Técnicas de Inteligencia artificial

Tema 3. Árboles de decisión I

Profesora: Magdalena Ruiz

unir LA UNIVERSIDAD EN INTERNET

00 – ¿De qué hablamos en nuestra clase anterior?

- Qué es la Inteligencia Artificial y sus aplicaciones
- Inteligencia Artificial, Minería de datos y Aprendizaje Automático
- Cómo aprenden los ordenadores
- Problema : Discernir entre aprendizaje supervisado y no supervisado

00 – Recordando

¿Qué es la Inteligencia artificial?

"Crea programas informáticos que ejecutan operaciones comparables a la las que realiza la mente humana como el aprendizaje o el razonamiento lógico" De esta definición se desprenden dos conceptos importantes:

- **Aprendizaje**: Capacidad de adquirir conocimientos y habilidades a partir de los datos.
- **Razonamiento lógico**: Capacidad de procesar información, seguir reglas y realizar inferencias lógicas que permitan tomar decisiones y resolver problemas.

istockphoto.com

00 – Recordando

¿Qué es la minería de datos?

Es un proceso que utiliza técnicas de inteligencia artificial sobre grandes cantidades de datos, con el objetivo de descubrir y describir patrones en los datos, a partir de los cuales se pueda obtener un beneficio. La técnica de Inteligencia artificial que utiliza es el aprendizaje automático.

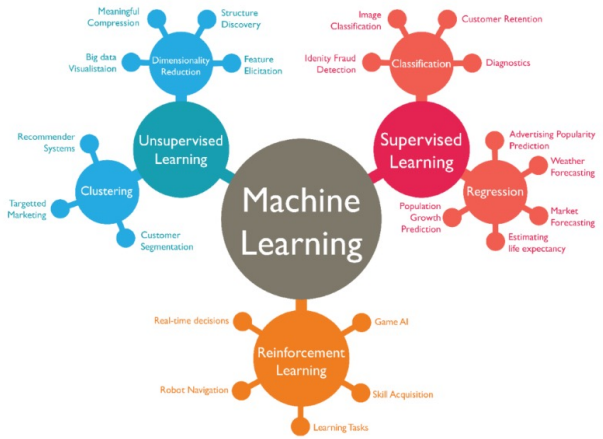
istockphoto.com

00 – Recordando

¿Qué es el aprendizaje automático?

Hace referencia a cómo aprenden las máquinas a través de la experiencia y los datos con el objetivo de generalizar comportamientos y encontrar patrones en los mismos. Existen dos grandes tipologías de aprendizaje automático:

- ✓ Aprendizaje supervisado.
- ✓ Aprendizaje no supervisado.

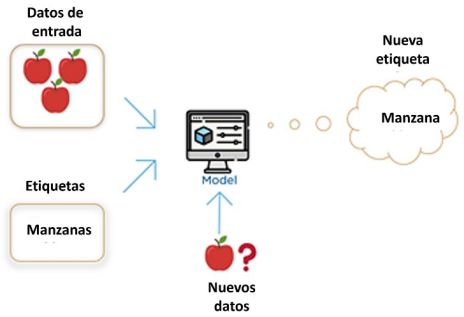


Tipos de aprendizaje automático. Fuente: Medium.com

00 – Recordando

Aprendizaje supervisado

- Se utiliza para **predecir** (o un valor numérico o una categoría).
- Disponemos de **ejemplos etiquetados** (aprendizaje guiado).
- Se aplica a dos tipos de problemas:
  - ✓ **Regresión:** L-2, K vecinos más cercanos, SVR..
  - ✓ **Clasificación:** Árboles de decisión, Random Forest, Naive Bayes.



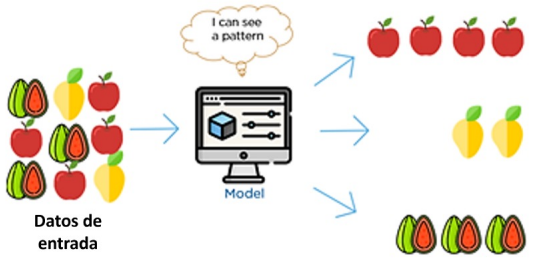
Aprendizaje supervisado. Fuente: keepcodding.com

00 – Recordando

Aprendizaje no supervisado

- Se utiliza para detectar **patrones** y **tendencias** en los datos.
- Para detectar esos patrones **no disponemos de ejemplos etiquetados.**
- Se aplica a tres tipos de problemas:

  ✓ **Agrupamiento (clustering):** K-means, DBSCAN…
  ✓ **Detección de anomalías:** LOF, Insolation Forest..
  ✓ **Reducción de la dimensión:** PCA, Análisis Factorial..



