

Ensayo de utilización de tres metodologías de inteligencia artificial aplicado a los inventarios pirenaicos del Sistema de Información de la Vegetación Ibérica y Macaronésica (SIVIM).

Raúl Pérez¹, Gwendolyn Peyre¹, Xavier Font^{1,2}

(1) Departament de Biologia Evolutiva, Ecologia i Ciències Ambientals, Universitat de Barcelona

(2) IRBIO. Universitat de Barcelona



INTRODUCCIÓN

Objetivo general

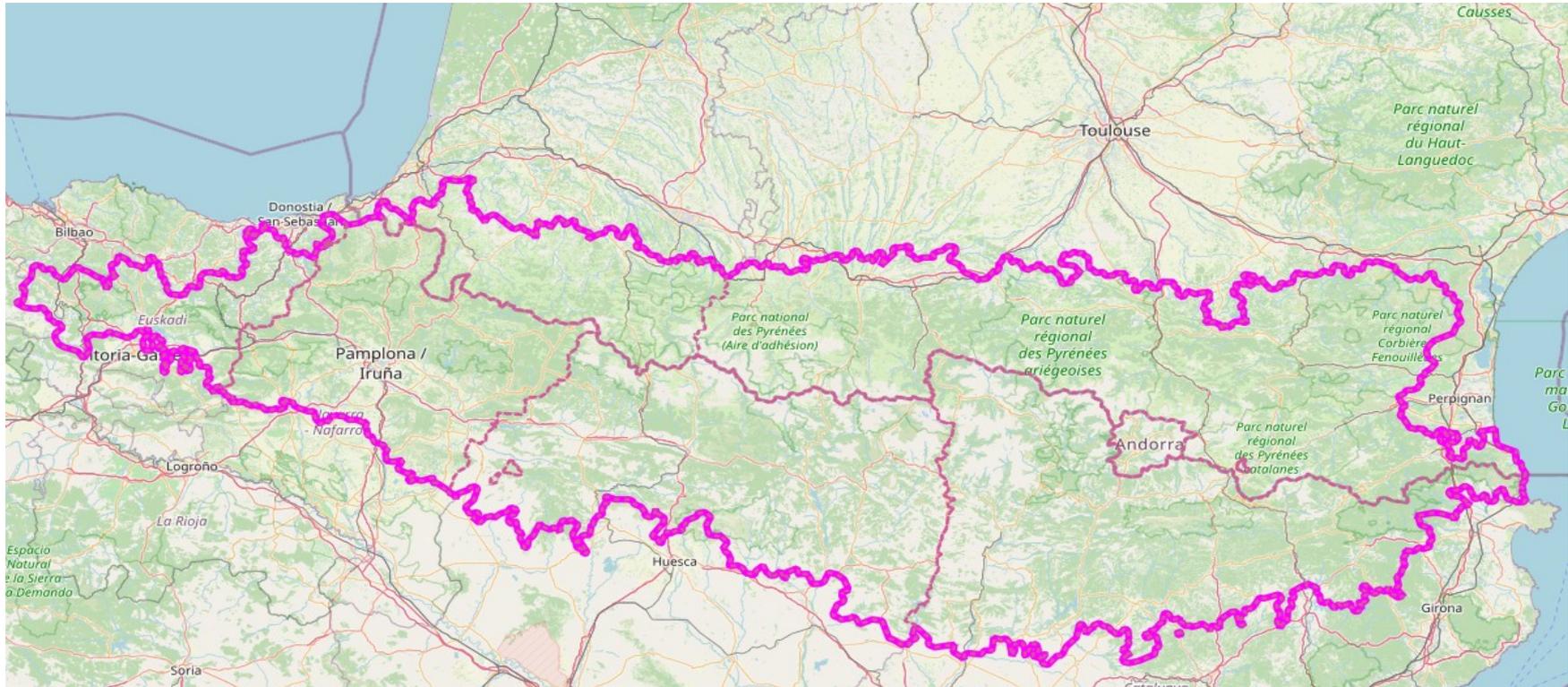
Aplicar tres técnicas de aprendizaje automático a la clasificación fitosociológica de inventarios de vegetación en los pirineos

Objetivo específico

- Predecir la pertenencia de los inventarios de vegetación de SIVIM a las alianzas fitosociológicas descritas para los Pirineos
- Comparar el rendimiento de tres algoritmos: Red Neuronal Perceptrón Multicapa (MLP) y dos tipos de árboles de decisiones (Gradient Boosting Trees, XGBoost y LightGBM), así como dos enfoques: disperso y reducido
- Revisar las alianzas "críticas"
- Determinar 3.457 inventarios de SIVIM que no disponen de alianza

MÉTODOS: Zona de Estudio

Definidas por el proyecto FLORAPYR y el Observatorio Pirenaico del Cambio Climático



INTRODUCCIÓN

- Clasificación fitosociológica → experiencia botánica → Braun-Blanquet, O. de Bolòs, S. Rivas-Martínez, etc.
- Limitaciones: diferentes escuelas y diferentes autores pueden llegar a conclusiones en parte diferentes

¿Podemos usar la inteligencia artificial para facilitar el trabajo de clasificar inventarios de vegetación?

