

# **Modelos Cart**

**Universidad Nacional de Colombia**

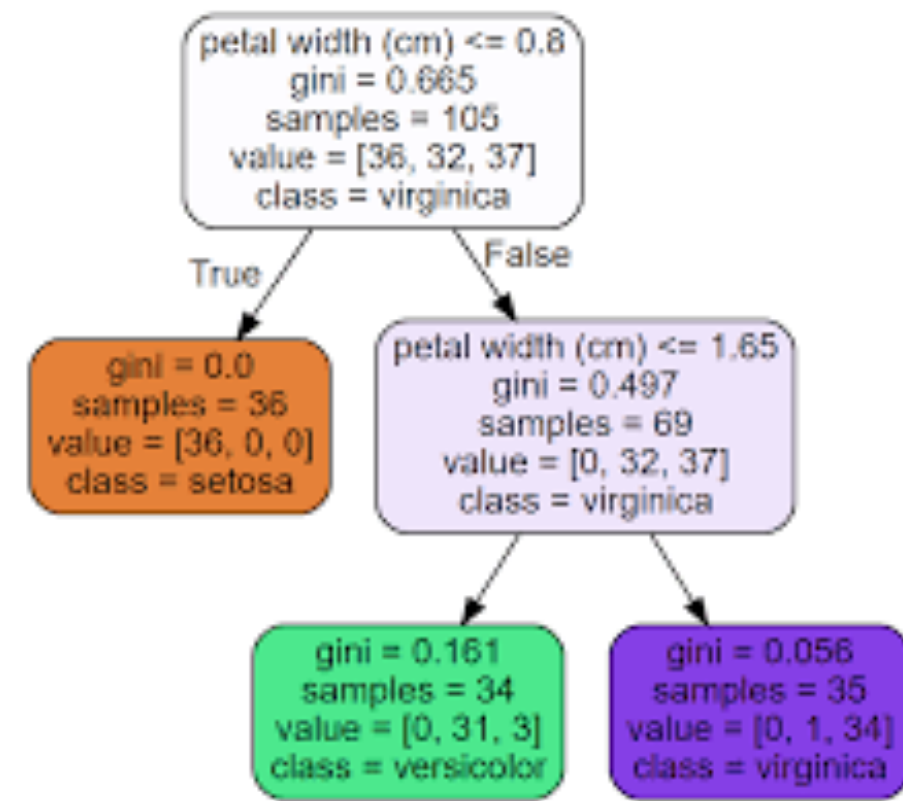
**Prof. Carlos Daniel Jiménez**

# Introducción a los Modelos CART

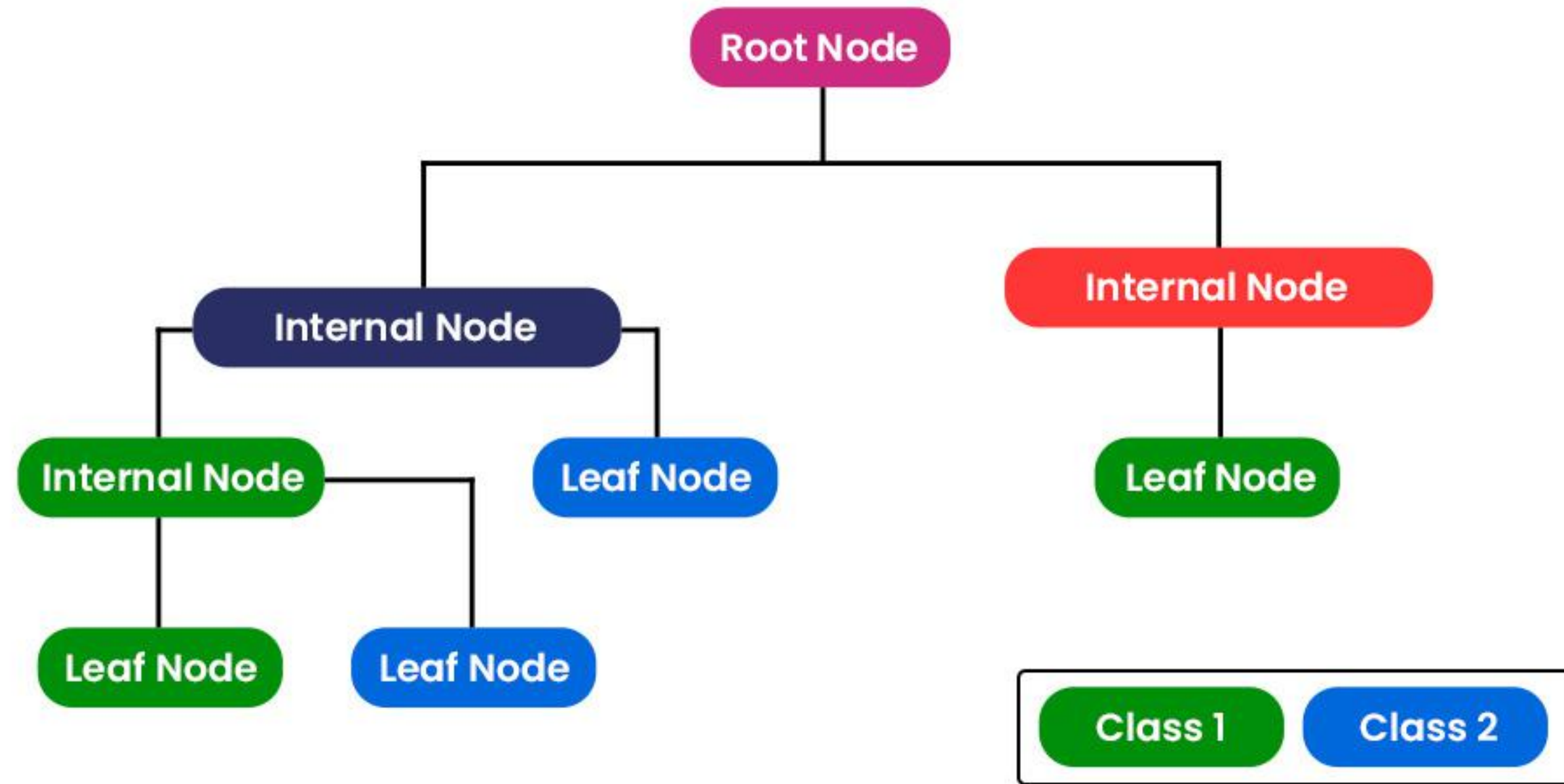
- **Definición:** CART (Classification and Regression Trees) es una técnica de aprendizaje automático utilizada tanto para tareas de clasificación como de regresión.
- **Propósito:** Se utiliza para construir modelos predictivos basados en la partición del espacio de características en regiones homogéneas.
- **Componentes:** Nodos, ramas y hojas.
- **Aplicación Ambiental:** Usado para predecir la calidad del aire, clasificación de tipos de suelo y análisis de deforestación.

