

Árboles de Decisión

Cristian López Del Alamo

clopezd@utec.edu.pe

IPRODAM3D - Research group

2022



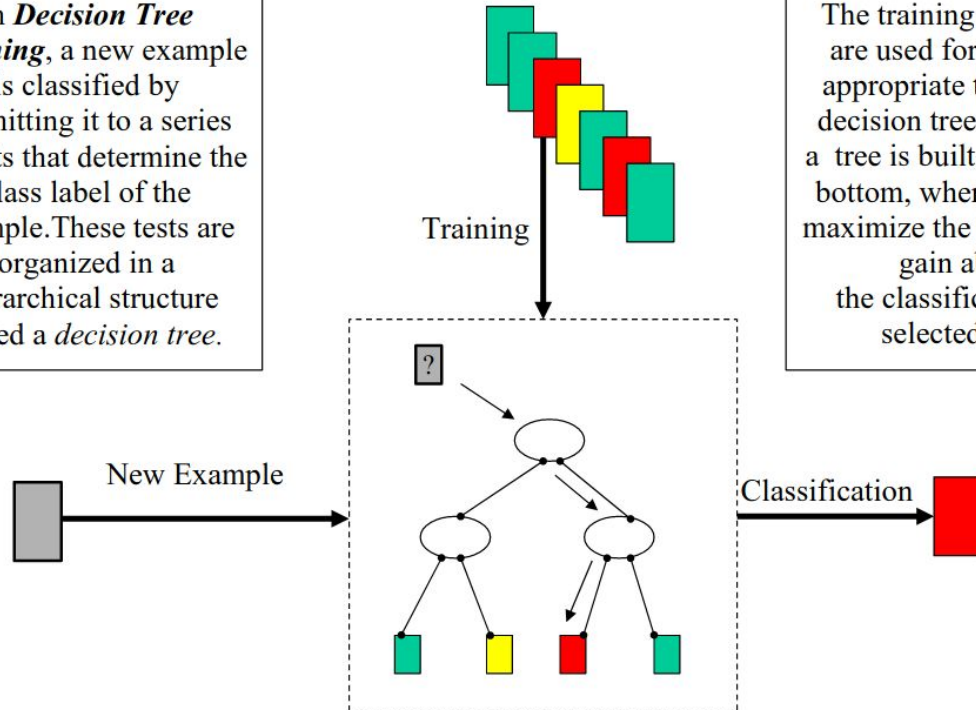
1

Árboles de Decisión



Decision Tree for PlayTennis

In **Decision Tree Learning**, a new example is classified by submitting it to a series of tests that determine the class label of the example. These tests are organized in a hierarchical structure called a *decision tree*.



The training examples are used for choosing appropriate tests in the decision tree. Typically, a tree is built from top to bottom, where tests that maximize the information gain about the classification are selected first.

Base de Datos: Vampiros

Oscuridad	Ajo	Tez	Acento	Vampiro
?	Si	Pálido	No	No
Si	Si	Rojo	No	No
?	No	Rojo	No	Si
No	No	Promedio	Fuerte	Si
?	No	Promedio	Extraño	Si
Si	No	Pálido	Fuerte	No
Si	No	Promedio	Fuerte	No
?	Si	Rojo	Extraño	No

Base de Datos: Vampiro

F1

F2

F3

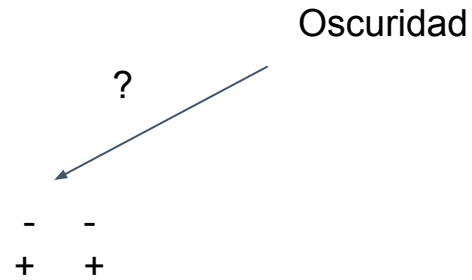
F4

Y

Oscuridad	Ajo	Tez	Acento	Vampiro
?	Si	Pálido	No	No
Si	Si	Rojo	No	No
?	No	Rojo	No	Si
No	No	Promedio	Fuerte	Si
?	No	Promedio	Extraño	Si
Si	No	Pálido	Fuerte	No
Si	No	Promedio	Fuerte	No
?	Si	Rojo	Extraño	No

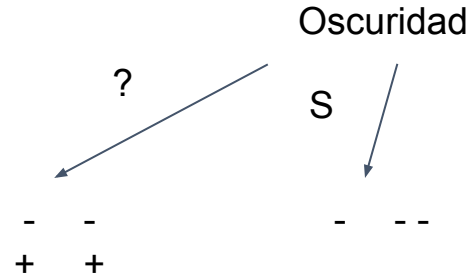
Analizamos F1 : Oscuridad

Oscuridad	Vampiro
?	No
Si	No
?	Si
No	Si
?	Si
Si	No
Si	No
?	No



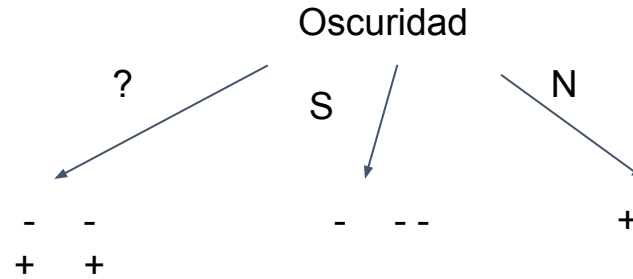
Analizamos F1 : Oscuridad

Oscuridad	Vampiro
?	No
Si	No
?	Si
No	Si
?	Si
Si	No
Si	No
?	No



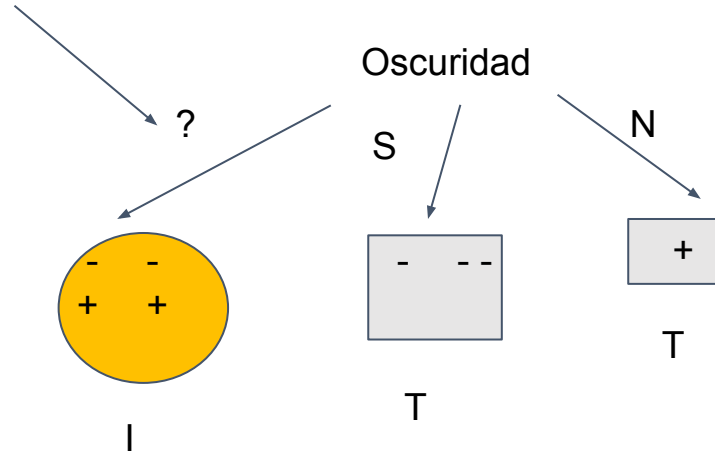
Analizamos F1 : Oscuridad

Oscuridad	Vampiro
?	No
Si	No
?	Si
No	Si
?	Si
Si	No
Si	No
?	No



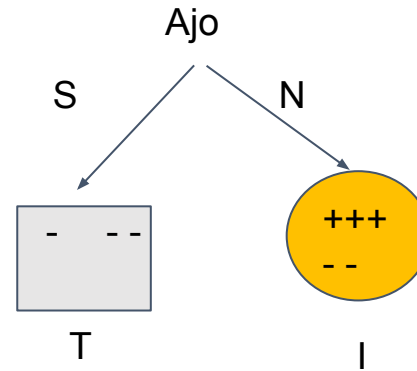
Analizamos F1 : Oscuridad

Oscuridad	Vampiro
?	No
Si	No
?	Si
No	Si
?	Si
Si	No
Si	No
?	No



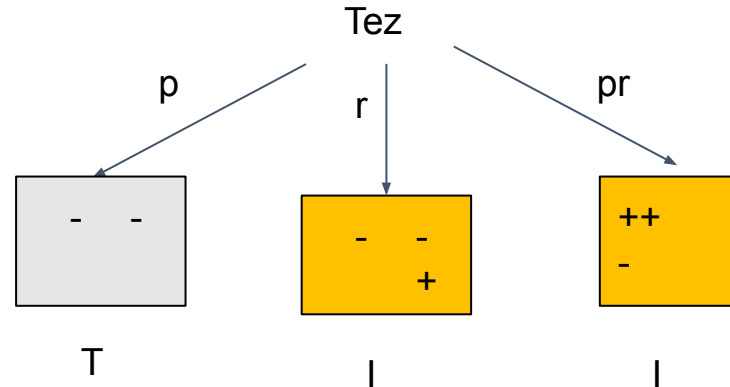
Analizamos F2 : Ajo

F2	Y
Ajo	Vampiro
Si	No
Si	No
No	Si
No	Si
No	Si
No	No
No	No
Si	No



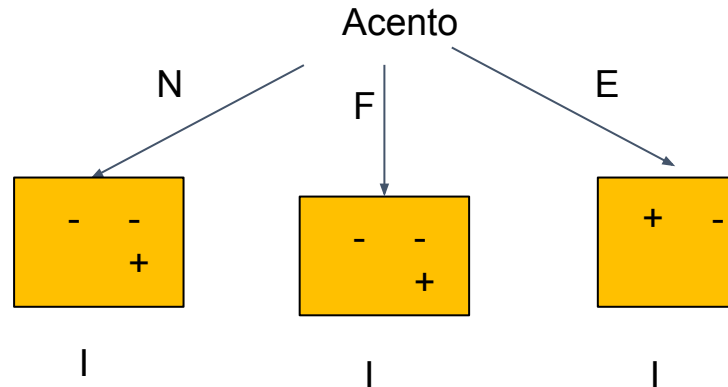
Analizamos F3 : Tez

F3	Y
Tez	Vampiro
Pálido	No
Rojo	No
Rojo	Si
Promedio	Si
Promedio	Si
Pálido	No
Promedio	No
Rojo	No

