Project Title: Wind Energy Production Forecasting Using Historical Wind

Prof. Dr. ZAFER ASLAN

DATA MINING SEN431

(Group 9)

MOHI EDDIN ALKADMANI	B2005.090146
MOHAMMAD HAMZA	B1805.090019
WASSIM IDALBI	B2005.090026
Ahmad miri	B2205.090058

1. Definition of the Topic, Problem, Aim of the Report

Topic:

Wind Energy Production Forecasting Using Historical Wind Speed and Direction Data for **FETHIYE** Region Over the Next Year and Validating Model's Predictions with Existing Data.

Problem:

Accurate forecasting of wind energy production is crucial for optimizing the operation and maintenance of wind farms, ensuring grid stability, and maximizing economic returns. However, variability in wind speed and direction poses significant challenges in predicting energy output and making informed decisions regarding turbine placement and resource allocation.

Aim:

To analyze a 4-year dataset of daily wind speed and direction to develop predictive models for wind speed and direction, using data to predict the future and compare the model's performance, while also applying data mining techniques to enhance operational efficiency through data-driven insights.

---

2. General Review on Literature

Overview of Wind Energy and Its Importance

Wind energy plays a significant role in the global renewable energy landscape. Saidur et al. (2011) highlight that wind energy is one of the healthiest and most environmentally friendly energy sources available, with a lower carbon footprint compared to other renewables like biomass, solar PV, and marine energy. They also note wind energy's potential to reduce air pollution and water consumption compared to fossil fuel power plants.

Ramírez et al. (2018) discuss how wind energy has been the renewable source that has contributed the most to renewable energy development in Europe. The authors analyze the substantial wind energy capacity in Spain, which had the second highest installed wind power in Europe as of 2016. Their research assesses how wind energy repowering and new installations can help Spain meet its 2020 renewable energy targets.

In summary, these studies underscore wind energy's environmental benefits, sustainability, and key position in enabling countries to achieve renewable energy goals, while acknowledging the need for careful planning to minimize local impacts.

Wind Data Analysis in Renewable Energy

Wind speed and direction data are crucial inputs for accurately forecasting wind power generation. Yatiyana et al. (2017) develop an autoregressive integrated moving average (ARIMA) model to predict both wind speed and direction based on historical time series data. Their analysis shows that considering wind direction in addition to just wind speed can significantly improve the accuracy of wind power forecasts. By modeling speed and direction together, their ARIMA approach reduces forecast errors to under 5% for speed and 16% for direction in case studies. This underscores the importance of factoring in wind direction when aiming to precisely estimate future wind power output.

Accurately predicting wind power production is essential for reliable grid operations and integration of wind energy. As Ernst et al. (2007) explain, electricity grids were historically designed around dispatchable fossil fuel and nuclear plants. In contrast, wind power availability depends on variable weather conditions. To balance electricity supply and demand at all times as more wind comes online, power system operators need to schedule other generators based on forecasted wind. Ernst et al. note that forecast accuracy directly impacts the amount of balancing reserves needed and the resulting costs of integrating wind power. Improved predictions reduce emissions and costs from reserve plants and increase the value of wind. The authors present examples of wind forecasting systems used by grid operators to optimally dispatch their fleets. This illustrates how rigorous wind prediction has become a key enabler for the growing penetration of wind power in electricity systems worldwide.

In summary, the two papers highlight the critical role that high-quality wind forecasting, using both speed and direction data, plays in the efficient and reliable operation of power grids with increasing shares of wind generation. By improving the accuracy of predicted wind power availability, system operators can more cost-effectively balance supply and demand, ultimately facilitating the continued growth of clean wind energy.

Data Mining Techniques in Wind Energy

Data mining techniques are emerging as powerful tools for analyzing the large volumes of data generated by wind turbines and farms. Astolfi et al. (2015) demonstrate several innovative ways to apply data mining to performance analysis of onshore wind farms. They process data from the Supervisory Control And Data Acquisition (SCADA) systems that record detailed operational information.

One application is formulating indexes to quantify turbine malfunctioning based on operating state dynamics. These non-dimensional metrics can universally assess operational quality across different turbine types. The authors also revisit power curve analysis, plotting power output against wind speed, to characterize individual turbine performance.

Further, they redefine the polar efficiency plot for onshore farms, mapping how efficiency varies with wind direction, to highlight wake losses and terrain effects. Finally, they introduce Stationarity and Misalignment Indexes to relate turbine nacelle behavior to power degradation from wakes. Together, these data-driven techniques enable detailed performance assessment to optimize operations.

In summary, Astolfi et al. (2015) showcase the power of creatively mining the vast SCADA datasets to extract key insights into the complex behavior of wind farms. Their work illustrates how data mining is becoming indispensable for wind farm monitoring, evaluation, and enhancement as the penetration of wind energy grows.

