

# GreenWest: inteligencia artificial para la predicción de créditos de carbono en proyectos de (re)forestación en España, y que corresponde al grupo B1 (GreenWest) de las líneas de investigación de DemIA.

Maider Araceli Urbón Jiménez<sup>1\*</sup>, Jaime Gabriel Vegas<sup>1</sup>, Ana de Luis Reboredo<sup>1</sup>, Belén Pérez Lancho<sup>1</sup> y Ana-Belén Gil-González<sup>1</sup>

Grupo B1, Equipo de investigación BISITE, Universidad de Salamanca  
Facultad de Ciencias, Salamanca, Castilla y León, España  
{murbon001, JaimeGabrielVegas, adeluis, lancho, abg}@usal.es

## Resumen

Este trabajo presenta **GreenWest**, un modelo de inteligencia artificial para predecir la cantidad de carbono capturado en proyectos de forestación y reforestación en España. El modelo se entrena con datos multifuente: registros del Inventario Forestal Nacional (**IFN2–IFN4**, MITECO), variables climáticas de **Copernicus/ERA5-Land** e **índices espectrales Landsat** (*Collection 2, Level 2, USGS*). Con ellos se construye una base de datos jerárquica por parcela, especie y clase diamétrica que permite predicciones a alta resolución espacial y taxonómica. Los resultados muestran un ajuste robusto ( $R^2 > 0,8$ ), con buena capacidad de generalización temporal. Este modelo permite estimar la absorción futura de CO<sub>2</sub> y optimizar la asignación de especies y terrenos para maximizar la fijación de carbono, contribuyendo a la gestión forestal sostenible y a la transición hacia una economía baja en emisiones.

## 1. Introducción

El cambio climático y el aumento del CO<sub>2</sub> atmosférico suponen una amenaza global [1, 2]. Los bosques, al actuar como sumideros naturales de carbono, son esenciales para su mitigación. En este contexto, **GreenWest** aplica inteligencia artificial para estimar la captura de carbono de cultivos forestales a partir de datos del **Inventario Forestal Nacional (IFN)**, **Copernicus/ERA5-Land** e imágenes **Landsat** [3, 4, 5].

El objetivo es desarrollar un modelo predictivo que estime la capacidad de absorción de CO<sub>2</sub> de los bosques españoles e integre información multifuente compatible con los marcos internacionales de créditos de carbono (CMNUCC, Protocolo de Kioto) [1, 2].

## 2. Metodología

La base de datos se organiza en un modelo relacional normalizado que garantiza la integridad y trazabilidad de la información. Su estructura jerárquica abarca distintos niveles de agregación: la **parcela** como unidad espacial base; el nivel **parcela–inventario**, que representa estados temporales discretos; el nivel **parcela–inventario–especie**, que describe la composición por especie; el nivel **parcela–inventario–especie–clase diamétrica**; y finalmente el nivel **parcela–inventario–estación**, que integra las variables climáticas e índices espectrales asociados.

---

\* Autora de correspondencia: murbon001@usal.es

Se implementan diversos modelos de regresión supervisada: *ensembles* basados en árboles, métodos de vecinos, redes neuronales, máquinas de soporte vectorial y modelos lineales bayesianos. Todos se entrenan en un *pipeline* con imputación, escalado y codificación *one-hot*, optimizados mediante **GridSearchCV** y métrica  $R^2$ . En líneas generales el mejor rendimiento se obtiene con **CatBoost**[6]. Se exploran cuáles de las variables recogidas tienen más impacto en la captura de carbono.

Los datos se dividen mediante una partición **hold-out** 80/20 por parcelas, aplicando validación cruzada **GroupKFold(5)** coherente con esta separación. Dado que los proyectos forestales abarcan varias décadas, se realizó una validación adicional por grupos de periodo, confirmando que el modelo generaliza adecuadamente a distintos intervalos temporales.

### 3. Resultados y conclusión

El modelo alcanzó un desempeño sólido, con un coeficiente de determinación de  $R^2 = 0,822$  y un error cuadrático medio ( $RMSE$ ) de 16,60 al predecir carbono en toneladas por hectárea (tC/ha), y un  $R^2 = 0,842$  con  $RMSE = 9,12$  al estimar el carbono bruto total por conjunto de árboles. Si se restringe la base de datos a aquellas entradas en las que la medición segunda de carbono en toneladas (variable objetivo) es superior a la primera se obtiene hasta un  $R^2 = 0,905$  con  $RMSE = 7,60$ , siendo que la evolución natural de una plantación es el crecimiento, se estaría dejando fuera parcelas en las que hayan sucedido incendios, talas y otros hechos excepcionales.

El proyecto da lugar a una **herramienta precisa para la predicción de créditos de carbono**, con un valor significativo en la toma de decisiones relacionadas con la gestión forestal y la certificación de proyectos de compensación climática. En trabajos futuros se plantea el desarrollo de un *frontend* interactivo que facilite la visualización y consulta de predicciones por parcela o región.

### Agradecimientos

Investigación financiada por la subvención **TSI-100933-2023-1** de la **Convocatoria de Cátedras Universidad-Empresa (Cátedras ENIA 2022)**, Ministerio de Transformación Digital y Función Pública de España, y el **Plan de Recuperación y Resiliencia de la UE (NextGenerationEU/PRTR)**.

### Referencias

- [1] United Nations. (1992). *United Nations Framework Convention on Climate Change (UNFCCC)*.
- [2] United Nations. (1998). *Kyoto Protocol to the UNFCCC*.
- [3] MITECO (España). *Inventario Forestal Nacional (IFN2, IFN3, IFN4)*: metodología y bases de datos. Ministerio para la Transición Ecológica y el Reto Demográfico.
- [4] Muñoz-Sabater, J., et al. (2021). *ERA5-Land: A state-of-the-art global reanalysis at land surfaces*. Copernicus Climate Data Store / ECMWF.
- [5] USGS (2021). *Landsat Collection 2 Level-2 Science Products: Surface Reflectance*. U.S. Geological Survey.
- [6] Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A.V., Gulin, A. (2018). *CatBoost: unbiased boosting with categorical features*. NeurIPS.