

# Árbol de Decisión

29 de marzo de 2025

## Introducción

Un Árbol de Decisión es un algoritmo supervisado de aprendizaje automático utilizado para clasificación y regresión. Su estructura en forma de árbol permite tomar decisiones lógicas de forma intuitiva y visual. Son muy útiles por su facilidad de interpretación y buena precisión en tareas con datos categóricos y numéricos.

## Metodología

Para esta actividad se utilizó el conjunto de datos `artists_billboard_fix3.csv`, el cual contiene información sobre artistas musicales y su aparición en el ranking Billboard. Los pasos realizados fueron:

1. Carga y exploración del conjunto de datos.
2. Visualización de variables categóricas como tipo de artista, género y estado de ánimo.
3. Cálculo de la edad de los artistas al momento de entrar al ranking, manejando valores nulos.
4. Entrenamiento de un Árbol de Decisión usando `scikit-learn`, validando con 10-fold cross-validation.

## Fragmento de código clave

```
from sklearn import tree
from sklearn.model_selection import KFold, cross_val_score

X = artists_billboard[['durationSeg', 'anioNacimiento', '
    edad_en_billboard']]
Y = artists_billboard['top']

kf = KFold(n_splits=10, shuffle=True, random_state=1)
modelo = tree.DecisionTreeClassifier(criterion='entropy',
    max_depth=5)
scores = cross_val_score(modelo, X, Y, cv=kf)
print("Precisi n promedio (cross-validation):", round(scores
    .mean(), 4))
```

## Resultados

Tras ejecutar el modelo de Árbol de Decisión con validación cruzada, se obtuvieron los siguientes resultados:

- Edad promedio de los artistas: **30.1 años**
- Desvío estándar de edad: **8.4 años**
- Precisión promedio del modelo: **86.14 %**

Se generaron diversas gráficas de distribución de datos y una visualización final del árbol entrenado.

## Conclusión

Esta actividad permitió comprender cómo se construye y evalúa un Árbol de Decisión en Python. El modelo logró una buena precisión (86.14 %) utilizando características como duración de la canción, año de nacimiento y edad del artista. Además, se reforzaron habilidades en limpieza de datos, visualización y validación de modelos.