

ÁRBOLES DE DECISIÓN: FINE, MEDIUM, COARSE

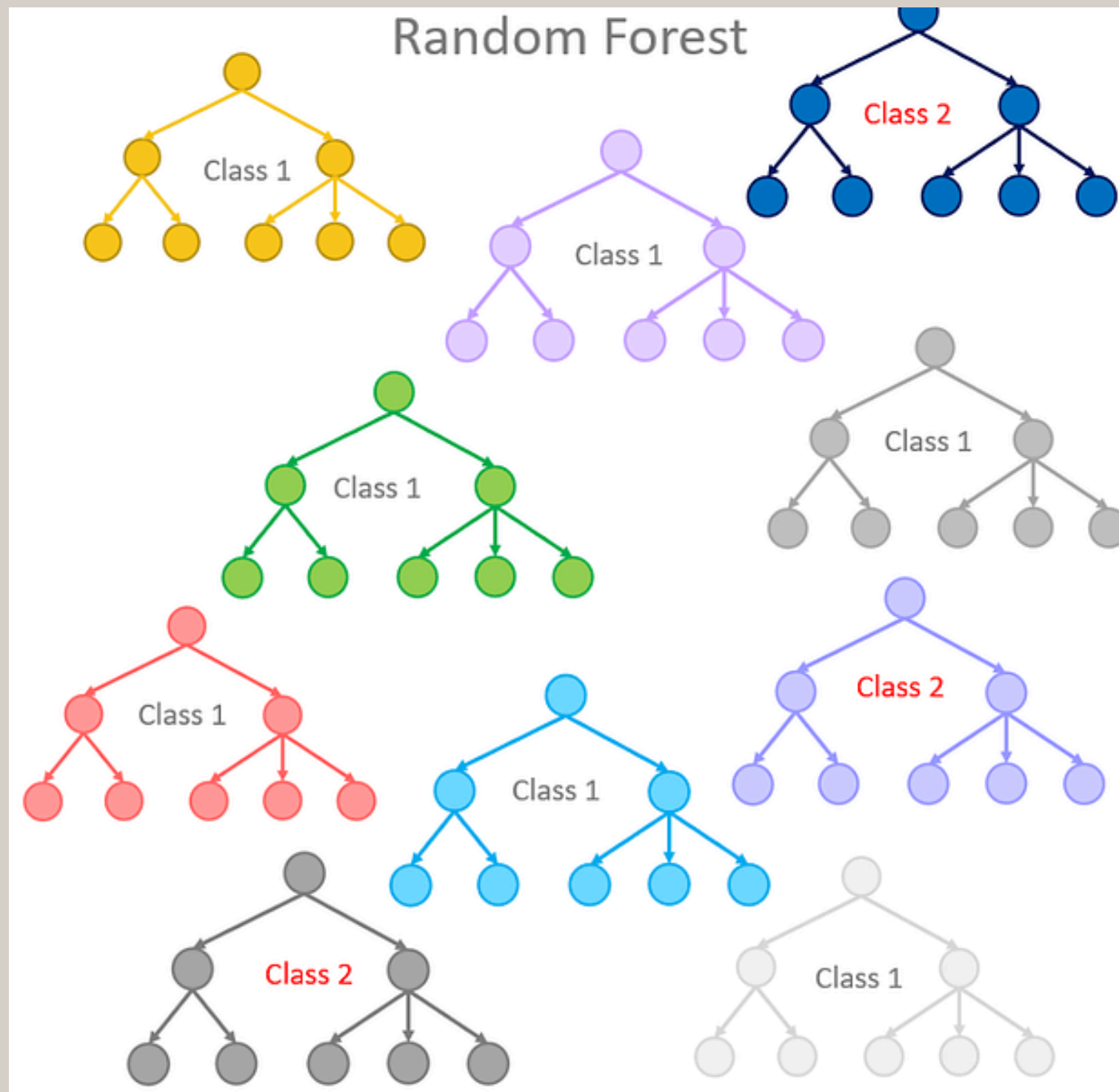
Procesamiento Digital de Imagenes II

Docente:
David Stephen Fernandez Mccann

Estudiantes: Noemi Alean Acuña
Pedro Andrés Viloria Colon

INTRODUCCIÓN A LOS ÁRBOLES DE DECISIÓN

Un árbol de decisión es una estructura de tipo gráfico que se asemeja a un árbol, donde cada nodo interno representa una pregunta o condición sobre los datos (en el caso de visión artificial, características de una imagen como color, textura, forma, etc.), las ramas corresponden a los resultados de esa pregunta, y las hojas finales representan las categorías o valores que se predicen.