- Existen ya trabajos recientes de clasificación de hábitats con IA (Leblanc et al., 2024)

MÉTODOS: Datos

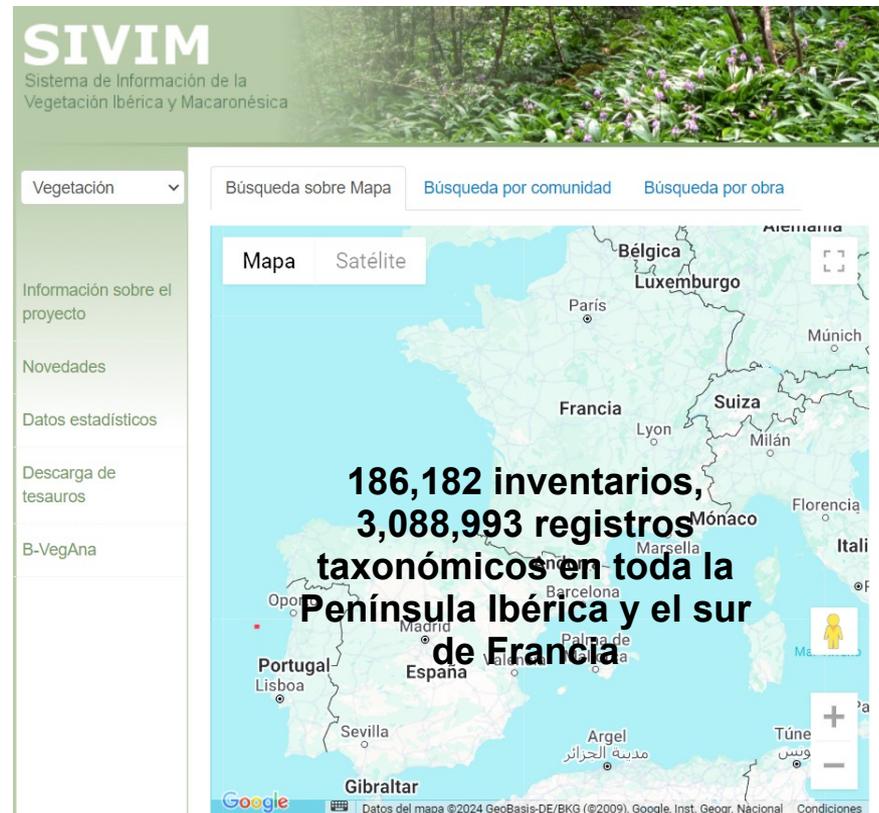
28.079 inventarios pirenaicos disponibles en SIVIM

Especies:

1. Cormófitos, a nivel específico
2. Briófitos para las comunidades hidrófilas

Inventarios:

3. Con alianza definida por el autor (267 en total)
4. Con mas de 20 inventarios



SIVIM
Sistema de Información de la Vegetación Ibérica y Macaronésica

Vegetación

Búsqueda sobre Mapa Búsqueda por comunidad Búsqueda por obra

Mapa Satélite

**186,182 inventarios,
3,088,993 registros
taxonómicos en toda la
Península Ibérica y el sur
de Francia**

Información sobre el proyecto
Novedades
Datos estadísticos
Descarga de tesauros
B-VegAna

Google | Datos del mapa ©2024 GeoBasis-DE/BKG (©2009), Google, Inst. Geogr. Nacional | Condiciones

