

Jornadas de Automática

Detección de anomalías en turbinas eólicas: un análisis comparativo

Marcos, D.^{a,*} Gómez-Silva, M. J.^a, Santos, M.^b, López-González, C. I.^c

^a Dpto. de Arquitectura de Computadores y Automática, Facultad de Ciencias Físicas, Universidad Complutense de Madrid, Plaza de Ciencias, 1, Ciudad Universitaria, 28040, Madrid, España.

^b Instituto de Ingeniería del Conocimiento, Universidad Complutense de Madrid, Facultad de Informática, 28040, Madrid, España.

^c Dpto. De Ingeniería del Software e Inteligencia Artificial, Facultad de Informática, Universidad Complutense de Madrid.

To cite this article: Marcos, D., Gómez-Silva, M.J., Santos, M., López-González, C. I. 2024. Anomaly detection in wind turbines: a comparative analysis. *Jornadas de Automática*, 45. <https://doi.org/10.17979/ja-cea.2024.45.10813>

Resumen

El contexto de emergencia climática y el agotamiento de los combustibles fósiles ha llevado a organizaciones privadas y a entidades públicas a aumentar los esfuerzos en el desarrollo e investigación de fuentes de energía alternativas y limpias, entre las que destaca la energía eólica. En este artículo se propone la implementación de diferentes algoritmos de aprendizaje automático para tareas de detección de anomalías en datos de aerogeneradores. Se ha realizado una batería de experimentos con diferentes algoritmos de aprendizaje automático para realizar un análisis comparativo del rendimiento de estos métodos. El objetivo final es contribuir con herramientas computacionales a mejorar la operación y mantenimiento de las turbinas eólicas tanto terrestres como marinas.

Palabras clave: Aprendizaje automático, detección de anomalías, diagnóstico de fallos, turbinas eólicas

Anomaly detection in wind turbines: a comparative analysis

Abstract

The context of climate emergency and the depletion of fossil fuels has led private organizations and public entities to increase efforts in the development and research of alternative energy sources to fossil fuels. Within this framework, wind energy emerges as a great alternative. This article aims at implementing and analysing different machine learning algorithms for anomaly detection tasks in data belonging to wind turbines. To achieve this, a battery of experiments has been conducted with various machine learning algorithms, with the goal of performing a comparative analysis of the performance of these methods. The ultimate goal is to contribute with computational tools to improve the operation and maintenance of both onshore and offshore wind turbines.

Keywords: Machine Learning, anomaly detection, fault diagnosis, wind turbines

1. Introducción

Abordar el agotamiento de los combustibles fósiles y la acuciante crisis climática actual ha llevado tanto a empresas privadas como a instituciones públicas a aumentar el esfuerzo y los recursos destinados a la investigación y desarrollo de métodos alternativos de obtención de energía, con el objetivo de satisfacer la creciente demanda sin comprometer aún más la situación del planeta. En este marco, la energía eólica emerge como una de las grandes candidatas para suplir parte

de la demanda energética del presente y del futuro (Zhou, et al., 2023).

La potencia generada por una turbina eólica depende desde los materiales empleados para su fabricación hasta de las condiciones climáticas (Muñoz-Palomeque, et al. 2023). Debido en gran parte a la dificultad de generar un modelo matemático que describa con exactitud la potencia generada en función de las diferentes magnitudes físicas que intervienen en el proceso, las técnicas basadas en datos (data-driven) han emergido en los últimos años como una opción para la

creación de modelos de detección de anomalías en diversos ámbitos (Sacie et al., 2022; Morrison et al., 2022). La detección de comportamientos anómalos en la producción de potencia de una turbina permite corregir de forma temprana el mal funcionamiento del sistema y, por tanto, optimizar su rendimiento (Pandit et al., 2022; Zhang et al., 2024). Este artículo presenta la aplicación de diferentes técnicas de aprendizaje automático (ML, por sus siglas en inglés de Machine Learning) a la detección de anomalías en datos provenientes de turbinas eólicas, con el objetivo de realizar un estudio comparativo entre los diferentes métodos de aprendizaje automático aplicados a las turbinas eólicas.

