



FIA – 4BM1
ESCOM – IPN



PRÁCTICA 10: ÁRBOLES DE DECISIÓN

Carmona Serrano Ian Carlo

Rojas Alarcón Sergio Ulises

4BM1

Fundamentos de IA

Macario Hernández Cruz

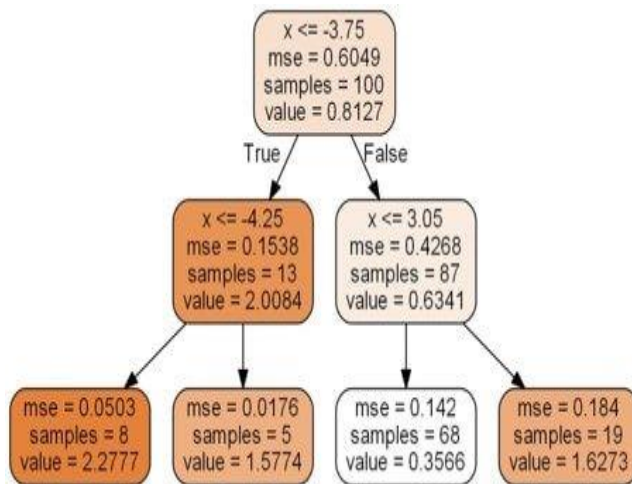
ESCOM



Introducción

Un árbol de decisión es un algoritmo de aprendizaje supervisado no paramétrico, que se utiliza tanto para tareas de clasificación como de

regresión. Tiene una estructura de árbol jerárquica, que consta de un nodo raíz, ramas, nodos internos y nodos hoja.



El aprendizaje del árbol de decisiones emplea una estrategia de divide y vencerás mediante la

realización de una búsqueda codiciosa para identificar los puntos de división óptimos dentro de un árbol. Este proceso de división se repite de forma recursiva de arriba hacia abajo hasta que todos o la mayoría de los registros se hayan clasificado bajo etiquetas de clase específicas.

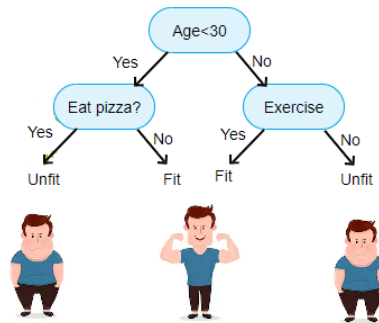
Un diagrama de árbol de decisiones permite evaluar mediante una representación gráfica los posibles resultados, costos y consecuencias de una decisión compleja. Este método es muy útil para analizar datos cuantitativos y tomar una decisión basada en números.



Dentro de un árbol de decisión distinguimos diferentes tipos de nodos:

- ⦿ Primer nodo o nodo raíz: en él se produce la primera división en función de la variable más importante.

- ⦿ Nodos internos o intermedios: tras la primera división encontramos estos nodos, que vuelven a dividir el conjunto de datos en función de las variables.



- ⦿ Nodos terminales u hojas: se ubican en la parte inferior del esquema y su función es indicar la clasificación definitiva.

Ventajas

Cuando se necesita tomar una decisión compleja y se debe evaluar una gran cantidad de datos, los árboles de decisiones pueden ayudar a visualizar con mayor claridad las posibles consecuencias o beneficios asociados a cada decisión.

- ➔ **Transparente:** la mayor ventaja de este método es que brinda un enfoque centrado en la toma de decisiones.
- ➔ **Eficiente:** los árboles de decisiones son eficientes ya que pueden ser creados en muy poco tiempo y con pocos recursos. Un árbol de decisiones es una forma sencilla y eficiente de tomar una decisión.
- ➔ **Flexible:** Es posible agregar ramificaciones para posibles resultados si se obtiene más información durante el análisis.



