

Universidad Autónoma del Carmen

Facultad
Facultad de Ciencias de la Información

Programa educativo
Ingeniería en sistemas computacionales

Nombre de los alumnos
Ángel Ernesto Bailón Itzá [170714]

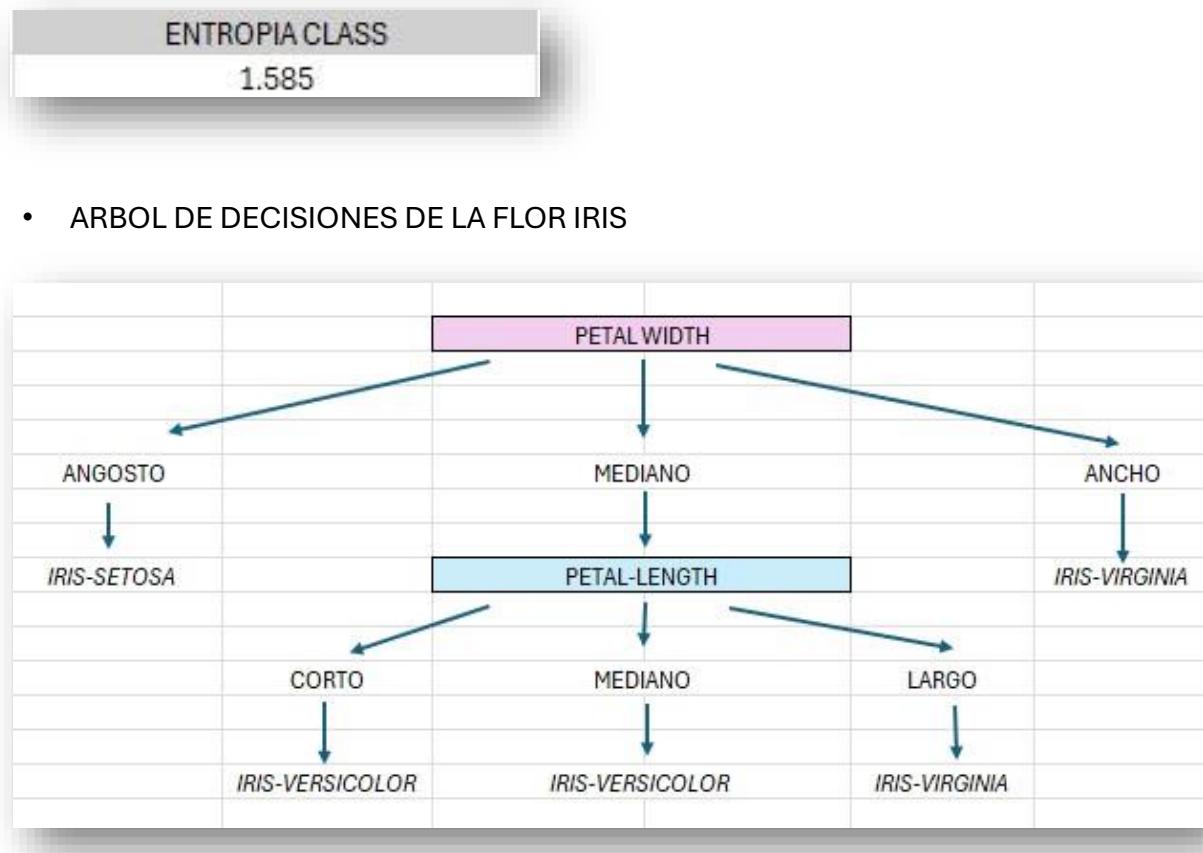
Curso
Programación avanzada

Horario
Martes y jueves de 1:00 pm – 3:00 pm

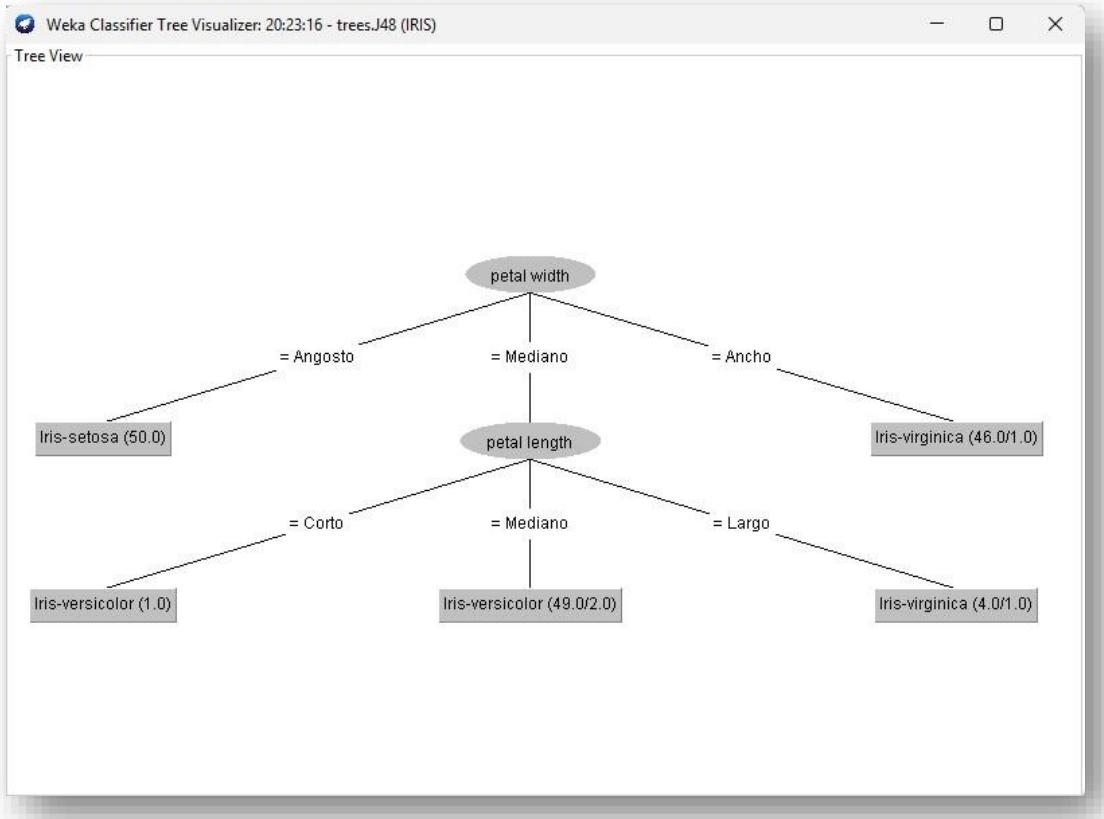
Nombre del docente
Jesús Alejandro Flores