2. Fundamentos

La detección de anomalías consiste en la identificación de una o varias muestras que se desvían del comportamiento esperado (Michelena et al., 2024). En series temporales de una única variable, las anomalías que pueden aparecer se clasifican en (Zamanzadeh Darban, et al., 2022):

- **Anomalías globales:** Muestran valores extremos comparados con el resto de las muestras de la serie temporal. Estableciendo un umbral de referencia, una muestra se considera anómala cuando $|x_t - \hat{x}_t| > threshold$, donde x_t es la muestra de la serie y \hat{x}_t es la predicción del modelo.
- **Anomalías contextuales:** Desviación de una muestra con respecto de sus vecinas. El umbral mediante el que se cataloga la muestra como una anomalía contextual se calcula teniendo en cuenta los vecinos.
- **Anomalías estacionales:** Es posible que la serie temporal presente cierta estacionalidad. Se considera que una serie temporal (o un fragmento de ella) constituye una anomalía estacional cuando $diss_s(S, \hat{S}) < threshold$, donde $diss_s$ representa la disimilitud entre la secuencia S y la estacionalidad esperada \hat{S} .

Atendiendo al tipo de técnica empleada, los métodos de detección de anomalías se pueden clasificar en:

- Métodos estadísticos tradicionales:** Su funcionamiento consiste en la comparación de la serie temporal a evaluar con un modelo estadístico que describe el comportamiento normal de la serie.
- Métodos de aprendizaje automático:** Emplean técnicas de ML para detectar anomalías en los datos. Algunos de los algoritmos más empleados son: Random Forest, Isolation Forest y K-nearest neighbours (KNN).

2.1 Aerogeneradores

Una turbina eólica es un dispositivo mecánico capaz de convertir la energía eólica en energía eléctrica. (Portillo, 2024). La potencia generada por una turbina eólica (P) viene dada por la ecuación (1) (Bilendo, et al., 2023), donde:

- ρ es la densidad del aire en kg/m^3 .
- A es el área barrida por las palas en m^2 .

- C_p es el coeficiente de potencia de la turbina.
- v es la velocidad del viento.

$$P(v) = \frac{1}{2} \rho A C_p v^3 \quad (1)$$

Considerando que la densidad del aire se mantiene constante, la potencia generada únicamente depende del cubo de la velocidad del viento y del coeficiente de potencia.

La curva de potencia, que es específica para cada turbina, se divide en cuatro regiones de operación atendiendo a la velocidad del viento (Figura 1):

- **Región 1 ($v < v_{cut\ in}$):** Si la velocidad del viento es menor que la velocidad de arranque ($v_{cut\ in}$), la turbina no será capaz de generar potencia alguna.
- **Región 2 ($v_{cut\ in} < v < v_{rated}$):** Cuando la velocidad del viento se encuentra comprendida entre los valores de $v_{cut\ in}$ y velocidad nominal (v_{rated}), la curva de potencia presenta un comportamiento cúbico dado por (1).
- **Región 3 ($v_{rated} < v < v_{cut\ off}$):** Una vez superado el valor de velocidad nominal, la turbina genera un valor de potencia constante para no dañar la estructura.
- **Región 4 ($v > v_{cut\ off}$):** Para velocidad del viento por encima de un cierto umbral $v_{cut\ off}$ generalmente se desactiva la turbina como medida de seguridad.

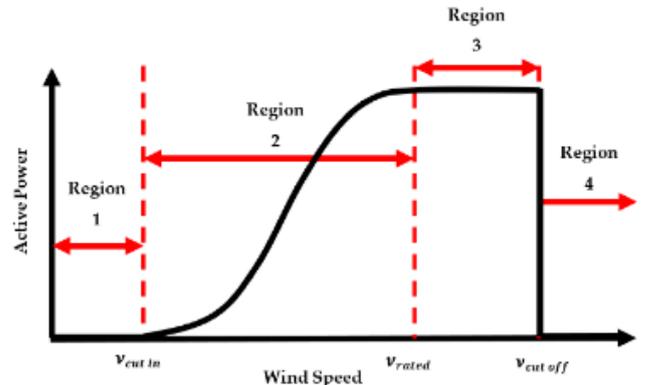


Figura 1: Regiones de operación de una turbina

3. Análisis exploratorio y preprocesamiento de los datos

Los datos empleados provienen de las medidas realizadas por un sistema SCADA de una turbina eólica situada en Turquía. Estos datos son de dominio público, y se encuentran disponibles en la página web Kaggle¹. Aunque, los datos empleados provienen de una turbina eólica terrestre, el enfoque presentado para la detección de anomalías y las conclusiones derivadas del análisis realizado son aplicables a turbinas eólicas tanto terrestres como marinas. Los datos corresponden a medidas tomadas, en intervalos de 10 minutos, de los siguientes parámetros:

1. **Date/Time:** Fecha y hora de la medida.
2. **LV ActivePower (kW):** Potencia generada por la turbina en el instante de la medida.