27.354 inventarios
167 Alianzas
3.760 especies

MÉTODOS: Programas utilizados



OPTUNA



LightGBM



TensorFlow



python™



pandas

XGBoost

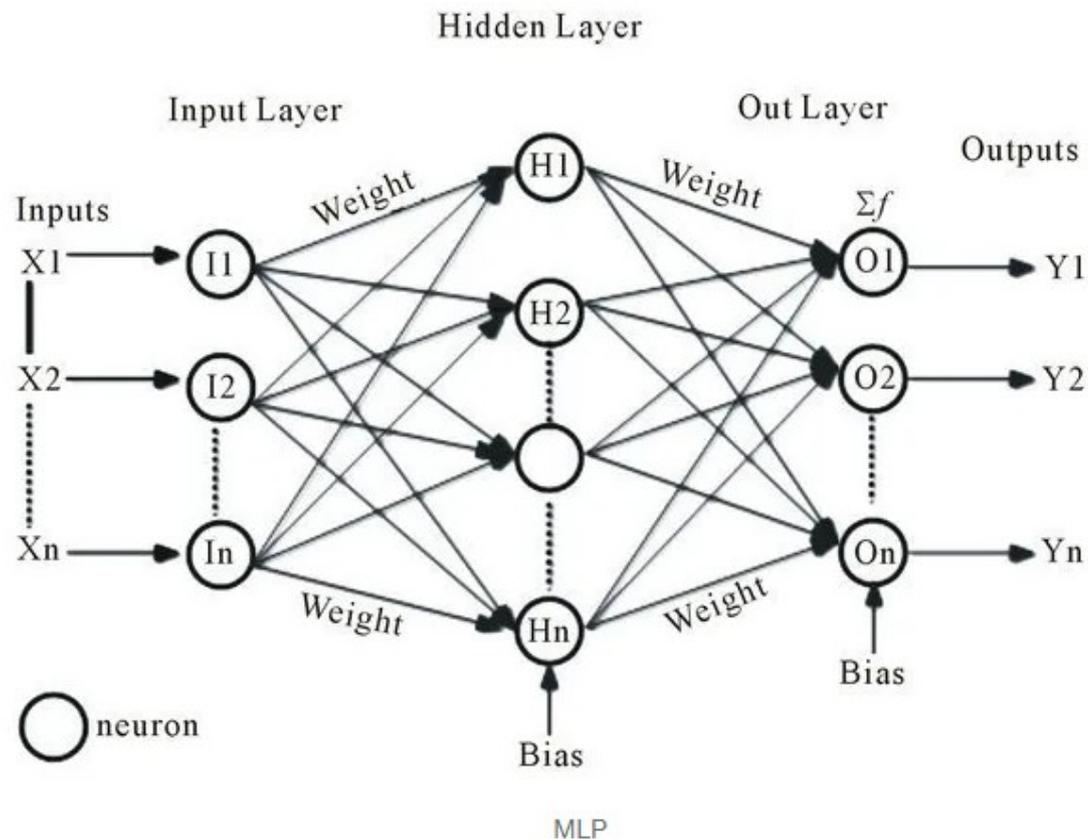


NumPy

MÉTODOS: Algoritmos

Red Neuronal (Perceptrón Multicapa, MLP)

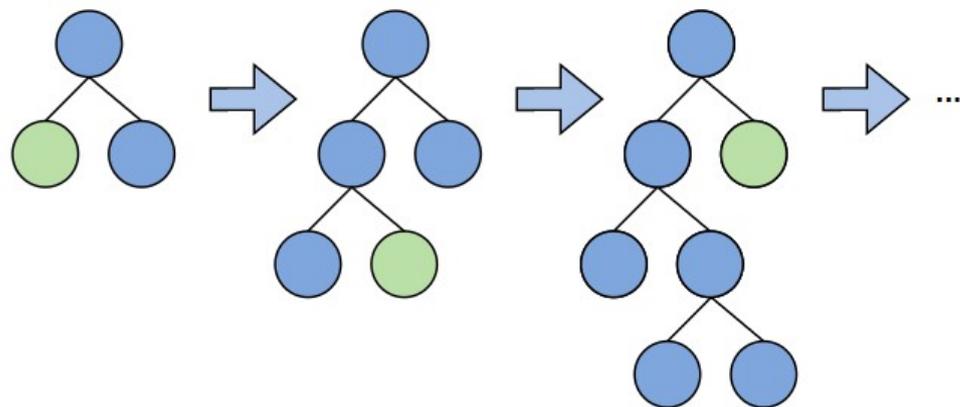
(McCulloch & Pitts 1943)



Métodos: Algoritmos

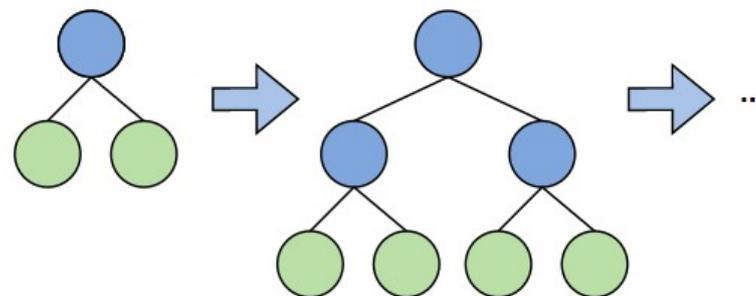
Gradient Boosting Trees

Leaf-Wise Growth



LightGBM
(Ke et al. 2017)

