

## Práctica No. 11

### Árboles de decisión

El módulo *sklearn.tree* incluye modelos basados en árboles de decisión para clasificación y regresión mediante el algoritmo CART.

Ejemplo de uso:

```

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

#Preprocesamiento (Transformar valores categóricos a numéricos)
outlook=LabelEncoder()
temp=LabelEncoder()
humidity=LabelEncoder()
windy=LabelEncoder()
play=LabelEncoder()

dataframe['outlook']=outlook.fit_transform(dataframe['outlook'])
dataframe['temp']=temp.fit_transform(dataframe['temp'])
dataframe['humidity']=humidity.fit_transform(dataframe['humidity'])
dataframe['windy']=windy.fit_transform(dataframe['windy'])
dataframe['play']=play.fit_transform(dataframe['play'])
  
```

Al ser variables categóricas se requiere transformarlas a datos numéricos, *LabelEncoder* sirve para ello, si los datos fueran numéricos no sería necesario este paso.

day	outlook	temp	humidity	windy	play
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Rain	Mild	High	Strong	No
D13	Overcast	Hot	Normal	Weak	Yes
D14	Overcast	Mild	High	Strong	Yes

Conjunto de entrada original

outlook	temp	humidity	windy	play
2	1	0	1	0
2	1	0	0	0
0	1	0	1	1
1	2	0	1	1
1	0	1	1	1
1	0	1	0	0
0	0	1	0	1
2	2	0	1	0
2	0	1	1	1
1	2	1	1	1
2	2	1	0	1
1	2	0	0	0
0	1	1	1	1
0	2	0	0	1

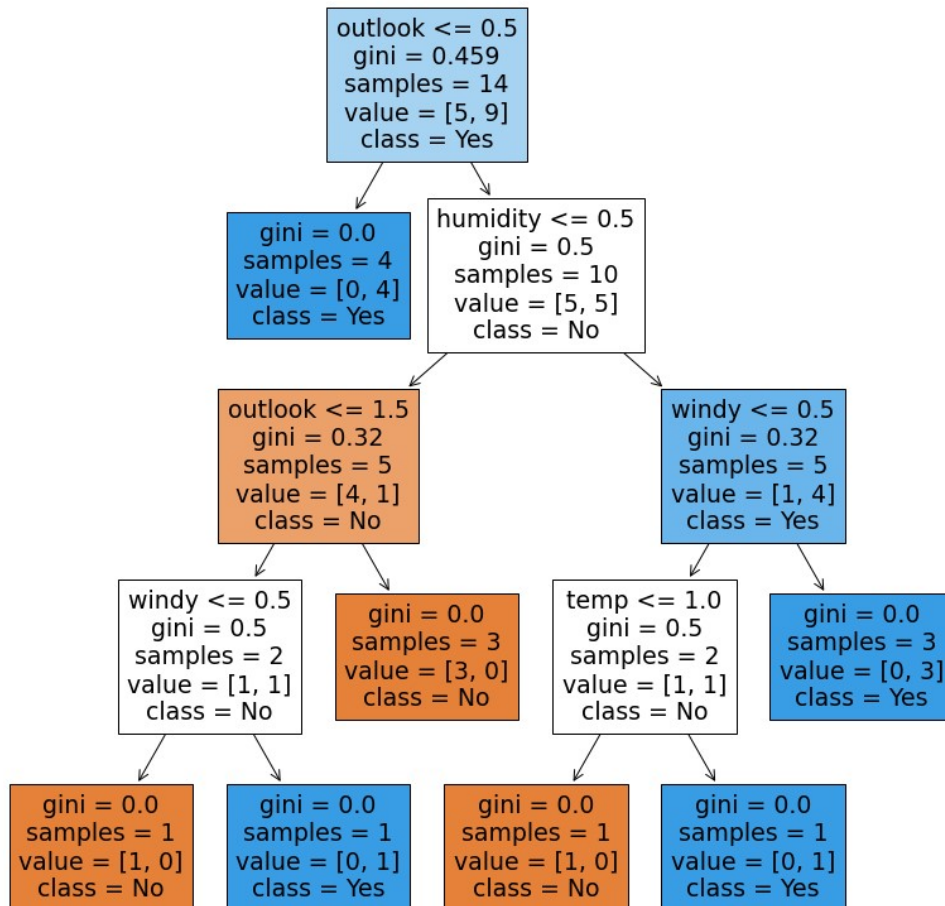
Datos transformados

```
#Prepara los datos
features_cols=['outlook','temp','humidity','windy']
X=dataframe[features_cols]
y=dataframe.play

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

#Visualización
px = 1/plot.rcParams['figure.dpi'] # Pixel in pulgadas
fig = plot.figure(figsize=(1000*px,1000*px))
_ = plot_tree(tree, feature_names=features_cols, class_names=['No','Yes'], filled=True)
```

Como resultado de aplicar el algoritmo CART, se construye un árbol binario.