¹ <https://www.kaggle.com/datasets/berkerisen/wind-turbine-scada-dataset>

3. **Wind Speed (m/s):** Velocidad del viento en el instante de la medida.
4. **Theoretical Power Curve (kWh):** Valor de la curva teórica correspondiente a la velocidad del viento medida.
5. **Wind Direction (°):** Dirección de la que proviene el viento en el instante de la medida.

En la Figura 2 (derecha) se representa la curva de potencia teórica de la turbina que se empleará para el entrenamiento de los algoritmos, la cual tiene una potencia nominal de 3600 kW. Si el comportamiento de la turbina fuese ideal, la potencia generada dada una determinada velocidad del viento sería la misma, independientemente del resto de factores. No obstante, tanto las no idealidades como las anomalías modificarán este valor, haciendo que la potencia generada por la turbina difiera con respecto del valor teórico correspondiente.

En la Figura 2 (izquierda) se muestra el diagrama de dispersión de la potencia real con respecto de la velocidad del viento. Se asemeja a la curva teórica, con la diferencia de que hay valores de potencia que no coinciden con el valor teórico.

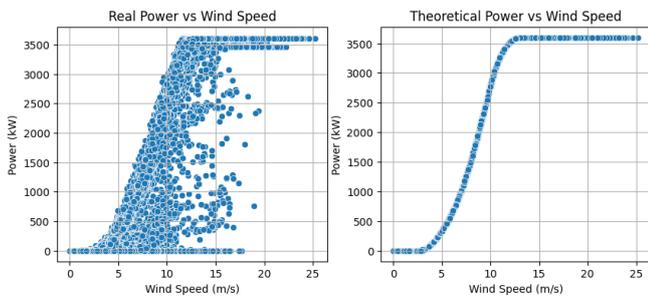


Figura 2: Diagramas de dispersión de la potencia real generada (izquierda) y la potencia teórica (derecha).

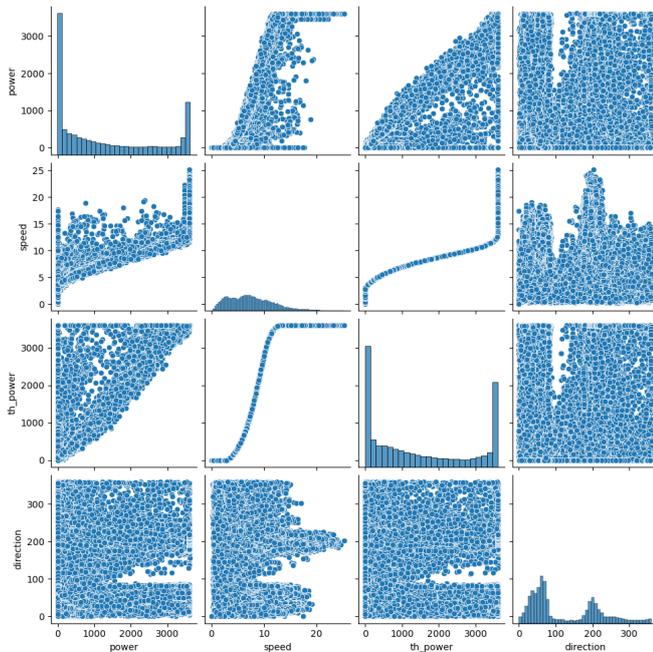


Figura 3: Diagramas de dispersión para las diferentes variables presentes en el conjunto de datos.

La Figura 3 muestra el diagrama de dispersión de las diferentes variables presentes en el conjunto de datos. En

ellos se observa que tanto la potencia como la curva teórica no dependen de la dirección del viento, sino que estas únicamente dependen de la velocidad del viento.