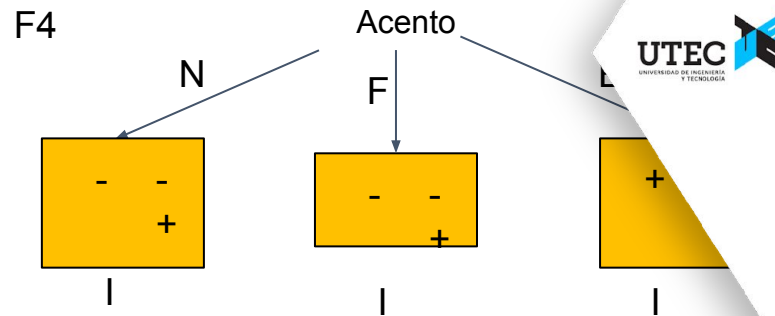
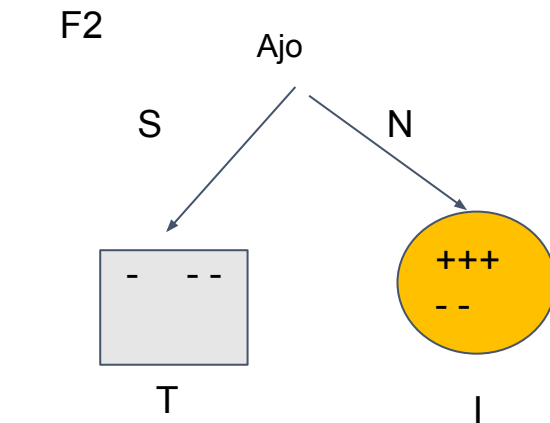
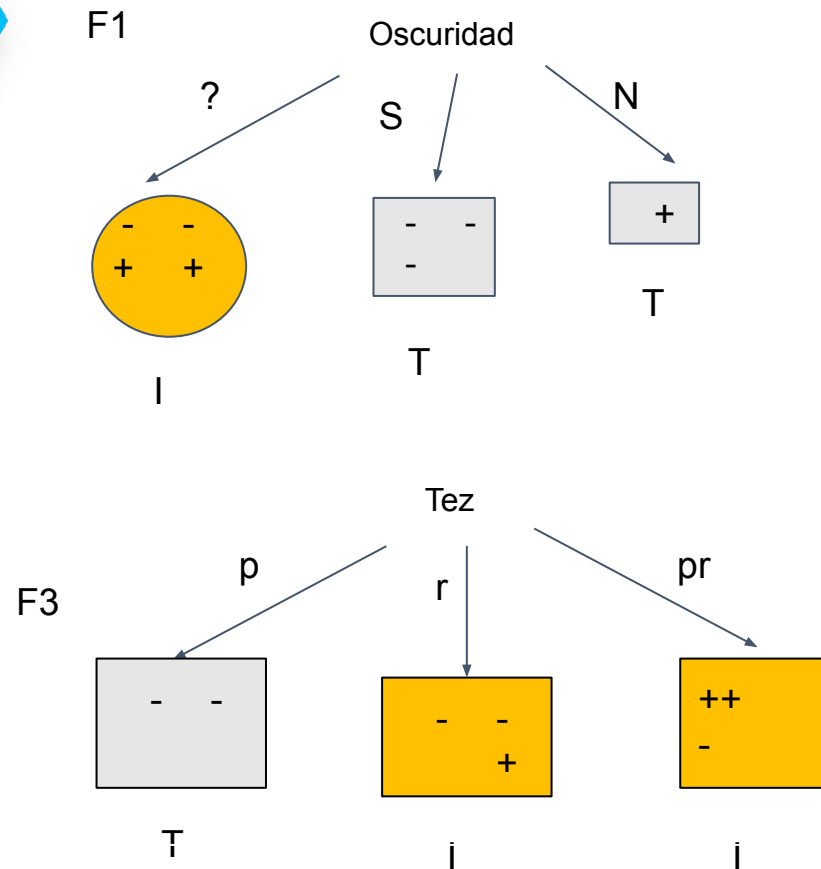


Analizamos F4 : Ajo

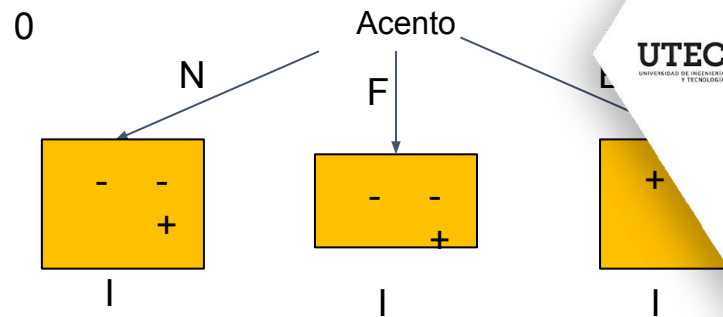
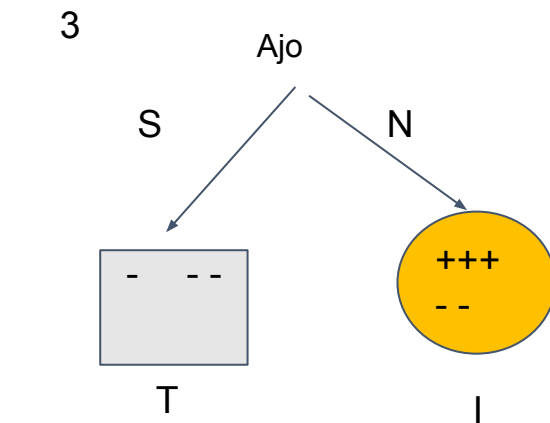
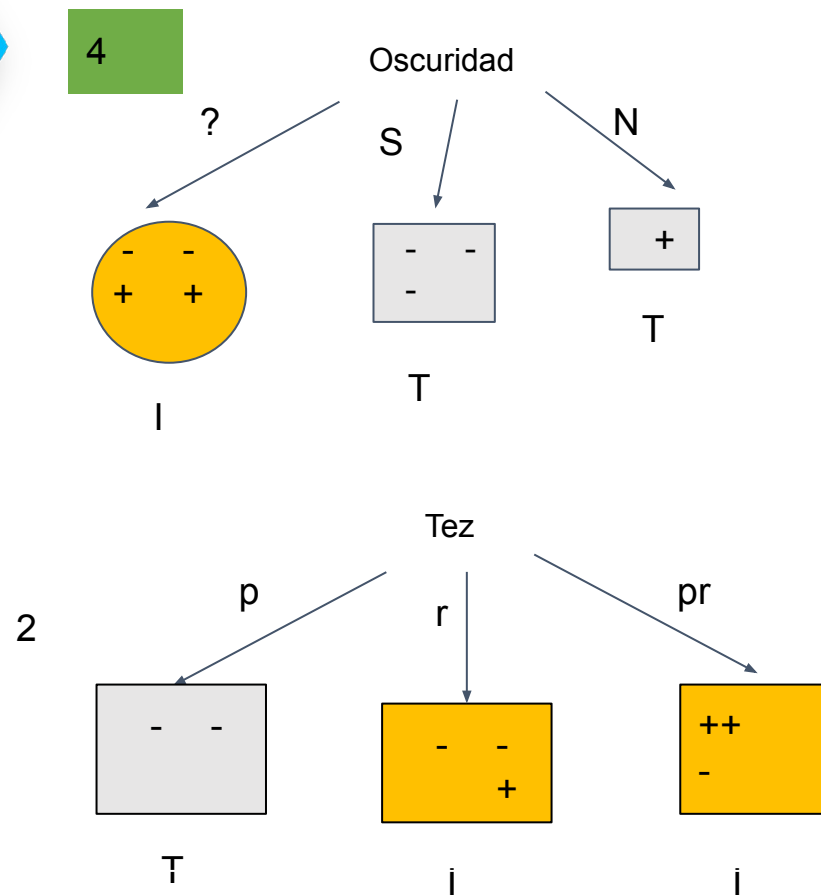
F4	Y
Acento	Vampiro
No	No
No	No
No	Si
Fuerte	Si
Extraño	Si
Fuerte	No
Fuerte	No
Extraño	No



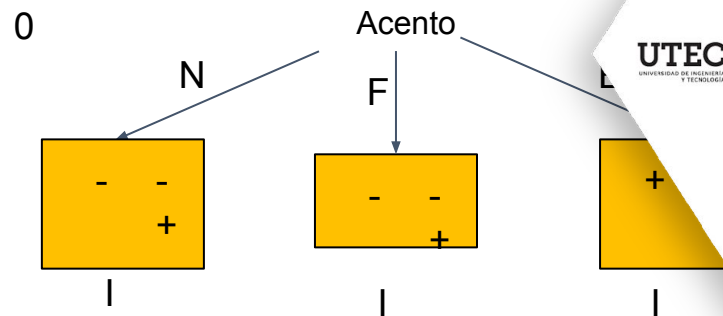
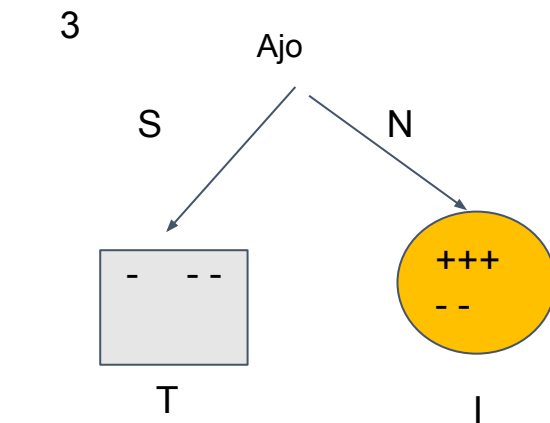
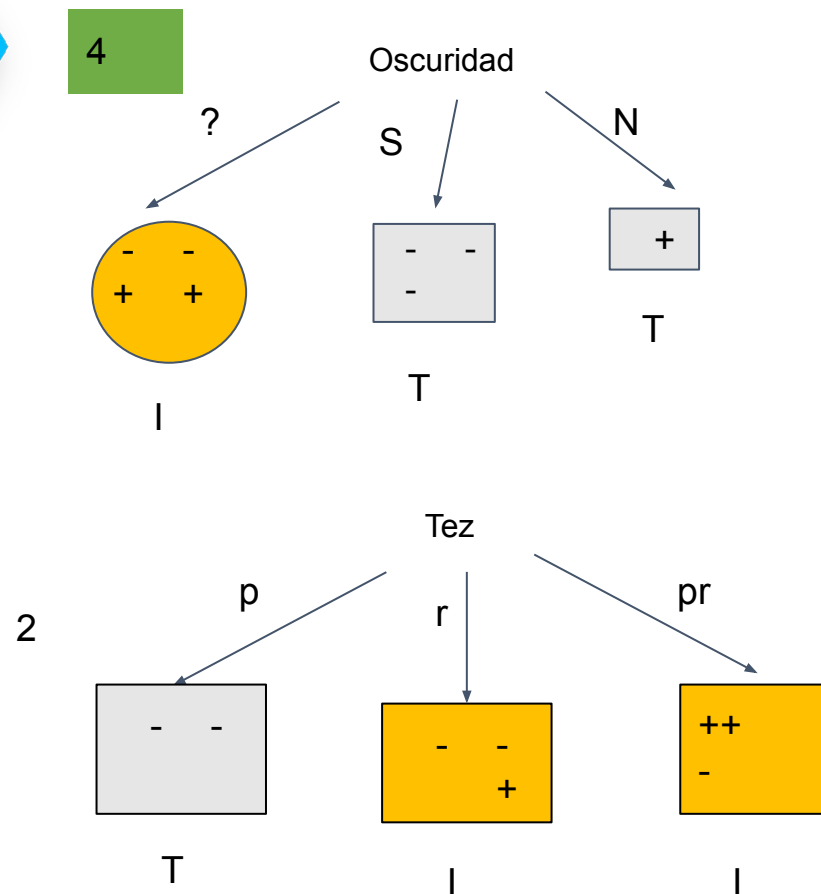
Analizando las 4 características



Analizando las 4 características



Analizando las 4 características



Base de Datos: Vampiro

F1

F2

F3

F4

Y

Oscuridad	Ajo	Tez	Acento	Vampiro
?	Si	Pálido	No	No
Si	Si	Rojo	No	No
?	No	Rojo	No	Si
No	No	Promedio	Fuerte	Si
?	No	Promedio	Extraño	Si
Si	No	Pálido	Fuerte	No
Si	No	Promedio	Fuerte	No
?	Si	Rojo	Extraño	No