Tomada de https://machinelearningparatodos.com/wp-content/uploads/2023/06/decision_tree.png

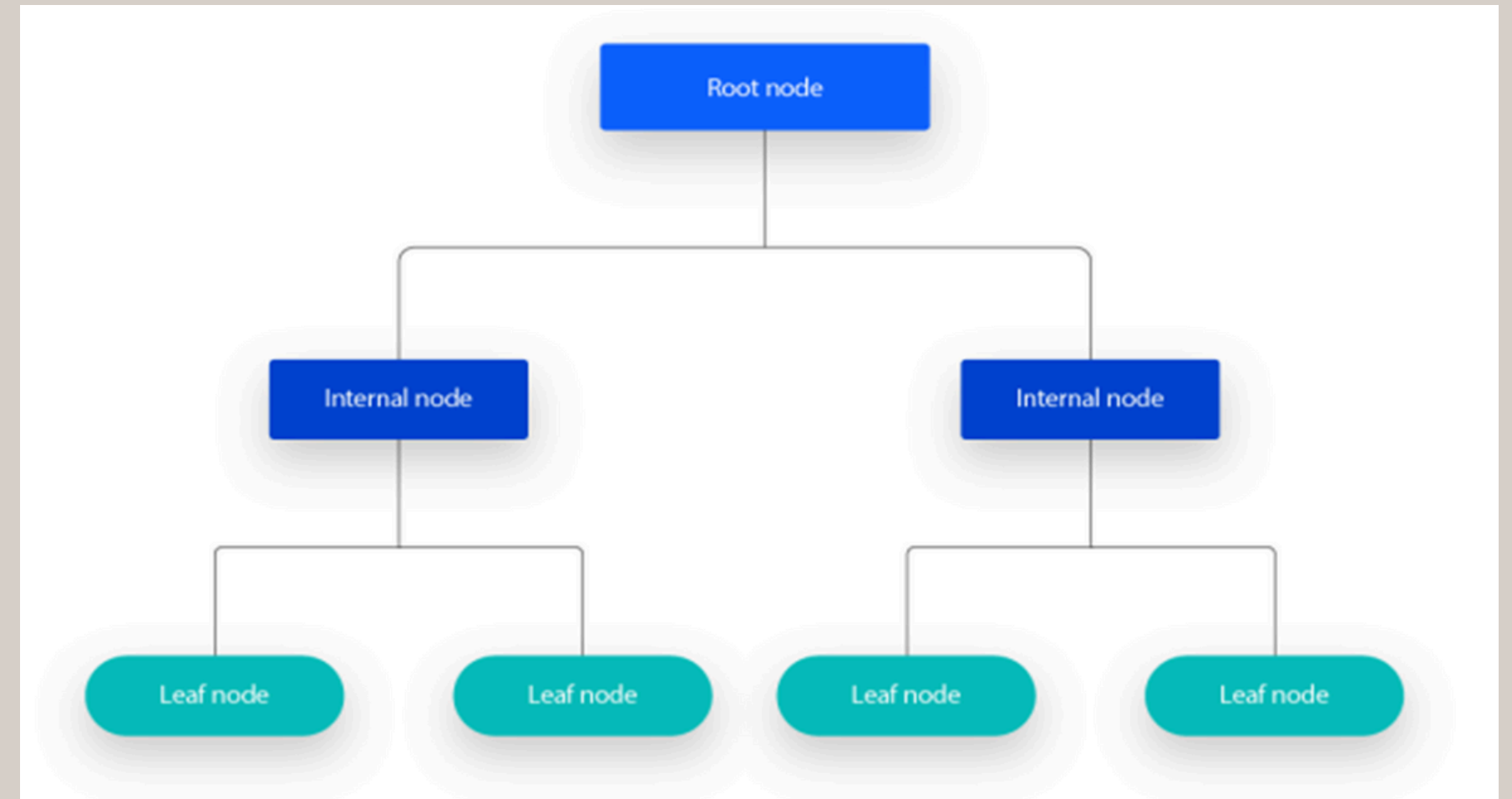
CONCEPTO CLAVE: DESCOMPOSICIÓN RECURSIVA