Level-Wise Growth

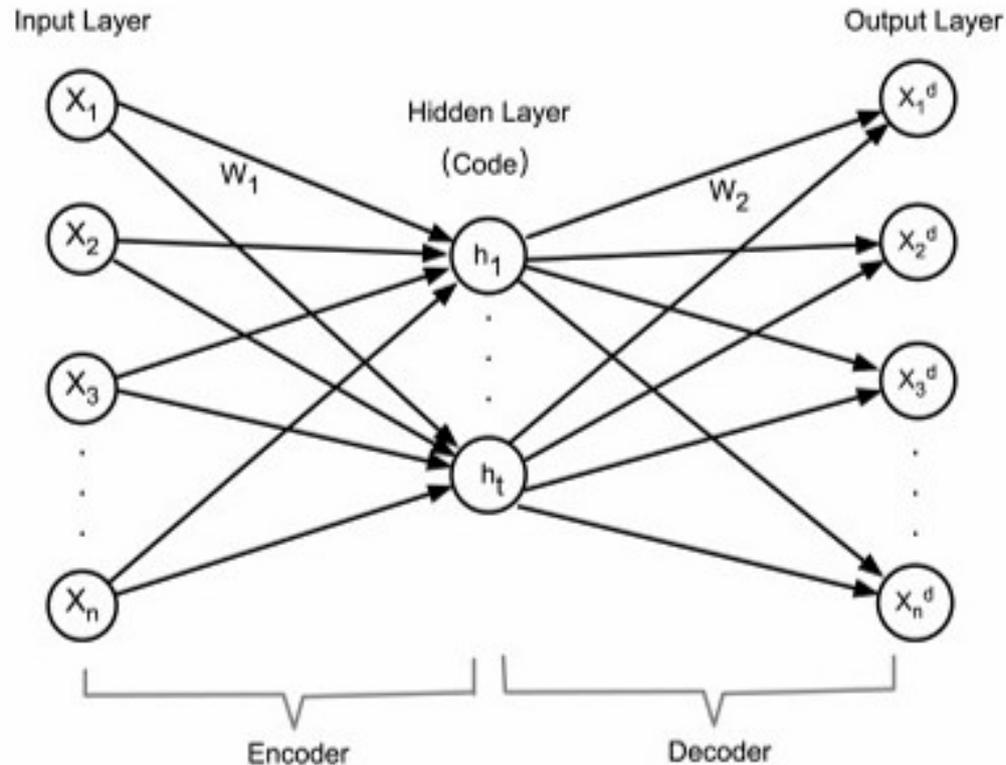


XGBoost
(Chen & Guestrin 2016)

MÉTODOS: Autocodificadores

Proceso de reducción de las dimensiones del espacio de las variables

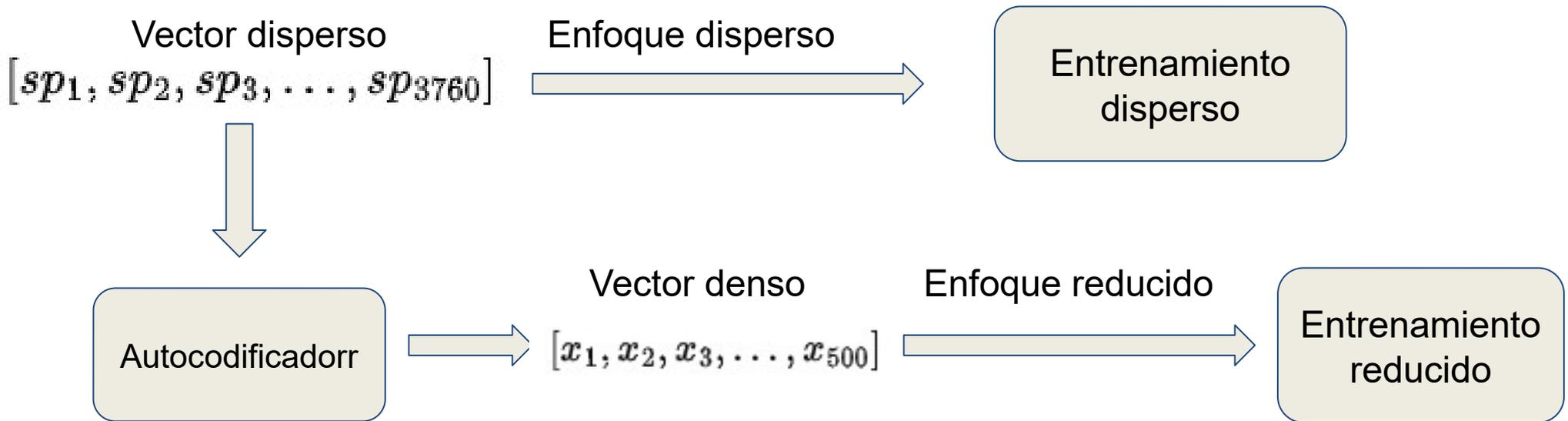
Autocodificador superficial (Li et al. 2023)



