Brandenburg
University of Technology
Cottbus - Senftenberg

WIND ENERGY FORECASTING PROJECT

Comprehensive Data Overview Report

A Detailed Analysis of Wind Energy Dataset

Table of Contents

1. Executive Summary

This comprehensive report provides an in-depth analysis of the wind energy dataset used for developing machine learning forecasting models. The dataset contains daily measurements of wind power generation, installed capacity, and temperature for Germany spanning from 2017 to 2019. This analysis serves as the foundational understanding required for building accurate and reliable forecasting models.

Key Highlights:

- Total dataset contains **1,094 daily records**, providing comprehensive coverage
- Time period spans **1094 days** from 2017-01-01 to 2019-12-30
- Average daily wind generation: **305,820 MW**
- Peak wind generation recorded: **998,899 MW**
- Average capacity utilization: **679.37%**

2. Introduction to Wind Energy Forecasting

Wind energy has emerged as one of the most important renewable energy sources in the global transition towards sustainable power generation. Accurate forecasting of wind power generation is crucial for energy grid management, economic planning, and ensuring reliable power supply.

What is Wind Energy Forecasting?

Wind energy forecasting involves predicting the amount of electrical power that will be generated by wind turbines over a specific time period. This prediction is based on historical patterns, weather conditions, and various meteorological factors. Accurate forecasts help energy operators plan power generation, manage grid stability, and optimize energy trading.

3. Dataset Overview and Characteristics

The dataset used in this project contains comprehensive daily measurements of wind energy generation in Germany. Understanding the dataset structure and characteristics is essential for developing effective forecasting models.

Attribute	Value	Explanation
Total Records	1,094	Number of daily observations in the dataset
Date Range Start	2017-01-01	First date of data collection
Date Range End	2019-12-30	Last date of data collection
Total Days	1094	Complete time span covered
Number of Variables	3	Total features/columns in the dataset
Data Frequency	Daily	Measurements taken once per day
Geographic Region	Germany	Country where data was collected

What Do These Numbers Mean?

The dataset contains **1,094 daily records**, which means we have wind energy measurements for approximately **3.0 years** of data. This is a substantial amount of historical data that allows machine learning models to learn patterns, seasonal trends, and relationships between variables. Having multiple years of data is crucial because it enables models to capture yearly seasonal patterns, which are important for accurate forecasting.

6. Statistical Summary and Descriptive Analysis

Statistical analysis helps us understand the distribution, central tendencies, and variability of our data. These statistics are fundamental for understanding what 'normal' values look like and identifying potential outliers or unusual patterns.

Statistic	wind_generation_actual	wind_capacity	temperature
count	1,094	1,094	1094.00
mean	305,819.77	45,066.02	10.05
std	205,728.86	4,315.84	7.74
min	16,482	37,149	-9.36
25%	148,553	41,448	3.58
50%	254,332	46,073	10.05
75%	412,570	49,203	16.57
max	998,899	50,452	28.24

Understanding These Statistics:

COUNT: Number of non-missing values - tells us how many data points we have

MEAN: Average value (sum of all values divided by count) - the typical value

STD: Standard deviation - measures how spread out the values are (higher = more variability)

MIN: Minimum value observed - the lowest value in the dataset

25%: First quartile - 25% of values are below this number

50%: Median - the middle value when data is sorted (50% above, 50% below)

75%: Third quartile - 75% of values are below this number

MAX: Maximum value observed - the highest value in the dataset

Key Observations from Statistics:

• Wind generation ranges from 16,482 MW to 998,899 MW, showing significant variability

• The median (254,332 MW) is lower than the mean (305,820 MW), indicating right-skewed distribution

• Standard deviation of 205,729 MW shows high day-to-day variability in wind generation

• Temperature ranges from -9.4°C to 28.2°C, typical for German climate

7. Data Quality Assessment

Data quality is crucial for building reliable machine learning models. We assess completeness, consistency, and validity of the data. High-quality data leads to better model performance.

Column Name	Missing Values	Percentage	Data Quality
Wind Generation Actual	0	0.00%	Excellent
Wind Capacity	0	0.00%	Excellent
Temperature	0	0.00%	Excellent

What Does Data Quality Mean?

Missing values represent data points that were not recorded or are unavailable. High percentages of missing data can reduce model accuracy because the model has less information to learn from. Our dataset shows excellent data quality with minimal or no missing values, which is ideal for machine learning. This means we can use almost all of our data for training models without having to remove many records.

Key Insights

- Average daily wind generation: 305,820 MW
- Maximum wind generation: 998,899 MW
- Minimum wind generation: 16,482 MW
- Average wind capacity: 45,066 MW
- Average temperature: 10.05 °C
- Capacity utilization (avg): 679.37%

Student: **Abdul Ghaffar Ansari**

Report Generated: January 13, 2026