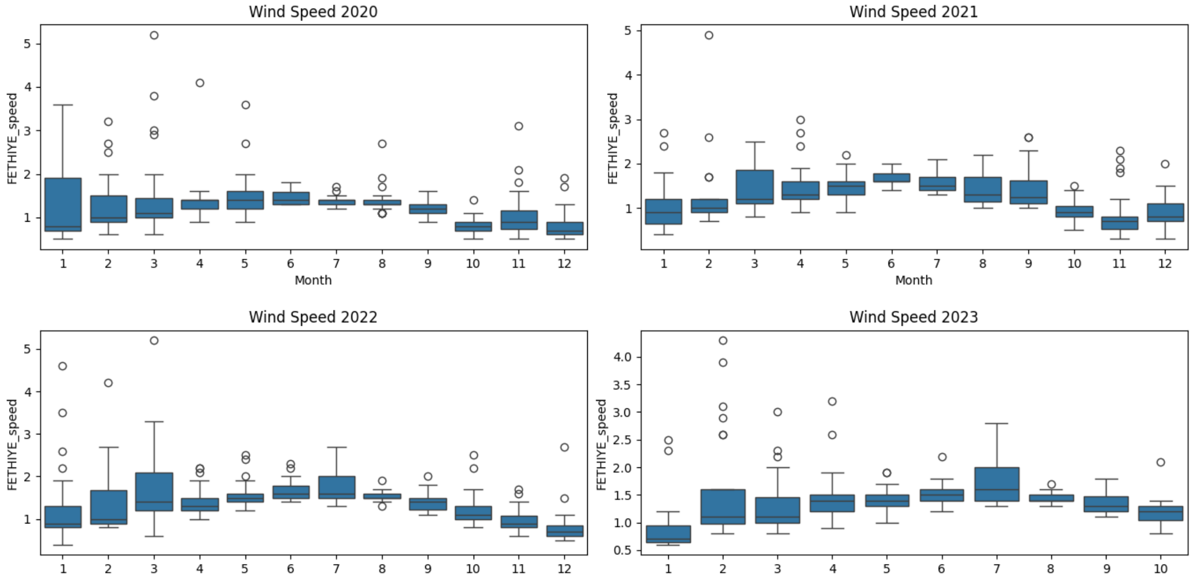
---

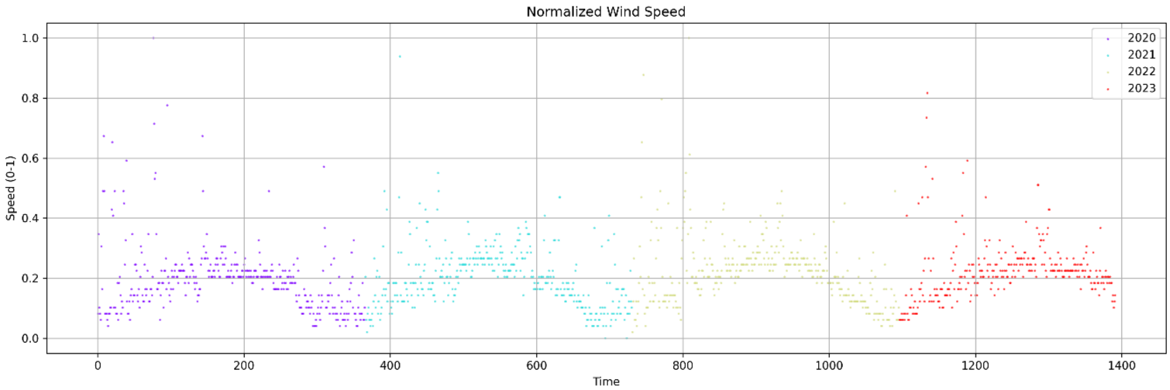
3. Methodology

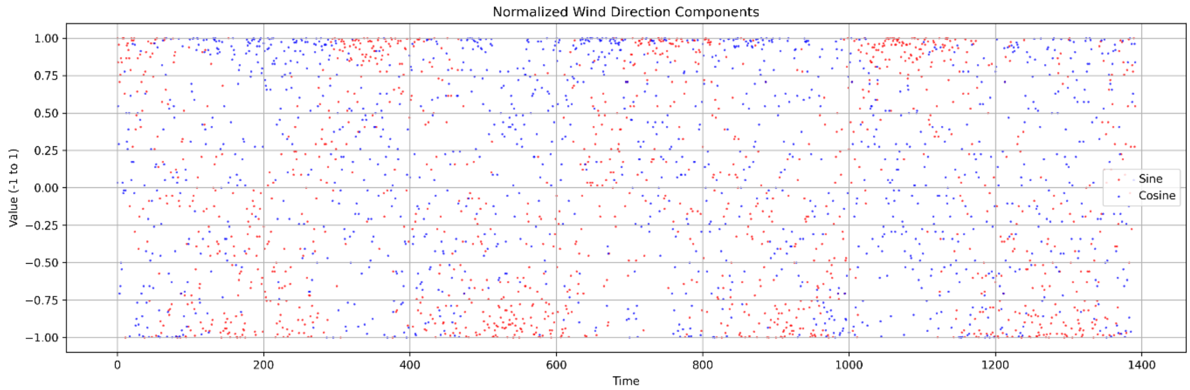
Data Preprocessing

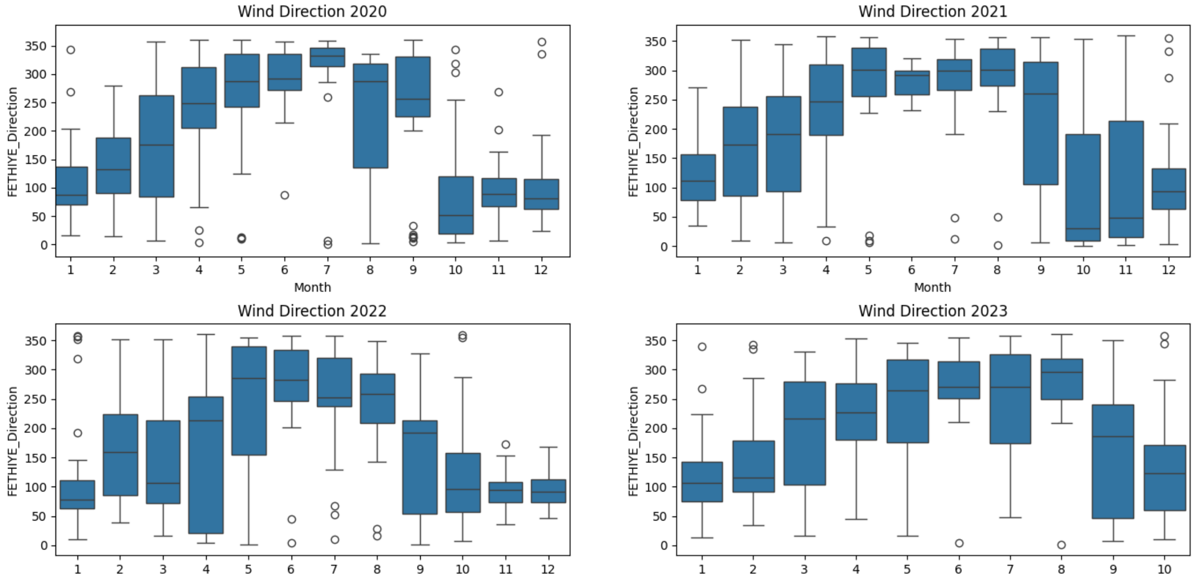
Data Cleaning

Our analysis of the FETHIYE wind dataset, which spans from 2020 to 2023 and contains 1,392 records, revealed remarkably clean data. We conducted thorough checks for missing values and inconsistencies, finding no instances of either. The data validation process confirmed that all wind speed and direction values fell within logical and possible ranges for the region.


Wind Speed 2020
Wind Speed 2021
Wind Speed 2022
Wind Speed 2023
FETHIYE_speed
Month


Normalized Wind Speed
Speed (0-1)
Time
2020
2021
2022
2023

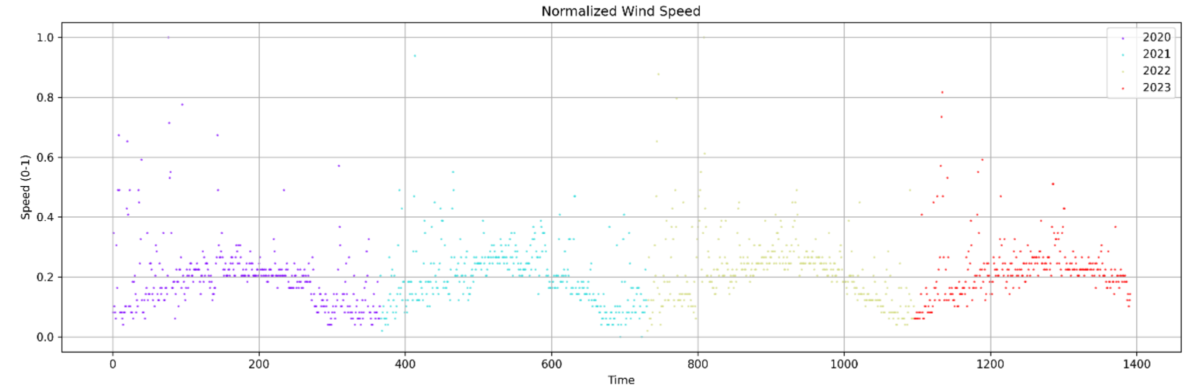

Normalized Wind Direction Components
Value (-1 to 1)
Time
Sine
Cosine

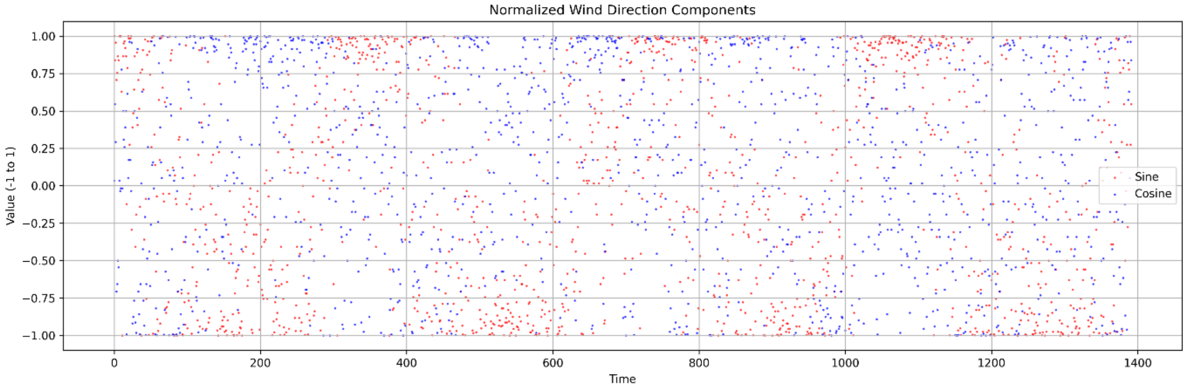


While our outlier analysis identified some extreme values in both wind speed and direction measurements, we made the strategic decision to retain these data points. This decision was driven by the nature of wind energy forecasting, where extreme values carry particular significance for understanding maximum potential energy production and planning for extreme weather conditions. We documented this analysis using dedicated Python scripts (outlier_speed.py and outlier_direction.py), which performed monthly IQR-based outlier detection and logged the findings for future reference.