MÉTODOS: Vectorización, Enfoques disperso y reducido

Inventarios se estructuran en un vector disperso (variables originales) de 3.760 variables (especies)

A cada variable les corresponde al coeficiente de recubrimiento transformado de 2 a 9 (Van der Maarel 1979), y 0 → ausencia



MÉTODOS: Validación de los modelos

- Validación entreno con el 70% de los datos
- Validación con un conjunto de prueba del 30% de los datos

RESULTADOS: Rendimiento de los algoritmos

Enfoque disperso (variables reducidas)

Modelo	Top 1 performance	Top 2 performance	Top 3 performance
MLP	0,850	0,947	0,971
XGBoost	0,830	0,935	0,966
LightGBM	0,803	0,914	0,949

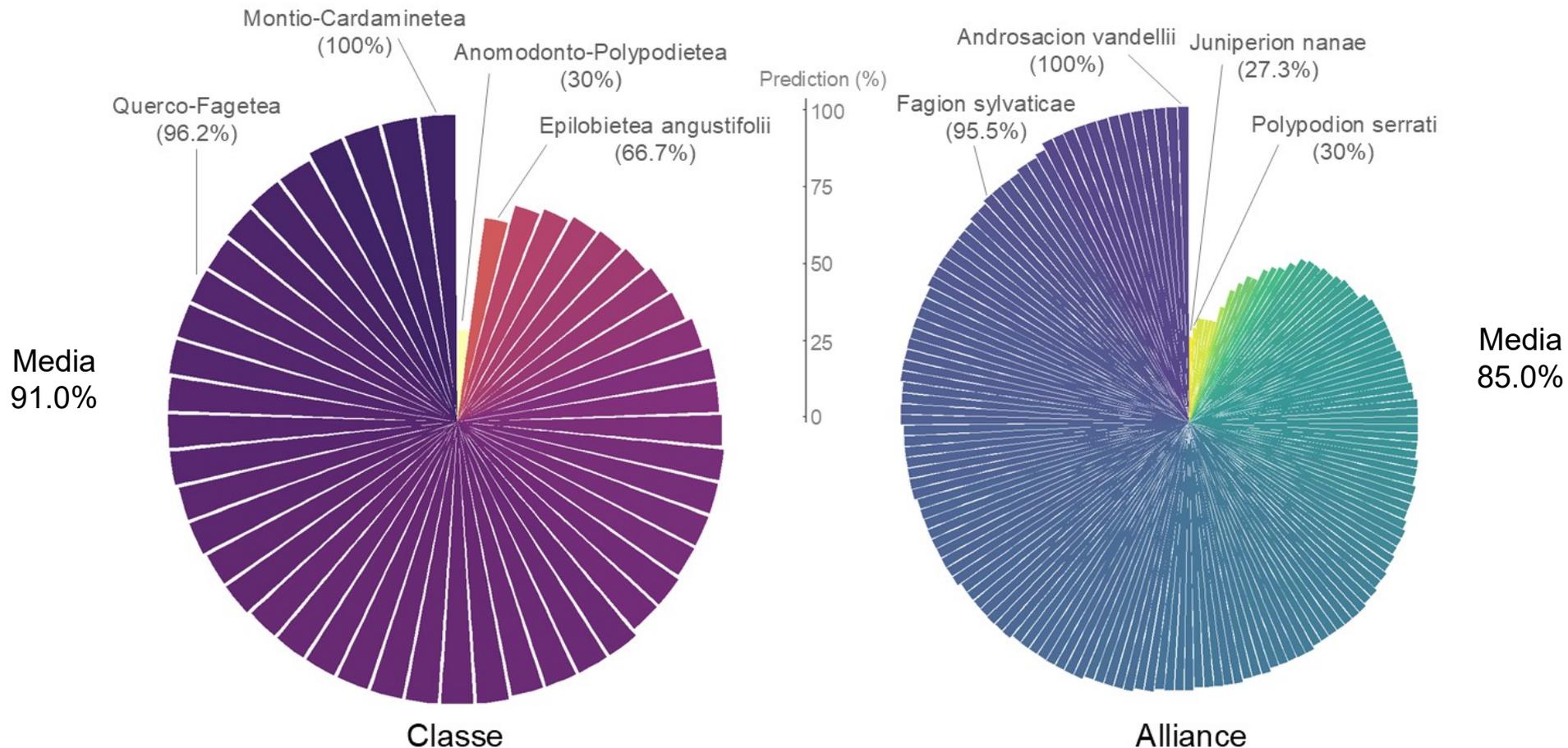
Enfoque reducido

Modelo	Top 1 performance	Top 2 performance	Top 3 performance
LightGBM	0,837	0,938	0,968
MLP	0,795	0,923	0,960
XGBoost	0,841	0,942	0,971