Base de Datos: Vampiro

F1

F2

F3

F4

Y

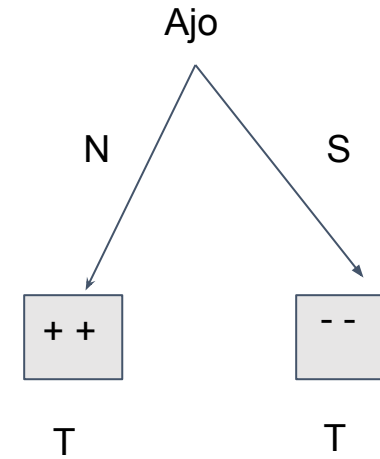
Oscuridad	Ajo	Tez	Acento	Vampiro
?	Si	Pálido	No	No
?	No	Rojo	No	Si
?	No	Promedio	Extraño	Si
?	Si	Rojo	Extraño	No

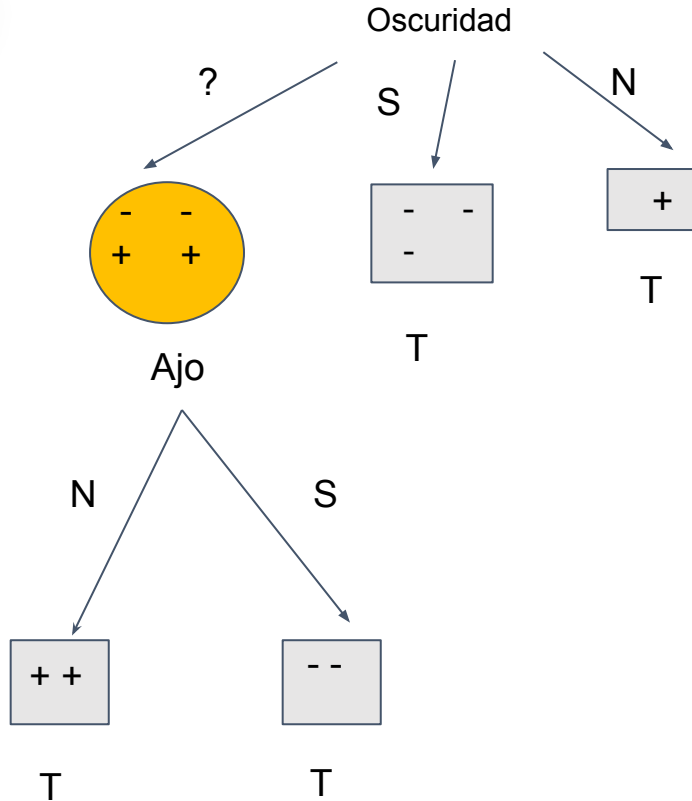
F2

Oscuridad	Ajo	Tez	Acento	Vampiro
?	Si	Pálido	No	No
?	No	Rojo	No	Si
?	No	Promedio	Extraño	Si
?	Si	Rojo	Extraño	No

Oscuridad	Ajo	Vampiro
?	Si	No
?	No	Si
?	No	Si
?	Si	No

Oscuridad	Ajo	Vampiro
?	Si	No
?	No	Si
?	No	Si
?	Si	No





```

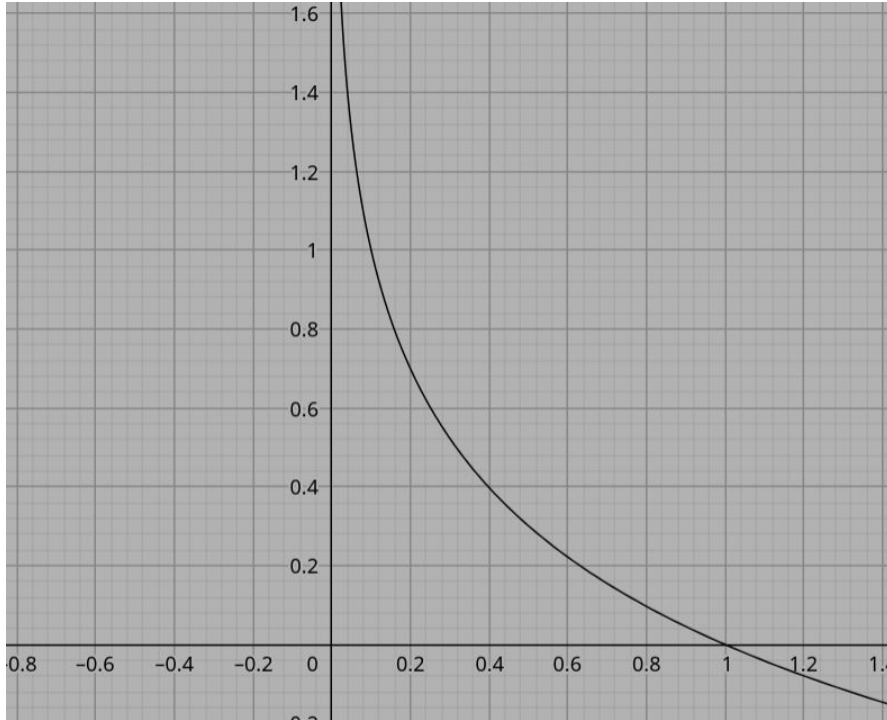
Si Oscuridad == S
    return No Vampiro
Si Oscuridad == N
    return Vampiro
Si Oscuridad == ?
    Si Ajo == N
        return Vampiro
    Sino
        return No Vampiro
  
```

¿Qué Problemas encuentra con este método?

2 Entropía

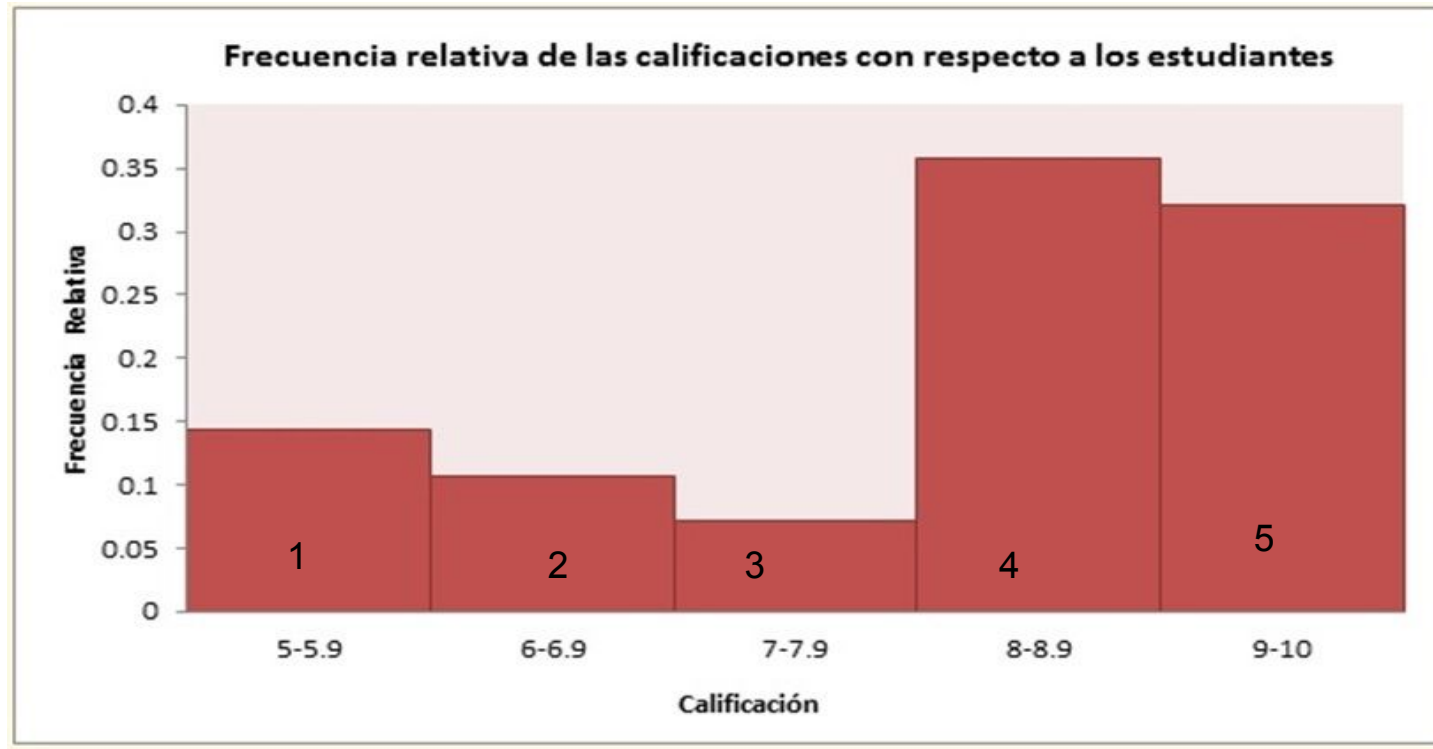


Información



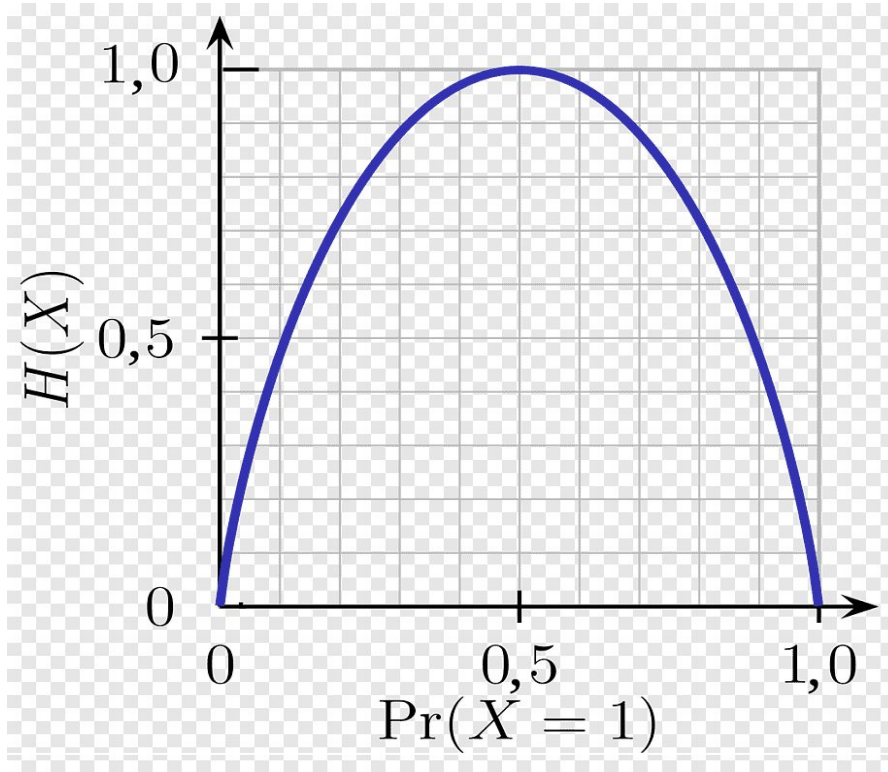
$$I(X) = \log(1/p(x)) = -\log(p(x))$$

¿Qué grupo tiene más información?



