

# Regresión

Fabián Villena

Enero 2024

Se le pide seleccionar el mejor modelo a la hora de predecir el índice de masa corporal de un paciente dados sus hábitos y condición física<sup>1</sup>.

En particular, trabajaremos con un conjunto de datos que contiene información sobre hábitos alimenticios y condición física recopilados a través de una encuesta electrónica en países latinoamericanos. Usted deberá predecir sobre el atributo BMI (*Body Mass Index*)

Los resultados esperados son que puedan seleccionar la mejor combinación de hiperparámetros y modelos utilizando técnicas de validación de modelos.

<https://github.com/fvillena/biocompu/blob/2023/data/obesity.csv>

## Preguntas

Responda las siguientes preguntas en un *Jupyter Notebook* con código desarrollado en el lenguaje de programación Python.

1. Verifique y preprocese si es necesario las características.
2. Separe el conjunto de datos en un subconjunto de entrenamiento y prueba.
3. Genere una grilla de hiperparámetros para el algoritmo *Random Forests*.
4. Optimice los hiperparámetros del algoritmo *Random Forests* sobre el subconjunto de entrenamiento utilizando la técnica de *Grid-Search* y *Cross-Validation* para validar los resultados.
5. Prediga sobre el subconjunto de prueba utilizando un modelo ajustado con los mejores hiperparámetros.
6. Calcule al menos dos métricas de rendimiento para las predicciones realizadas anteriormente.

---

<sup>1</sup><https://doi.org/10.1016/j.dib.2019.104344>