- Son fáciles de construir, interpretar y visualizar.
- Selecciona las variables más importantes y en su creación no siempre se hace uso de todos los predictores.
- Si faltan datos no podremos recorrer el árbol hasta un nodo terminal, pero sí podemos hacer predicciones promediando las hojas del sub-árbol que alcancemos.
- No es preciso que se cumplan una serie de supuestos como en la regresión lineal (linealidad, normalidad de los residuos, homogeneidad de la varianza, etc.).
- Sirven tanto para variables dependientes cualitativas como cuantitativas, como para variables predictoras o independientes numéricas y categóricas. Además, no necesita variables dummies, aunque a veces mejoran el modelo.
- Permiten relaciones no lineales entre las variables explicativas y la variable dependiente.
- Nos podemos servir de ellos para categorizar variables numéricas.

Desventajas

El árbol de decisiones también tiene sus desventajas por lo que no es una herramienta de toma de decisiones perfecta.

- **Complejo:** si bien los árboles de decisiones suelen tener puntos finales bien definidos, pueden volverse complejos si se agregan demasiadas decisiones al diagrama. Si el árbol se ramifica en muchas direcciones, posiblemente se presenten dificultades para hacer un seguimiento de todo y calcular los valores esperados. La mejor manera de usar esta herramienta es mantenerlo lo más simple posible para que no cause confusión.



- **Inestable:** es importante asegurarse de que los valores en el árbol de decisiones sean precisos para que las ecuaciones sean más exactas. Si se cambia incluso una pequeña parte de los datos, los datos más grandes pueden presentar errores.
- **Riesgoso:** el árbol de decisiones se basa en un algoritmo de probabilidades, por lo que el valor esperado se calcula es una estimación, no una predicción exacta de cada resultado. Si no se evalúan lo suficiente las probabilidades y los beneficios de los resultados, es probable que se asuma un riesgo muy alto con la decisión que elijas.

→

Desarrollo

Lo que hace esta parte del código es importar las librerías que vamos a utilizar vamos a leer el csv de los datos después vamos a codificar los datos a números si su contenido es igual se le asigna el mismo numero pero si es diferente se le crea otro número después esos datos ya con números los vamos a entrenar, después vamos a declarar 2 variables las cuales van a ser 'X' los datos con los que se entrenaron y 'y' van a ser los datos resultados



```
import pandas as pd
from matplotlib import pyplot as plt
import sklearn.tree as skl
from sklearn.tree import plot_tree
from sklearn.preprocessing import LabelEncoder

# Adquiere los datos desde un archivo csv usando la biblioteca PANDAS
dataframe = pd.read_csv('datos.csv', encoding='ISO-8859-1')

# Preprocesamiento (Transformar valores categóricos a numéricos)
Edad = LabelEncoder()
Sexo = LabelEncoder()
PresionSanguinea = LabelEncoder()
Colesterol = LabelEncoder()
Medicamento = LabelEncoder()

dataframe['Edad'] = Edad.fit_transform(dataframe['Edad'])
dataframe['Sexo'] = Sexo.fit_transform(dataframe['Sexo'])
dataframe['PresionSanguinea'] = PresionSanguinea.fit_transform(dataframe['PresionSanguinea'])
dataframe['Colesterol'] = Colesterol.fit_transform(dataframe['Colesterol'])
dataframe['Medicamento'] = Medicamento.fit_transform(dataframe['Medicamento'])

# Prepara los datos
features_cols = ['Edad', 'Sexo', 'PresionSanguinea', 'Colesterol']
X = dataframe[features_cols]
y = dataframe.Medicamento
```