Los árboles de decisión emplean un proceso conocido como **descomposición recursiva**. Esto significa que, en cada nodo de decisión, se selecciona una característica para dividir el conjunto de datos en dos o más subgrupos. Este proceso continúa hasta que se cumplen ciertos criterios de parada (profundidad máxima, número mínimo de observaciones por hoja, etc.).

1. Raíz

2. Nodos Intermedios

3. Hojas



https://www.ibm.com/content/dam/connectedassets-adobe-cms/worldwide-content/cdp/cf/ul/g/df/de/Decision-Tree.component.complex-narrative-lts-l693403600692.png/content/adobe-cms/es/es/topics/decision-trees/jcr:content/root/table_of_contents/body/content_section_styled/content-section-body/complex_narrative/items/content_group_1423241468/image

LA MATEMÁTICA DETRÁS DE LOS ÁRBOLES DE DECISIÓN

El objetivo principal de un árbol de decisión es **minimizar la impureza** en los subgrupos generados por las decisiones. Hay varias métricas que se usan para medir esta impureza

$$\text{Entropía} = - \sum_{i=1}^n p_i \log_2(p_i)$$

$$\text{Ganancia de Información} = I(p) - \frac{N_L}{N} I(p_L) - \frac{N_R}{N} I(p_R)$$

$$\text{Índice Gini} = 1 - \sum_{i=1}^n p_i^2$$

TIPOS DE ÁRBOLES DE DECISIÓN:

Dependiendo de la profundidad del árbol y el número de divisiones en cada nodo, se pueden clasificar los árboles de decisión en tres tipos principales

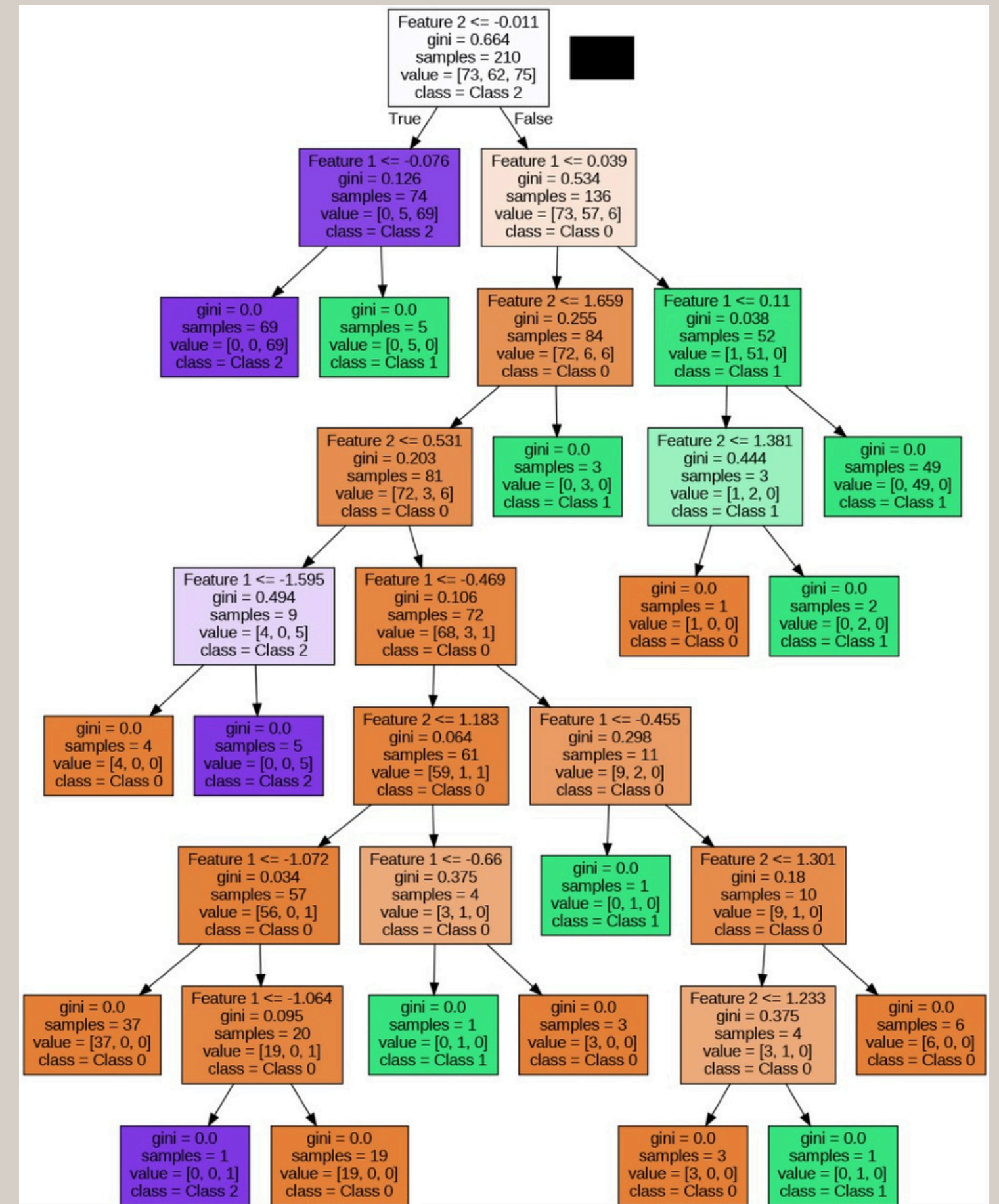
Fine Tree (Árbol Fino)