Se observan también valores de potencia generada muy por debajo de la potencia teórica, lo que indica la presencia de datos anómalos.

2.2. Preprocesamiento

Para la limpieza de los datos se realizará una sustitución de los valores anómalos de potencia generada por valores de funcionamiento normales. Para llevar a cabo la catalogación de los datos en “anómalos” y “normales” se compararán los valores de potencia real generada con los valores de la curva teórica de potencia.

El comportamiento ideal de la turbina tendría una potencia generada igual que la potencia teórica. Se considerará como anómalo aquel valor cuya distancia con respecto al valor teórico sea mayor que 0.45 veces la desviación típica de la distribución de la curva de potencia.

Los datos anómalos se han sustituido por el valor teórico correspondiente con una componente de ruido gaussiano de media nula y desviación típica de 200 kW. El resultado de esta limpieza de datos se muestra en la Figura 4. Los datos coloreados en azul muestran un comportamiento normal de la turbina, mientras que los naranjas son los considerados como anómalos. En la Figura 4, derecha, se muestra la curva tras haber sustituido las muestras catalogadas como anómalas por su correspondiente valor de funcionamiento normal.

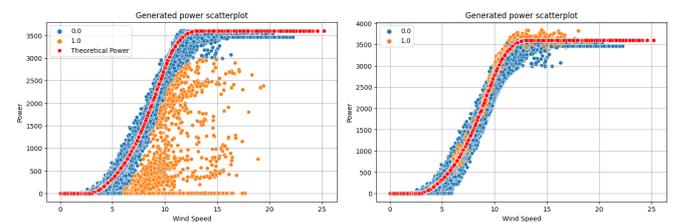


Figura 4: Potencia generada en comparación con la curva teórica (izquierda). En azul, los datos normales. En naranja, los datos catalogados como anómalos. Sustitución de los datos anómalos por datos de un funcionamiento normal de la turbina (derecha).

Cabe la posibilidad de que los datos presenten un cierto comportamiento periódico que, en caso afirmativo, podría ser de utilidad para realizar predicciones. Para comprobar si es así se realiza un análisis de la transformada de Fourier de todas las señales disponibles, mostrado en la Figura 5.

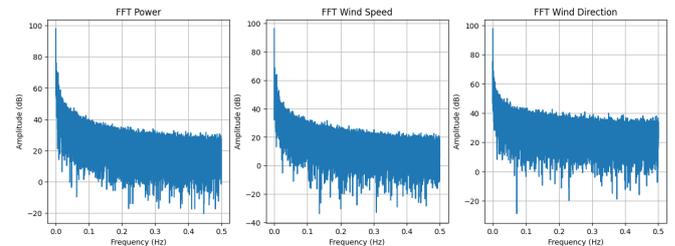


Figura 5: Transformadas de Fourier de las tres señales de estudio.

Las representaciones de las transformadas de Fourier muestran una clara dominancia de la frecuencia 0 sobre el resto de las frecuencias en las tres variables, por lo que queda descartado cualquier componente periódico.

El conjunto de datos cuenta con un total de 50530 muestras de comportamiento normal de la turbina, mientras que únicamente consta de 2923 muestras anómalas (un 5.78% del total).

3. Resultados experimentales de las técnicas de ML

3.1. Detección de anomalías con KNN

El algoritmo de los K-Vecinos, KNN (del inglés, K-Nearest Neighbors), es un algoritmo de aprendizaje automático de clasificación basado en clustering (Uddin, et al., 2022). Dado un conjunto de datos $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ con n muestras y dos o más clases $l = [l_1, l_2, \dots, l_m]$, el algoritmo de los K-vecinos determina qué etiqueta corresponde a cada uno de los valores del conjunto de datos. El término "K" hace referencia al número de vecinos utilizados para clasificar la muestra. Determinar el valor óptimo de este parámetro resulta clave para obtener un buen rendimiento del algoritmo.

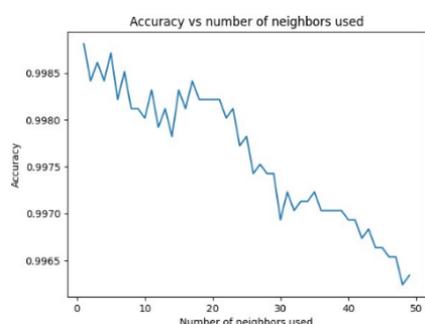


Figura 5: Precisión del algoritmo K-NN en función del número de vecinos utilizado.