Esta parte del código lo que hace es imprimir las columnas y las filas en una tabla para que mostráramos como se imprimen forma de una tabla después declararemos una función con la que guardaremos la imagen que se generó y también preguntaremos los datos con los que queremos predecir el medicamento

```
# Imprimir la tabla completa de X y y
pd.set_option('display.max_columns', None) # Muestra todas las columnas
pd.set_option('display.width', None) # Sin límites de ancho
pd.set_option('display.max_rows', None) # Muestra todas las filas
df_full = pd.concat([X, y], axis=1)
print(df_full.to_string(index=False))

# Entrenamiento
tree = skl.DecisionTreeClassifier(criterion='gini')
tree.fit(X, y)

# Visualización y guardar imagen
px = 1 / plt.rcParams['figure.dpi'] # Pixel en pulgadas
fig = plt.figure(figsize=(1000 * px, 1000 * px))
_ = plot_tree(tree, feature_names=features_cols, class_names=['No', 'Yes'], filled=True)
fig.savefig('arbol_decision.png', dpi=plt.rcParams['figure.dpi']) # Guardar imagen

# Probar el Modelo
edad = int(input('Ingresa la edad (Joven - 0), (Mediana-Edad - 1), (Senior - 2): '))
sexo = int(input('Ingresa el sexo (F - 0), (M - 1): '))
presionsanguinea = int(input('Ingresa la presión sanguínea (Alta - 0), (Baja - 1), (Normal - 2): '))
colesterol = int(input('Ingresa el colesterol (Alto - 0), (Normal - 1): '))
```

Después de los datos con los que preguntamos los vamos a meter en el predictor y lo vamos a mandar a llamar después imprimiremos los datos



que elegimos después imprimimos la recomendación de los datos y guarda la imagen en la computadora

```
dfprueba = pd.DataFrame()
dfprueba['Edad'] = [edad]
dfprueba['Sexo'] = [sexo]
dfprueba['PresionSanguinea'] = [presionsanguinea]
dfprueba['Colesterol'] = [colesterol]

prediccion = tree.predict(dfprueba)

print('\n\nResultado de la prueba')
print('*****')
print('Con los datos')
print(dfprueba)
print('\nSe recomienda:')
if prediccion[0] == 0:
    print('Se recomienda la medicina A')
else:
    print('Se recomienda la medicina B')
print('*****')
print('Imagen del árbol de decisión generada: arbol_decision.png')
```

Discusión de Resultados (Tabla de los datos codificados)

Edad	Sexo	PresionSanguinea	Colesterol	Medicamento
0	0	0	1	0
0	0	0	0	0
1	0	0	1	1
2	0	2	1	1
2	1	1	1	1
2	1	1	0	0
1	1	1	0	1
0	0	2	1	0
0	1	1	1	1
2	1	2	1	1
0	1	2	0	1
1	0	2	0	1
1	1	0	1	1
2	0	2	0	1

(Pedimos los datos y los mostramos)



```
Ingresar la edad (Joven - 0), (Mediana-Edad - 1), (Senior - 2): 1
Ingresar el sexo (F - 0), (M - 1): 0
Ingresar la presión Sanguínea (Alta - 0), (Baja - 1), (Normal - 2): 2
Ingresar el Colesterol (Alto - 0), (Normal - 1): 1
```

Resultado de la prueba

Con los datos

Edad	Sexo	PresiónSanguínea	Colesterol
0	1	0	2
1	0	2	1

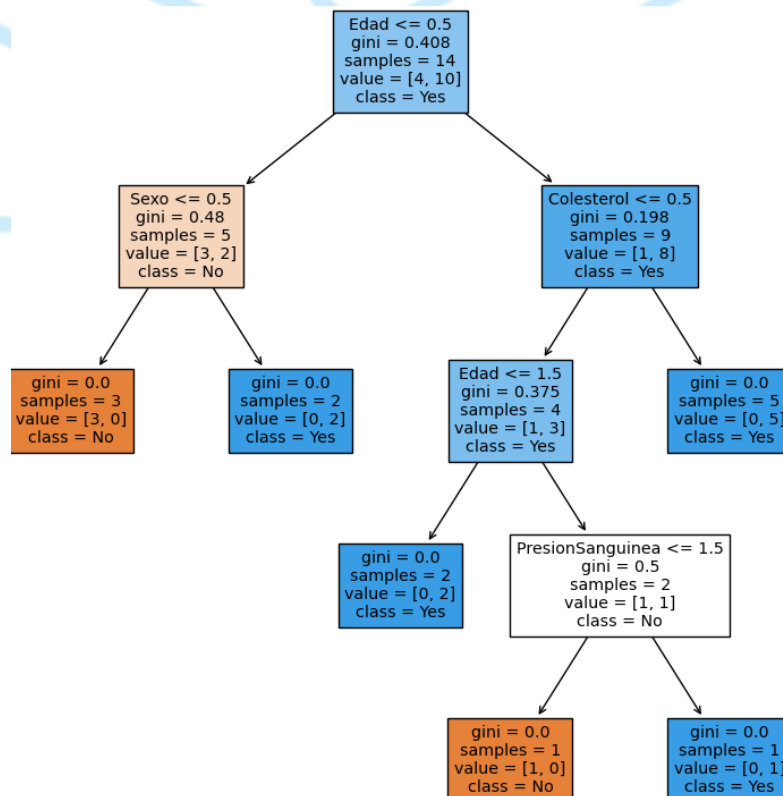
(Mostramos los resultados de la predicción y en este caso nos recomienda la medicina B y genera la imagen)

