

Tarea 3

Introducción

La empresa Irlandesa **Dole Food Company** busca desarrollar una máquina capaz de distinguir entre plátanos de buena calidad (*Good Quality*) y aquellos de mala calidad (*Bad Quality*). Se ha encargado a un equipo del departamento de robótica de la empresa la creación de esta máquina, y han contratado sus servicios para desarrollar un modelo de Minería de Datos que permita llevar a cabo esta distinción.

Datos

El dataset se puede encontrar en WebCursos y contiene los siguientes atributos:

- **Size**: tamaño de la fruta
- **Weight**: peso de la fruta
- **Sweetness**: dulzura de la fruta
- **Softness**: suavidad de la fruta
- **HarvestTime**: cantidad de tiempo transcurrido desde la cosecha de la fruta
- **Ripeness**: madurez de la fruta
- **Acidity**: acidez de la fruta
- **Quality**: calidad de la fruta (Buena o Mala)

Entregables

- En Webcursos, usted debe subir el código Python utilizado, en formato **Jupyter Notebook (ipynb)**.
- El nombre del Archivo debe ser **Tarea3_iniciales1_iniciales2.ipynb**
 - Ejemplo, el Notebook de Alfonso Tobar y Sofía Álvarez se debe entregar como **Tarea3_AT_SA.ipynb**.

Código

El código Python escrito en Jupyter Notebook debe:

1. Abrir el archivo (.csv). Además se debe incluir un EDA básico que muestre los rasgos principales del dataset.
2. Utilizar una estrategia Holdout separando los datos en conjuntos de **train** (70%) y **test** (30%) usando la función [train_test_split\(\)](#) de Scikit-Learn.

3. Normalizar las columnas numéricas usando [MinMaxScaler\(\)](#). Tenga cuidado con realizar el procedimiento de manera correcta para no generar Data Leakage.
4. Utilice como modelo Baseline un clasificador que prediga la calidad (*Quality*), usando el algoritmo KNN con $k=1$. Calcule las predicciones dadas por el modelo y el **Accuracy** comparando etiquetas reales y predichas. Explique si la métrica es aceptable y comente sobre el ajuste del modelo.
5. Volver a realizar el punto 5, pero esta vez usted elija el mejor K. Debe elegir el mejor K en función de tres métricas de evaluación (*Accuracy*, Precisión y *Recall*) variando K entre 1 y 30. Grafique y argumente.
6. Volver al punto 2, y ahora utilice [StandardScaler\(\)](#). Luego, realice el paso 5 con esta nueva normalización. Indique si el mejor K cambia o no.

El Jupyter Notebook debe explicar:

- Todos las decisiones y métodos utilizados en el código **en detalle**.
 - o Si se entrega un código sin su explicación, entonces será evaluado con nota 1.
 - o Si se entrega código con código explicado parcialmente, solo las partes explicadas serán evaluadas.