**INFORME FINAL**

**Reporte Final del Proyecto - Predicción del Riesgo Operativo de Aerogeneradores**

**Autor: Sabrina Villegas**

**Materia: Aprendizaje Automático**

**Fecha: Junio 2025**

**1. Análisis Exploratorio de Datos (EDA)**

El análisis exploratorio se realizó sobre un dataset unificado compuesto por variables meteorológicas y datos operativos de turbinas eólicas del parque de Río Cullen, Tierra del Fuego. Se cargaron 319 registros con 11 variables, incluyendo: potencia activa, velocidad y dirección del viento, temperaturas y datos derivados.

Estadísticas generales:

* **LV ActivePower (kW)**: media 1317, mínimo 0, máximo 3602 kW.
* **Wind Speed (m/s)**: media 7.58 m/s, mínimo 1.72 m/s, máximo 19.3 m/s.
* **Temperatura promedio (tavg)**: media 5.55ºC, variando entre -7.8ºC y 17.3ºC.

Hallazgos clave:

* Se detectaron valores atípicos en potencia activa y velocidad del viento, evidenciados mediante boxplots.
* Las distribuciones muestran un sesgo hacia valores bajos en potencia activa y viento, lo que es consistente con las condiciones extremas y frías de la región.
* Se creó una variable compuesta de riesgo (riesgo\_score) combinando: velocidad del viento, ráfagas (wspd) y temperatura mínima. Se categorizó en niveles bajo, medio y alto.
* La clase "bajo riesgo" representa el 93% de las observaciones, revelando un fuerte desbalance de clases.

**2. Creación de Variable Objetivo**

La variable riesgo\_operativo se definió según un índice compuesto:

riesgo\_score = 0.6 \* Wind Speed + 0.2 \* wspd\_m\_s - 0.1 \* tmin

Donde:

* Vientos más fuertes aumentan el riesgo.
* Temperaturas más bajas lo agravan.
* Se añadió una variable binaria viento\_oeste (1 si la dirección está entre 225º y 315º).

**3. Entrenamiento de Modelos de Clasificación**

Se aplicó una metodología sistemática:

* División: 80% entrenamiento, 20% test, estratificada.
* Imputación de valores faltantes por media.
* Balanceo con SMOTE para corregir la desproporción de clases.
* Normalización con StandardScaler para SVM y K-NN.

Modelos evaluados:

* Árbol de Decisión
* Random Forest
* SVM (Support Vector Machine)
* K-Nearest Neighbors (K-NN)

Métricas comparadas (macro promedio):

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Modelo** | **Accuracy** | **Precision** | **Recall** | **F1-Score** |
| Decision Tree | 0.95 | 0.65 | 0.56 | 0.59 |
| Random Forest | 0.95 | 0.65 | 0.56 | 0.59 |
| **SVM** | 0.92 | **0.77** | 0.87 | **0.75** |
| K-NN | 0.86 | 0.60 | **0.95** | 0.68 |

**Elección del mejor modelo:**

El modelo SVM fue seleccionado como el más robusto por su equilibrio entre precisión y recall, especialmente destacando en la identificación de clases minoritarias.

**4. Conclusión**

El modelo de clasificación desarrollado permite predecir el nivel de riesgo operativo de aerogeneradores en condiciones climáticas extremas, alcanzando un rendimiento destacado con el modelo SVM (F1-score macro: 0.75).

Gracias a este enfoque:

* Se facilita la detección temprana de escenarios críticos.
* Se habilita la planificación de mantenimiento predictivo, reduciendo costos y paradas inesperadas.
* Se mejora la seguridad operativa del parque eólico.

Este proyecto demuestra que es posible aprovechar técnicas de Machine Learning para anticipar condiciones de riesgo a partir de datos meteorológicos y operativos, contribuyendo a la sostenibilidad energética en regiones con condiciones desafiantes como Tierra del Fuego.

El siguiente paso podría enfocarse en:

* Incluir más años de datos.
* Ajustar los umbrales de riesgo con expertos.
* Explorar modelos secuenciales o de tiempo real.