RESULTADOS: Rendimiento de los algoritmos, sintaxonomía

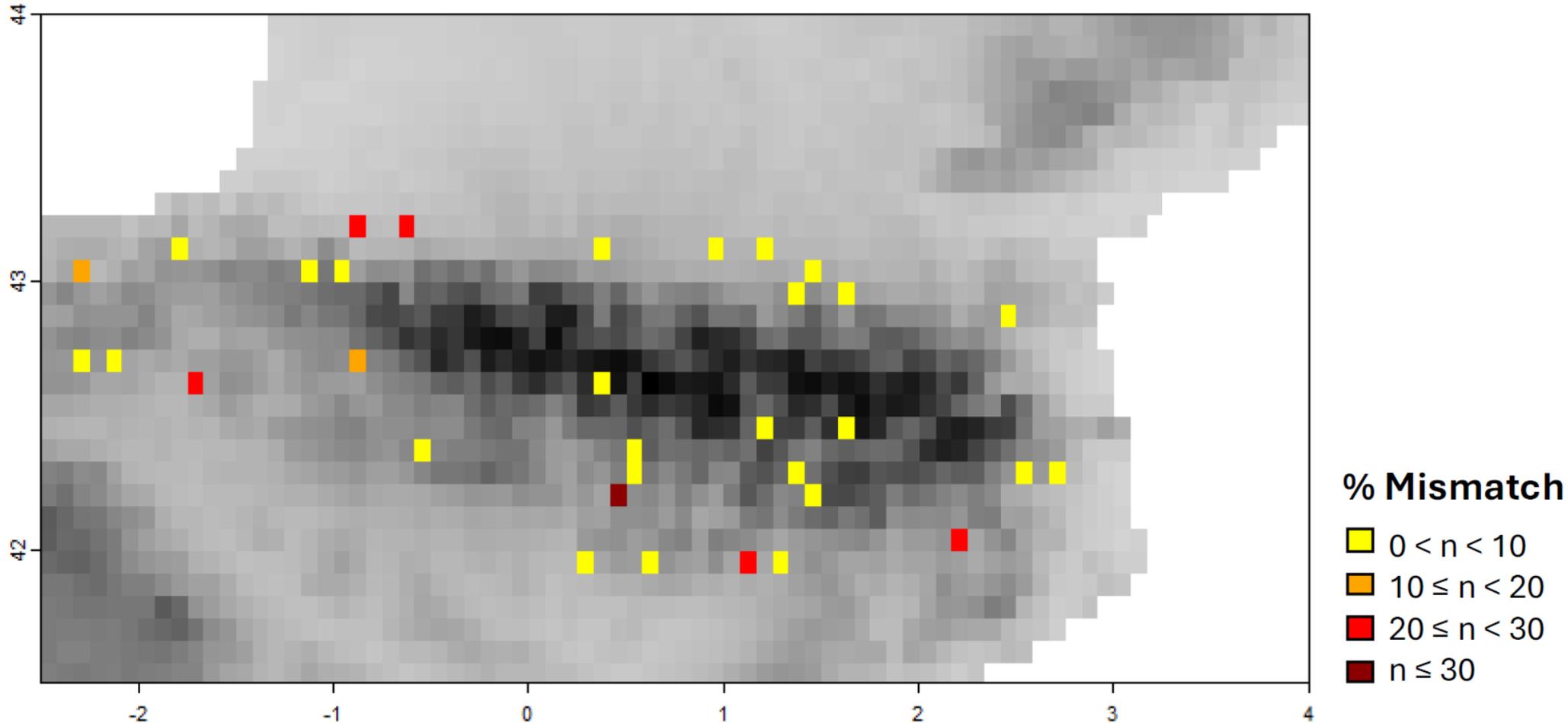
	Alianza	Orden	Clase
GBM reduced	84	87	90
GBM sparse	80	84	88
XGBoost reduced	84	87	91
XGBoost sparse	83	86	90
MLP reduced	80	84	88
MLP sparse	85	88	91