Aprendizaje no supervisado. Fuente: keepcodding.com

00 – Recordando

Conceptos clave del aprendizaje supervisado

- **Entrenamiento y test o** *hould-out.*
- **Sobreajuste** u overfiting.
- Técnica de **validación cruzada.**

Un concepto muy importan en aprendizaje supervisado (en aprendizaje automático en general) es la aleatoriedad

istockphoto.com

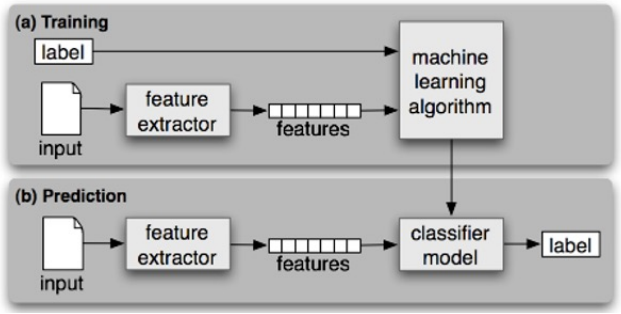
00 – Recordando

¿En qué consiste el proceso de entrenamiento y test (*hold out*)?

- Es el primer paso que damos a la hora de aplicar aprendizaje supervisado.
- Consiste en dividir los datos en dos grupos:

  ✓ Entrenamiento.
  ✓ Test o prueba.

Con los datos de entrenamiento entrenamos el algoritmo y con los de test lo probamos



Entrenamiento de un modelo de aprendizaje automático. Fuente: (nltk.org, 2023)

00 – Recordando

Paso 1. Entrenamiento

- Se toma aleatoriamente un % de características del dataset.
- Ese % suele ser de entre un 75% y un 80% de los datos.
- Se aplica el algoritmo sobre ese % de los datos.
- Se evalúa el rendimiento del algoritmo.

Paso 2. Test

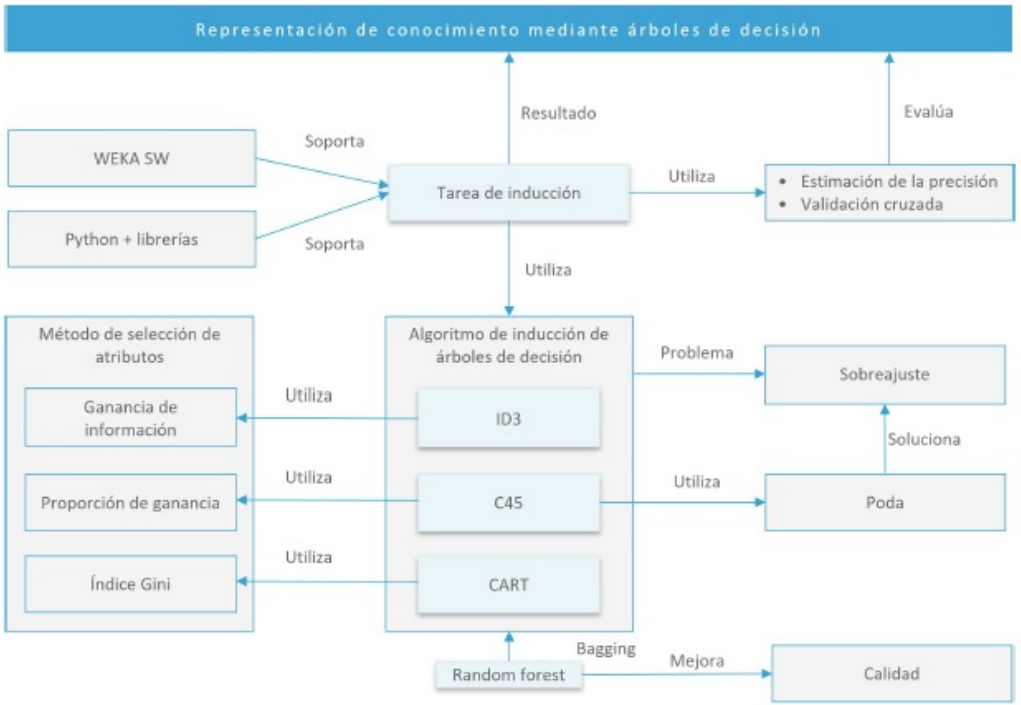
- Se toma el resto de datos del dataset .
- Si hemos tomado el 80% tomamos el otro 20% por ejemplo.
- Se aplica el modelo entrenado sobre esos datos.
- Se evalúa el rendimiento del algoritmo.