[Fuente: click](#)

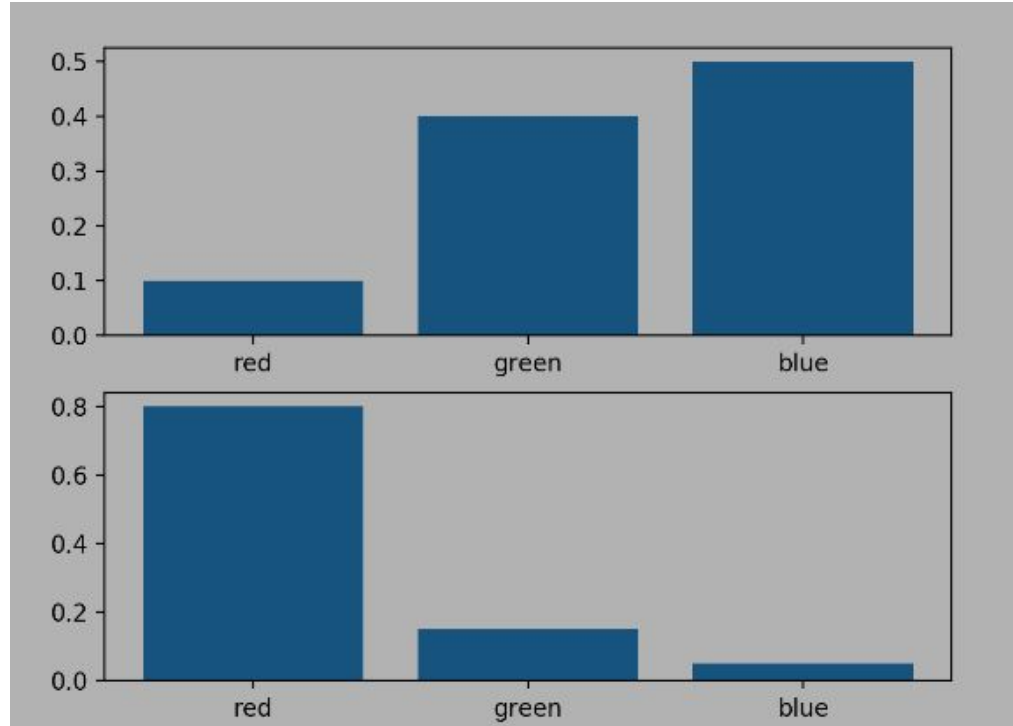
Entropía



$$H(X) = E[I(X)] = \sum_{i=1}^n p(x_i) I(x_i) = - \sum_{i=1}^n p(x_i) \log(x_i)$$

¿Cuál de los dos histogramas tiene mayor entropía?

1

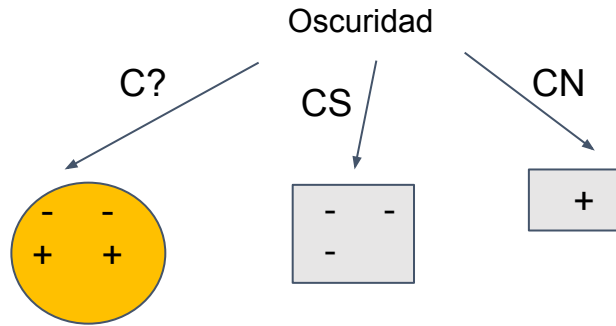


2

Nivel de Desorden de un conjunto o una clase

$$D_s = - \sum_{c \in C} P_c \log_2 P_c$$

¿Cuál es el nivel de desorden de las 3 clases en Oscuridad?

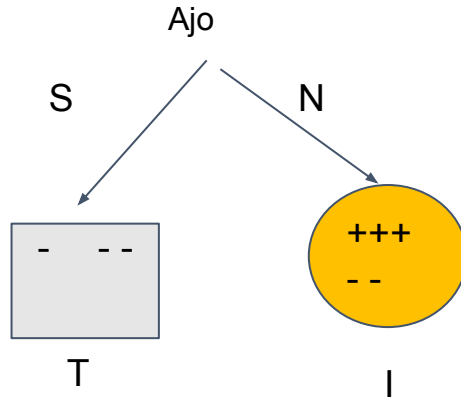


$D(C?) =$

$D(CS) =$

$D(CN) =$

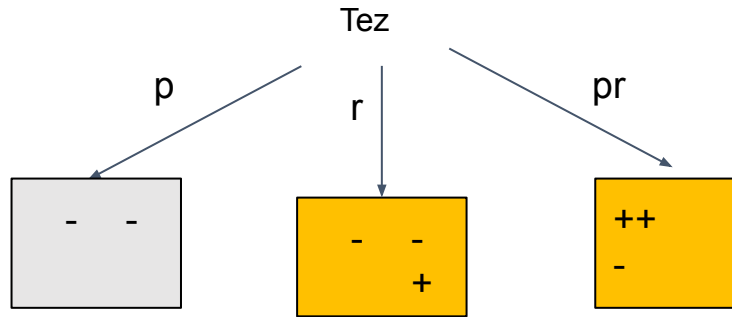
¿Cuál es el nivel de desorden de las 3 clases en Ajo?



$D(CS) =$

$D(CN) =$

¿Cuál es el nivel de desorden de las 3 clases en Ajo?

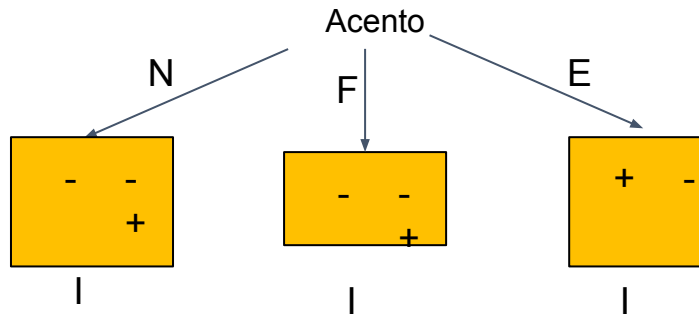


$D(C_p) =$

$D(C_r) =$

$D(C_{pr}) =$

¿Cuál es el nivel de desorden de las 3 clases en Ajo?

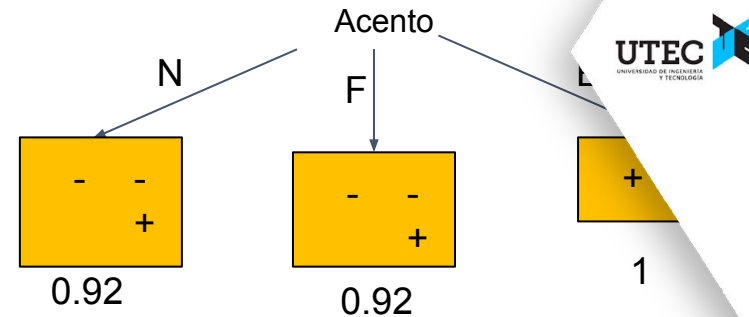
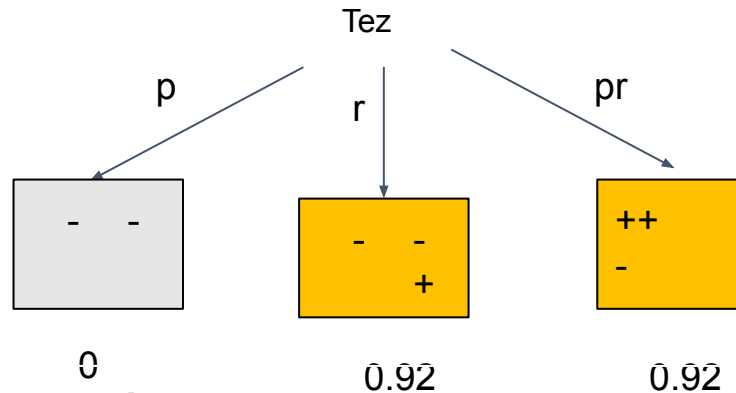
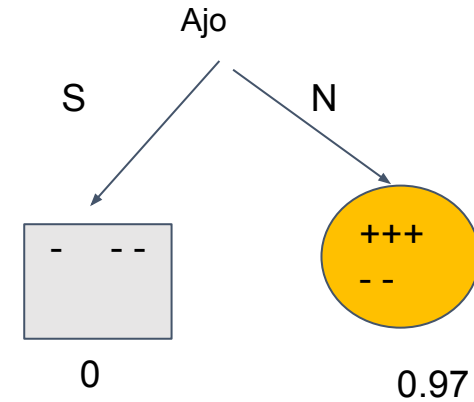
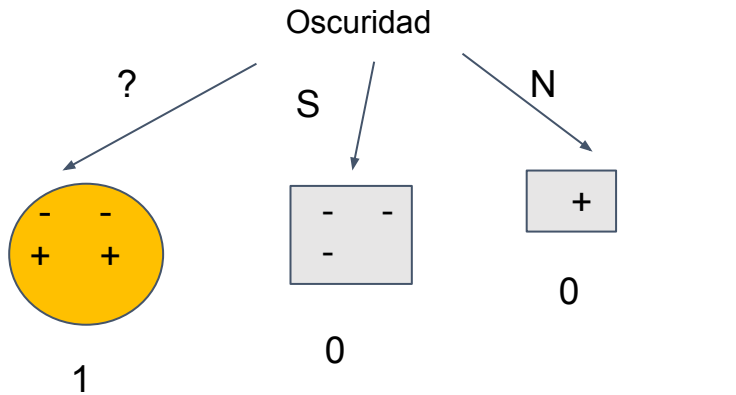