RESULTADOS: Clasificación de las alianzas

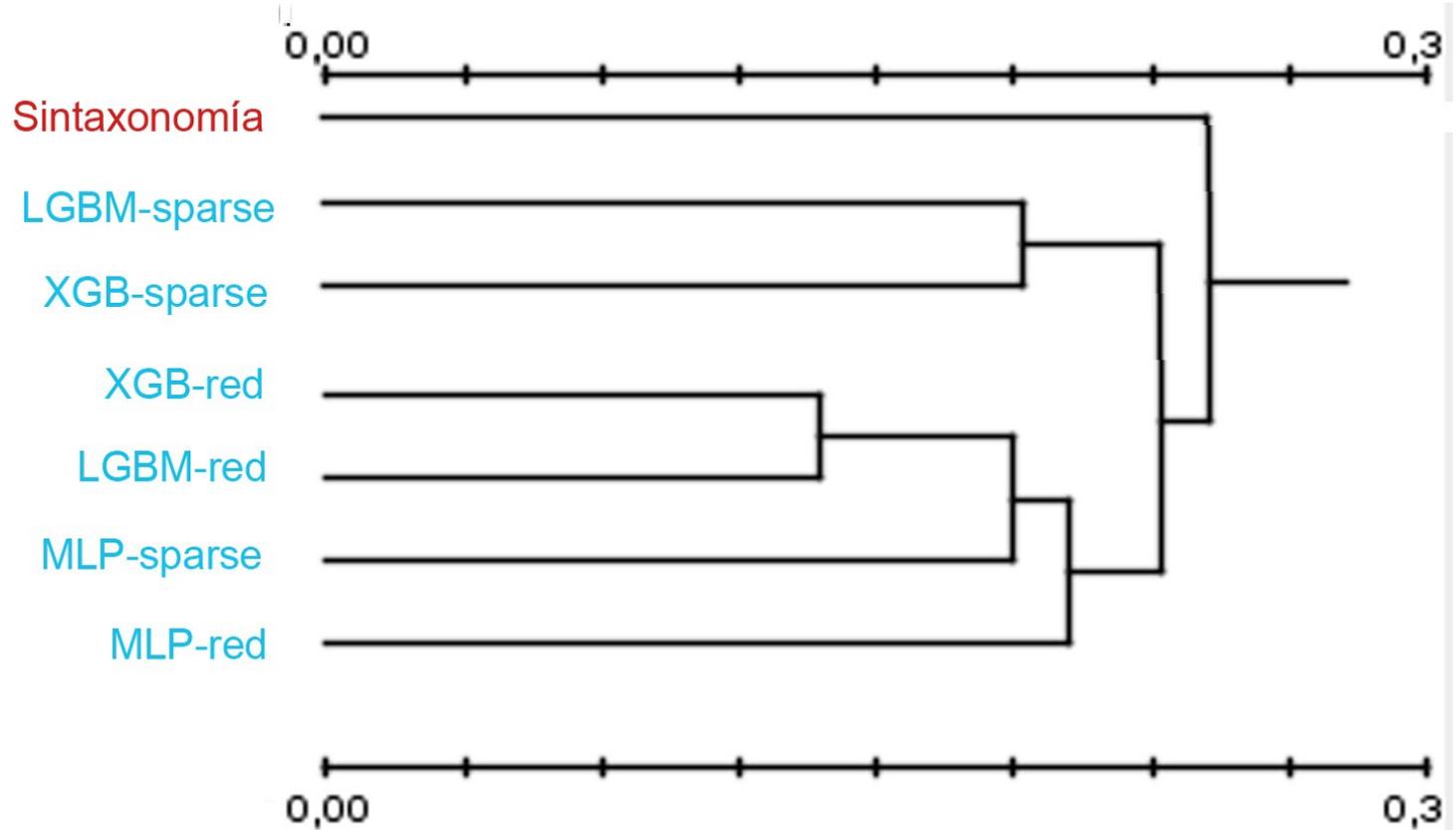


Clasificación de las 167 alianzas y sus 49 clases respectivas mediante el algoritmo MLP, enfoque disperso