Normalization

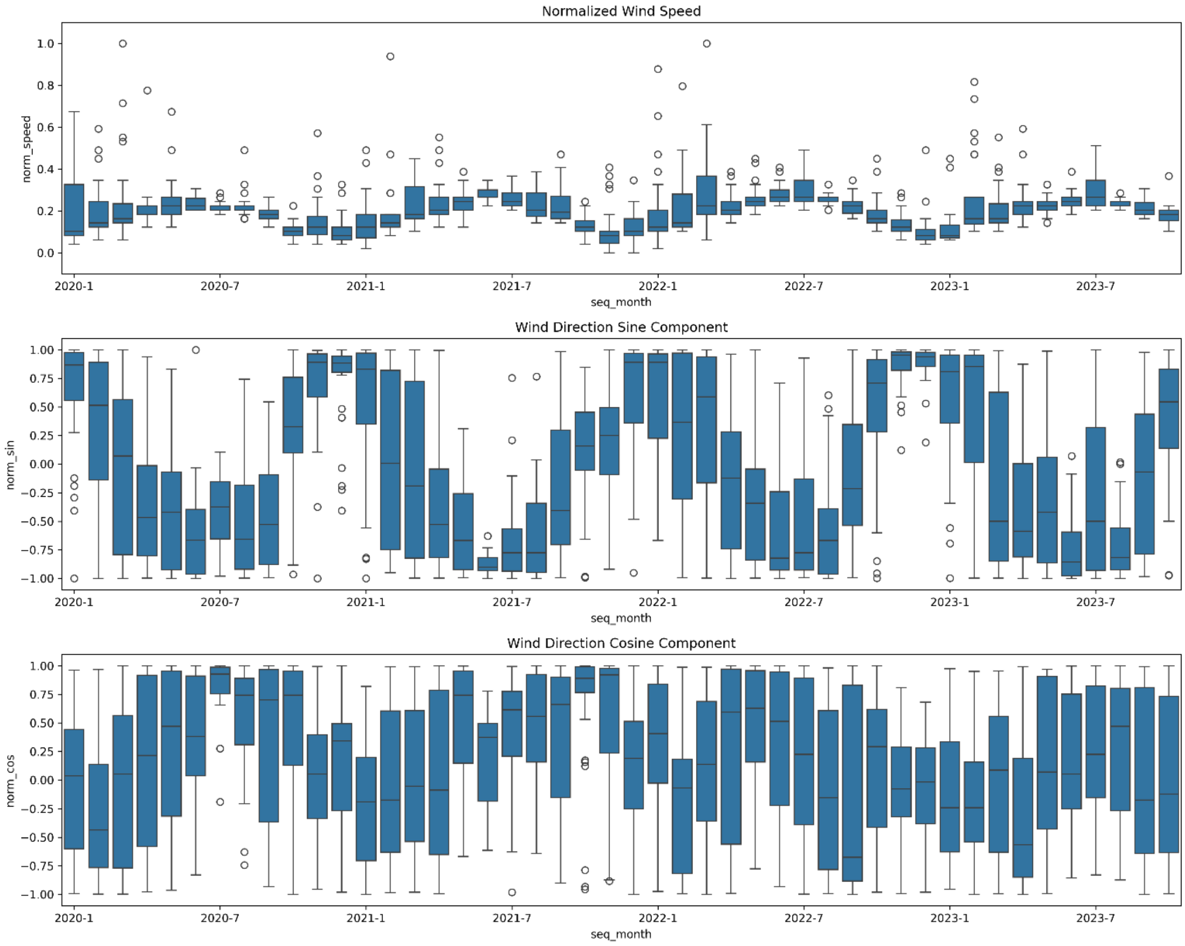
We implemented a comprehensive normalization strategy to standardize our measurements while preserving their analytical value. For wind speed, we applied min-max normalization to scale all values to the range [0,1], using the formula: (value - min) / (max - min). This approach maintains the relative relationships between speed measurements while making them more comparable across the dataset.





Wind direction normalization required a more sophisticated approach due to its circular nature (where 360° equals 0°). We addressed this by implementing two complementary normalization methods:

1. Converting the directions to vector components using sine and cosine transformations
2. Standardizing all measurements to the 0-360° range



Feature Engineering

We developed several engineered features to enhance the dataset's analytical potential:

Cyclical Month Components

We created month_sin and month_cos features using sinusoidal transformations of the month values. These features capture the cyclical nature of annual patterns, which is crucial for time series forecasting models, particularly ARIMA and SARIMA.

Rolling Statistics

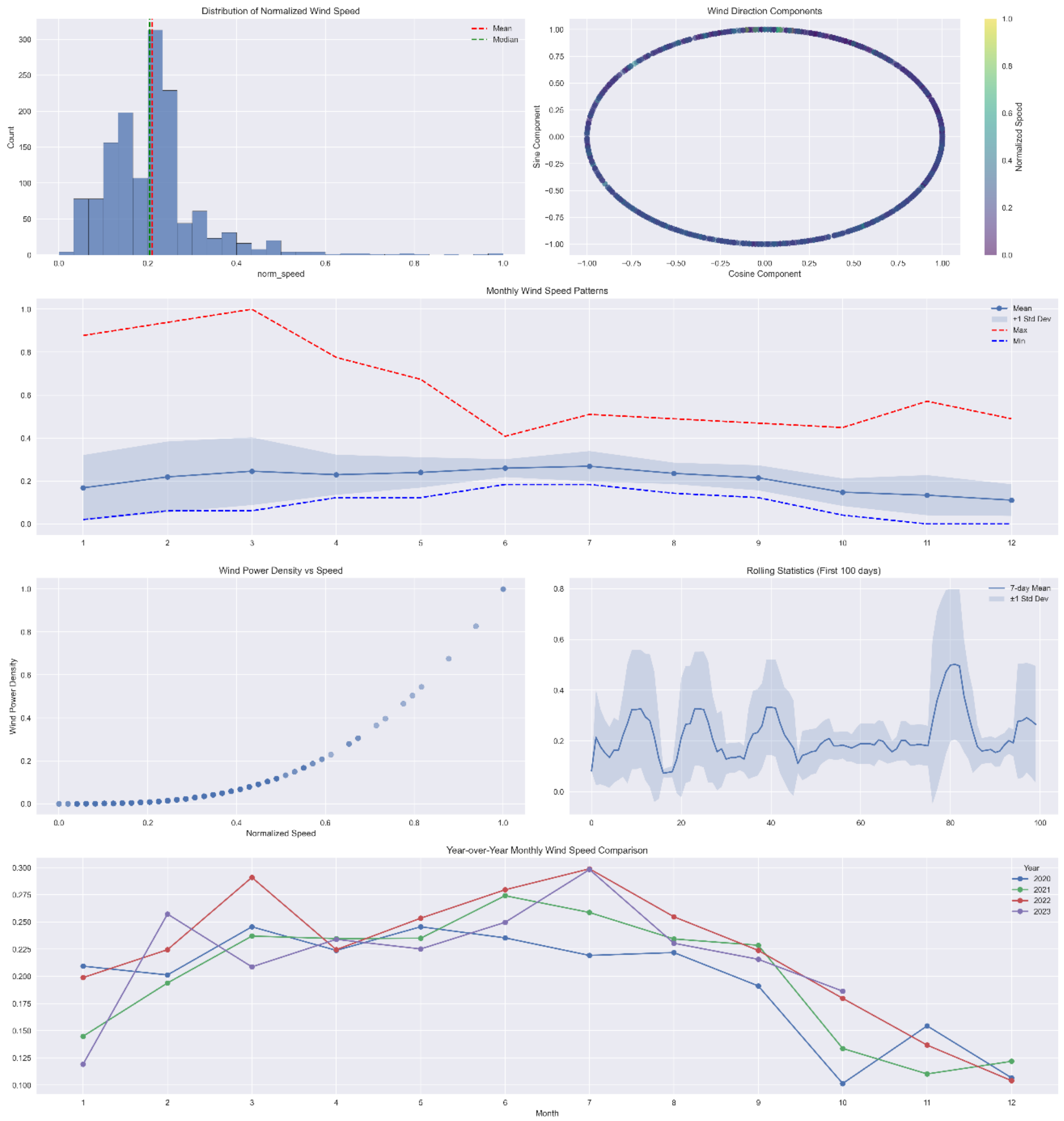
We implemented 7-day rolling windows to calculate:

- Speed moving averages (speed_7day_mean)
- Standard deviations (speed_7day_std)

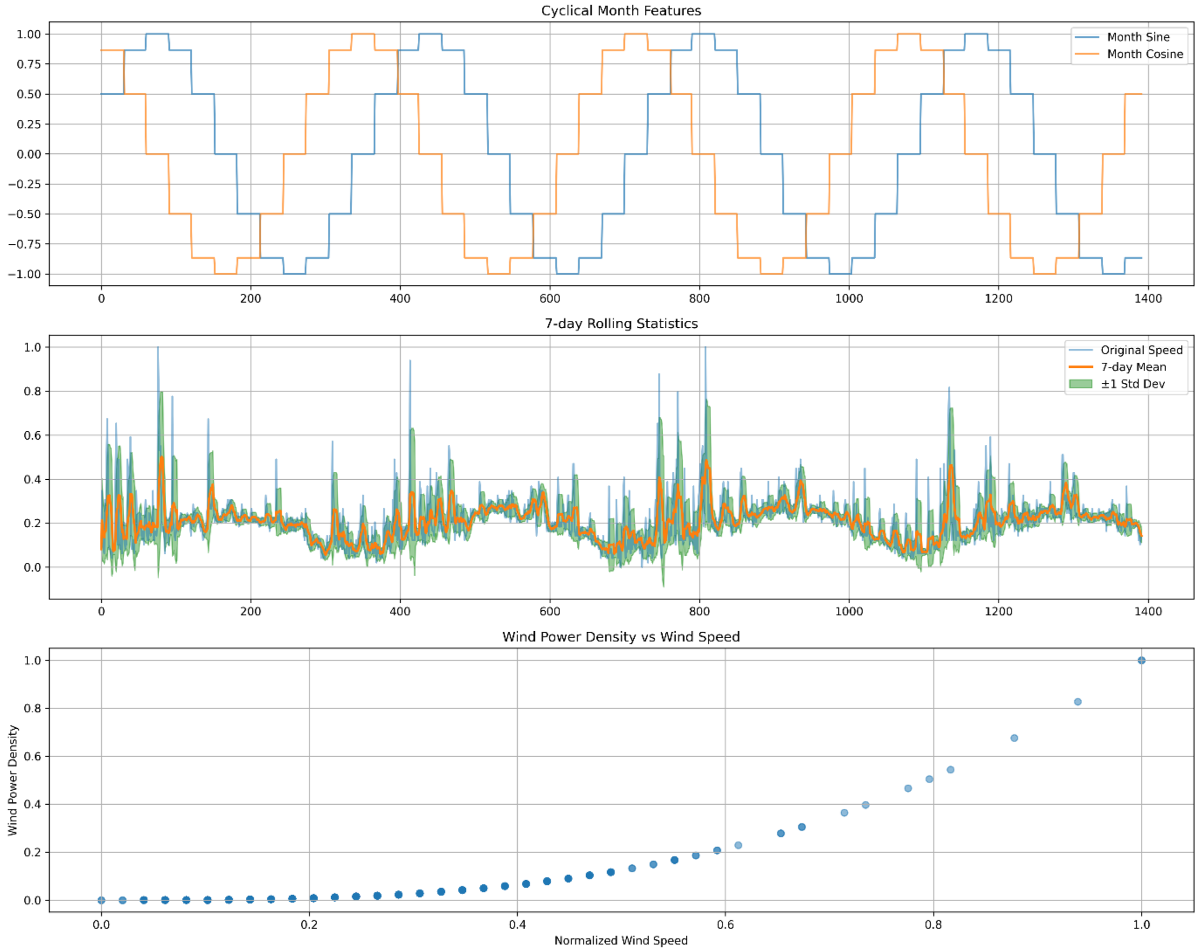
These metrics provide insights into short-term wind patterns and stability, which are essential for energy production classification and near-future forecasting.