istockphoto.com

00 – ¿De qué vamos a hablar hoy?

- Representación del conocimiento mediante árboles de decisión
- Cómo aprenden los árboles de decisión
- El Algoritmo básico de árboles de decisión: ID3
- Problema : Resolución de un problema de clasificación con árboles de decisión

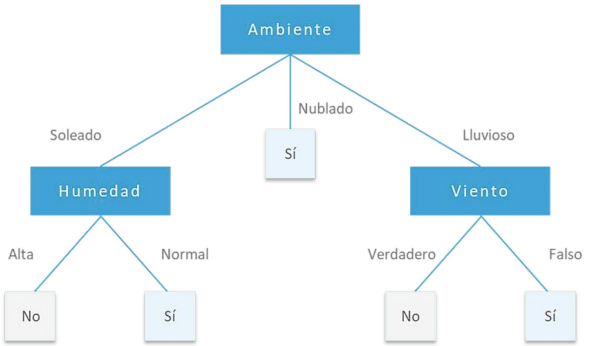
00 – Visión general del tema



01 – Representación del conocimiento mediante árboles

Aprendizaje de árboles de decisión

- Es una de las técnicas **de aprendizaje supervisado** más utilizadas.
- Permite clasificar instancias cuya clase sea desconocida.
- Se utiliza tanto para **regresión** como para **clasificación**.
- Funciona mejor con **variables categóricas**.



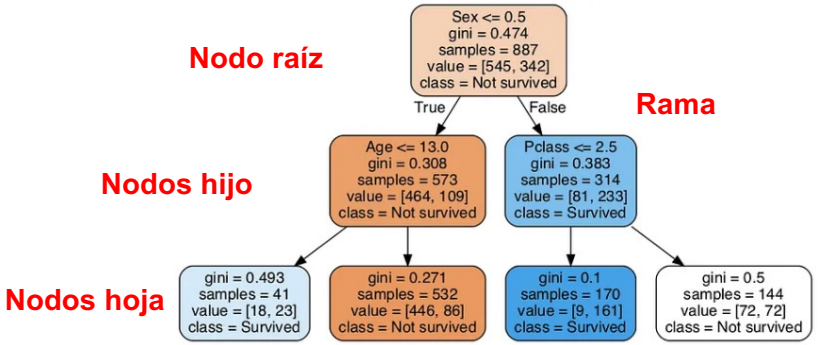
Ejemplo de árbol de decisión para el aprendizaje del concepto «Jugar al aire libre».

Se utilizan habitualmente para tareas de clasificación

01 – Representación del conocimiento mediante árboles

¿Cómo es un árbol de decisión?

- Están compuestos por **nodos** y **ramas**.
- Un nodo representa un conjunto o subconjunto de datos.
- Los datos de cada nodo han de ser lo más homogéneos posible (misma clase).
- Una rama representa la condición **True/False.**



Ejemplo de un árbol de decisión. Fuente: towardsdatascience.com

02 – ¿Cómo aprenden los árboles de decisión?

La tarea de inducción

- **Extraer conocimiento general** a partir de observaciones y experiencias particulares.
- Un elemento fundamental es la **selección de atributos.**
- El criterio para para seleccionar esos atributos dependerá de:
  - ✓ Si el atributo es discreto.
  - ✓ Si el atributo es numérico.

Si es un atributo discreto se crea una rama por cada valor conocido. Si es un atributo numérico se crea un umbral

Creación de ramas del árbol en función del tipo de atributo.

02 – ¿Cómo aprenden los árboles de decisión?

Elementos clave en los árboles de decisión

- **Selección de atributos** para generar las distintas ramas del árbol.
- Elegir el **método de selección** de atributos.
- Evitar el **sobreajuste** u *overfitting*.
- Manejo **prepoda** y **postpoda** de árboles.