RESULTADOS: Clasificación de las alianzas

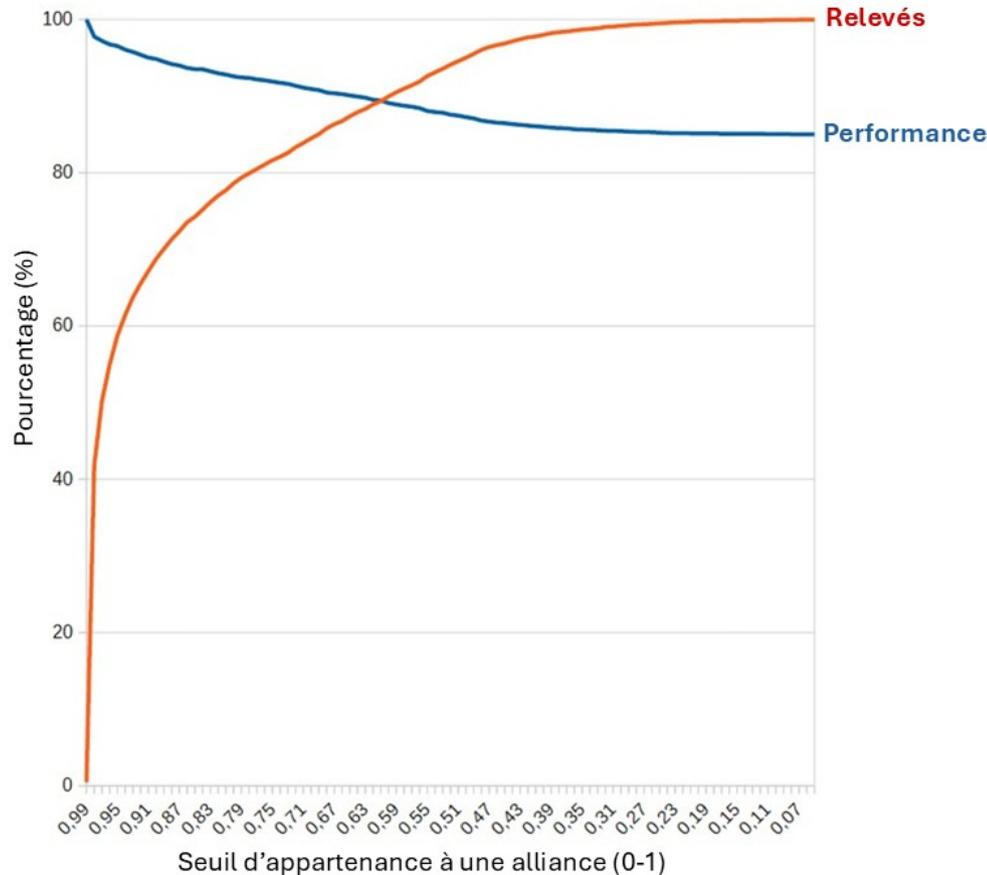


Variación geográfica de las predicciones erróneas por el algoritmo MLP, enfoque disperso.



Comparación árbol UPGM (basado en el índice de Rand [1971]) de la clasificación fitosociológica y de los tres algoritmos con los dos enfoques

RESULTADOS: Clasificación de los inventarios sin alianzas



Tendencia del rendimiento en relación con el número de inventarios de test y en función del umbral de pertenencia a la Alianza (Algoritmo MLP, enfoque disperso).

Clasificación de los 3.457 inventarios sin alianza

DISCUSIÓN

- El potencial de utilización de la Inteligencia Artificial para clasificar y revisar clasificaciones fitosociológicas es grande.
- El algoritmo de Red Neuronal Perceptrón Multicapa (MLP) es más eficiente, al igual que el enfoque disperso.
- Alianzas **mal clasificadas**
 - Detectar errores en el proceso de informatización de datos
 - Pocas especies diagnósticas, briofitos y niveles infraespecíficos.
 - Poco representadas en la zona de estudio.
 - Transición florística (sucesión).
 - Organización fitosociológica.
- Es prometedor para realizar nuevas clasificaciones,
- → PERO es imposible automatizar sin revisión experta en la actualidad.

IDEAS DE FUTURO

Utilizar otras variables además de las especies vegetales, por ejemplo coordenadas geográficas y altitud.

A partir de esta experiencia se podría implementar en SIVIM la técnica que resulte más idónea y esto permitirá la determinación automática de la alianza de vegetación de cualquier lista de plantas que se introduzca en el servidor.

REFERENCIAS

- Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. <https://doi.org/10.48550/ARXIV.1603.02754>
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). LightGBM: a highly efficient gradient boosting decision tree. Proceedings of the 31st International Conference on Neural Information Processing Systems, 3149–3157
- Leblanc, C., Bonnet, P., Servajean, M., Chytrý, M., Ačić, S., Argagnon, O., ... & Joly, A. (2024). A deep-learning framework for enhancing habitat identification based on species composition. Applied Vegetation Science, 27(3), e12802.
- Li, P., Pei, Y., & Li, J. (2023). A comprehensive survey on design and application of autoencoder in deep learning. Applied Soft Computing, 138, 110176.
- Maleki, F., Muthukrishnan, N., Ovens, K., Reinhold, C., & Forghani, R. (2020). Machine Learning Algorithm Validation. Neuroimaging Clinics of North America, 30(4), 433–445.
- McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. The bulletin of mathematical biophysics, 5, 115-133.
- Rand, W.M. (1971). Objective Criteria for the Evaluation of Clustering Methods. JASA 66: 846-850.
- Van Der Maarel, E. (1979) Transformation of Cover-Abundance Values in Phytosociology and Its Effects on Community Similarity. Vegetatio, 39, 97-114.