

# Cluster Feature Analysis Report for 400 turbines based on Nearest Neighbors algorithm selected in Denmark

This report interprets the characteristics of each cluster from the federated clustering results. Clusters are evaluated based on feature statistics such as power output, variation, ramp behavior, and downtime.

## Cluster Summary:

- **Cluster 0** represents a high-performing group with elevated power and volatility — a valuable segment for performance optimization. (51)
- **Cluster 1** contains near-fault turbines with almost full shutdowns and extreme coefficient of variation — classic outliers. (2)
- **Cluster 2** shows high ramping behavior, potentially indicating turbines in unstable environments. (4)
- **Cluster 3** is the dominant and stable group. (251)
- **Cluster 4** indicate mid-level risk, moderate shutdowns and lower-than-average output (76)
- **Cluster 5** appears promising but volatile, possibly reflecting new deployments or special configurations. (6)
- **Cluster 6** mildly unstable group, frequent ramp-up and occasional downtime (10)

Conclusion: Cluster 0,2,3,4,5,6 will come into prediction stage. (Cluster 1 is outlier)

## Evaluation Metrics for Wind Turbine Clustering

1. **mean\_power\_scaled\_mean**: Overall Average Power (Standardized)

- $> 0 \rightarrow$  Higher-than-average power output, good performance
- $< 0 \rightarrow$  Lower-than-average power, possibly due to environmental or performance issues

2. **std\_power\_scaled\_mean**: Power Output Fluctuation (Standardized)

- $> 0 \rightarrow$  Large variations in power, possibly caused by wind changes or control system behavior
- $< 0 \rightarrow$  Smaller fluctuations, more stable output and easier to forecast

3. **cv\_scaled\_mean**: Coefficient of Variation (Relative Instability)

- $> 0 \rightarrow$  High variability per unit of power, potentially less reliable
- $< 0 \rightarrow$  Relatively stable output and more consistent operation

4. **zero\_ratio\_mean**: Proportion of Downtime

- **Close to 0** → High availability, turbines run continuously
- **Close to 1** → Long shutdown periods or faults, affecting generation

5. **ramp\_mean\_scaled\_mean**: Average Power Change Trend

- $> 0 \rightarrow$  Ramp-up trend, possibly due to increasing wind speed or gradual engagement
- $< 0 \rightarrow$  Ramp-down trend, possibly due to decommissioning, maintenance, or performance decline

6. **ramp\_std\_scaled\_mean**: Volatility of Power Changes

- $> 0 \rightarrow$  Sharp or frequent power swings, harder to predict
- $< 0 \rightarrow$  Smoother changes, more stable and easier to model

7. Cluster Standard Deviation (After Scaling)

- $std > 1 \rightarrow$  The group's variability is greater than average
- $std < 1 \rightarrow$  The group's variability is lower than average

## Cluster Statistics

### Cluster Mean Statistics

Cluster	Count	mean_power_scaled_mean	std_power_scaled_mean	cv_scaled_mean	zero_ratio_mean	ramp_mean_scaled_mean	ramp_st
0	51	2.3383	2.3408	-0.1774	0.0672		-1.1569
1	2	-0.7299	-0.8549	13.2276	0.9987		0.2984
2	4	1.7114	1.8085	-0.1317	0.1102		4.7725
3	251	-0.3049	-0.2922	-0.0830	0.1465		0.0177
4	76	-0.6317	-0.7048	0.0607	0.3157		0.2485
5	6	0.5266	0.8178	-0.0903	0.0900		0.7233
6	10	-0.3271	-0.2911	-0.0114	0.2397		1.1653

### Cluster Standard Deviation Statistics

Cluster	mean_power_scaled_std	std_power_scaled_std	cv_scaled_std	zero_ratio_std	ramp_mean_scaled_std	ramp_std_scaled_std
0	0.9186	0.8308	0.0209	0.0207	1.9099	0.8828
1	0.0000	0.0018	6.5609	0.0011	0.0000	0.0083
2	1.5020	1.0891	0.0810	0.0611	1.2589	0.9328
3	0.0971	0.1045	0.0484	0.0820	0.3357	0.0973
4	0.0736	0.1091	0.1783	0.1942	0.1157	0.0919
5	0.0610	0.1199	0.0168	0.0183	0.5113	0.1440
6	0.1615	0.1489	0.1734	0.2085	0.3848	0.1384

## Cluster Details

### Cluster 0 — High Power Group

- **Count:** 51 turbines
- **Feature Highlights:**
  - `mean_power_scaled`: 2.34 — significantly high output
  - `std_power_scaled`: 2.34 — high output variation
  - `ramp_std_scaled`: 2.35 — strong ramp fluctuations
  - `zero_ratio`: 0.067 — very low downtime
  - `ramp_mean_scaled_std`: 1.91
- **Interpretation:** Represents turbines with excellent productivity and consistent operation, though possibly facing operational stress due to ramp volatility. Highly diverse cluster, turbines behave quite differently.

### Cluster 1 — Abnormal Group (Shutdown/Faulty)

- **Count:** 2 turbines
- **Feature Highlights:**
  - `cv_scaled`: 13.23 — extremely unstable
  - `zero_ratio`: 0.9987 — almost fully offline
- **Interpretation:** Non-operational or severely malfunctioning turbines.
- **Action:** Exclude from power forecasting; valuable for failure detection only.

### #Cluster 2 — Ramp-Dominated Group

- **Count:** 4 turbines
- **Feature Highlights:**
  - `ramp_mean_scaled`: 4.77 — extremely high ramp-up
  - `ramp_std_scaled`: 1.69 — strong variability
  - `mean_power_scaled_std`: 1.50
- **Interpretation:** These turbines may operate in turbulent or unstable environments. Highly diverse cluster, turbines behave quite differently.

### Cluster 3 — Baseline Stable Group

- **Count:** 251 turbines
- **Feature Highlights:**
  - All metrics are near average
  - `cv_scaled`: -0.08 → stable output
  - `zero_ratio`: 0.15 → moderate uptime
  - Most stds < 0.1, excellent internal consistency
- **Interpretation:** Majority group with reliable and balanced operation.

### Cluster 4 — Mid-Risk Group

- **Count:** 76 turbines
- **Feature Highlights:**
  - `mean_power_scaled`: -0.63 — lower-than-average output
  - `zero_ratio`: 0.32 — relatively frequent downtime
  - Most stds around 0.1, internally consistent — stable and predictable
- **Interpretation:** Turbines in lower wind zones or with more interruptions.

### Cluster 5 — Potentially Promising Group

- **Count:** 6 turbines
- **Feature Highlights:**

- Positive power output (`0.53`) with high `ramp` activity
  - Most stds around 0.1, internally consistent — stable and predictable
  - **Interpretation:** Possibly newly installed or unique operational strategies.
- 

#### Cluster 6 — Mildly Unstable Group

- **Count:** 10 turbines
- **Feature Highlights:**
  - `ramp_mean_scaled`: 1.17 — frequent ramp-up
  - `zero_ratio`: 0.24 — occasional downtime
  - std is slightly higher than average stds across features
- **Interpretation:** Semi-stable turbines with moderate risk.