- TABLA DISCRETIZADA

- ENTROPIA DEL ATRIBUTO CLASS



- ARBOL HECHO POR EL SOFTWARE WEKA



Ejemplos de realización (Iris):

Para la tabla de iris se implemento la formula de naive para determinar la clase más probable donde se encuentre el petalo, determinado por la probabilidad evidenciada.

Formula Naive Bayes

Dado que $P(x_1, \dots, x_n)$ es constante, Tomemos $p_x = P(x_1, \dots, x_n)$

Vemos que al comparar 2 valores (a/p_x) y (b/p_x) , como p_x es positivo y constante $(a/p_x) > (b/p_x)$ implica que $a > b$ dada la entrada. Por tanto como:

- $P(X)$ es la **probabilidad de la evidencia** (los datos de entrada)
- Es la **misma** para todas las clases C_j que estamos comparando
- $P(x)$ No afecta la comparación relativa entre clases y lo podemos quitar de la formula

$$P(c|x_1, \dots, x_n) = P(c) * \prod_{i=1}^n P(x_i|c)$$

Seguido de eso, se implementa la formula del calculo de las frecuencia de la clase en la que se encuentra.

$$P(c) = \frac{n_c}{N}$$

Realizado:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
145	145	mediano	chico	mediano	mediano	Iris-versicolor								
146	146	mediano	mediano	mediano	mediano	Iris-versicolor								
147	147	mediano	mediano	mediano	mediano	Iris-versicolor								
148	148	mediano	mediano	mediano	mediano	Iris-versicolor								
149	149	mediano	mediano	mediano	mediano	Iris-versicolor								
150	150	mediano	mediano	mediano	mediano	Iris-versicolor								
151	150	mediano	mediano	mediano	mediano	Iris-versicolor								
152														
153														
154														
155														
156														
157														
158														
159														
160														
161														
162														
163														
164														

Probabilidades de clase P(C)

Datos	150	p
setosa	50	p(setosa)
versicolo	50	p(versicolor)
virginica	50	p(virginica)

Frecuencia de valores de sepalw

sepall	ch	m	g	
clase	setosa	45	5	0
	versicolor	6	30	14
	virginica	1	12	37

Frecuencia de valores de sepalw

sepalw	ch	m	g
	0	0	0
	1	0	0
	2	0	0

Resultados:

a priori	sepall	sepalw	petall	petalw		
0.33	0.90	0.04	1.00	1	0.012	1%
0.33	0.90	0.04	1.00	0	0.000	0%
0.33	0.90	0.04	1.00	0	0.000	0%
0.33	0.90	0.04	0	1	0.000	0%
0.33	0.90	0.04	0	0	0.000	0%
0.33	0.90	0.04	0	0	0.000	0%
0.33	0.90	0.04	0	1	0.000	0%
0.33	0.90	0.04	0	0	0.000	0%
0.33	0.90	0.04	0	0	0.000	0%
0.33	0.90	0.32	1.00	1	0.096	10%
0.33	0.90	0.32	1.00	0	0.000	0%
0.33	0.90	0.32	1.00	0	0.000	0%
0.33	0.90	0.32	0	1	0.000	0%
0.33	0.90	0.32	0	0	0.000	0%
0.33	0.90	0.32	0	0	0.000	0%
0.33	0.90	0.32	0	1	0.000	0%
0.33	0.90	0.32	0	0	0.000	0%
0.33	0.90	0.32	0	0	0.000	0%
0.33	0.90	0.64	1.00	1	0.192	19%
0.00	0.00	0.00	0.00	0	0.000	0%

Referencias:

- <https://github.com/jflorespampano/actividades-ia/blob/main/se-actividades/actividad01.arboles/arboles/Naive.Bayes.pdf>