Wind Power Density

We introduced a wind_power_density feature calculated as the cube of normalized wind speed. This feature directly relates to potential energy production and helps identify high-value periods for energy generation.



Statistic	Wind Speed	Wind Power Density
Count	1392.0	1392.0
Mean	0.20993138634764247	0.019698479931977618
Median	0.2040816326530612	0.008499859752314
Std Dev	0.1108718856791104	0.060678314132532475
Skewness	1.9907884091162587	10.582948138292851
Kurtosis	6.532231623472327	140.05573780460796
Min	0.0	0.0
25%	0.1428571428571428	0.0029154518950437
75%	0.2448979591836734	0.0146877576519987
Max	1.0	1.0



Implementation Details

All these preprocessing steps were implemented through a series of Python scripts, with normalize.py handling the core normalization process and Feature Engineering Script.py managing the feature engineering aspects. The results were visualized using various plotting scripts to validate the transformations and provide insights into the data distribution across different time scales.

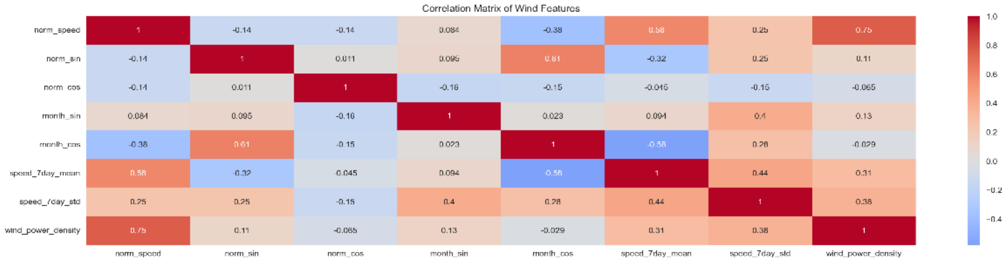
The preprocessed data has been saved in structured CSV formats, making it ready for subsequent modeling stages while maintaining full transparency and reproducibility of the preprocessing pipeline.

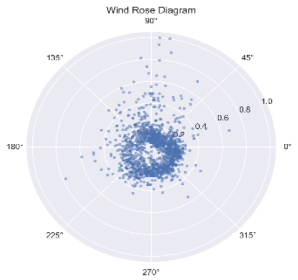
- The visualize.py script created box plots of the normalized wind speed, sine component of direction, and cosine component of direction, grouped by sequential months across the years. This allows for visualizing trends and distributions over time.
- The normalize.py script plotted the normalized wind speed over time, color-coded by year, as well as the sine and cosine components of the normalized direction. This provides an overview of the data after normalization.

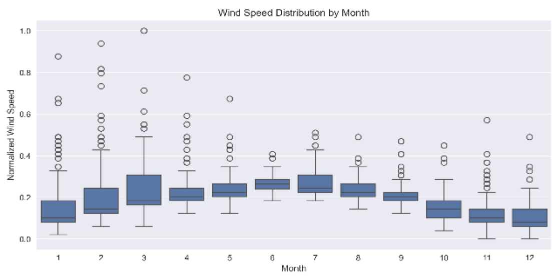
- The boxplot_wind_speed.py and boxplot_wind_direction.py scripts created box plots of the raw wind speed and direction data respectively, broken down by year and month. These visualizations help identify any seasonal patterns or outliers in the original data.

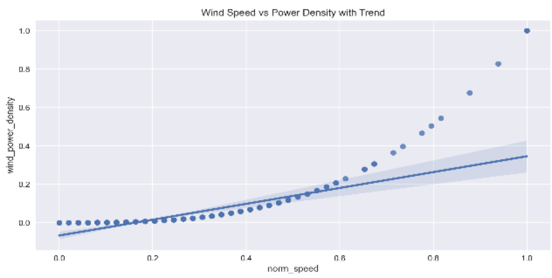
Descriptive Statistics:

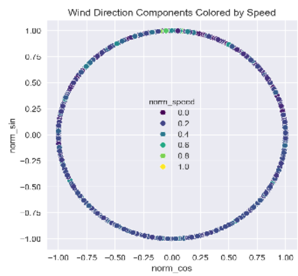
Our analysis of the wind data followed a structured approach combining descriptive statistics, correlation analysis, and comprehensive visualization techniques. We implemented this analysis through two Python scripts that processed data from the feature-engineered data.