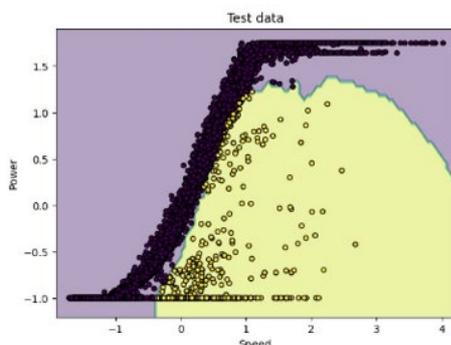


Figura 6: Resultados de la clasificación de los datos de prueba. Las muestras moradas son con comportamiento "normal" y las amarillas "anómalas".

La Figura 5 muestra el valor de precisión obtenida en función del número de vecinos utilizados. En la gráfica se observa que, pese a que la precisión es siempre cercana al 100%, un menor número de vecinos conlleva una ligera mejora del rendimiento. Por tanto, se utilizará un único vecino para clasificar las muestras. El resultado de la división del espacio de características realizada por el algoritmo se muestra en la Figura 6. La región morada corresponde a regiones donde los valores de potencia generada y velocidad del viento se consideran normales, mientras que en la región amarilla se sitúan los datos anómalos.

3.2. Detección de anomalías con Random Forest

El algoritmo de bosques aleatorios (Random Forest) es un algoritmo de aprendizaje automático supervisado (Gislason, et al., 2006). Consiste en la combinación de varios árboles de clasificación. Para la creación de dichos árboles se emplea la técnica de *bootstrapping*. Dado un nuevo conjunto de características, el algoritmo lo introducirá a todos los árboles que conforman el bosque, y cada uno clasificará el ejemplo con una etiqueta. La etiqueta que más se repita será asignada al nuevo ejemplo.

Para la implementación del algoritmo de Random Forest se calculará el rendimiento para distintos números de árboles de decisión (Figura 7). Los mejores resultados se obtienen utilizando 60 árboles de decisión. Empleando dicho número de árboles, los resultados obtenidos se representan en la Figura 8. En dicha Figura, la zona morada representa la región del espacio de características asociado a un comportamiento normal de la turbina, mientras que la zona amarilla representa la región donde se encuentran los datos anómalos.

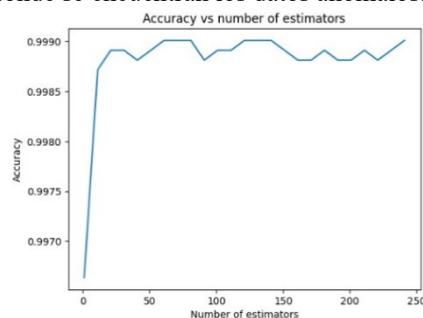


Figura 7: Precisión del algoritmo RF en función del número de árboles de decisión.

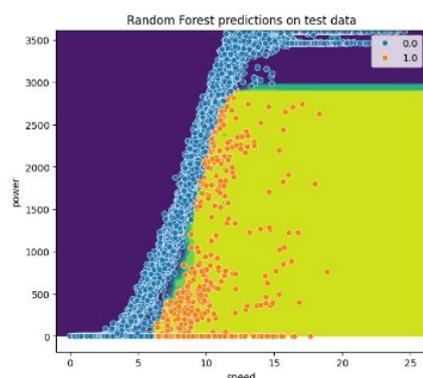


Figura 8: Resultados de la clasificación de las muestras con Random Forest. Los datos azules representan un comportamiento normal de la turbina, mientras que los naranjas son considerados anómalos.

3.3. Detección de anomalías con Isolation Forest

A diferencia de los métodos basados en clústeres, *Isolation Forest* (iForest) no utiliza ningún tipo de métrica de distancias para detectar los datos anómalos (Liu et al, 2008). El algoritmo consiste en una combinación de árboles de aislamiento.

La Figura 9 (izquierda) muestra la precisión obtenida por el algoritmo en función del número de árboles de aislamiento empleados. En ella se aprecia que la máxima precisión se obtiene cuando el número de árboles de aislamiento empleados es 61.