# Estructura de un Árbol CART

- **Nodos Internos:** Representan decisiones basadas en características de entrada.
- **Ramas:** Conectan nodos y representan posibles resultados de decisiones.
- **Hojas:** Representan las predicciones finales (clases en clasificación, valores en regresión).
- **Raíz del Árbol:** Nodo inicial que contiene todo el conjunto de datos.
- **Aplicación Ambiental:** Clasificación de áreas de riesgo ambiental basadas en características geográficas y meteorológicas.



Iris



Funcionalidad interna

# Construcción de Árboles CART

- Proceso de Construcción:
  - **Selección de Característica:** Elegir la característica que mejor divide los datos.
  - **Criterio de División:** Usar métricas como Gini, Entropía (para clasificación) o MSE (para regresión).
  - **División Recursiva:** Repetir el proceso de división en los subconjuntos resultantes.
  - **Condición de Parada:** Determinar cuándo detener la división (profundidad del árbol, número mínimo de muestras en una hoja, etc.).

# Construcción de Árboles CART

- **Métricas de Clasificación**

- **Precisión:** Proporción de verdaderos positivos entre los positivos predichos.
- **Recall (Sensibilidad):** Proporción de verdaderos positivos entre todos los casos reales positivos.
- **F1-Score:** Media armónica de precisión y recall.

**Aplicación Ambiental:** Predicción de la calidad del aire basada en características como emisiones de CO<sub>2</sub>, temperatura y humedad.

# Ejemplos de Modelos CART

- **Árbol de Clasificación:**
  - **Clasificación de la Calidad del Aire:** Predecir si la calidad del aire es buena, moderada o mala.
  - **Detección de Tipos de Suelo:** Clasificar diferentes tipos de suelo basados en sus características químicas y físicas.

# Estructura de un Árbol CART

- **Métricas de Evaluación**

- **Error Cuadrático Medio (MSE):** Promedio de los errores al cuadrado. Penaliza más los errores grandes.
- **Raíz del Error Cuadrático Medio (RMSE):** Raíz cuadrada del MSE. Facilita la interpretación ya que está en la misma escala que los valores de la variable objetivo.
- **Error Absoluto Medio (MAE):** Promedio de los errores absolutos. Menos sensible a los valores atípicos.
- **Coeficiente de Determinación ( $R^2$ ):** Proporción de la varianza en la variable dependiente que es predecible a partir de las variables independientes.

Aplicación en temas ambientales : Modelado de la Contaminación: Medir el rendimiento de modelos que predicen niveles de contaminación basados en factores industriales y meteorológicos.



# Estructura de un Árbol CART

- **Árbol de Regresión:**
  - **Predicción de Temperaturas Futuras:** Estimar las temperaturas futuras basadas en datos históricos y condiciones actuales.
  - **Modelado de la Deforestación:** Predecir la cantidad de deforestación en una región específica basada en variables económicas y ambientales.

# Ventajas y Desventajas de los Árboles CART

Ventajas	Desventajas
Fácil de entender y visualizar.	Tienden a sobreajustar si no se poda adecuadamente.
Puede manejar tanto tareas de clasificación como de regresión.	Pequeños cambios en los datos pueden resultar en árboles diferentes.
Captura relaciones no lineales entre características	Pueden volverse ineficaces con muchas características y alta dimensionalidad.

# Algos de Cart

- **Decision Tree Classifier:** Algoritmo de clasificación que divide los datos en subconjuntos basados en el valor de las características.
  - **Usos Ambientales:** Clasificación de la calidad del aire, detección de áreas con riesgo de incendios forestales.
- **Decision Tree Regressor:** Algoritmo de regresión que predice valores continuos dividiendo los datos en subconjuntos homogéneos
  - **Usos Ambientales:** Predicción de la temperatura, estimación de niveles de contaminación.

# Algos Cart

- **Random Forest:** Conjunto de múltiples árboles de decisión que mejora la precisión y reduce el sobreajuste.
  - **Usos Ambientales:** Modelado de la biodiversidad, predicción de patrones de precipitación.
- **Gradient Boosting Trees:** Conjunto de árboles de decisión entrenados secuencialmente para corregir los errores de los árboles anteriores
  - **Usos Ambientales:** Predicción de eventos climáticos extremos, estimación de emisiones de gases de efecto invernadero.