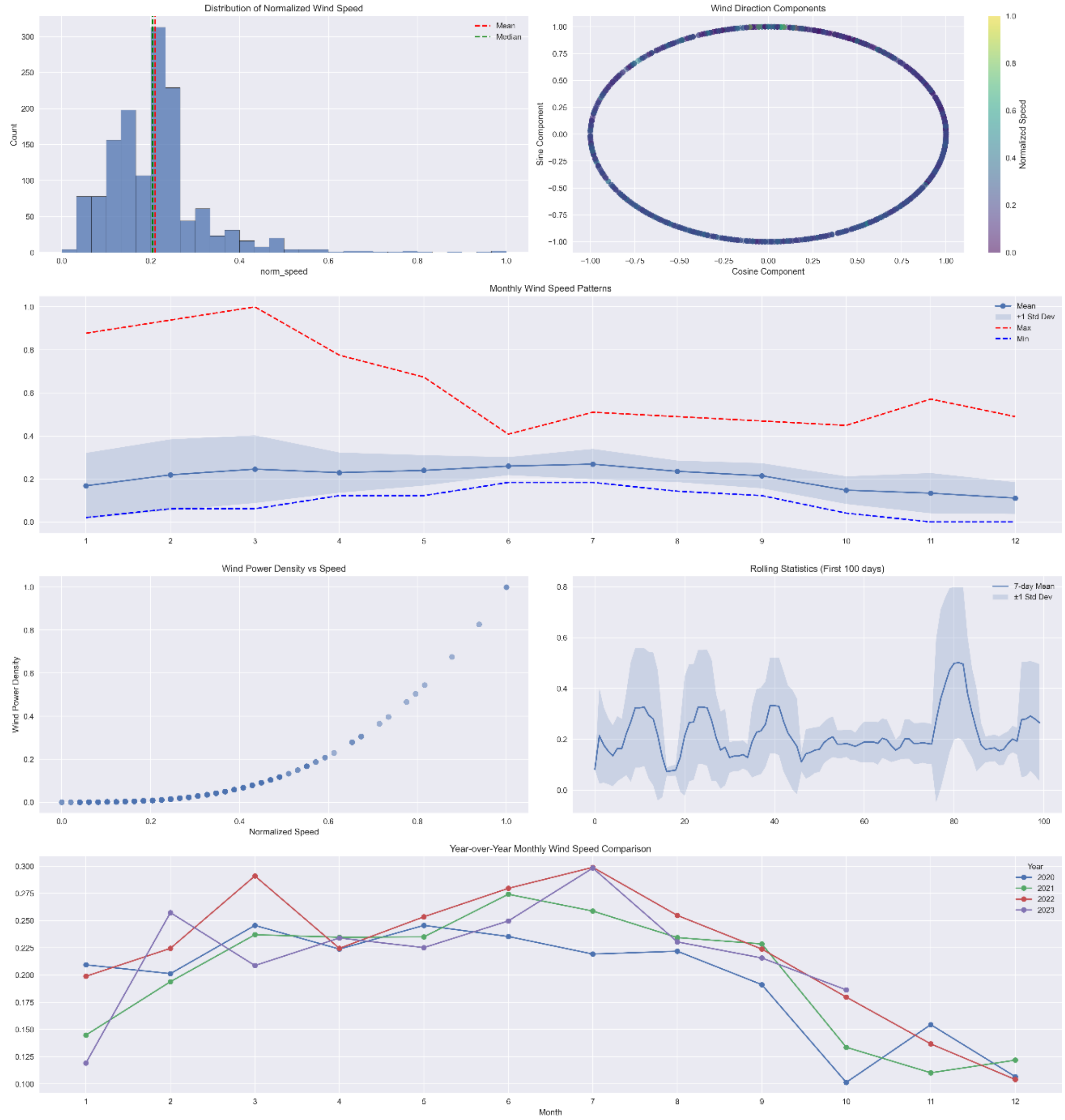




Variables	Pearson r	Pearson p-value	Spearman ρ	Spearman p-value
norm_speed vs speed_7day_mean	0.583	1.445e-127	0.658	3.813e-172
norm_speed vs speed_7day_std	0.253	7.581e-22	0.039	1.425e-01
norm_speed vs wind_power_density	0.754	2.536e-256	1.000	0.000e+00
norm_sin vs norm_speed	-0.141	1.198e-07	-0.332	3.607e-37
norm_sin vs speed_7day_mean	-0.321	8.349e-35	-0.371	1.209e-46
norm_sin vs speed_7day_std	0.252	1.254e-21	0.246	1.268e-20
norm_sin vs wind_power_density	0.110	4.032e-05	-0.332	3.607e-37
norm_cos vs norm_speed	-0.141	1.392e-07	-0.153	6.917e-09
norm_cos vs norm_sin	0.011	6.796e-01	0.030	2.661e-01
norm_cos vs speed_7day_mean	-0.045	9.259e-02	-0.032	2.327e-01
norm_cos vs speed_7day_std	-0.146	4.685e-08	-0.177	3.235e-11
norm_cos vs wind_power_density	-0.065	1.601e-02	-0.153	6.917e-09
month_sin vs norm_speed	0.084	1.707e-03	0.059	2.869e-02
month_sin vs norm_sin	0.095	3.998e-04	0.034	2.098e-01
month_sin vs norm_cos	-0.162	1.258e-09	-0.146	4.621e-08
month_sin vs speed_7day_mean	0.094	4.522e-04	0.108	5.188e-05
month_sin vs speed_7day_std	0.398	5.297e-54	0.375	1.014e-48
month_sin vs wind_power_density	0.130	1.073e-06	0.059	2.869e-02
month_cos vs norm_speed	-0.381	1.971e-49	-0.589	6.258e-131
month_cos vs norm_sin	0.615	1.617e-145	0.600	4.167e-137
month_cos vs norm_cos	-0.154	7.359e-09	-0.167	3.612e-10
month_cos vs month_sin	0.023	3.949e-01	0.032	2.370e-01
month_cos vs speed_7day_mean	-0.581	1.500e-126	-0.653	7.530e-170
month_cos vs speed_7day_std	0.282	7.470e-27	0.316	1.322e-33
month_cos vs wind_power_density	-0.029	2.832e-01	-0.589	6.258e-131
speed_7day_mean vs speed_7day_std	0.441	1.931e-67	0.243	3.316e-20
speed_7day_mean vs wind_power_density	0.307	8.143e-32	0.658	3.813e-172
speed_7day_std vs wind_power_density	0.375	7.619e-46	0.039	1.425e-01

Exploratory Data Analysis (EDA)

The first phase of our analysis focused on understanding the fundamental characteristics of the wind data. We examined the distribution, central tendency, and variability of both wind speed and direction. Our analysis revealed that the normalized wind speed follows a right-skewed distribution with a mean of 0.21 and a median of 0.20. The data spans the complete range from 0 to 1, with the interquartile range between 0.14 and 0.24,



Statistic	Wind Speed	Wind Power Density
Count	1392.0	1392.0
Mean	0.20993138634764247	0.019698479931977618
Median	0.2040816326530612	0.008499859752314
Std Dev	0.1108718856791104	0.060678314132532475
Skewness	1.9907884091162587	10.582948138292851
Kurtosis	6.532231623472327	140.05573760460796
Min	0.0	0.0
25%	0.1428571428571428	0.0029154518950437
75%	0.2448979591836734	0.0146877576519987
Max	1.0	1.0

indicating moderate wind conditions as the norm. For wind direction, we analyzed both sine and cosine components, finding mean values of -0.046 and 0.163 respectively, suggesting slightly predominant directional tendencies.

Correlation Analysis

We conducted a comprehensive correlation analysis to understand the relationships between various wind characteristics. The analysis employed both Pearson and Spearman correlation coefficients, accompanied by their respective p-values to ensure statistical significance. The most significant finding was the strong correlation (0.754) between normalized wind speed and wind power density. We also discovered meaningful correlations between monthly patterns and wind direction (0.615), as well as between current wind speeds and their 7-day moving averages (0.583), indicating temporal consistency in wind patterns.

Visualization Techniques

Our visualization strategy employed multiple complementary approaches to represent the wind data effectively. We created a correlation matrix heatmap to visualize relationships between all wind features, and developed a wind rose diagram to show the distribution of wind speeds and directions. The analysis included monthly box plots demonstrating wind speed distributions, scatter plots with trend lines for wind speed versus power density, and joint plots showing directional components colored by speed. Additionally, we generated histograms to visualize the distribution of normalized wind speed and created temporal plots to show patterns over time. All these visualizations were combined into comprehensive figures saved as ‘EDA.png’ and ‘correlation_analysis.png’, providing a complete visual representation of our findings.

Time Series Analysis:

Trend Identification

The Trend Identification sorts data into its fundamental components: trend, seasonality, and residuals. We utilized the classical additive decomposition method, which assumes that these components combine additively to form the observed data.

The analysis pipeline included:

1. Creation of a proper datetime index for temporal analysis
2. Implementation of multiple moving averages (7-day, 30-day, and 90-day) for trend visualization
3. Classical additive decomposition with a 365-day period to capture annual cycles
4. Calculation of trend and seasonal strength metrics
5. Analysis of monthly seasonal effects
6. Detection of significant trend changes

Key Findings

Component Strengths

- **Trend Strength**: 0.282 (28.2%)
  - This relatively low value indicates that the long-term trend component explains about 28% of the variation in the data
  - Suggests that wind speeds in FETHIYE don't follow strong long-term trends
- **Seasonal Strength**: 0.492 (49.2%)
  - The seasonal component accounts for nearly half of the data's variation
  - Indicates strong seasonal patterns in wind speed behavior
  - Suggests that seasonal factors should be given significant weight in forecasting models

Seasonal Patterns

Monthly seasonal effects revealed distinct patterns:

- **Peak Wind Season** (June-July):
  - Highest positive effects: June (0.0669) and July (0.0554)
  - Indicates consistently higher wind speeds during early summer
- **Low Wind Season** (October-December):
  - Strongest negative effects: December (-0.0919), November (-0.0693), October (-0.0617)
  - Shows significantly reduced wind speeds during winter months
- **Transition Periods**:
  - Spring (March-May): Moderate positive effects (0.0177 to 0.0397)
  - Late Summer/Early Fall (August-September): Declining positive effects (0.0333 to 0.0115)

Significant Trend Changes

Analysis of trend changes revealed several notable patterns:

1. **Temporal Distribution**:

- Most significant changes occurred during summer months (July-August)
- Secondary cluster of changes in September-October
- Fewer significant changes during winter months

2. **Change Magnitude**:

- Typical significant changes ranged from 0.0008 to 0.0023 in absolute terms
- Largest changes:
  - Positive: 0.0023 (September 18, 2021)
  - Negative: -0.0023 (August 19, 2021)

3. **Annual Patterns**:

- 2020: Concentrated changes in July-October
- 2021: More evenly distributed throughout the year
- 2022: Similar pattern to 2020, with focus on summer months

This decomposition analysis provides crucial insights for both operational planning and forecasting model development, suggesting that a combination of seasonal and trend-based approaches would be most effective for wind speed prediction in the FETHIYE region.