A lo largo de este tema hablaremos de entropía, indice de gini, ganancia de Información, poda de árboles

02 – ¿Cómo aprenden los árboles de decisión?

La selección de atributos

- Busca hacer la mejor división (*Best Split*).
- Esta será la que separe los datos en subconjuntos lo más homogéneos posibles.
- Se utilizan las siguientes técnicas:

  ✓ **Entropía.**
  ✓ **Índice de Gini.**
  ✓ **Ganancia de información.**

$$Gini(E) = 1 - \sum_{i=1}^{n} p_i^2 \qquad (3)$$

Cálculo del índice de Gini . Fuente: UNIR 2025.

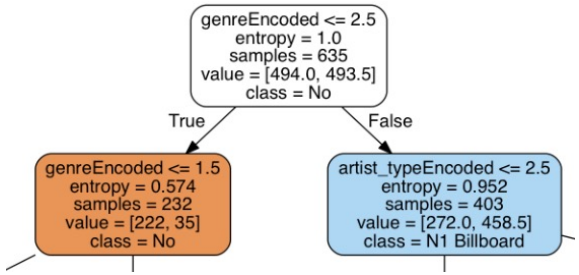
02 – ¿Cómo aprenden los árboles de decisión?

La entropía

- Mide el grado de **incertidumbre** o **impureza** de los datos.
- Se utiliza para medir la homogeneidad de los datos:

✓ Una entropía **cercana a 0** indica que la muestra es **homogénea**.
✓ Una entropía **cercana a 1** indica que la muestra es **heterogénea.**

Una entropía cercana a 0 implica que los datos de ese nodo "son de la misma clase" y una cercana a 1 que "son de distintas clases"



Entropia. Fuente:aprendemachinelearning.com/

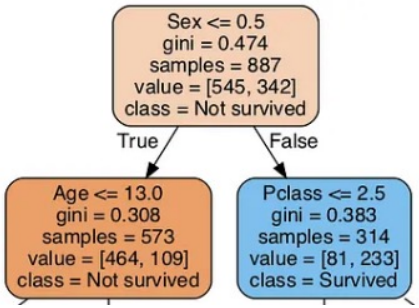
02 – ¿Cómo aprenden los árboles de decisión?

Gini index

- Mide la probabilidad de que una característica se clasifique incorrectamente.
- Se utiliza para medir la homogeneidad de los datos:

✓ Un Gini index **cercano a 0** indica que la muestra es **homogénea**.
✓ Un Gini index **cercano a 1** indica que la muestra es **heterogénea.**

En el índice de Gini al igual que la entropía es una medida de impureza del conjunto de datos



Ejemplo de un árbol de decisión. Fuente: towardsdatascience.com

02 – ¿Cómo aprenden los árboles de decisión?

Ganancia de información

- Mide la **impureza** de los datos (entropía) después de haber hecho la división.
- Mide cómo se reduce la entropía antes y después de dividir los datos.

✓ Una Ganancia de información **alta** indica que los datos son homogéneos
✓ Una Ganancia de información **baja** indica que los datos son heterogéneos.

La ganancia de Información mide cuanta Información proporciona una variable sobre la variable objetivo

$$InfoGain(F) = Entropy(S_1) - Entropy\ (S_2)$$

Fórmula de la Ganancia de Información. Fuente Unir 2024

02 – ¿Cómo aprenden los árboles de decisión?

¿Qué medida elegir?

- Gini index trabaja mejor con distribuciones de datos grandes.
- Gini index es más eficiente computacionalmente.
- Gini index tiende a favorecer las divisiones que aíslan la clase más frecuente.
- Ganancia de información crea nodos “más puros”.
- Ganancia de información es más costosa computacionalmente.

Estas medidas vienen incorporadas en las implementaciones de los algoritmos

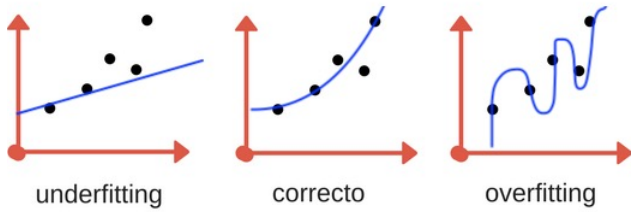
istockphoto.com

02 – ¿Cómo aprenden los árboles de decisión?