Medium Tree (Árbol Medio)

Coarse Tree (Árbol Grueso)

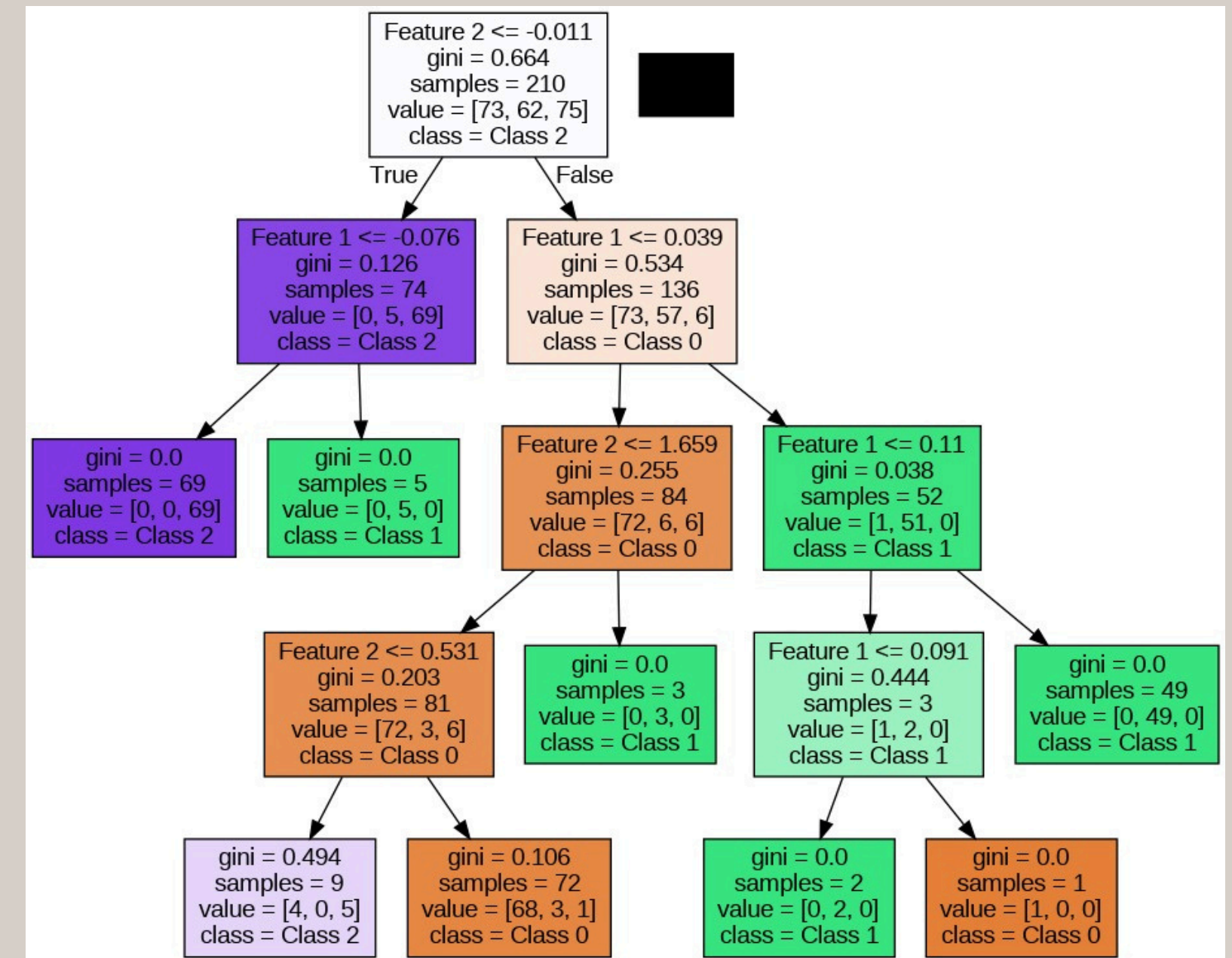
FINE TREE (ÁRBOL FINO)