Al consultar el árbol debe considerar los valores numéricos, por ejemplo, para **Outlook** la tabla de equivalencia es:

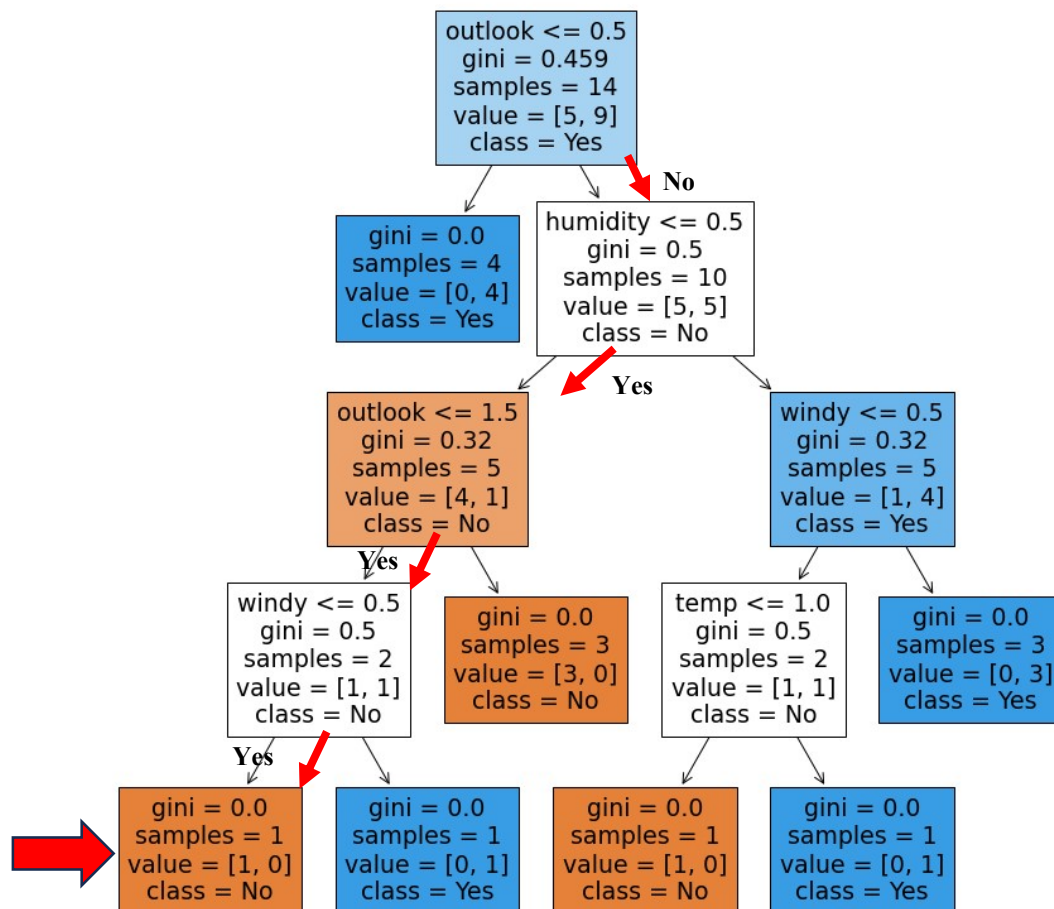
Valor	Significado
0	Overcast
1	Rain
2	Sunny

Por tanto, al consultar el nodo raíz se evalúa si **Outlook**  $\leq 0.5$ , si es verdadero el valor corresponde a “Overcast” y por tanto el resultado es “Yes” para jugar Volley Ball (rama izquierda).

Al pedirle predicciones al modelo también deben hacerse las transformaciones correspondientes, por ejemplo, si el día está lluvioso, con temperatura fría, humedad alta y con viento, ¿podemos hacer deporte?

Los datos de consulta son:

**Outlook = 1 (Rain)**  
**Temp = 0 (Cool)**  
**Humidity = 0 (High)**  
**Windy = 0 (Strong)**



De acuerdo con el árbol de decisión, con las condiciones mencionadas no es recomendable hacer deporte.

## Tarea

Construya un árbol de decisión con el siguiente conjunto de datos, considerando al atributo Medicamento como variable objetivo.

Paciente	Edad	Sexo	PresionSanguinea	Colesterol	Medicamento
p1	Joven	F	Alta	Normal	A
p2	Joven	F	Alta	Alto	A
p3	Mediana-Edad	F	Alta	Normal	B
p4	Senior	F	Normal	Normal	B
p5	Senior	M	Baja	Normal	B
p6	Senior	M	Baja	Alto	A
p7	Mediana-Edad	M	Baja	Alto	B
p8	Joven	F	Normal	Normal	A
p9	Joven	M	Baja	Normal	B
p10	Senior	M	Normal	Normal	B
p11	Joven	M	Normal	Alto	B
p12	Mediana-Edad	F	Normal	Alto	B
p13	Mediana-Edad	M	Alta	Normal	B
p14	Senior	F	Normal	Alto	B

De acuerdo con el árbol construido *¿qué medicamento se le debe recomendar a una chica de mediana edad, con presión sanguínea baja y colesterol normal?*

Elabore su reporte con:

- Introducción
- Desarrollo
- Discusión de resultados
- Conclusiones

Súbalo en formato PDF.