Sobreajuste u overfitting

- El modelo se ajusta demasiado a los datos de entrenamiento.
- Técnicas para evitar este sobreajuste:

  ✓ Reducción de la **dimensionalidad**.
  ✓ Utilizar **algoritmos más simples**.
  ✓ Utilizar la técnica de **validación cruzada**.



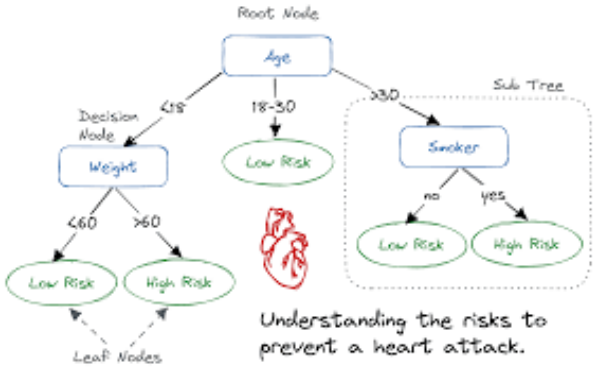
Ejemplo de overfitting. Fuente: aprendemachinelearning.com

02 – ¿Cómo aprenden los árboles de decisión?

Pre-poda

- Se aplica durante la construcción del árbol.
- Consiste en establecer criterios para limitar la expansión del árbol.
- Estos límites se pueden establecer configurando **hiperparámetros**:

  - ✓ Min_samples_leaf.
  - ✓ Min_samples_Split.
  - ✓ Max_features.
  - ✓ Max_Depth.



02 – ¿Cómo aprenden los árboles de decisión?

¿Qué hiperparámetros podemos ajustar en un árbol de decisión?

- **Min_samples_leaf**: Nº mínimo de observaciones de un nodo hoja (terminal).
- **Min_samples_split**: Nº mínimo de observaciones para dividir un nodo.
- **Max_features**: Número máximo de variables para dividir un nodo.
- **Max_Depth**: Número máximo de niveles del árbol (profundidad).

Los hiperparámetros qué más se suelen configurar son Max_Depth y Min_samples_split

istockphoto.com

02 – ¿Cómo aprenden los árboles de decisión?

Post-Poda

- Se aplica **después de crear el árbol.**
- Consiste en eliminar nodos.
- Para elegir el mejor subárbol a podar se utiliza la validación cruzada.
- Existen diferentes técnicas:

  ✓ Poda de costo-complejidad (CCP).
  ✓ Poda de error reducido (REP).
  ✓ Poda basada en el error mínimo (MEP).

El algoritmo CART (Classification and Regression Trees) utiliza CCP

```
[ ]  from sklearn.tree import DecisionTreeClassifier

     # Creamos un árbol de decisión sin poda
     clf = DecisionTreeClassifier()

     # Entrenamos el árbol
     clf.fit(X, y)

     # Podamos el árbol utilizando la regla de Cost Complexity
     ccp_alpha = 0.01
     clf_pruned = DecisionTreeClassifier(ccp_alpha=ccp_alpha)
     clf_pruned.fit(X, y)

     # Visualizamos el árbol original
     plot_tree(clf)

     # Visualizamos el árbol podado
     plot_tree(clf_pruned)

     # Compararamos la precisión de los árboles
     print("Precisión del árbol original:", clf.score(X_test, y_test))
     print("Precisión del árbol podado:", clf_pruned.score(X_test, y_test))
```

Ejemplo de post-poda python.Fuente: Elaboración propia

03 – Algoritmo básico ID3

Algoritmo ID3

- Construye los árboles top-down.
- Utiliza la ganancia de información como método de selección de atributos.
- El atributo mejor clasificador se convierte en el nodo raíz.

