

## **Ejercicio de clase**

### **KNN y Random-Forest**

#### **Parte 1**

Se tiene un set de datos de información médica de pacientes, para intentar predecir a partir de esta información, el padecimiento de diabetes. De este modo la etiqueta corresponde a tener o no diabetes. (1 para sí, 0 para no).

Se desea encontrar un modelo que permita clasificar lo mejor posible pacientes diabéticos de no diabéticos con la información que se cuenta.

1. Realice una clasificación usando KNN. Para esto deberá determinar el mejor K con ayuda de una curva de validación.
2. Evalúe el desempeño de este modelo. ¿Podría usarse como generador de características? De ser así, genere una nueva característica.
3. Intente un árbol de clasificación. ¿Cómo es su desempeño?
4. Ahora intente un modelo *Random Forest*. ¿Cómo es su desempeño? Determine las 4 características más importantes, y ajuste un nuevo modelo solo con estas características, ¿mejora el desempeño? Proponga una alternativa para mejorar el desempeño del modelo aún más en caso de que no sea satisfactorio.

#### **Parte 2.**

Intente un modelo Random Forest de regresión para encontrar un modelo que describa los counts en términos del ángulo de difracción, usando los datos de DRXCruudo.xlsx Grafique para mostrar el desempeño del modelo.

#### **Revisar la documentación de las funciones**

Documentación [RandomForestRegressor](#)

Documentación [RandomForestClassifier](#)