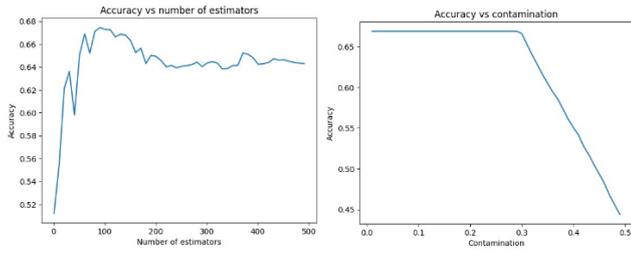


Figura 9: Precisión del algoritmo en función del número de árboles de aislamiento empleados (izquierda) y en función del parámetro de contaminación (derecha).

También se ha estudiado el impacto del parámetro de contaminación del algoritmo. La Figura 9 (derecha) muestra la precisión del algoritmo en función de este parámetro. Los resultados muestran que a partir de una contaminación de 0.3, la precisión del algoritmo se reduce drásticamente. Por tanto, la combinación de parámetros que optimizan la precisión del algoritmo puede ser 61 árboles de aislamiento y una contaminación entre 0.01 y 0.3.

La Figura 10 muestra el diagrama de dispersión de los datos, graficado sobre las regiones creadas por el algoritmo. Las regiones del espacio de características coloreadas en azul se corresponden con zonas que cuyos datos obtienen un valor de anomalía elevado. Las regiones marcadas en verde son aquellas en las que el valor de anomalía calculado es bajo.

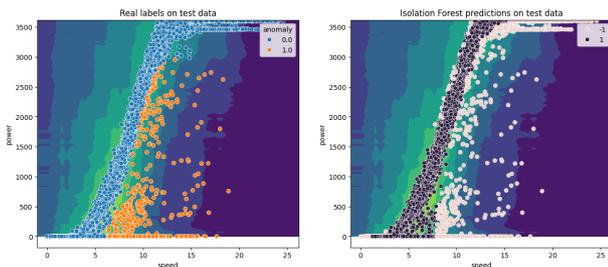


Figura 10: Etiquetas reales de los datos de prueba (izquierda) y clasificación realizada por el algoritmo (derecha). La etiqueta "-1" corresponde a datos anómalos y "1" para comportamientos normales.

4. Análisis y comparación de los resultados

Para el análisis comparativo de los métodos de detección de anomalías se comenzará con los métodos basados en clustering, para luego analizar el Isolation Forest.

4.1. Métodos basados en clustering (KNN y RF)

Los métodos de ML basado en clústeres dividen el espacio de características en dos regiones: comportamiento normal y anómalo. La catalogación de una nueva muestra dependerá de dónde se sitúe en el espacio de características. La Figura 11 muestra las regiones definidas por los dos algoritmos basados en clústeres usados, K-NN y RF.

Ambos algoritmos son capaces de detectar los datos anómalos con una alta precisión. No obstante, la limitación que presentan radica en la imposibilidad de, una vez entrenados, ser extrapolados a datos provenientes de otras turbinas, a no ser que se conozca la potencia nominal de la nueva turbina. La Figura 12 muestra un ejemplo de aplicación de los algoritmos a otra turbina si no se conoce la potencia nominal de la misma. Se observa que, puesto que la nueva curva de potencia es diferente a la empleada en el

entrenamiento, la división del espacio de características no se realiza correctamente.

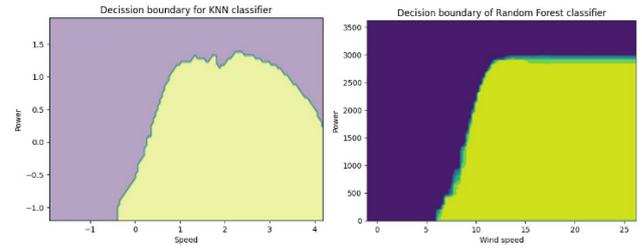


Figura 11: Regiones de comportamiento normal (morado) y anómalo (amarillo) en datos pertenecientes a una curva de potencia.

