

## Primera entrega Deep Learning

**Darwin Agudelo Hernández**

**Juliana Carvajal Guerra**

**José Alejandro Urrego Pabón**

## Contexto de aplicación

Este proyecto busca predecir la velocidad del viento a partir de datos históricos proporcionados por la plataforma NASA POWER, una herramienta valiosa no solo para estudios en energías renovables, sino también para otros ámbitos. La velocidad del viento es un factor clave para la operación de turbinas eólicas, las predicciones del clima, el diseño de infraestructuras seguras y la seguridad energética. Contar con un modelo predictivo eficiente permitiría optimizar la generación de energía eólica y mejorar la planificación en sectores que dependen de este recurso natural.

## Objetivo de machine learning

El objetivo de este proyecto es crear un modelo de aprendizaje profundo que pueda predecir la velocidad máxima del viento a 10 metros de altura en varias ubicaciones de Medellín, utilizando datos históricos diarios desde 2014-01-01 hasta 2023-12-31. Dado que la velocidad del viento es un fenómeno que varía con el tiempo y tiene dependencias temporales, se emplearán redes neuronales recurrentes (RNN) y/o LSTM (Tipo de red neuronal recurrente usada para procesar y predecir secuencias de datos) para captar esas relaciones en los datos. El modelo se entrenará y evaluará con distintos horizontes de predicción, explorando la capacidad de prever la velocidad del viento entre 7 y 15 días hacia el futuro.

Las RNN al ser redes neurales diseñadas para trabajar con series temporales o de texto, tienen una conexión o memoria que hace influyente el resultado anterior, es decir, el resultado obtenido en un tiempo  $t$  influye sobre el resultado obtenido en un tiempo  $t+1$ .

## Dataset

Los datos utilizados en este proyecto provienen de la plataforma [NASA Prediction Of Worldwide Energy Resources \(POWER\)](#), que fue creada originalmente para apoyar aplicaciones en energías renovables y modelado climático. Esta plataforma proporciona información sobre diversos parámetros meteorológicos y de superficie, disponible en diferentes resoluciones: horaria, diaria, mensual y anual.

El conjunto de datos utilizado en este proyecto incluye una única variable objetivo para el modelado (Velocidad del viento a una altura de 10 metros de altura), para 6 ubicaciones de Medellín. Junto con información adicional sobre las coordenadas geográficas de las ubicaciones:

- Variable objetivo: WS10M\_MAX (Velocidad máxima del viento a 10 metros), que se refiere a la velocidad máxima diaria del viento medida a 10 metros sobre el suelo, expresada en metros por segundo (m/s). Esta es la variable que se intentará predecir utilizando datos históricos.
- Coordenadas geográficas: Cada conjunto de datos está vinculado a un punto geográfico cercano a Medellín, con las coordenadas (latitud y longitud) específicas proporcionadas al descargar la información desde la API de NASA POWER. Las ubicaciones fueron seleccionadas estratégicamente para capturar diferentes patrones de viento dentro de la región:
  - o Punto 1: Comuna 1, latitud 6.297515, longitud -75.545754.
  - o Punto 2: Comuna 6, latitud 6.302866, longitud -75.578610.
  - o Punto 3: Comuna 9, latitud 6.234145, longitud -75.552422.
  - o Punto 4: Comuna 12, latitud 6.256823, longitud -75.602741.
  - o Punto 5: Comuna 14, latitud 6.206440, longitud -75.572283.

- Punto 6: Comuna 16, latitud 6.230899 longitud -75.602435.



El conjunto de datos abarcará un período de 10 años, desde el 2014-01-01 hasta el 2023-12-31. Este intervalo permitirá capturar patrones estacionales y tendencias a largo plazo en las velocidades del viento en las diferentes ubicaciones seleccionadas. Con 6 ubicaciones y un valor diario de la velocidad del viento para cada ubicación, se tienen 21913 registros en total que, en formato CSV, ocuparían un tamaño aproximado en disco de ~0.83 MiB.