Árbol con muchas divisiones y profundidad alta. Esto significa que cada nodo está más detallado y genera subgrupos muy pequeños.



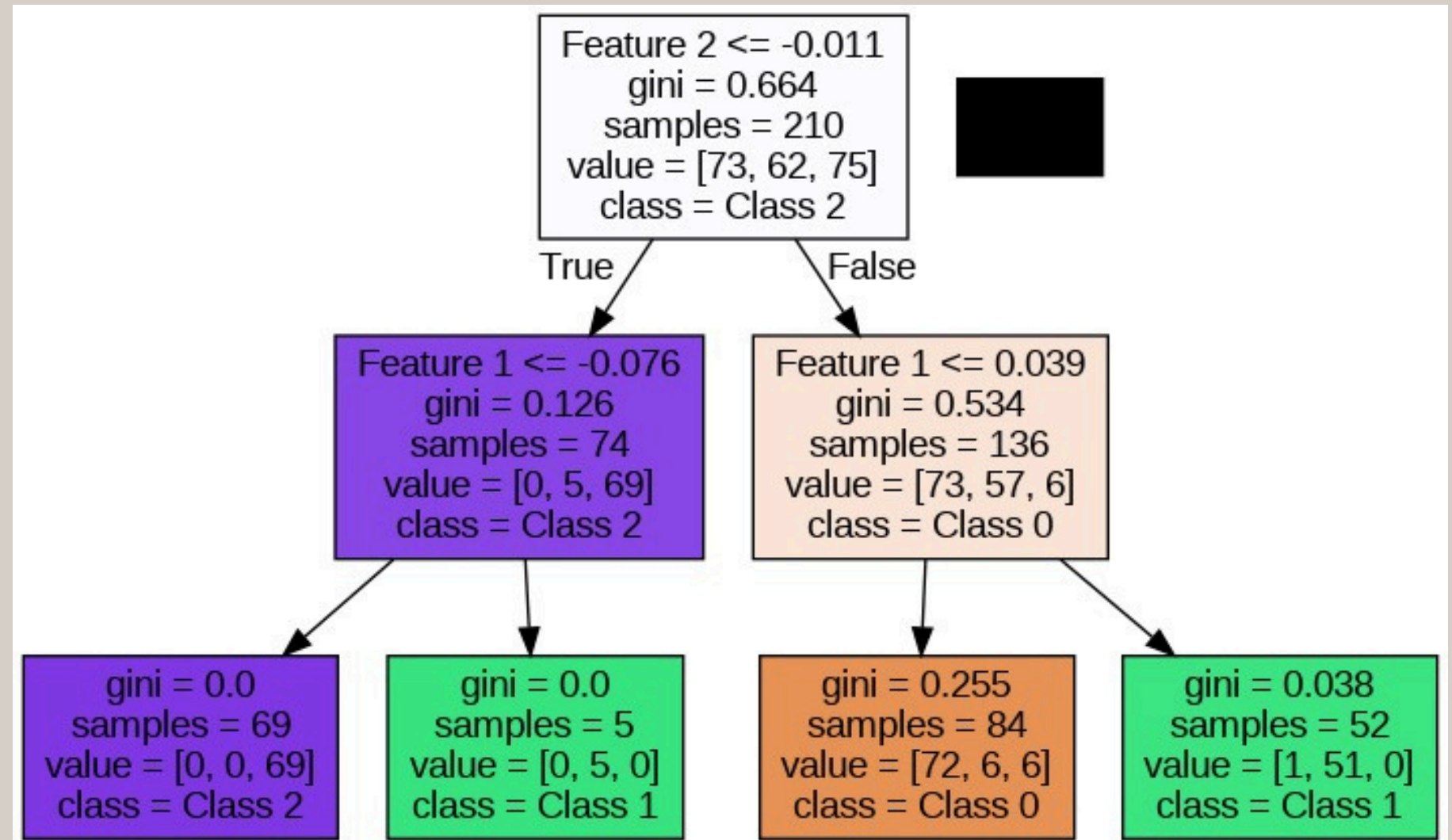
MEDIUM TREE (ÁRBOL MEDIO)

