

Predicción de la desatención de demanda en España debido a la transición energética

*Giraldo G. Edward, Ingeniero electricista y Acevedo Mafredy, Ingeniero electricista
Universidad de Antioquia*

Resumen— Este proyecto aborda la predicción de la brecha entre la generación de energía mediante combustibles fósiles y fuentes renovables en España. En él, se realizó una exploración de datos, seguida de imputaciones, tratamiento de valores atípicos y reducción de características. Seguidamente, se aplicaron técnicas de regresión lineal y no lineal, destacando el modelo Random Forest Regressor como el más efectivo. Las métricas evaluadas, con puntajes máximos de 0,32 para R^2 y 4315 para RMSE, corresponden a la máxima precisión obtenida mediante el modelo. El enfoque se centró en predecir el déficit energético utilizando variables climáticas específicas de cada ciudad. Este estudio proporciona información relevante para conocer la energía necesaria para suplir el déficit relacionado a la aleatoriedad del recurso renovable en España, con el objetivo de mejorar la confiabilidad y sostenibilidad del suministro eléctrico del país.

Palabras claves: Déficit energético, Energías Renovables, Transición energética, Python, Pandas, Machine Learning, Aprendizaje Supervisado

I. INTRODUCCIÓN

En los últimos años, la sostenibilidad de las fuentes de energía se ha convertido en una preocupación crítica a nivel mundial. La dependencia de fuentes basadas en combustible fósil para la generación de electricidad plantea desafíos para garantizar un suministro de energía constante y respetuoso con el medio ambiente. Como respuesta a esta problemática, los gobiernos están invirtiendo cada vez más en infraestructuras de energía renovable.

A medida que se acelera la transición energética, surgen problemáticas relacionadas con la generación de energía no convencional. Una de ellas es el déficit de suministro energético que puede ocurrir debido a la naturaleza aleatoria de los recursos renovables, como la disponibilidad del viento y el sol [2]. En contraste, la generación de energía térmica basada en combustibles fósiles que se pretende reemplazar, tiene una disponibilidad más constante, siempre y cuando haya un suministro adecuado de combustibles. Este déficit energético es un desafío que debe abordarse en conjunto con la transición energética [3][6].

Siguiendo la tendencia, el gobierno de España está considerando la expansión de sus recursos de energía renovable. Para respaldar este proceso de toma de decisiones, es esencial comprender las tendencias y patrones de las fuentes de energía renovable respecto a la generación de energía mediante combustibles fósiles en el país.

Este estudio tiene como objetivo desarrollar un modelo

predictivo que estime la brecha entre la energía generada por los combustibles fósiles y diversas fuentes renovables en España. Mediante el análisis de características climáticas específicas de cada ciudad, como la presión atmosférica, la velocidad del viento y la humedad, buscamos pronosticar el déficit diario. Los resultados de este estudio proporcionarán información valiosa para el gobierno de España a la hora de tomar decisiones informadas sobre la expansión de la infraestructura de energía renovable y mejorar la sostenibilidad y confiabilidad general del suministro energético del país. Se tomará como punto de partida la información suministrada para el desafío Kaggle realizado por Explore Data Science Academy llamado “Electricity Shortfall Challenge” [1]

II. MARCO TEÓRICO

A. Sostenibilidad y confiabilidad

1) Sostenibilidad

La sostenibilidad se refiere a la capacidad de mantener un sistema eléctrico eficiente, de manera tal que se minimicen los impactos ambientales negativos y se maximicen los beneficios socio-económicos para sus usuarios.

Actualmente, a pesar de la creciente penetración de las energías renovables en los Sistemas Eléctricos de Potencia (SEP) en las últimas dos décadas, la energía generada a base de combustibles fósiles representa más del 70% de la energía producida a nivel mundial [4].

2) Confiabilidad

La confiabilidad en sistemas eléctricos se refiere a la capacidad de suministrar energía eléctrica de manera constante y segura, sin interrupciones o cortes inesperados, evitando problemas como apagones o fluctuaciones de voltaje que puedan afectar a los usuarios y a los dispositivos eléctricos conectados.