Forecasting Models:

1. Data Loading and Preparation

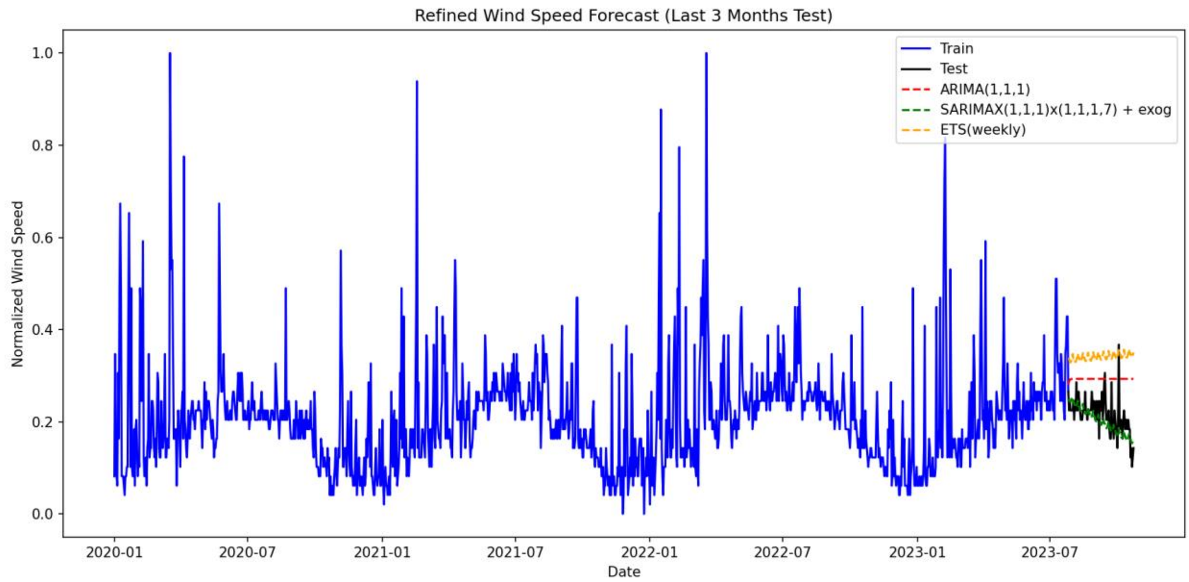
- Loaded normalized wind speed data from normalized_data.csv
- Converted to datetime index and focused on norm_speed column
- Ensured daily frequency and filled missing values by interpolation

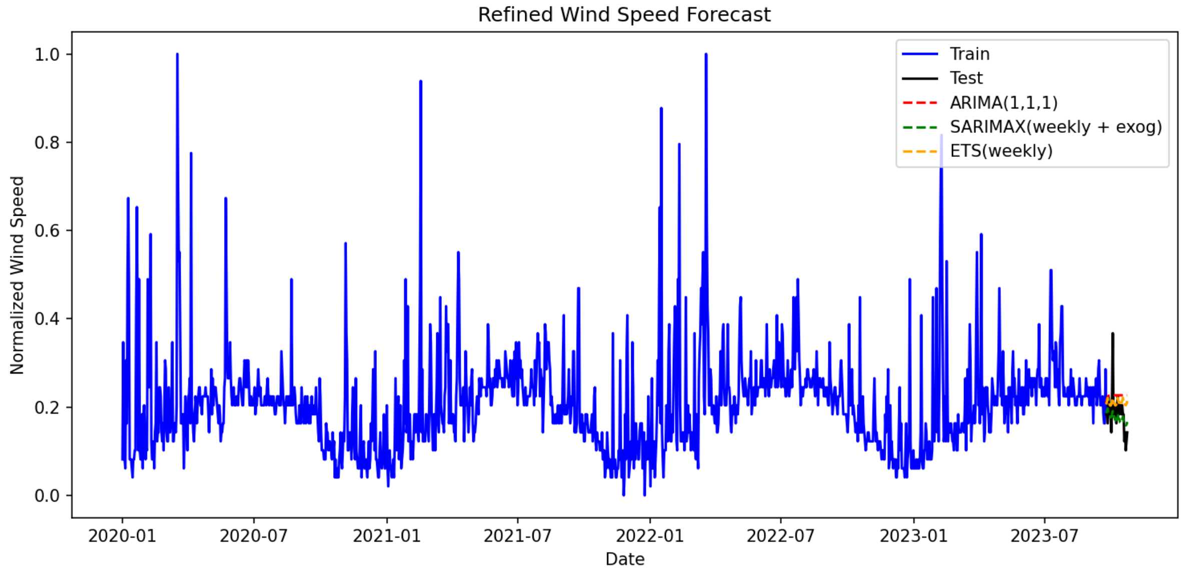
2. Feature Engineering for Seasonality

- Created day-of-year sine and cosine features to capture annual seasonality
  - Normalized for 365-day cycle
  - Stored in exog_df DataFrame for use in SARIMAX model

3. Train-Test Split

- Split data into train and test sets
  - Last 30 days used as test set in test.py
  - Last 3 months used as test set in test2.py


Refined Wind Speed Forecast (Last 3 Months Test)
Train
Test
ARIMA(1,1,1)
SARIMAX(1,1,1)x(1,1,1,7) + exog
ETS(weekly)
Normalized Wind Speed
Date
1.0
0.8
0.6
0.4
0.2
0.0
2020-01
2020-07
2021-01
2021-07
2022-01
2022-07
2023-01
2023-07


Refined Wind Speed Forecast
Train
Test
ARIMA(1,1,1)
SARIMAX(weekly + exog)
ETS(weekly)
Normalized Wind Speed
Date
1.0
0.8
0.6
0.4
0.2
0.0
2020-01
2020-07
2021-01
2021-07
2022-01
2022-07
2023-01
2023-07

4. Model Fitting

- Fit three models:
1. ARIMA(1,1,1) as baseline without seasonality
2. SARIMAX(1,1,1)x(1,1,1,7) with exogenous features (day-of-year sine/cosine) to capture weekly and annual seasonality
3. ETS (Holt-Winters) with weekly seasonality (seasonal_periods=7) for comparison

5. Model Evaluation

- Evaluated models using Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE)
- Results for last 30 days test set (test.py):
  - ARIMA(1,1,1):
    - RMSE: 0.0596
  - SARIMAX(1,1,1)x(1,1,1,7) + exog(doy_sin, doy_cos):
    - RMSE: 0.0427
    - MAE: 0.0295
  - ETS (weekly):
    - RMSE: 0.0492
    - MAE: 0.0348
- Results for last 3 months test set (test2.py):
  - ARIMA(1,1,1):
    - RMSE: 0.0879
    - MAE: 0.0801
  - SARIMAX(1,1,1)x(1,1,1,7) + exog(doy_sin, doy_cos):
    - RMSE: 0.0378
    - MAE: 0.0270
  - ETS (weekly):
    - RMSE: 0.1331
    - MAE: 0.1268

6. Results Visualization

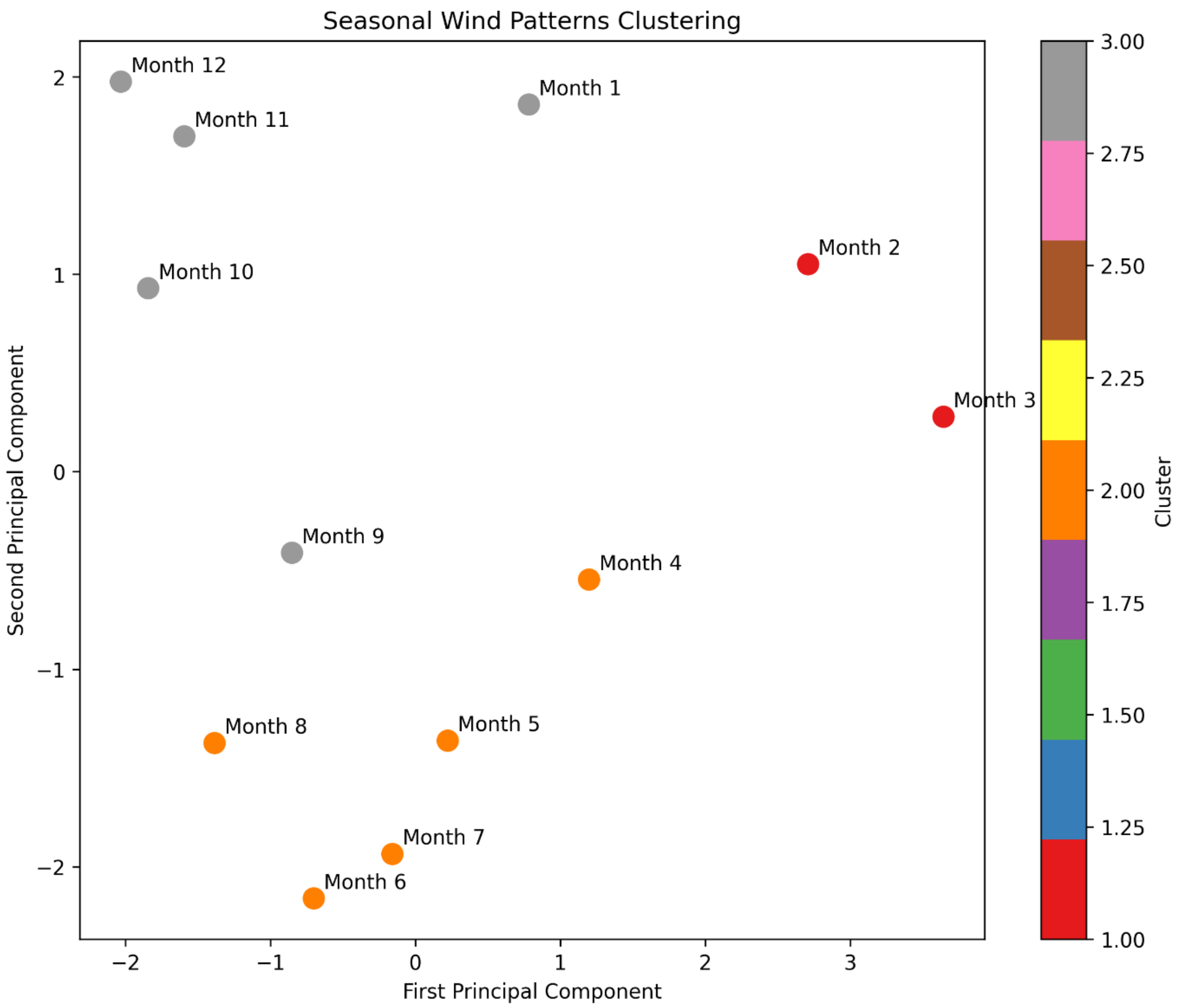
- Plotted train data, test data, and forecasts from each model
- Saved plots as refined_forecast.png and refined_forecast_3months_test.png