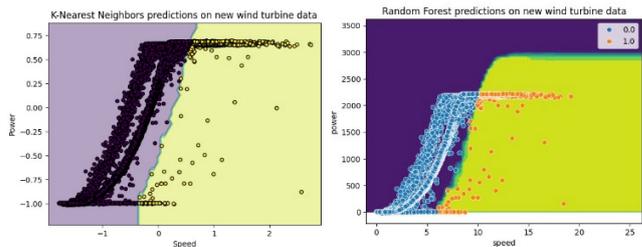


Figura 12: Aplicación de KNN y Random Forest a datos provenientes de otra turbina.

4.2. Isolation Forest

La detección de anomalías con el algoritmo de Isolation Forest es similar al de los métodos basados en clústeres. La diferencia está en que Isolation Forest calcula un valor que representa cómo de probable es que una determinada muestra sea anómala. El principal problema que presenta este algoritmo es que clasifica, de forma errónea, los datos de la tercera región de la curva como anómalos. Esto es debido a que, cuando se realizan las divisiones del espacio de características, los valores de esta región de la curva están alejados de la segunda región, lo que hace que resulten fáciles de aislar. Puesto que estos datos se aíslan con facilidad, el algoritmo interpretará que son anómalos, cuando realmente no lo son. Además, el algoritmo también presenta la limitación de los dos algoritmos anteriores: es necesario conocer la potencia nominal de la turbina. La Figura 13, izquierda, muestra la curva de potencia trazada por el algoritmo donde se ve claramente que la región 3 de la curva es considerada como anómala. La Figura 13, derecha, muestra la aplicación del algoritmo a una turbina de potencia nominal desconocida.

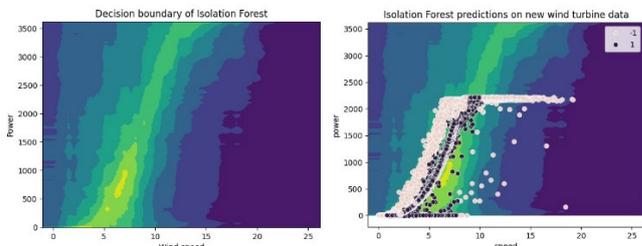


Figura 13: Curva de potencia obtenida por el algoritmo de Isolation Forest (izquierda) y aplicación del algoritmo a datos de otra turbina desconocida.

La Tabla 1 muestra, a modo de resumen, las principales ventajas y desventajas de los diferentes métodos aplicados. Los métodos basados en clústeres son la mejor alternativa para la detección de anomalías cuando se conocen los parámetros

que definen la curva de potencia. Por otra parte, el algoritmo de Isolation Forest únicamente es capaz de detectar anomalías que se encuentren dentro de las 2 primeras regiones de la curva de potencia (siempre y cuando se conozcan los parámetros de la turbina).

Tabla 1: Ventajas y desventajas de los métodos analizados.

	K-NN, Random Forest	Isolation Forest
Ventajas	Fáciles de aplicar. Regiones de decisión gráficamente representables.	Fácil de aplicar. Regiones determinadas por el algoritmo gráficamente representables.
Desventajas	Necesario conocer datos acerca de la curva de potencia de la turbina.	Necesario conocer datos acerca de la curva de potencia de la turbina Solo trabaja en las regiones 1 y 2 de la curva de potencia.

La Tabla 2 muestra los resultados obtenidos. Para la evaluación de los algoritmos se han empleado tres métricas:

1. **Precisión:** Indica qué proporción de muestras catalogadas como “anómalas” lo eran realmente.
2. **Recall:** Indica la proporción de anomalías detectadas.
3. **F1:** Se define como la media armónica de los parámetros precisión y recall:

$$F1 = 2 \cdot \frac{\text{precisión} \cdot \text{recall}}{\text{precisión} + \text{recall}} \quad (2)$$

En la Figura 11 se observa que Isolation Forest consigue detectar todas las muestras anómalas (recall=1), pero a costa de obtener una gran cantidad de falsos positivos (precisión=0.018). Por otra parte, los algoritmos basados en clustering (KNN y Random Forest) presentan valores de las tres métricas muy cercanas a la unidad, lo que indica que realizan una clasificación casi perfecta.

Tabla 2: Resultados obtenidos en los datos de prueba para los diferentes algoritmos.