$$D(CN) =$$

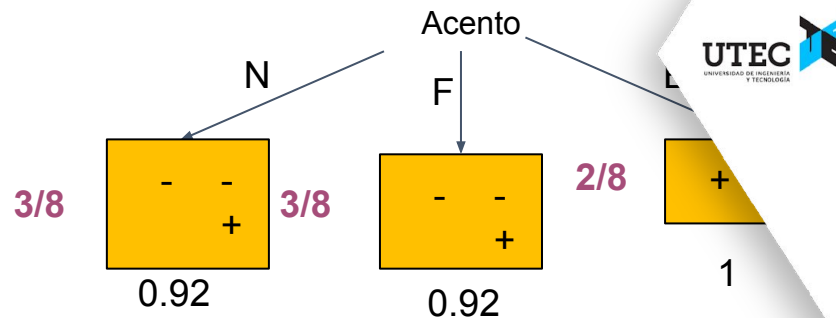
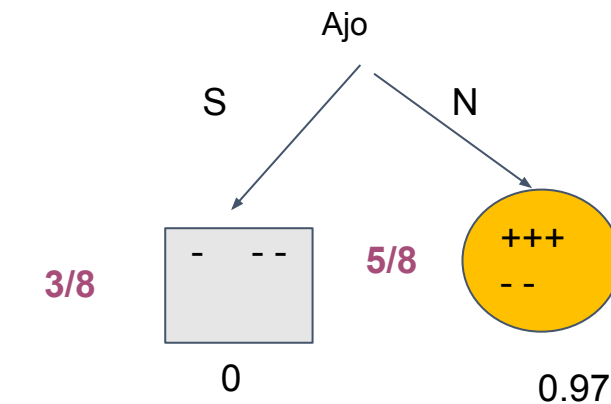
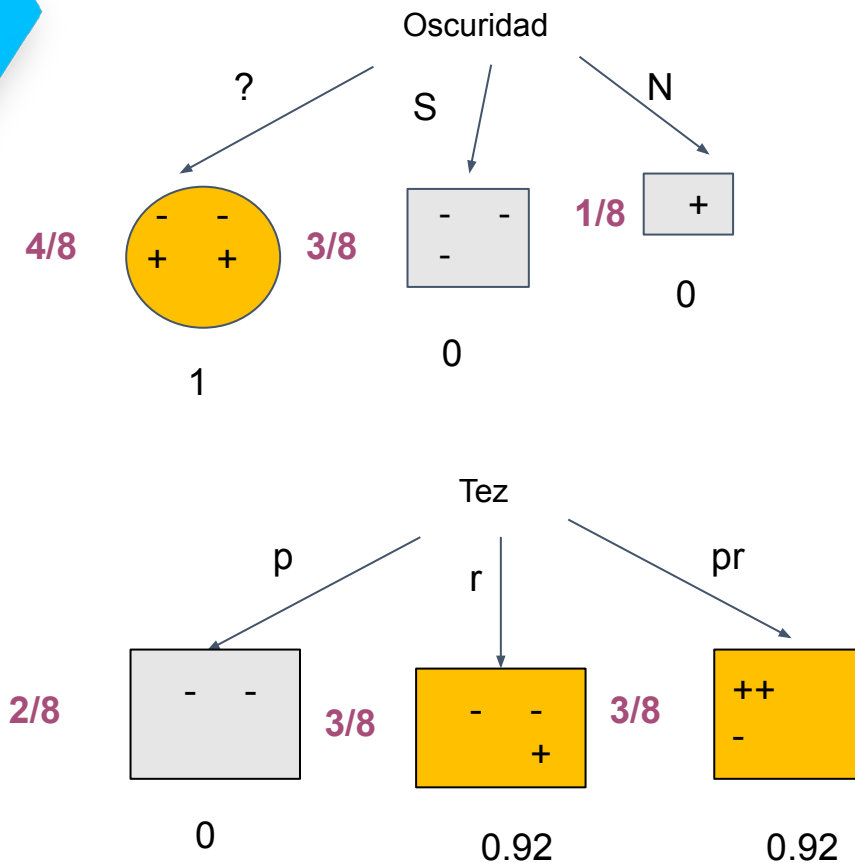
$$D(CF) =$$

$$D(CE) =$$

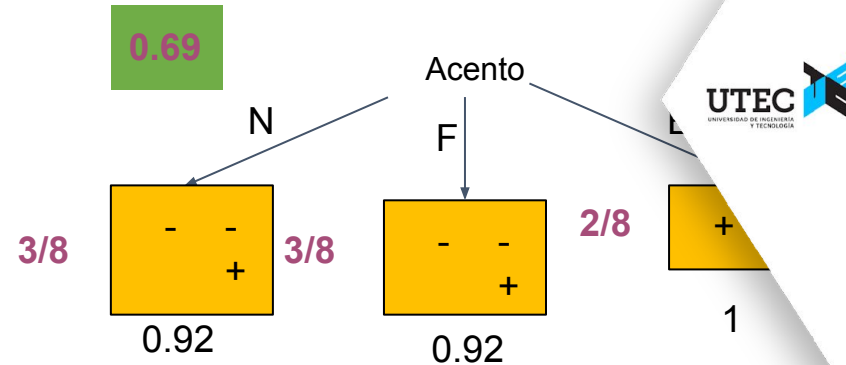
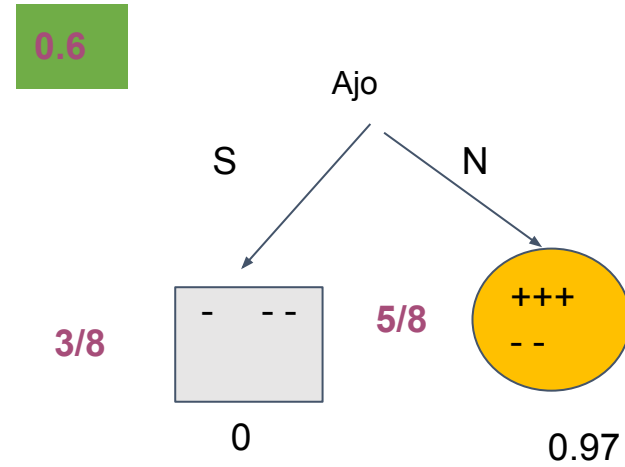
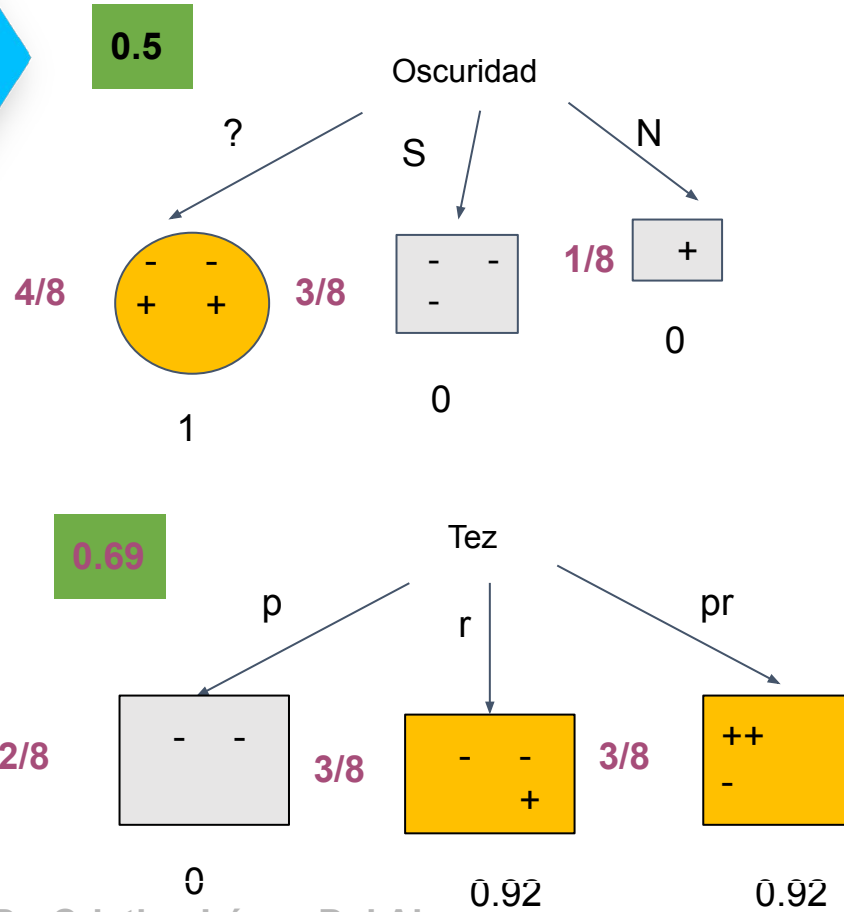
Calculando la cantidad de entropía de cada conjunto



Pesos de cada clase



Pesos de cada clase



3

Information Gain



Ganancia de Información en cada Feature

$$Gain(Vampiro) = H(Vampiro) - \sum_{e \in E} \frac{|e|}{|E|} H(e)$$

Ganancia de Información en cada Feature

Entropía de predicción.

Vampiro
No
No
Si
Si
Si
No
No
No

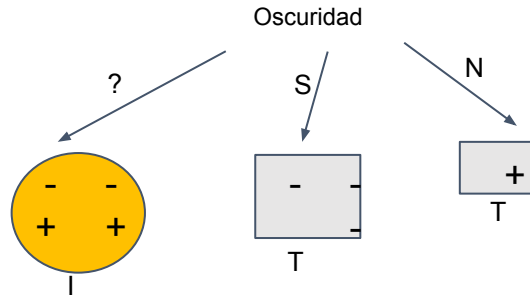
$$H(Vampiro) = -\frac{5}{8} \log \frac{5}{8} - \frac{3}{8} \log \frac{3}{8}$$

