



## 1. DATOS BÁSICOS DEL TFG:

**Título:** Aplicación de técnicas de Machine Learning sobre conjuntos de datos relacionados con la biología

**Descripción general (resumen y metodología):**

La aplicación de técnicas de Machine Learning (aprendizaje automático) para el análisis de datos permite la obtención de modelos (basados en reglas, en árboles, en redes neuronales, etc.) que ayudan a la predicción (aprendizaje predictivo mediante clasificación o regresión) o que permiten entender la naturaleza de los datos en forma de información interpretable y útil (aprendizaje descriptivo mediante reglas de asociación o clustering). En ocasiones ambos escenarios se pueden dar a la vez, como en el caso de los modelos basados en reglas, en los que el modelo puede ser interpretado desde el punto de vista del usuario final, a la vez que permiten realizar una predicción acertada para un nuevo caso/objeto (normalmente desconocida y/o difícil de medir para ciertos problemas). El manejo y comprensión de éstas herramientas de la IA es de especial interés en el área de la biología, la química, la medicina, etc...

**Tipología:** Trabajos experimentales, de toma de datos de campo o de laboratorio.

**Objetivos planteados:**

El objetivo del presente TFG es familiarizar al alumno/a con dicho tipo de técnicas mediante su aplicación para resolver algunos problemas relacionados con la biología, ya sean de clasificación, regresión, asociación y/o clustering. Para ello se propone el siguiente plan de trabajo:

- El alumno/a debe estudiar y comprender el funcionamiento de los algoritmos básicos más conocidos que sean de aplicación a los datos biológicos con los que se trabaje.
- El alumno/a debe familiarizarse con herramientas como KEEL, Weka, Knime, librerías de R y/o Python, etc., para aplicar los algoritmos existentes en los conjuntos de datos biológicos anteriormente mencionados: Preparación de los datos (excel, etc.), carga de datos, ejecución, posibles comparativas entre algoritmos, etc.
- El alumno/a debe aplicar los algoritmos más prometedores para resolver uno o varios problemas del mundo real.
- El alumno/a debe analizar los modelos obtenidos para ver si se puede extraer conocimiento útil y/o tener cierta validación o justificación biológica.

**Bibliografía básica:**

Greener, J. G., Kandathil, S. M., Moffat, L., & Jones, D. T. (2022). A guide to machine learning for biologists. *Nature Reviews Molecular Cell Biology*, 23(1), 40-55. <https://doi.org/10.1038/s41580-021-00407-0>

J. Alcalá-Fdez, L. Sánchez, S. García, M.J. del Jesus, S. Ventura, J.M. Garrell, J. Otero, C. Romero, J. Bacardit, V.M. Rivas, J.C. Fernández, F. Herrera. KEEL: A Software Tool to Assess Evolutionary Algorithms to Data Mining Problems. *Soft Computing* 13:3 (2009) 307-318, doi: 10.1007/s00500-008-0323-y.

J. Alcalá-Fdez, A. Fernandez, J. Luengo, J. Derrac, S. García, L. Sánchez, F. Herrera. KEEL Data-Mining Software Tool: Data Set Repository, Integration of Algorithms and Experimental Analysis Framework. *Journal of Multiple-Valued Logic and Soft Computing* 17:2-3 (2011) 255-287.

Raschka, S., & Mirjalili, V. (2019). *Python machine learning: Machine learning and deep learning with Python, scikit-learn, and TensorFlow 2*. Packt publishing Ltd.

**Recomendaciones y orientaciones para el estudiante:**

Es recomendable tener cierto conocimiento (básico) sobre el manejo de ficheros de datos (uso de Excel, etc.)

Es recomendable tener cierto conocimiento (básico al menos) sobre programación. Por ejemplo, saber algo de programación en Python o cualquier otro lenguaje (aunque sea a nivel principiante).

**Plazas:** 1

## **2. DATOS DEL TUTOR/A:**

**Nombre y apellidos:** RAFAEL ALCALÁ FERNÁNDEZ

**Ámbito de conocimiento/Departamento:** CIENCIAS DE LA COMPUTACIÓN E INTELIGENCIA ARTIFICIAL

**Correo electrónico:** rcalca@ugr.es

## **3. COTUTOR/A DE LA UGR (en su caso):**

**Nombre y apellidos:**

**Ámbito de conocimiento/Departamento:**

**Correo electrónico:**

## **4. COTUTOR/A EXTERNO/A (en su caso):**

**Nombre y apellidos:**

**Correo electrónico:**

**Nombre de la empresa o institución:**

**Dirección postal:**

**Puesto del tutor en la empresa o institución:**

**Centro de convenio Externo:**

## **5. DATOS DEL ESTUDIANTE:**

**Nombre y apellidos:**

**Correo electrónico:**