B. Déficit energético

El déficit energético ocurre cuando el consumo de energía eléctrica (d) en un instante de tiempo t es más grande que la energía generada (p) en ese mismo instante [5]:

$$p(t) - d(t) \leq 0$$

Normalmente, los SEP se planean con antelación para evitar esta brecha entre consumo y generación de energía [5], la mayor cantidad de tiempo posible. No obstante, este déficit puede ocurrir por diversas situaciones durante algunos instantes en un día. Es de interés para el presente análisis, la diferencia

1) *BD completa*

Se genera la primera BD completa (todas las características) para entrenar y probar los modelos como resultado de este tratamiento de datos.

2) BD por correlación

Usando un criterio de correlación mayor a 0,7 y menor a -0,7 se eliminaron las características altamente relacionadas (temperaturas). Se obtiene una segunda BD reducida a 27 características (Figura 4)

C. Comparación de modelos de regresión

Con el fin de obtener la mejor estimación se realizó una comparación de las diferentes métricas seleccionadas para evaluar los modelos, los resultados se detallan a continuación:

Tabla 4. Valores de RMSE y R^2 para los modelos implementados

Modelo	RMSE	R^2
Linear Regressor	4862,52	0,13
Random Forest	4621,65	0,21
Random Forest + Randomized Search	4315,4	0,32
StackingRegressor	4542,1	0,24
Máquina de soporte vectorial (SVM)	-4597,2	0,21
Arboles de decisión	-4896,1	0,11

De la Tabla 4 se identifica que el modelo con los mejores resultados es el modelo de *Random Forest* sintonizado con la técnica *Randomized Search*. En la Figura 6 se observa el comportamiento de la variable de salida estimada con el mejor modelo de regresión.

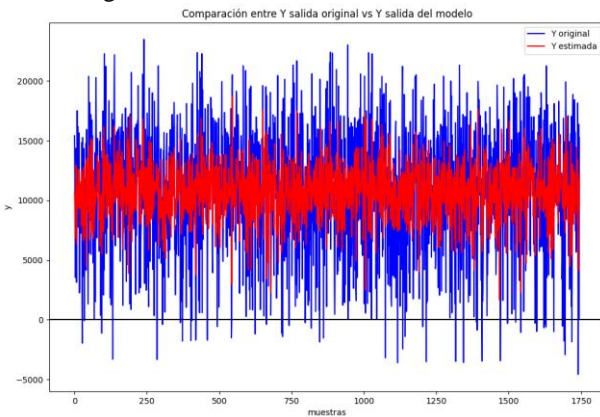


Figura 6. Comparación de la variable de salida original y la estimada por el modelo de *Random Forest + Randomized Search*.

Se puede observar que, si bien se mejoran las métricas para los diferentes modelos, estas siguen siendo bastante bajas, lo que explica el comportamiento de la variable de salida estimada, la cual no es capaz de seguir muy de cerca la variable original, esto puede deberse a una carencia presente en el tratamiento de la BD, a una alta presencia de datos atípicos o a la misma calidad de los datos.

V. CONCLUSIONES

Este ejercicio permitió predecir la demanda de energía que podría quedar insatisfecha debido a las fluctuaciones en la generación de energía a través de fuentes no convencionales. Se utilizó modelamiento basado en técnicas de Aprendizaje Automático Supervisado utilizando Python como lenguaje de programación, junto con información sobre las condiciones climáticas de algunas ciudades de España.

La metodología empleada se diseñó considerando la aplicación de técnicas de análisis de datos, adaptadas a las características de la información disponible y la variable a predecir. Durante la exploración de los datos, se encontró que era necesario realizar una preparación previa de la información antes de implementar los modelos. Se eliminaron variables consideradas irrelevantes para el modelo predictivo, así como variables suministradas mediante categorías cuya interpretación no era factible en ese momento. Además, se requirió la imputación de datos faltantes

en algunas columnas y la eliminación de valores atípicos que no concordaban con datos climatológicos reales como referencia. Por otro lado, el análisis de datos reveló altas correlaciones entre las variables de entrada y la posibilidad de reducir características mediante el Análisis de Componentes Principales (PCA). A partir de esto, se construyeron tres bases de datos: una completa, otra reducida por correlación y una reducida mediante PCA, con el fin de evaluar el impacto de eliminar variables de la base de datos.

Se aplicaron técnicas de Aprendizaje Automático a estas bases de datos, incluyendo modelos de regresión lineal y no lineal. Los primeros ensayos indicaron que el desempeño de los modelos era muy sensible a la eliminación de variables de entrada, obteniéndose peores resultados al suprimirlas. En consecuencia, se decidió realizar exploraciones más amplias utilizando la base de datos con todas las variables de entrada.