Analizamos F1 : Oscuridad

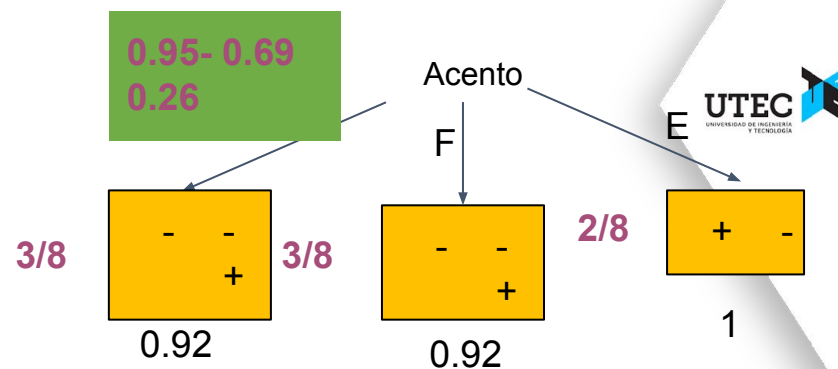
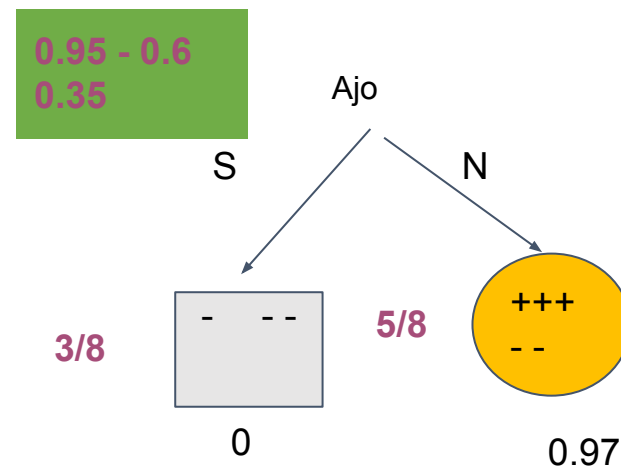
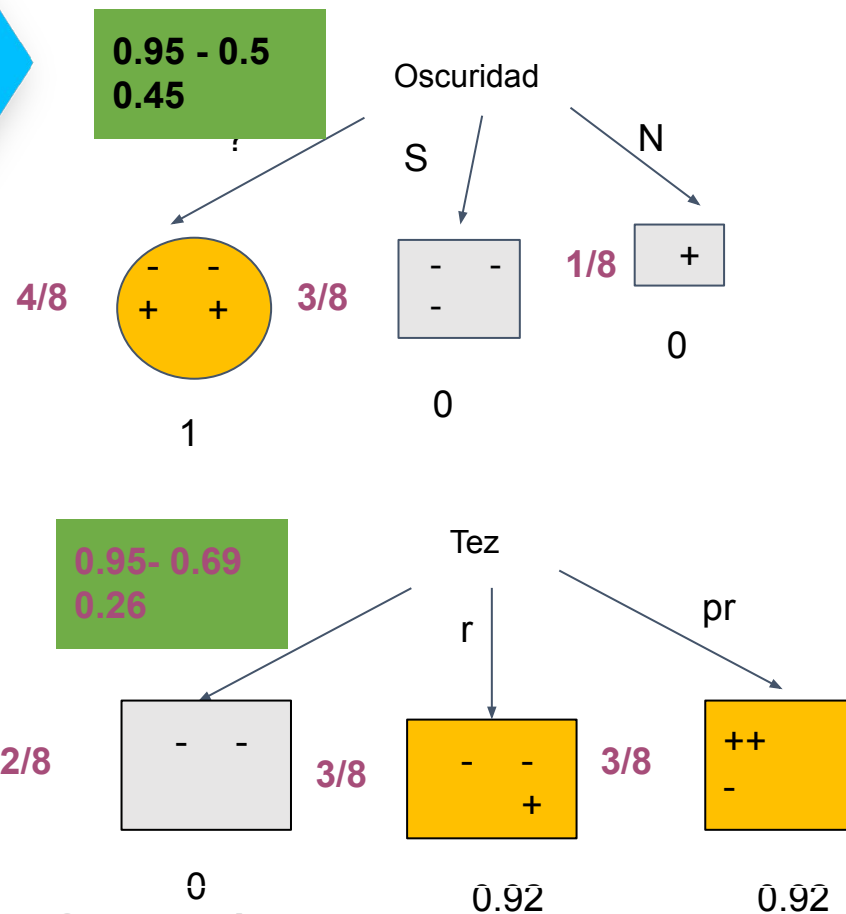
$$Gain(Vampiro) = H(Vampiro) - \sum_{e \in E} \frac{|e|}{|E|} H(e)$$

$$E = \{?, Si, No\}$$

Oscuridad	Vampiro
?	No
Si	No
?	Si
No	Si
?	Si
Si	No
Si	No
?	No



Pesos de cada clase



4

Gini Index



Ganancia de Información en cada Feature

$$Gini(x) = 1 - \sum_j [p(j|t)]^2$$

$p(j|t)$ es la frecuencia relativa de la clase j en el nodo t

Ejemplo del cálculo de Gini

C1	0
C2	6
Gini =	

C1	1
C2	5
Gini =	

C1	2
C2	4
Gini =	

C1	3
C2	3
Gini =	

$p(j|t)$ es la frecuencia relativa de la clase j en el nodo t

Ejemplo del cálculo de Gini

C1	0
C2	6
Gini = 0	

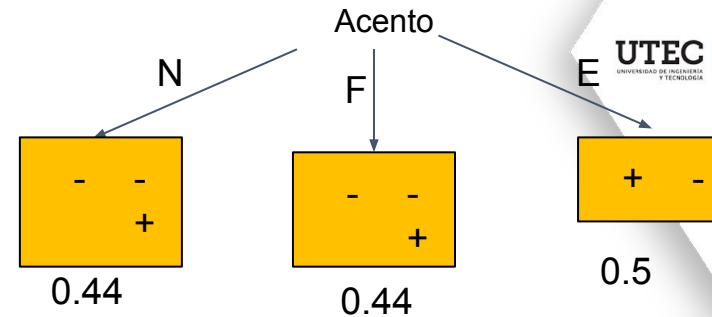
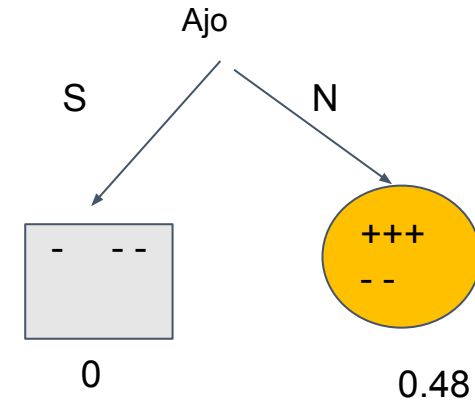
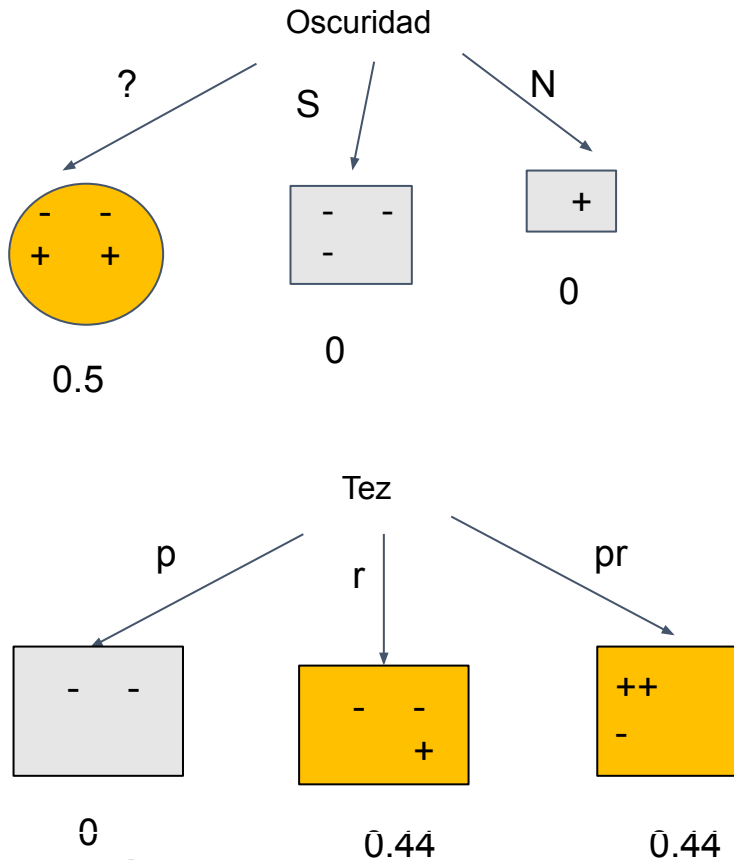
C1	1
C2	5
Gini = 0.28	

C1	2
C2	4
Gini = 0.44	

C1	3
C2	3
Gini = 0.5	

$p(j|t)$ es la frecuencia relativa de la clase j en el nodo t

Calculando la cantidad de entropía de cada conjunto





5

Gini Split

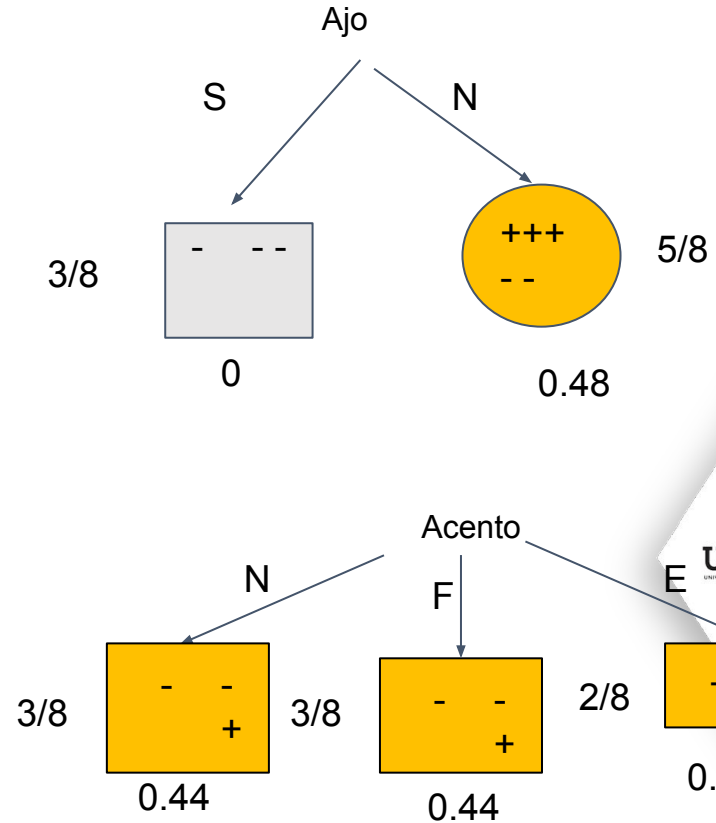
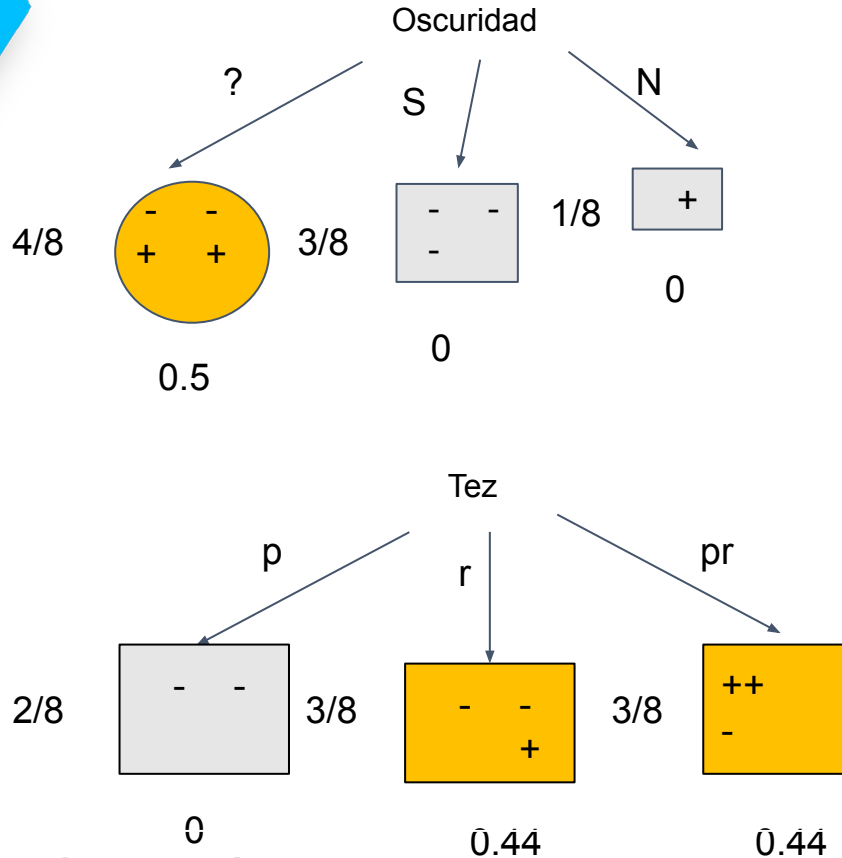
UTEC