Un término medio entre complejidad y simplicidad. Mantiene un buen balance entre precisión y capacidad de generalización. Tiene menos divisiones que un Fine Tree, pero aun así es capaz de modelar adecuadamente relaciones importantes entre las variables.



COARSE TREE (ÁRBOL GRUESO)

Árbol con pocas divisiones, lo que resulta en un modelo simple y más fácil de interpretar. Debido a la menor complejidad, este tipo de árbol corre menos riesgo de sobreajuste.



COMPARACIÓN DE FINE, MEDIUM, COARSE EN MACHINE LEARNING

Tipo de árbol	Ventajas	Desventajas
Fine Tree	Alta precisión, captura patrones complejos	Mayor riesgo de sobreajuste
Medium Tree	Buen equilibrio entre precisión y simplicidad	Puede no ser tan preciso en problemas complejos
Coarse Tree	Rápido y fácil de interpretar	Menor precisión, más propenso a subajuste

ALGORITMOS POPULARES EN MACHINE LEARNING PARA ÁRBOLES DE DECISIÓN

En Machine Learning, los árboles de decisión se construyen a partir de diferentes algoritmos, que varían en la manera en que seleccionan las divisiones:

•**ID3 (Iterative Dichotomiser 3)**: Utiliza la **ganancia de información** y la **entropía** como métricas para seleccionar las características más importantes para dividir los datos. Este algoritmo fue desarrollado por Ross Quinlan.

•**C4.5**: Es una evolución de ID3 que también emplea la **ganancia de información**, pero con mejoras para manejar datos continuos y la **razón de ganancia** para evitar problemas con muchas categorías.

•**CART (Classification And Regression Trees)**: Introducido por Leo Breiman, CART utiliza la **impureza de Gini** para seleccionar la mejor división en los nodos. CART puede manejar tanto problemas de clasificación como de regresión, lo que lo hace versátil en Machine Learning.

