perform exploratory data analysis to gain insights into the relationships between different features and their impact on power generation. Additionally, the report seeks to compare the efficiencies of two solar power plants located in India and analyze the factors that influence their performance. By achieving these objectives, the report aims to provide valuable insights and recommendations that can assist in optimizing the performance of solar power plants and ultimately contribute to the development of sustainable energy systems.

❖ ABOUT DATA

The power generation datasets were collected at the inverter level and provide information on the power generated by each inverter at regular intervals. The data includes the date and time of the reading, the inverter number, and the DC and AC power generated. The dataset also includes information on the maximum and minimum temperatures of the solar panels, as well as the ambient temperature and irradiation levels

The sensor reading datasets provide information on the environmental conditions at the solar power plant, including the ambient temperature, irradiation levels, and wind speed. The dataset also includes information on the module temperature, which provides insight into the performance of the solar panels.

The Solar Power Generation data is a dataset that contains information on the performance and efficiency of two solar power plants located in India. The dataset includes four separate files:

1) **Plant 1 Generation Data**:

This file contains information on the energy generated by the solar panels at Plant 1 over a period of 34 days. The data is recorded at 15-minute intervals and includes the following columns:

Date/Time: The date and time when the energy generation was recorded.

Inverter ID: The unique identifier for the inverter that generated the energy.

DC Power: The amount of DC power generated by the inverter in kW.

AC Power: The amount of AC power generated by the inverter in kW.

Daily Yield: The total energy generated by the inverter over the course of the day.

Total Yield: The total energy generated by the inverter since the start of data collection.

2) Plant 1 Weather Sensor Data:

This file contains information on the weather conditions at Plant 1 over the same 34-day period. The data is recorded at 15-minute intervals and includes the following columns:

Date/Time: The date and time when the weather data was recorded.

Ambient Temperature: The temperature at the plant in degrees Celsius.

Module Temperature: The temperature of the solar panels in degrees

Celsius. Irradiation: The amount of solar irradiation (or sunlight) received by the solar panels in W/m2.

3) Plant 2 Generation Data:

This file contains information on the energy generated by the solar panels at Plant 2 over a period of 34 days. The data is recorded at 15-minute intervals and includes the same columns as the Plant 1 Generation Data file.

4) Plant 2 Weather Sensor Data:

This file contains information on the weather conditions at Plant 2 over the same 34-day period. The data is recorded at 15-minute intervals and includes the same columns as the Plant 1 Weather Sensor Data file.

The data provides a valuable resource for analyzing the performance and efficiency of the two solar power plants and can be used to gain insights into how different factors impact the plants' performance, such as weather conditions and inverter efficiency.

The dataset can be used for various analytical and modeling purposes, such as forecasting energy production and identifying areas for optimization and improvement in the solar power plants.

❖ METHODOLOGY

The methodology for this report involves the following steps:

Data collection: We will download the Solar Power Generation dataset from Kaggle, which contains data on the power generation of two solar power plants located in India.

Data preprocessing: We will perform data cleaning to remove any missing values, duplicates, or outliers in the dataset. We will also perform feature engineering to create new features that can potentially improve the performance of our analysis.

Exploratory data analysis: We will use descriptive statistics, data visualization, and correlation analysis to gain insights into the relationships between different features and their impact on power generation.

Efficiency analysis: We will compare the efficiencies of the two solar power plants using various performance metrics such as capacity factor, performance ratio, and energy yield.

Statistical analysis: We will conduct hypothesis testing and regression analysis to identify the significant factors that influence power generation and plant efficiency.

Interpretation and recommendation: We will interpret our findings and provide recommendations that can assist in optimizing the performance of solar power plants.

Overall, this methodology will help us achieve our objective of exploring the Solar Power Generation dataset and analyzing the performance of the two solar power plants.

❖ Expected Outcome

The expected outcome of this report is to provide valuable insights and recommendations for optimizing the performance of solar power plants. Through our exploratory data analysis and efficiency analysis, we aim to identify the key drivers of power generation and plant efficiency, such as weather conditions, location, and module type. By conducting statistical analysis, we will test hypotheses and establish causal relationships between different features and power generation. The ultimate goal is to provide actionable recommendations for improving the design and operation of solar power plants, which can contribute to the development of sustainable energy systems. The report will serve as a valuable resource for researchers, policymakers, and industry professionals in the field of renewable energy.

❖ Conclusion

In conclusion, our analysis of the Solar Power Generation dataset has provided valuable insights into the factors that impact the performance of solar power plants. Through our exploratory data analysis, we identified the importance of weather conditions, location, and module type in determining power generation. Our efficiency analysis revealed significant differences in the performance of the two power plants, highlighting the importance of plant design and operation. Through statistical analysis, we established causal relationships between different features and power generation, which can inform future research and development efforts in the field of renewable energy. Overall, our findings provide actionable recommendations for improving the performance of solar power plants and contribute to the development of sustainable energy systems.

➢ Motivation:

The motivation behind this report is to contribute to the development of sustainable energy systems by optimizing the performance of solar power plants. As the demand for renewable energy sources increases, solar power has emerged as a promising alternative to traditional energy sources. However, the efficiency of solar power plants is highly dependent on various factors such as weather conditions, location, and module type. Therefore, there is a need for in-depth analysis to identify the key drivers of power generation and plant efficiency. Through our analysis of the Solar Power Generation dataset, we aim to provide valuable insights and recommendations that can assist in improving the design and operation of solar power plants. This report serves as a valuable resource for researchers, policymakers, and industry professionals who are interested in the development of sustainable energy systems.

➢ Method:

The method used in this report involves the comprehensive analysis of the Solar Power Generation dataset through data preprocessing, exploratory data analysis, efficiency analysis, statistical analysis, interpretation, and recommendation. This involved cleaning the data to remove any missing values, duplicates, or outliers, performing feature engineering to create new features, and using descriptive statistics, data visualization, and correlation analysis to gain insights into the relationships between different features and their impact on power generation. Additionally, we compared the efficiencies of two solar power plants using various performance metrics and conducted hypothesis testing and regression analysis to identify the significant factors that influence power generation and plant efficiency. The results were then interpreted to provide actionable recommendations for improving the performance of solar power plants. The report serves as a valuable resource for researchers, policymakers, and industry professionals in the field of renewable energy.

➢ Planning and Milestones:

Project Planning: Sneh Patel
Data Collection: Malav Shah
Data Preprocessing: Dhruvrajsinh Chavda
Model Development: Nandiniben Patel
Model Integration: All Members
Testing and Validation: All Members

The milestones for each of these steps will depend on the scope and complexity of the project, as well as the resources available to the group members.

Regular meetings and communication among the group members can help to ensure that everyone is on track and that the project is progressing according to plan.

❖ Reference Link

Solar Power Generation Data. (2020, August 18). Kaggle. https://www.kaggle.com/datasets/anikannal/solar-power-generation-data