A continuación, se presenta un ejemplo de cómo estará organizado el dataset en formato tabular, donde cada fila corresponde a una fecha específica y la velocidad del viento para cada punto geográfico:

Date	Location	WS10M_MAX (m/s)
1/01/2014	Punto 1	X
2/01/2014	Punto 1	X
3/01/2014	Punto 1	X
...	...	...
31/12/2023	Punto 1	X
1/01/2014	Punto 2	X
2/01/2014	Punto 2	X
31/08/2024	Punto 6	X

## Métricas de desempeño

La métrica de desempeño elegida para la predicción de las velocidades del viento en varios puntos de la ciudad será el MAE (Error Absoluto Medio) que nos permite comparar y cuantificar la diferencia entre las velocidades del viento reales (Las reportadas por la NASA) y las obtenidas mediante la red neuronal recurrente **Shabbir et al 2021**, cabe aclarar que las velocidades del viento en esta zona al ser relativamente bajas (2-4 m/s), se podrían tolerar diferencias hasta de una unidad, es decir:

- ✓ MAE<1 (Excelente) si las magnitudes de las variables estudiadas son pequeñas
- ✓ MAE entre 1 y 5 (Aceptable) dependiendo de las magnitudes
- ✓ MAE mayor a 5: motivo de revisión y mejora

## Negocio

El análisis predictivo realizado se podría implementar en regiones con un recurso eólico potencial, es decir, entre 5 y 12 m/s para garantizar la viabilidad del proyecto partiendo de la permanencia de este recurso eólico por varios años más.

## Resultados previos

El uso de redes neuronales recurrentes (RNN) para predecir la velocidad del viento ha demostrado ser una técnica muy eficaz para entender los patrones meteorológicos. Varios estudios han utilizado estas redes para predecir el comportamiento del viento en diferentes situaciones, mostrando buenos resultados en comparación con los métodos más tradicionales.

**Balluff et al. (2017)** aplicaron un modelo de RNN para analizar datos de velocidad del viento de la base de datos MERRA de la NASA. Su estudio mostró que las RNN pueden hacer predicciones precisas basadas en datos históricos, aunque mencionaron que todavía se puede mejorar la precisión y robustez del sistema.

Por otro lado, **Mohandes et al. (2021)** estudiaron la capacidad de predicción del viento a diferentes alturas utilizando RNN y las compararon con modelos de perceptrón multicapa (MLP). Descubrieron que las RNN eran más precisas, especialmente a mayores alturas, lo que refuerza la idea de que son útiles para hacer predicciones de viento en distintas escalas.

Otro estudio, realizado por **Guan (2022)**, propuso un modelo híbrido que combinaba descomposición espectral singular (SSA) con RNN, mejorando la estabilidad y precisión de las predicciones al reducir la volatilidad en los datos.

Finalmente, **Cao et al. (2012)** compararon modelos ARIMA, que son más tradicionales, con las RNN para predecir la velocidad del viento. Ellos encontraron que las RNN eran consistentemente más precisas, tanto en escenarios simples como en aquellos más complejos.

## Referencias

- Balluff, S., Bendfeld, J., & Krauter, S. (2017). Meteorological data forecast using RNN. *International Journal of Grid and High Performance Computing (IJGHPC)*, 9(1), 61-74.
- Cao, Q., Ewing, B. T., & Thompson, M. A. (2012). Forecasting wind speed with recurrent neural networks. *European Journal of Operational Research*, 221(1), 148-154.
- Guan, P. (2022, December). Short-term wind speed prediction method based on deep learning. In *2022 IEEE Conference on Telecommunications, Optics and Computer Science (TOCS)* (pp. 467-471). IEEE.
- Hyndman, R.J., & Koehler, A.B. (2006). "Another look at measures of forecast accuracy". *International Journal of Forecasting*, 22(4), 679-688
- Mohamed, M., Rehman, S., Nuha, H., Islam, M. S., & Schulze, F. H. (2021). Accuracy of wind speed predictability with heights using Recurrent Neural networks. *FME Transactions*, 49(4), 909.