Hoy vamos a ver cómo funciona un árbol de decision con un ejemplo sencillo: Un problema de clasificación

03 –Algoritmo básico ID3

Implementaciones ID3 y otros algoritmos de árboles de decisión

- ID3, *Quinlan, J. R. 1986. Induction of Decision Trees. Mach.Learn. 1, 1 (Mar. 1986), 81 106*
- C4.5, *Quinlan, J. R. C4.5: Programs for Machine Learning .Morgan Kaufmann Publishers, 1993.*
- CART *Breiman,Friedman,Olshen,Stone : Classification andDecision Trees, Wadsworth, 1984*

Implementaciones en librerías: Scikit learn: https: //scikit-learn.org/stable/modules/tree.html y Tensor Flow: https://www.tensorflow.org/decision_forests/api_docs/python/tfdf/keras/CartModel

04 – Ejemplo del funcionamiento de un árbol de decisión

Vamos a verlo con un ejemplo sencillo

Para ver cómo aprende un árbol de decisión lo haremos en el contexto de un problema, concretamente un **problema de clasificación binaria**: Predecir si una persona sobrevive o no a la tragedia del Titanic.

En base a las características de los pasajeros como edad, sexo, el camarote en el que se alojaban o su clase social podremos predecir la posibilidad de sobrevivir al hundimiento. Para hacerlo disponemos de un conjunto de datos con 887 instancias y 12 variables:

- **PassengerId** = identificador único de cada pasajero
- **Name** = nombre del pasajero
- **Sex** = factor, con niveles (masculino y femenino)
- **Age** = edad del pasajero
- **Pclass** = clase eb la que viajaba el pasajero embarked = lugar en el que embarcó el pasajero
- **Ticket** = número de ticket del pasajero (na para la tripulación)
- **Fare** = precio del ticket (na para la tripulación, musicos, empleados y otros)
- **SibSp** = número de hermanos/familiares
- **Cabin** = cabina que ocupa cada pasajero
- **Parch** = número de padres e hijos a bordo
- **Survived** = especifica si el pasajero sobrevivió al hundimiento

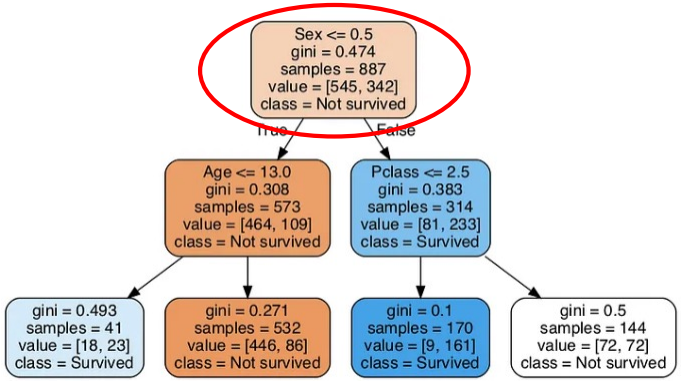
istockphoto.com

04 – Ejemplo del funcionamiento de un árbol de decisión

¿Qué representa el nodo raíz?

- **Sex:** Es la variable que mejor separa los datos y los separa en hombre-mujer.
- **Gini:** Grado de pureza del conjunto de datos (0,474)
- **Samples:** El número de observaciones totales del nodo (887)
- **Value:** Número de instancias de cada clase (**NO** sobrevivieron 545, **SI** 342)
- **Class:** Clase con más ejemplos.

El nodo raiz representa la pregunta ¿Ha sobrevivido al naufragio o no ha sobrevivido?



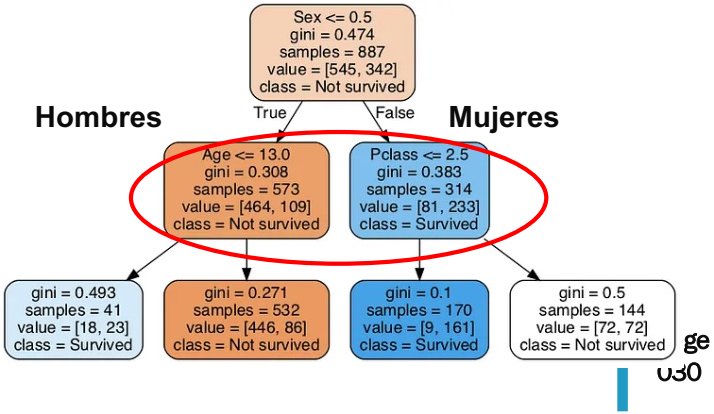
Nodo raiz. Fuente: towardsdatascience.com

04 – Ejemplo del funcionamiento de un árbol de decisión

¿Qué representan los nodos hijo?

