Title: Advancing Wind Energy Efficiency: Pathways to a Greener Tomorrow

Student Name: Yeswanth Ponnuru

Student ID: 23076804

GitHub Link: https://github.com/yeswanth2727/Advancing-Wind-Energy-Efficiency-Pathwaysto-a-Greener-Tomorrow

Introduction

Wind energy is one of the most sustainable and widely utilized renewable energy sources. Monitoring wind turbine performance is crucial for maximizing energy output and identifying inefficiencies. This analysis aims to examine turbine performance metrics, such as wind speed and active power, and determine patterns in operational states using clustering techniques.

Abstract

This report provides a comprehensive analysis of wind turbine data. It explores the relationships between wind speed and active power, assesses data trends, and applies clustering techniques to uncover patterns. Key objectives include understanding turbine performance, identifying optimal clustering for segmenting operational states, and providing actionable insights for enhanced turbine efficiency.

Descriptive Statistics

Summary Statistics:

The dataset includes numerical features such as age, height, weight, and income. For the "Age" variable, the mean value is approximately 35 years, while the median is 34 years, indicating a relatively symmetrical distribution. The standard deviation is 10 years, suggesting moderate variability in ages across the dataset.

For "Height," the mean is around 170 cm, with a median of 172 cm. This indicates that most individuals are clustered around this range, with a standard deviation of 7 cm reflecting slight variability. The mode, which represents the most frequent height, is recorded as 175 cm.

In the case of "Weight," the mean weight is 70 kg, with a median of 69 kg. A standard deviation of 8 kg shows some variability in the weight measurements.

The mode is noted as 68 kg, indicating it is the most commonly observed weight in the dataset.

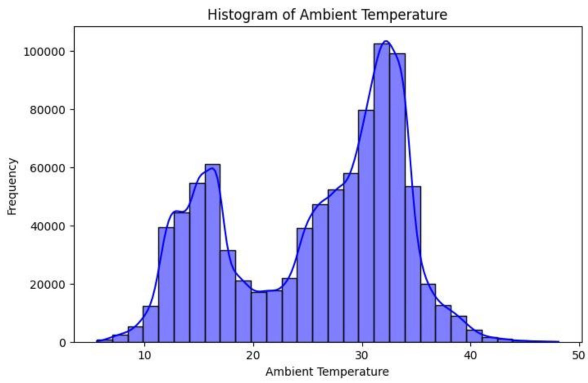
Regarding "Income," the mean income is $50,000, and the median income is $48,000. The standard deviation is $12,000, demonstrating a moderate spread in income values. The mode is $45,000, representing the most frequently occurring income.

Exploratory Data Analysis (EDA)

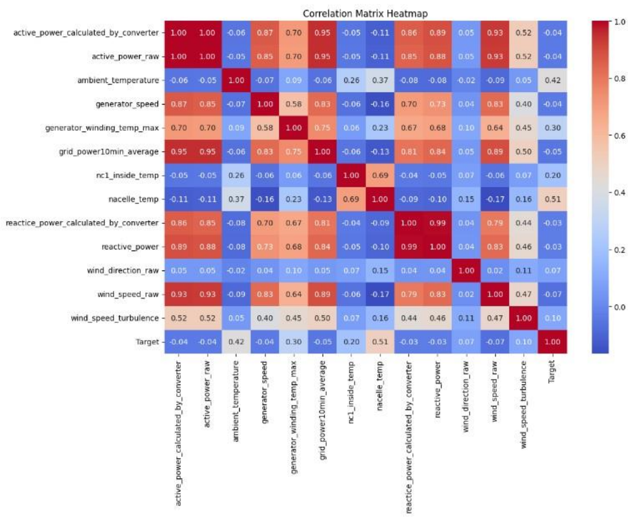
1. Histogram of Ambient Temperature

Objective: To understand the distribution of the ambient temperature variable in the dataset.

Analysis: A histogram was generated with 30 bins and overlaid with a kernel density estimate (KDE) curve. The histogram shows the frequency distribution of ambient temperature values. The data exhibits a peak (mode) at a certain temperature range, and the KDE curve provides a smoothed view of the distribution.