ALGORITMO	PRECISIÓN	RECALL	F1
KNN	0.988	0.991	0.989
Random Forest	0.999	0.979	0.989
Isolation Forest	0.018	1.000	0.036

4. Conclusiones

En este artículo se han implementado diferentes algoritmos de Machine Learning para tareas de detección de anomalías en datos reales de una turbina eólica. Dos de los métodos se basan en clústeres (KNN y Random Forest) y uno de ellos en árboles de aislamiento (Isolation Forest). Se ha comprobado que los algoritmos de clustering presentan un muy buen rendimiento en tareas de detección de anomalías, mientras que Isolation Forest es incapaz de catalogar

correctamente las anomalías presentes en la tercera región de la curva de potencia.

Los métodos comparados se basan en el análisis de las medidas tomadas en cada instante como muestras independientes. Por ello, los resultados obtenidos son susceptibles de mejorarse mediante el análisis no sólo de muestras puntuales, sino de secuencias temporales, de manera que la consistencia temporal y el modo de evolución de los parámetros medidos sean tenidos en cuenta. Para realizar esta mejora futura se precisa de métodos más complejos, como los basados en redes neuronales, que a su vez requieren de un volumen mayor de datos de entrenamiento.

Los resultados de este estudio muestran la potencial aplicación de estos métodos en tareas de monitorización de turbinas eólicas, tanto marinas como terrestres, para la detección temprana de comportamientos anómalos y, por tanto, la optimización de la obtención de energía.

Agradecimientos

Este trabajo ha sido realizado parcialmente gracias al apoyo del proyecto PID2021-123543OB-C21 de MICIU/AEI/ y FEDER Una manera de hacer Europa

Referencias

- Bilendo, F., A. M., Badihi, H., Lu, N., Cambron, P., & Jiang, B. (2023). Applications and Modelling Techniques of Wind Turbine Power Curve for Wind Farms - A Review. *Energies*, 16.
- Gislason, P. O., Benediktsson, J. A., & Sveinsson, J. R. (2006). Random Forest for land cover classification. *Science Direct*.
- Michelena, Á., Zayas-Gato, F., Jove, E., Casteleiro-Roca, J. L., Quintián, H., Fontenla-Romero, Ó., & Luis Calvo-Rolle, J. (2024). Novel adaptive approach for anomaly detection in nonlinear and time-varying industrial systems. *Logic Journal of the IGPL*, jzae070.
- Muñoz-Palomeque, E., Sierra-García, J. E., & Santos, M. (2023). Técnicas de control inteligente para el seguimiento del punto de máxima potencia en turbinas eólicas. *Revista Iberoamericana de Automática e Informática industrial*.
- Morrison, R., Liu, X., & Lin, Z. (2022). Anomaly detection in wind turbine SCADA data for power curve cleaning. *Renewable Energy*, 184, 473-486.
- Pandit, R., Astolfi, D., Hong, J., Infield, D., & Santos, M. (2023). SCADA data for wind turbine data-driven condition/performance monitoring: A review on state-of-art, challenges and future trends. *Wind Engineering*, 47(2), 422-441.
- Portillo, G. (15 de enero de 2024). *Renovables Verdes*. Obtenido de <https://www.renovablesverdes.com/turbina-eolica/>
- Sacie, M., Santos, M., López, R., Pandit, R. (2022). Use of state-of-art machine learning technologies for forecasting offshore wind speed, wave and misalignment to improve wind turbine performance. *Journal of Marine Science and Engineering* 10(7), 938.
- Zamanzadeh Darban, Z., Webb, G. L., Pan, S., Aggarwal, C. C., & Salehi, M. (2022). Deep Learning for Time Series Anomaly Detection: A Survey. *ACM*, 43.
- Zhang, S., Robinson, E., & Basu, M. (2024). Wind turbine condition monitoring based on three fitted performance curves. *Wind Energy*, 27(5), 429-446.
- Zhou, B., Zhang, Z., Li, G., Yang, D., & Santos, M. (2023). Review of key technologies for offshore floating wind power generation. *Energies*, 16(2), 710.
- Uddin, S., Haque, I., Lu, H., Moni, M., Gide & Gide, E. (2022). Comparative performance analysis of K-nearest neighbour (KNN) algorithm and its different variants for disease prediction. *Nature*.
- Liu, F., Ting, M. & Zhou, Z. (2008). Isolation Forest. 2008 eight ieeec international conference on data mining. *IEEE*.