Se recomienda:

Se recomienda la medicina B

Imagen del árbol de decisión generada: arbol_decision.png

(Imagen Resultado)





Podemos ver que los resultados son buenos pero con más información serian mucho mejores también le podemos poner mas datos y modificar la practica para que acepte números y nos dé una predicción más personalizada

Conclusiones

Los árboles de decisión son una herramienta de aprendizaje supervisado no paramétrico ampliamente utilizada en diversas aplicaciones. Los árboles de decisión tienen una estructura jerárquica en forma de árbol, que consta de nodos raíz, nodos internos y nodos hoja.

Se utilizan en la toma cotidiana de decisiones, evaluación de oportunidades de crecimiento empresarial, aprendizaje automático (machine learning), planificación de productos, decisiones comerciales generales, diseño de software, aprobación de préstamos y clasificación de datos.

Algunas ventajas de los árboles de decisión incluyen su transparencia, eficiencia, flexibilidad y facilidad de construcción, interpretación y visualización. Estos modelos seleccionan las variables más importantes y no requieren que se cumplan supuestos específicos como en la regresión lineal. Además, son adecuados para variables dependientes cualitativas y cuantitativas, así como para variables predictoras numéricas y categóricas. También permiten capturar relaciones no lineales entre las variables explicativas y la variable objetivo.

Sin embargo, los árboles de decisión también presentan desventajas. Pueden volverse complejos si se agregan demasiadas decisiones, lo que dificulta su seguimiento y cálculo de los valores esperados. Además, son sensibles a cambios en los datos y pueden ser inestables. Es esencial evaluar adecuadamente las probabilidades y beneficios de los resultados, ya que se basan en estimaciones y conllevan cierto nivel de riesgo.



En resumen, los árboles de decisión son una poderosa herramienta en el análisis y la toma de decisiones en una amplia gama de aplicaciones. Sin embargo, es importante comprender tanto sus ventajas como sus limitaciones al aplicarlos en situaciones reales, considerando cuidadosamente el contexto específico de cada caso.

Referencias

- Ferrero, R. (2020, mayo). QUÉ SON LOS ÁRBOLES DE DECISIÓN Y PARA QUÉ SIRVEN. MáximaFormación. Recuperado 16 de junio de 2023, de <https://www.maximaformacion.es/blog-dat/que-son-los-arboles-de-decision-y-para-que-sirven/>
- IBM. (s. f.-b). ¿Qué es un árbol de decisión? Recuperado 16 de junio de 2023, de <https://www.ibm.com/mx-es/topics/decision-trees>
- Lucidspark. (s. f.). Cómo hacer y usar árboles de decisiones. Recuperado 16 de junio de 2023, de <https://lucidspark.com/es/blog/como-hacer-arboles-de-decisiones>
- Miro. (s. f.). Árbol de decisiones. Recuperado 16 de junio de 2023, de <https://miro.com/es/diagrama/que-es-arbol-decisiones/>
- Team Asana. (2021, 6 diciembre). ¿Qué es un árbol de decisiones? Un análisis de 5 pasos para tomar mejores decisiones. Asana. Recuperado 16 de junio de 2023, de <https://asana.com/es/resources/decision-tree-analysis>

ESCOM