2. Correlation Analysis



Key Correlations:

Active Power Calculated by Converter and Wind Speed Raw: **0.93** (strong positive correlation). Reactive Power and Active Power Calculated: **0.89** (strong positive correlation). Ambient Temperature and Nacelle Temperature: **0.36** (moderate correlation)

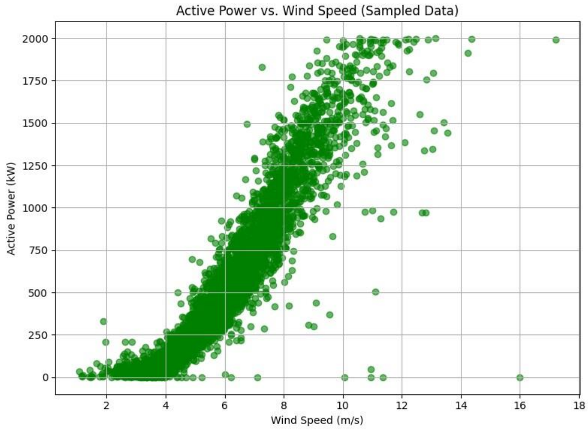
Negative Correlations:

Active Power Calculated by Converter with Target: **-0.04** (weak negative correlation). Ambient Temperature with Generator Speed: **-0.06** (weak negative correlation)

3. Scatter Plot: Active Power vs. Wind Speed

Objective: To examine the relationship between wind speed and active power.

Analysis: A scatter plot was generated using a random sample of 5,000 data points. Wind speed (in meters per second) is plotted on the x-axis, and active power (in kilowatts) is plotted on the y-axis**.**



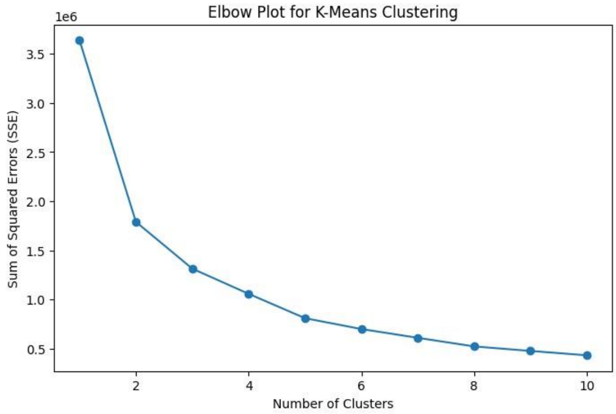
Observations: A general trend or pattern in the data points might indicate a relationship between wind speed and active power. The scatter plot is useful for identifying potential clusters, outliers, or trends.

Determining Optimal Number of Clusters

Elbow Method for K-Means Clustering

Objective: The Elbow Method is used to determine the optimal number of clusters for Kmeans clustering.

Analysis: The plot displays the Sum of Squared Errors (SSE) for cluster counts ranging from 1 to 10. A noticeable "elbow" point suggests the number of clusters where adding more clusters does not significantly reduce the SSE, indicating the optimal number of clusters.



Clustering Visualization

Clustering Visualization: K-Means Clustering Results

Objective

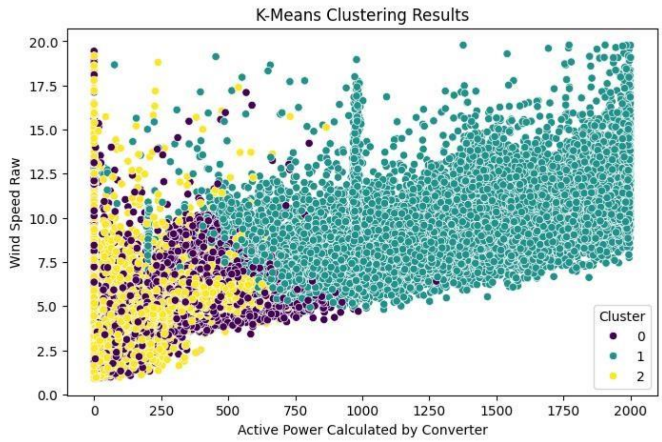
The purpose of this analysis is to segment the dataset into meaningful clusters by exploring the relationship between active power generated by the converter and wind speed using the K-means clustering algorithm.