Key Findings

- SARIMAX model with both weekly and annual seasonality captured via exogenous features performed the best on both test sets

- ETS model with only weekly seasonality performed better than ARIMA on 30-day test but worse on 3-month test
- Incorporating both short-term (weekly) and long-term (annual) seasonality is crucial for accurate wind speed forecasting

Clustering and Classification:



Optimal Clustering

- Algorithm identified 3 distinct seasonal clusters
- Best silhouette score: 0.357 (indicating moderate cluster separation)

Cluster Characteristics

Cluster 1: Early Spring (February-March)

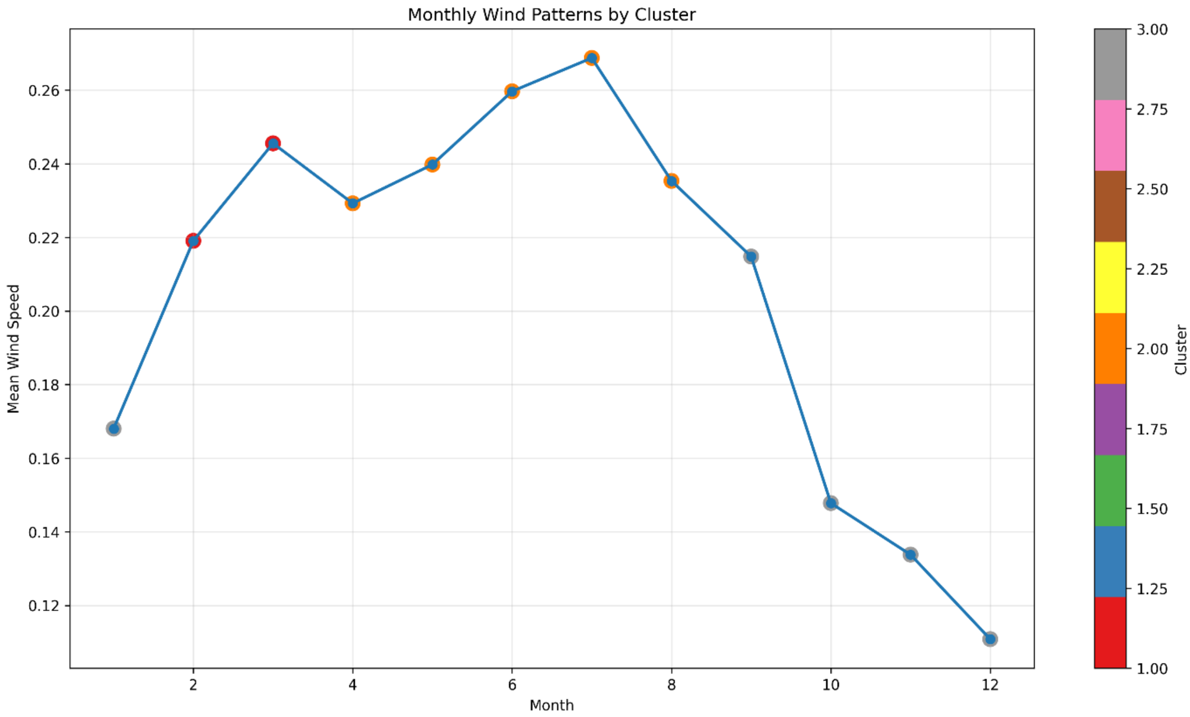
- Months: 2-3
- Moderate wind speeds (mean: 0.232)
- Highest wind power density (0.040)
- Highest variability (std: 0.161)
- Low directional stability (0.164)

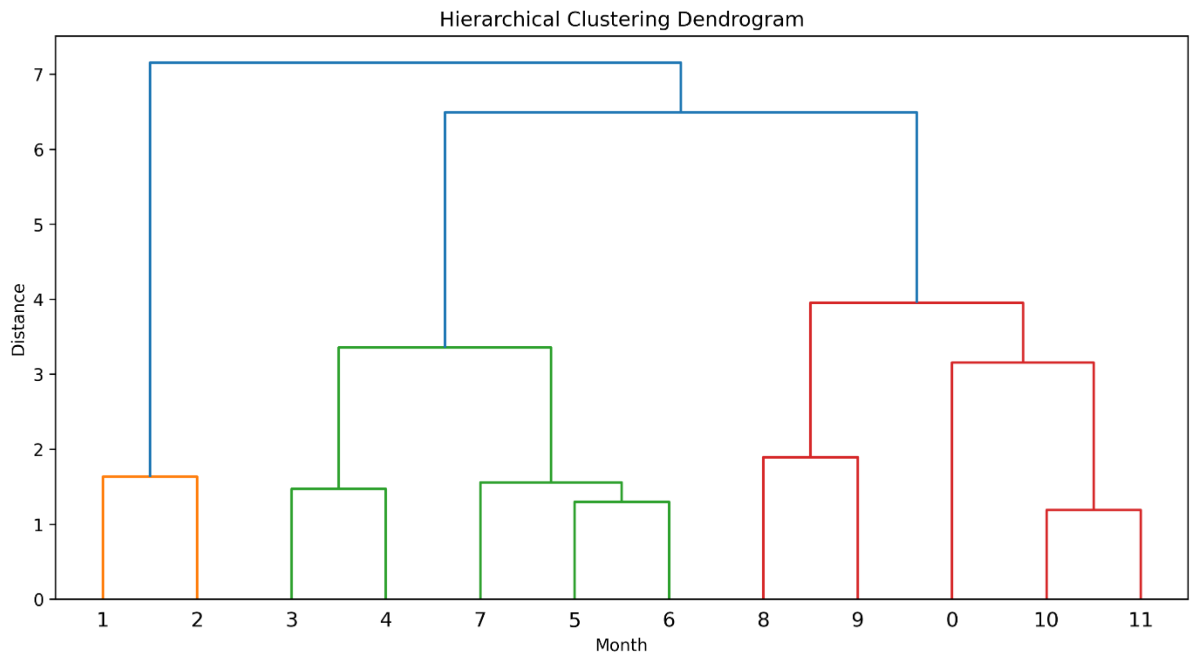
Cluster 2: Peak Season (April-August)

- Months: 4-8
- Highest mean wind speeds (0.247)
- Most consistent speeds (std: 0.065)
- Strong directional stability (0.574)
- Moderate power density (0.019)

Cluster 3: Off-Peak Season (September-January)

- Months: 1, 9-12
- Lowest wind speeds (0.155)
- Moderate variability (std: 0.088)
- Lowest power density (0.011)
- Good directional stability (0.524)