Entre los modelos evaluados, destaca el modelo *Random Forest Regressor*, con un R^2 máximo de 0.32 y un RMSE mínimo de 4315. Esto, junto con las bajas métricas obtenidas mediante el método *Linear Regressor*, indica que nos enfrentamos a un problema de regresión no lineal.

Además, se destaca la implementación de la técnica de búsqueda *Randomized Search*, la cual, fue fundamental para la obtención de mejores métricas debido a la amplia búsqueda que puede llegar a hacer, llegando a proporcionar hiperparámetros más acotados a los mejores a la hora de implementar técnicas de rejilla como *GridSearch*.

En resumen, aunque las métricas obtenidas no alcanzaron los valores ideales, este ejercicio resulta satisfactorio ya que se implementaron diversas técnicas de análisis de datos, desde la exploración y preparación de los datos hasta la implementación de múltiples modelos y técnicas de búsqueda. Esto permitió identificar las necesidades de la base de datos utilizada y su afinidad con técnicas específicas, generando ideas que se aplicarán en futuros trabajos para mejorar las métricas de la predicción realizada por el modelo final.

VI. RECOMENDACIONES PARA TRABAJOS FUTUROS

Al no conocer con claridad cuáles son las unidades de las características presentes en la BD se dificulta su análisis, por lo tanto, como trabajo futuro se propone revisar con más detalle la BD, con el fin de no descartar las características “*Seville pressure*” y “*Valencia wind deg*” ya que pueden ser de utilidad en los modelos.

Explorar con más detalle los datos, buscando formas de eliminar datos atípicos, agrupar variables y pulir las características.

Se propone ampliar la revisión de los modelos mediante el *RandomizedSearch* al ser este el que nos presenta mayor facilidad para revisar una mayor cantidad de hiperparámetros para luego ajustar esa selección mediante el *GridSearch*, reduciendo de esta forma el costo computacional.

Utilizar otros métodos de aprendizaje supervisado con el fin de encontrar resultados que aporten mejores métricas a las encontradas en el presente estudio.

Revisar las métricas utilizadas para la evaluación de los modelos, ya que puede darse el caso, de que estas no sean las más adecuadas debido al comportamiento de los datos.

REFERENCIAS

- [1] Kaggle. (2021). EDSA Individual Electricity Shortfall Challenge. [Online]. Disponible en: <https://www.kaggle.com/competitions/edsa-individual-electricity-shortfall-challenge>.
- [2] K. van der Wiel, et al. (2019). Meteorological conditions leading to extreme low variable renewable energy production and extreme high energy shortfall. *Renewable and Sustainable Energy Reviews*, Volume 111, Pages 261-275, ISSN 1364-0321, <https://doi.org/10.1016/j.rser.2019.04.065>.
- [3] Rafat Aljarrah, Hesamoddin Marzoughi & Vladimir Terzija. (2023). Mitigating the impact of fault level shortfall in future power systems with high penetration of converter-interfaced renewable energy sources. *International Journal of Electrical Power & Energy Systems*, Volume 149, 109058, ISSN 0142-0615, <https://doi.org/10.1016/j.ijepes.2023.109058>.
- [4] Weichen Li, Xusheng Ren, Qiaoting He, Di Xu, Lichun Dong, Sustainability assessment of power generation systems under the objective consideration of criteria interactions, *Journal of Cleaner Production*, Volume 395, 2023, 136423, ISSN 0959-6526, <https://doi.org/10.1016/j.jclepro.2023.136423>.
- [5] B. François, H.D. Puspitarini, E. Volpi & M. Borga. (2022). Statistical analysis of electricity supply deficits from renewable energy sources across an Alpine transect. *Renewable Energy*, Volume 201, Part 1, Pages 1200-1212, ISSN 0960-1481, <https://doi.org/10.1016/j.renene.2022.10.125>.
- [6] S. Zaidi, M. K. Abbas, S. Saleem, B. M. Khan and S. Haider. (2021). Renewable Tidal Power Generation Significance & Challenges. 2021 International Bhurban Conference on Applied Sciences and Technologies (IBCAST), Islamabad, Pakistan, 2021, pp. 1-6, doi: 10.1109/IBCAST51254.2021.9393280.