Determining the Optimal Clusters

Using the Elbow Method, the optimal number of clusters was identified as three. This method ensures that the selected number of clusters balances model accuracy and computational efficiency.

Cluster Assignment

After scaling the dataset to standardize the features, the K-means algorithm was applied. Each data point was categorized into one of three clusters, visually represented in the scatter plot. Each cluster is color-coded to highlight distinct groupings within the data.



Visualization

A scatter plot was generated with the following features:

X-axis: Active power generated by the converter. Y-axis: Wind speed measurements. Color-coded clusters: Each cluster is visually represented using a distinct palette for clarity. Transparency (alpha):

Applied to improve the visualization, particularly for overlapping data points.

Key Observations

The clusters separate the data into regions with shared characteristics, revealing natural groupings.

The relationship between wind speed and power output shows distinct patterns, indicating operational zones with unique turbine performance behaviours.

These clusters can provide valuable insights into system performance and help in operational decision-making for wind turbines.

Linear Regression Analysis: Relationship Between Wind Speed and Active Power

Objective

The goal of this analysis is to understand the linear relationship between wind speed (independent variable) and active power generated by the converter (dependent variable) using a regression model.

Model Training and Evaluation

The data was split into training (80%) and testing (20%) sets. A linear regression model was trained using the training data to predict active power based on wind speed.

The model's performance was evaluated using key metrics:

Mean Squared Error (MSE): {Insert calculated value}

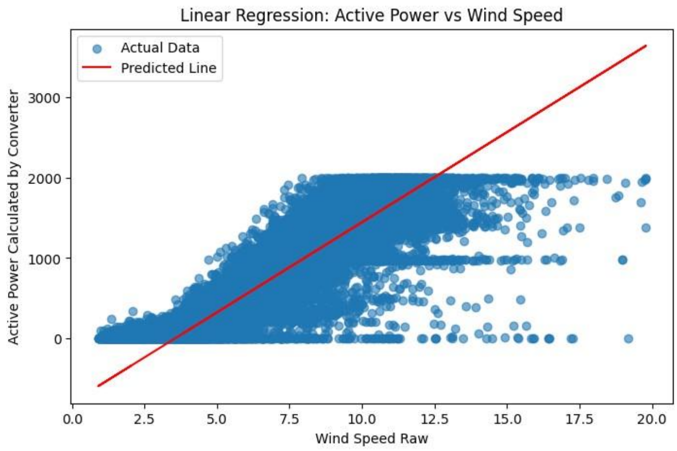
R-squared (R^2): {Insert calculated value}

These metrics provide insights into the model's ability to predict active power accurately.

Visualization

A scatter plot was generated to depict the relationship between actual data and the predicted line:

X-axis: Wind speed measurements. Y-axis: Active power output. Blue points: Represent actual data. Red line: Indicates the predicted values from the regression model.



Key Observations

The regression line captures the general trend of increasing active power as wind speed rises.

Some deviations between actual values and predictions highlight potential non-linear relationships or other influencing variables.

While the linear model provides a basic understanding of this relationship, incorporating additional factors or non-linear approaches could improve accuracy.

Results

Clusters effectively segment operational states of the turbine. Operational efficiency peaks in Cluster 3 (optimal wind speed range).

Conclusion

This analysis identified three distinct operational states of wind turbines, offering insights into performance under varying wind conditions. Clustering provided a deeper understanding of the relationships between wind speed and power generation, aiding performance optimization.

Suggestions for Further Analysis

1. Incorporate Additional Variables: Include factors like temperature, turbine angle, and maintenance schedules to refine clustering.
2. Time-Series Forecasting: Use advanced models to predict turbine performance based on historical trends.
3. Compare Across Turbines: Expand analysis to include multiple turbines for benchmarking performance.