Ganancia de Información en cada Feature

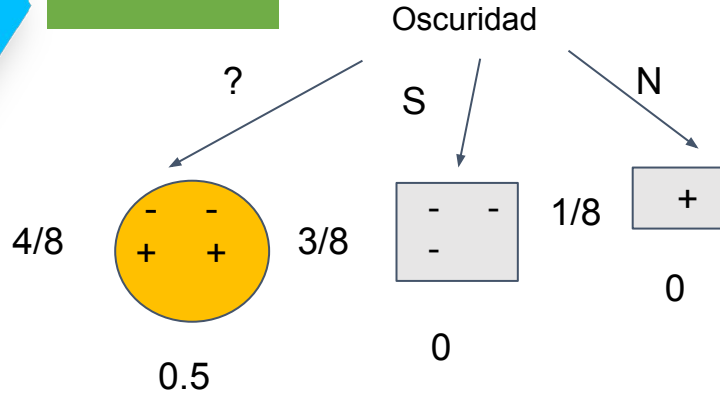
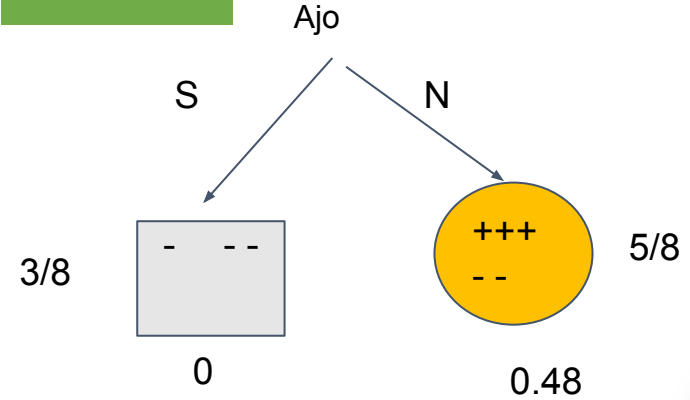
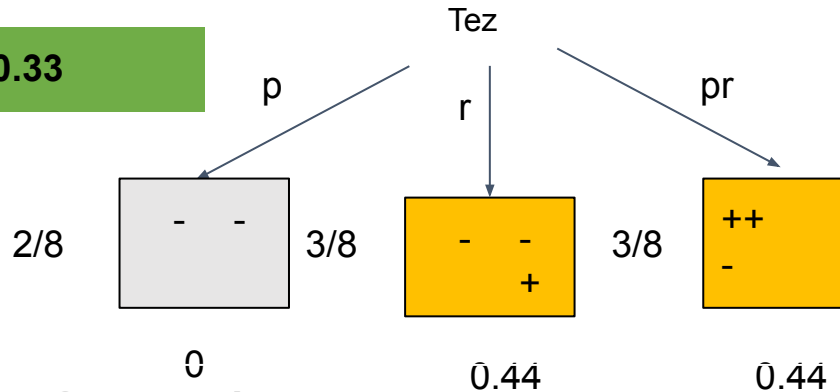
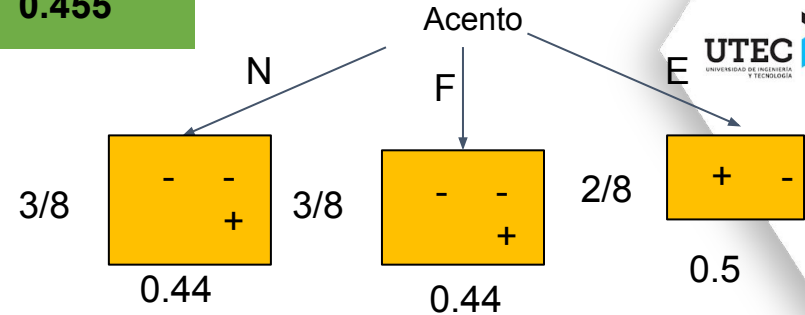
$$Gini_{split} = \sum_{i=1}^k \frac{n_i}{n} Gini(i)$$

n_i es el número de registros en el hijo i , n es el número de registros en el nodo p

Calculando la cantidad de entropía de cada conjunto



Calculando la cantidad de entropía de cada conjunto

0.25**0.3****0.33****0.455**

Function ID3

- **Input:** Example set S
- **Output:** Decision Tree DT
- If all examples in S belong to the same class c
 - return a new leaf and label it with c
- Else
 - i. Select an attribute A according to some heuristic function
 - ii. Generate a new node DT with A as test
 - iii. For each Value v_i of A
 - (a) Let S_i = all examples in S with $A = v_i$
 - (b) Use ID3 to construct a decision tree DT_i for example set S_i
 - (c) Generate an edge that connects DT and DT_i

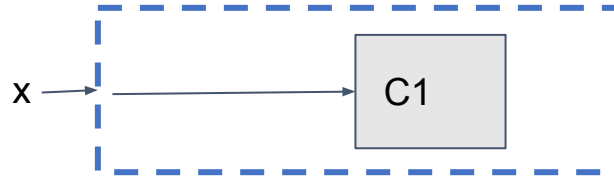
6

Bagging y Random Forest

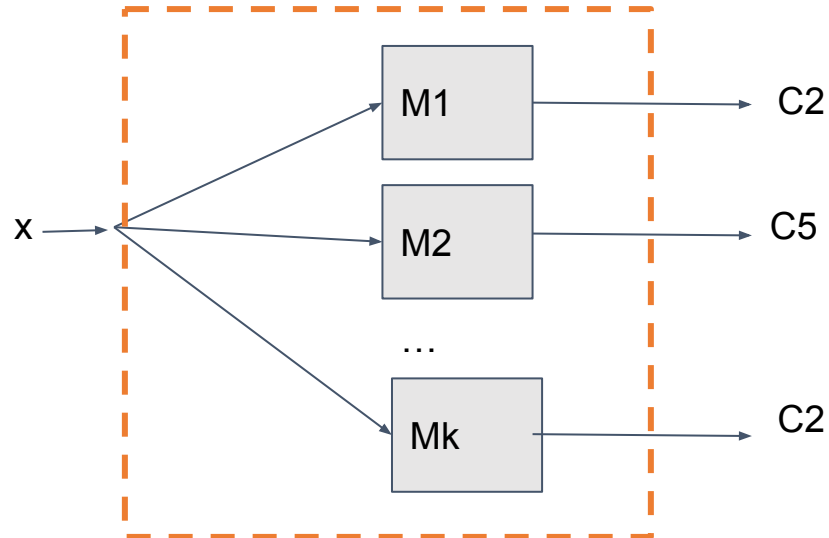


Ensembling: Utiliza varios predictores de aprendizaje de máquinas para hacer un mejor predictor

No Ensembling



With Ensembling



Bagging: Bootstrap aggregating

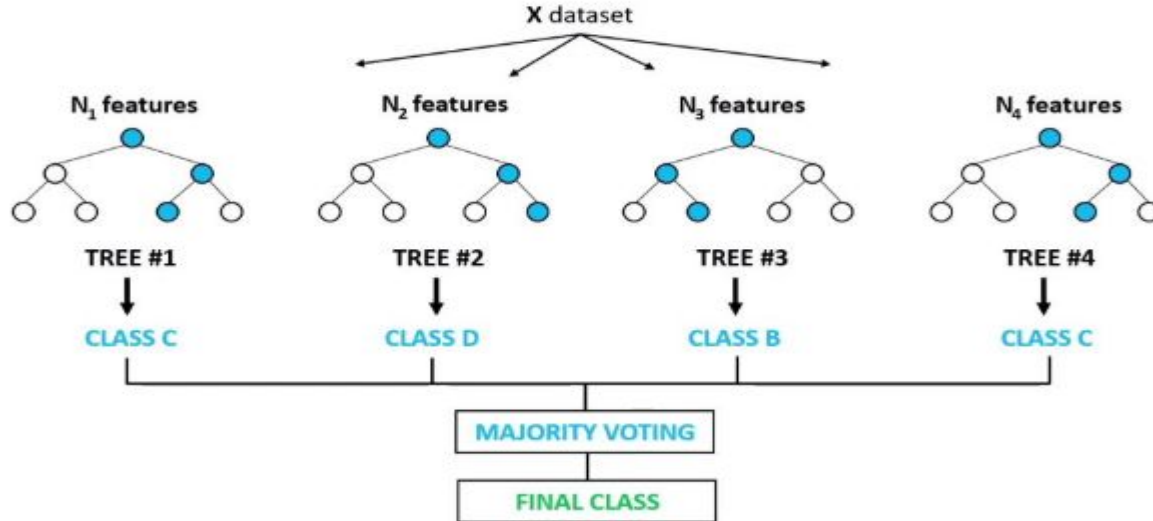
Bootstrap: Selección aleatoria de datos, pero con reemplazo (1990). Reduce la varianza

```
def train(DB, Number_Trees):  
    for i in range(Number_Trees):  
        x_train, y_train = SelectRandomReplace(DB, 70)  
        T[i] = new D_Tree  
        T[i].train(x_train, y_train)  
    return T
```

```
def predict(Trees, x_i):  
    for i in range(Number_Trees):  
        clases = [T.predict(x_i) for T in Trees]  
    return majority_vote(clases)
```