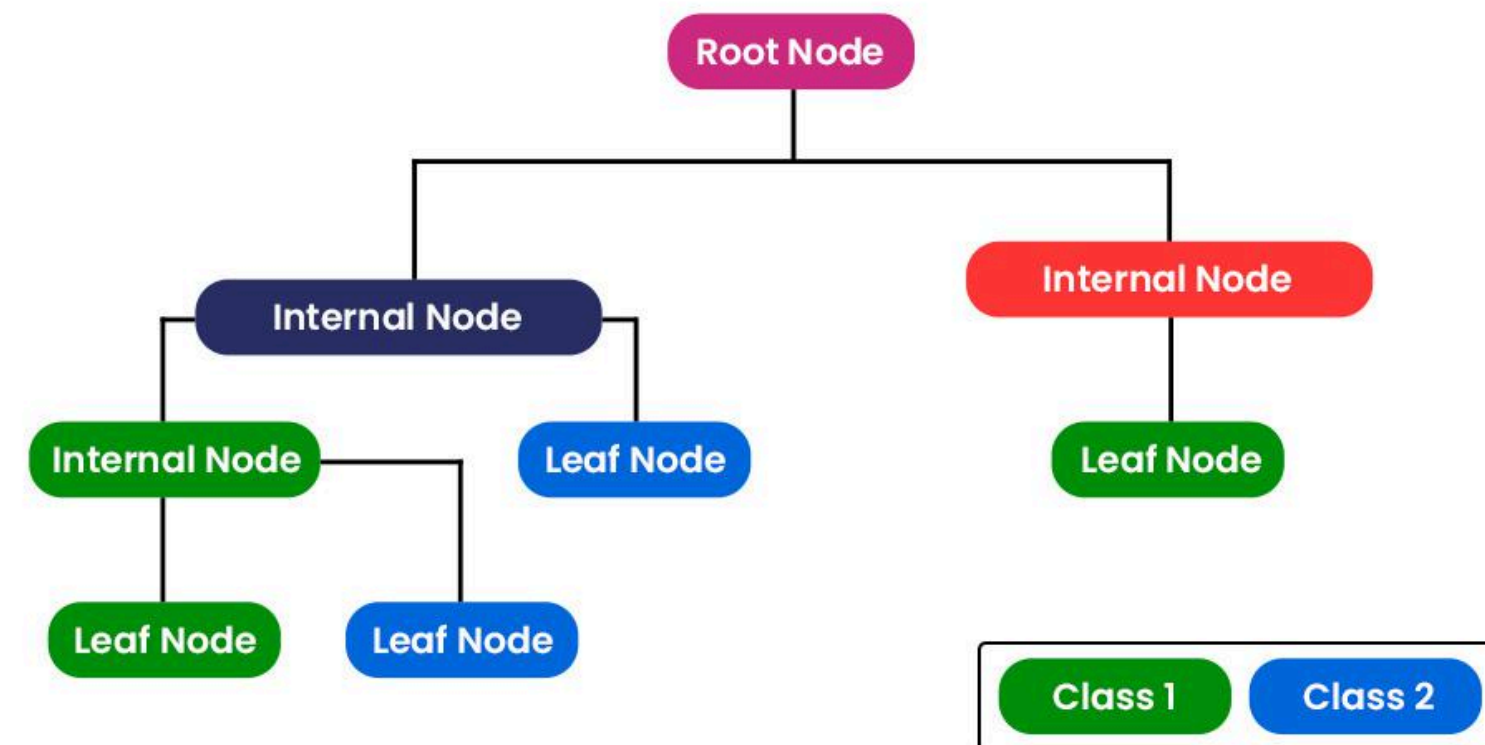
ALGORITMO CART

Selección de la Mejor División:

En cada nodo, se prueba cada característica y cada valor para encontrar la división que minimice la impureza.

Poda del Árbol: Después de que el árbol es completamente desarrollado, se puede podar, eliminando ramas que no proporcionen mejoras significativas.

Criterios de Parada: Se define una profundidad máxima o un número mínimo de muestras por hoja para evitar sobreajuste.



Tomada de <https://www.geeksforgeeks.org/cart-classification-and-regression-tree-in-machine-learning/>

APLICACIONES COMUNES

- **Fine Tree:** Se utiliza en problemas donde se necesita alta precisión, como la detección de fraudes o problemas médicos complejos, aunque con cuidado del sobreajuste.
- **Medium Tree:** Aplicable en áreas como marketing o análisis de riesgo, donde se requiere un balance entre precisión y velocidad.
- **Coarse Tree:** Se puede aplicar en sistemas embebidos o cuando se necesita un modelo rápido y sencillo, por ejemplo, en dispositivos móviles o IoT.



REFERENCIAS

1. <https://www.ibm.com/es-es/topics/decision-trees>
2. https://cienciadedatos.net/documentos/py07_arboles_decision_python
3. <https://medium.com/@alexdiandro/%C3%A1rboles-de-decisi%C3%B3n-960aebd7eeda>
4. <https://www.codificandobits.com/blog/clasificacion-arboles-decision-algoritmo-cart>
5. <https://www.mathworks.com/help/stats/train-decision-trees-in-classification-learner-app.html>



MUCHAS GRACIAS