Key Insights

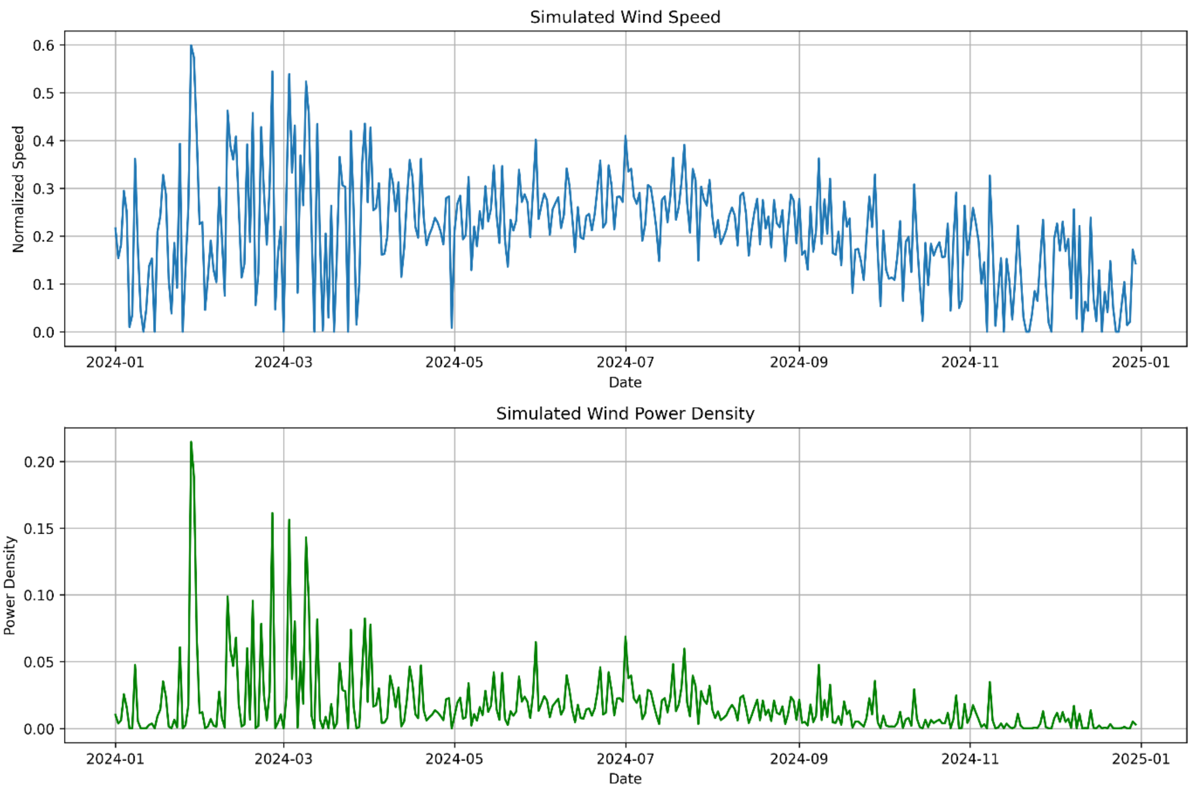
1. Clear seasonal patterns with three distinct periods
2. Power density peaks in early spring despite higher speeds in summer
3. Wind direction is most stable during peak and off-peak seasons
4. Transition months (Feb-Mar) show highest variability but best power generation potential

Implications for Wind Energy

- Optimal power generation window: February-March
- Most reliable operations: April-August
- Reduced output expected: September-January
- Planning maintenance during off-peak months (Cluster 3) recommended

Future Simulations:

Our Monte Carlo simulation framework generated synthetic wind data for FETHIYE region through 2024, combining historical patterns with stochastic variations. The simulation results revealed:



1. Wind Speed Distribution

- Mean normalized speed: 0.210
- Standard deviation: 0.112
- Range: 0.000 to 0.599
- Typical operating range (IQR): 0.148 to 0.277

2. Directional Components

- Minor southward bias (mean sin: -0.013)

- Slight eastward tendency (mean cos: 0.166)
- Wide directional variation (std: ~0.67)

3. Power Generation Potential

- Mean power density: 0.017
- Maximum density: 0.215
- High variability (std: 0.026)
- 75% of values below 0.021

The results suggest moderate but stable wind energy potential, with significant variability in both speed and direction throughout the year. Power density values indicate optimal generation periods occurring during peak wind speeds, though with notable intermittency.

---

4. Conclusion

Summary of Findings:
Our comprehensive analysis of wind data from the Fethiye region between 2020 and 2023 revealed several significant insights that can substantially impact wind energy operations and planning. The normalized wind speed data showed a right-skewed distribution with a mean of 0.21, indicating predominantly moderate wind conditions. Our time series decomposition identified a strong seasonal component accounting for 49.2% of wind speed variation, while the trend component explained 28.2% of the variation.

The SARIMAX model incorporating both weekly and annual seasonality emerged as the most accurate forecasting approach, achieving an RMSE of 0.0378 and MAE of 0.0270 over a three-month test period. This significantly outperformed both the baseline ARIMA model and the ETS approach, demonstrating the importance of capturing multiple seasonal patterns in wind speed prediction.

Our clustering analysis successfully identified three distinct seasonal patterns in the wind data, each with unique characteristics that have important implications for wind energy production. The clustering showed clear delineation between early spring, peak season (April-August), and off-peak season (September-January), with silhouette scores indicating moderate but meaningful cluster separation.

Implications for Wind Energy Management:
The identification of distinct seasonal clusters and accurate forecasting models enables more efficient operational planning and resource allocation. The strong

performance of our SARIMAX model suggests that wind farm operators in Fethiye can reliably forecast wind conditions up to three months ahead, allowing for better maintenance scheduling and grid integration planning. The clustering analysis revealed that February-March offers the highest power generation potential despite not having the highest mean wind speeds, suggesting that these months should be prioritized for maximum energy capture.

Recommendations:
Based on our analysis, we recommend the following strategies for wind farm operators in the Fethiye region:

1. Schedule major maintenance activities during the identified off-peak season (September-January) when wind speeds are consistently lower and more predictable.

2. Implement the SARIMAX forecasting model with both weekly and annual seasonality components for operational planning, as it demonstrated superior predictive accuracy.

3. Optimize turbine operations for the distinct characteristics of each seasonal cluster, particularly during the high-variability period of February-March when power generation potential is highest.

4. Develop contingency plans for the peak season (April-August) when wind conditions are most stable but energy demand may be higher due to seasonal factors.

5. Consider implementing real-time monitoring systems that integrate our predictive models for dynamic adjustment of turbine operations based on forecasted conditions.

Limitations:
While our analysis provides valuable insights, several limitations should be considered:

1. The dataset spans only four years (2020-2023), which may not capture longer-term climate patterns or extreme weather events that could affect wind conditions.

2. The analysis focuses solely on wind speed and direction, without incorporating other meteorological variables that might influence wind patterns.

3. While our forecasting models show strong performance, they may not fully capture extreme events or sudden weather changes that could impact wind energy production.

4. The clustering analysis, while revealing clear patterns, shows moderate cluster separation (silhouette score of 0.357), indicating some overlap between seasonal patterns that operators should consider in their planning.

Future research could address these limitations by incorporating additional meteorological variables, extending the temporal scope of the analysis, and developing more sophisticated models that better account for extreme weather events. Despite these constraints, our findings provide a robust foundation for improving wind energy operations in the Fethiye region through data-driven decision-making and strategic planning.

5. References

- **Academic Papers:**

1. Saidur, R., Rahim, N. A., Islam, M. R., & Solangi, K. H. (2011). Environmental impact of wind energy. *Renewable and sustainable energy reviews*, *15*(5), 2423-2430.
2. Ramírez, F. J., Honrubia-Escribano, A., Gómez-Lázaro, E., & Pham, D. T. (2018). The role of wind energy production in addressing the European renewable energy targets: The case of Spain. *Journal of Cleaner Production*, *196*, 1198-1212.
3. Yatiyana, E., Rajakaruna, S., & Ghosh, A. (2017, November). Wind speed and direction forecasting for wind power generation using ARIMA model. In *2017 Australasian universities power engineering conference (AUPEC)* (pp. 1-6). IEEE.
4. Ernst, B., Oakleaf, B., Ahlstrom, M. L., Lange, M., Moehrlen, C., Lange, B., … & Rohrig, K. (2007). Predicting the wind. *IEEE power and energy magazine*, *5*(6), 78-89.
5. Astolfi, D., Castellani, F., Garinei, A., & Terzi, L. (2015). Data mining techniques for performance analysis of onshore wind farms. *Applied Energy*, *148*, 220-233.

- **Tools and Libraries Documentation:**
- Python 3. *Primary Programming Language*.
- Pandas. *Data manipulation and analysis library*.
- NumPy. *Numerical computing library*.
- Matplotlib. *Data visualization library*.
- Seaborn. *Statistical data visualization library*.
- Scikit-learn. *Machine Learning library*.
  - Used for clustering, preprocessing, and metrics calculations
- Statsmodels. *Statistical modeling library*.
  - Used for time series analysis (ARIMA, SARIMAX)
- SciPy. *Scientific computing library*.
  - Used for statistical tests and hierarchical clustering
- Warnings. *Python built-in library*.
  - Used for warning management

- Datetime. *Python built-in library*.
  - Used for date and time handling
- Math. *Python built-in library*.
  - Used for mathematical operations
- ExponentialSmoothing. *Time series smoothing library*.
  - Used for ETS (Holt-Winters) modeling
- Principal Component Analysis (PCA). *Dimensionality reduction technique*.
  - Used for visualization of clustering results