Random Forest: Bagging + decision trees + extra randomness

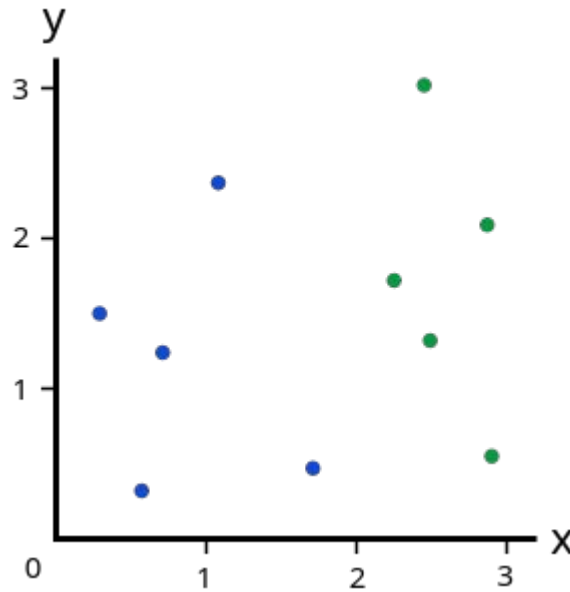
Random Forest Classifier



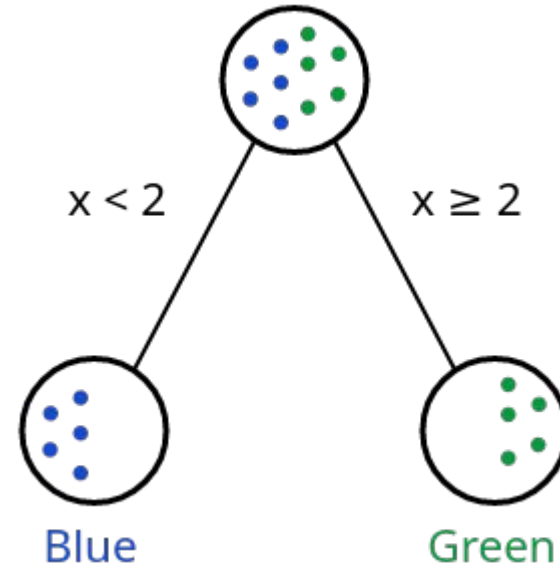
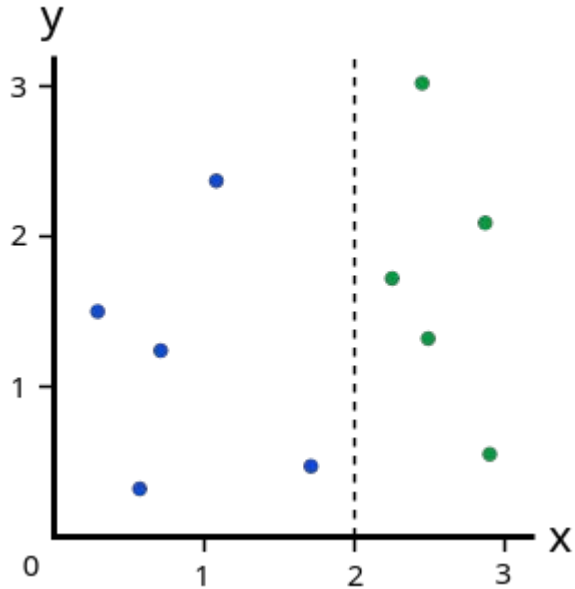
Gracias



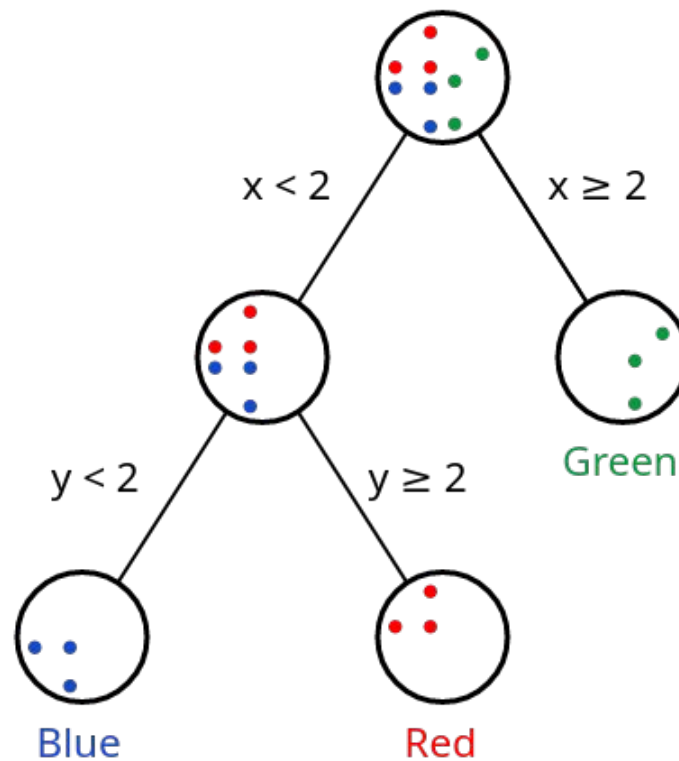
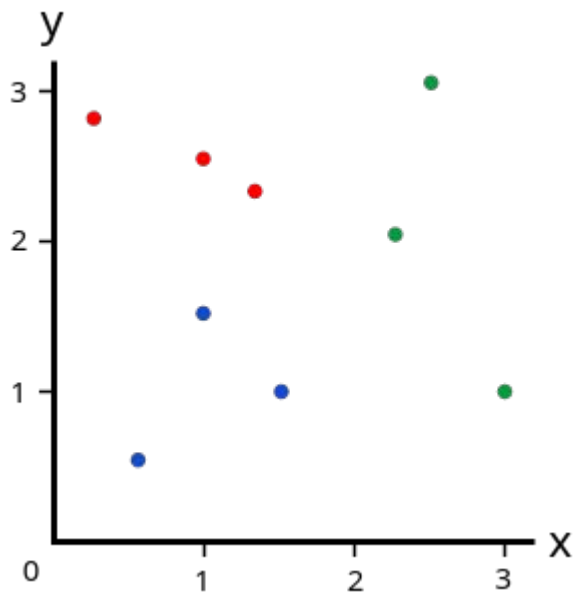
Árbol de Decisión: Regresión



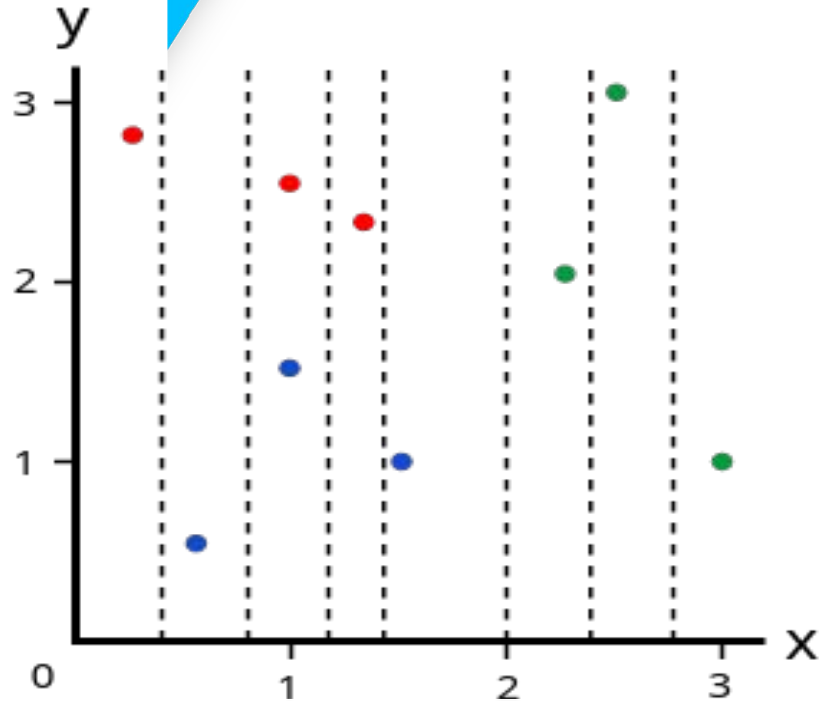
Árbol de Decisión



Árbol de Decisión

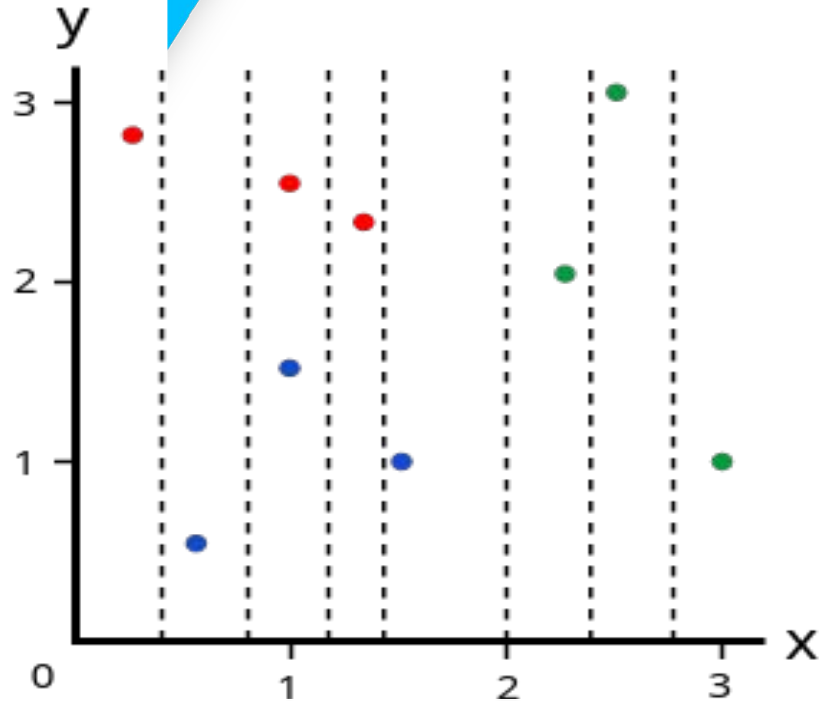


Árbol de Decisión



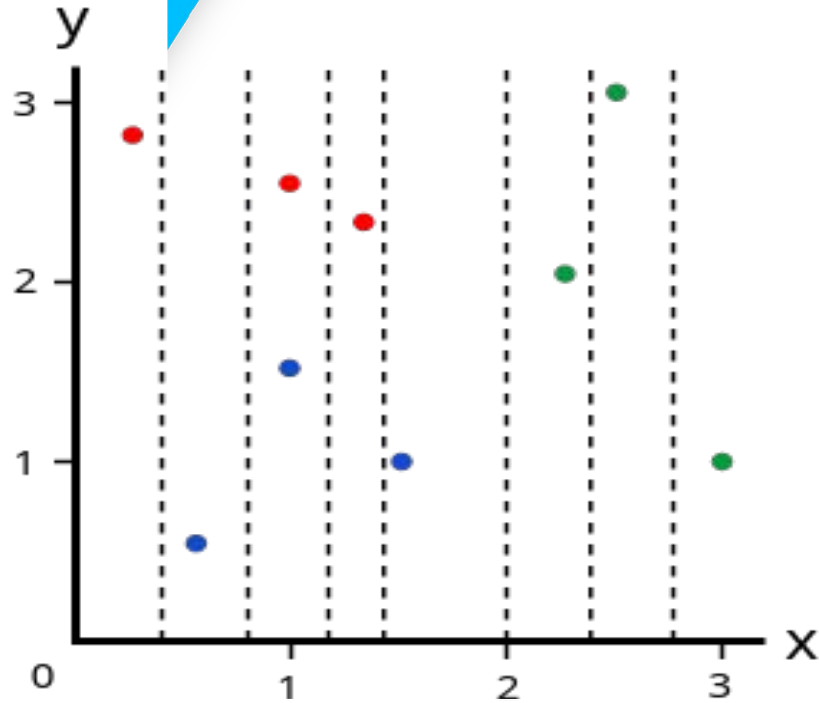
¿Cómo puedes el mejor split?

Árbol de Decisión



¿Cómo puedes el mejor split?

Árbol de Decisión



¿Cómo puedes el mejor split?