- **Age** : Es la variable que mejor separa si son hombres (**<= 13** y **>= 13**)
- **Gini:** 0,0308 indica la homogeneidad del conjunto de datos.
- **Samples**: Contiene 573 observaciones (pasajeros).
- **Value**: De esos 573 pasajeros **NO** sobrevivieron, 109 **SI** sobrevivieron.
- **Class**: Clase con más ejemplos: Los que no sobrevivieron.

¿Podríais explicar que representa el nodo Pclass?

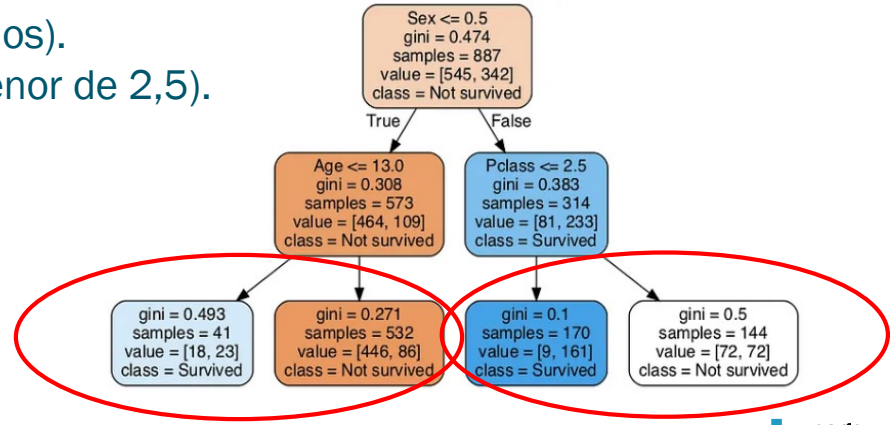


Nodos hijo. Fuente: towardsdatascience.com

04 – Ejemplo del funcionamiento de un árbol de decisión

¿Qué representan los nodos hoja?

- Representan las **predicciones finales.**
- Para hacerlo ha ido dividiendo los datos en función de:
  - ✓ Si era hombre o mujer.
  - ✓ Si era hombre su edad (mayor o menor de 13 años).
  - ✓ Si era mujer la clase en que viajaba (mayor o menor de 2,5).



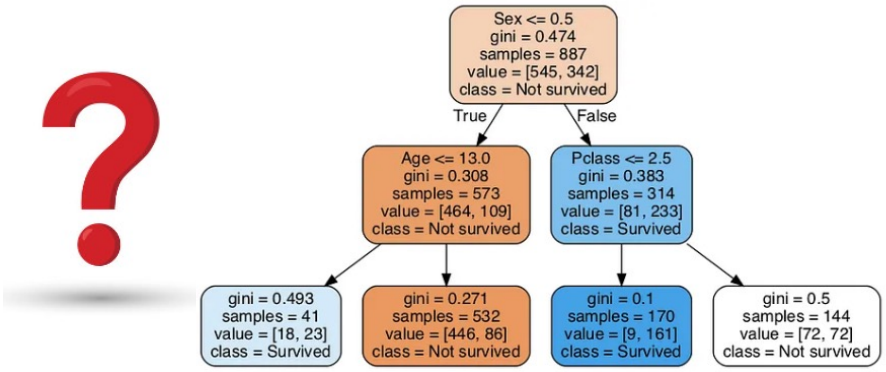
Nodos hijo. Fuente: towardsdatascience.com

04 – Ejemplo del funcionamiento de un árbol de decisión

Interpretación de los resultados del árbol

- Los pasajeros que no sobrevivieron fueron más que los que sobrevivieron.
- De entre que sobrevivieron, sobrevivieron más mujeres que hombres.
- De entre los hombres sobrevivieron más los menores de 13 años.
- De las mujeres sobrevivieron más las que iban en 1ª clase y 2ª clase.

¿Quién tiene una mayor probabilidad de sobrevivir y quién no?



Nodos hijo. Fuente: towardsdatascience.com

00 – ¿De qué vamos a hablar en nuestra próxima clase?

- Medidas de precisión
- Poda de árboles de decisión
- Ensamble Learning y Random Forest
- Problema: Resolución de un problema de clasificación utilizando ensamble learning
- Presentación Actividad 1. Árboles de decisión, reglas y ensamble learning

Esta clase tendrá una duración de 90 minutos

Muchas gracias por vuestra atención

unir LA UNIVERSIDAD EN INTERNET

www.unir.net