

4) a)

Proporción de los ejemplos de cada clase:

$$P(SI) = 9/16$$

$$P(NO) = 7/16$$

$$\text{Entropía}(E) = - (9/16) * \log_2(9/16) - (7/16) * \log_2(7/16) = 0.9886$$

RAÍZ

Cálculos para el título universitario

	P(SI)	P(NO)
SI	3/3	0/3
NO	6/13	7/13

$$\text{Entropía}(E_{SI}) = -(3/3) * \log_2(3/3) = 0$$

$$\text{Entropía}(E_{NO}) = -(6/13) * \log_2(6/13) - (7/13) * \log_2(7/13) = 0.9957$$

$$P(SI) = 3/16$$

$$P(NO) = 13/16$$

$$\text{Entropía}(E, \text{Título universitario}) = 3/16 * 0 + 13/16 * 0.9957 = 0.8090$$

Cálculos para experiencia en el cargo

	P(SI)	P(NO)
BAJA	3/5	2/5
MEDIA	5/6	1/6
ALTA	1/5	4/5

$$\text{Entropía}(E_{baja}) = -(3/5) * \log_2(3/5) - (2/5) * \log_2(2/5) = 0.9709$$

$$\text{Entropía}(E_{media}) = -(5/6) * \log_2(5/6) - (1/6) * \log_2(1/6) = 0.6500$$

$$\text{Entropía}(E_{alta}) = -(1/5) * \log_2(1/5) - (4/5) * \log_2(4/5) = 0.7219$$

$$P(baja) = 5/16$$

$$P(media) = 6/16$$

$$P(alta) = 5/16$$

Entropía (E, Experiencia en el cargo) = $5/16 * 0.9709 + 6/16 * 0.6500 + 5/16 * 0.7219 = 0.7219$

Cálculos para trabaja actualmente

	P(SI)	P(NO)
SI	3/3	0/3
NO	6/13	7/13

Entropía(Esi) = $-(1/5) * \log_2(1/5) - (4/5) * \log_2(4/5) = 0.7219$

Entropía(Eno) = $-(8/11) * \log_2(8/11) - (3/11) * \log_2(3/11) = 0.8453$

P(SI) = 5/16

P(NO) = 11/16

Entropía (E, Trabaja actualmente) = $5/16 * 0.7219 + 11/16 * 0.8453 = 0.8067$

Cálculos para cantidad de trabajos anteriores

Corte = 7

	P(SI)	P(NO)
<=7	4/9	5/9
>7	5/7	2/7

Entropía(E>7) = $-(5/7) * \log_2(5/7) - (2/7) * \log_2(2/7) = 0.8631$

Entropía(E<=7) = $-(4/9) * \log_2(4/9) - (5/9) * \log_2(5/9) = 0.9910$

P(>7) = 7/16

P(<=7) = 9/16

Entropía (E, Cantidad de trabajos anteriores) = $7/16 * 0.8631 + 9/16 * 0.9910 = 0.9350$

Corte = 8.5

	P(SI)	P(NO)
<=8.5	8/13	5/13
>8.5	1/3	2/3

Entropía(E>8.5) = $-(1/3) * \log_2(1/3) - (2/3) * \log_2(2/3) = 0.9612$

Entropía(E<=8.5) = $-(8/13) * \log_2(8/13) - (5/13) * \log_2(5/13) = 0.9182$

$$P(>7) = 13/16$$

$$P(\leq 7) = 3/16$$

$$\text{Entropía (E, Cantidad de trabajos anteriores)} = 13/16 * 0.9612 + 3/16 * 0.9182 = 0.9531$$

Corte	Valor entropía
7	0.9350
8.5	0.9531

El valor del corte 7 es menor, utilizaremos ese punto para el corte.

$$\text{Ganancia (E, Título universitario)} = 0.9886 - 0.8090 = 0.1796$$

$$\text{Ganancia (E, Experiencia en el cargo)} = 0.9886 - 0.7727 = 0.2159$$

$$\text{Ganancia (E, Cantidad de Trabajos Anteriores)} = 0.9886 - 0.9350 = 0.0536$$

$$\text{Ganancia (E, Trabaja actualmente)} = 0.9886 - 0.8067 = 0.1819$$

El atributo con más ganancia es la experiencia en el cargo, el cual hará de raíz.

	Entropía(E,A)	Ganancia(E,A)
Título universitario	0.8090	0.1796
Experiencia en el cargo	0.7727	0.2159
Cantidad de trabajos anteriores	0.9350	0.0536
Trabaja actualmente	0.8067	0.1819

Resultado parcial:



RAMA ALTA

$$P(SI) = 1/5$$

$$P(NO) = 4/5$$

$$\text{Entropía}(E) = - (1/5) * \log_2(1/5) - (4/5) * \log_2(4/5) = 0.7219$$

Cálculos para el título universitario

	P(SI)	P(NO)
SI	1/1	0/1
NO	0/4	4/4

$$\text{Entropía}(E_{SI}) = -(1/1) * \log_2(1/1) = 0$$

$$\text{Entropía}(E_{NO}) = -(4/4) * \log_2(4/4) = 0$$

$$P(SI) = 1/5$$

$$P(NO) = 4/5$$

$$\text{Entropía}(E, \text{Título universitario}) = 1/5 * 0 + 4/5 * 0 = 0$$

Cálculos para trabaja actualmente

	P(SI)	P(NO)
SI	0/2	2/2
NO	1/3	2/3

$$\text{Entropía}(E_{SI}) = -(2/2) * \log_2(2/2) = 0$$

$$\text{Entropía}(E_{NO}) = -(1/3) * \log_2(1/3) - (2/3) * \log_2(2/3) = 0.9182$$

$$P(SI) = 2/5$$

$$P(NO) = 3/5$$

$$\text{Entropía}(E, \text{Trabaja actualmente}) = 2/5 * 0 + 3/5 * 0.9182 = 0.5509$$

Cálculos para cantidad de trabajos anteriores

Corte = 7.5

	P(SI)	P(NO)
≤ 7.5	1/4	3/4
> 7.5	0/1	1/1

$$\text{Entropía}(E > 7.5) = -(1/1) * \log_2(1/1) = 0$$

$$\text{Entropía}(E \leq 7.5) = -(1/4) * \log_2(1/4) - (3/4) * \log_2(3/4) = 0.8112$$

$$P(>7.5) = 1/5$$

$$P(\leq 7.5) = 4/5$$

$$\text{Entropía (E, Cantidad de trabajos anteriores)} = 4/5 * 0.8112 + 1/5 * 0 = 0.6490$$

Cálculos de entropía y ganancia con esta fórmula:

$$\text{Ganancia (E, Título universitario)} = 0.7219 - 0 = 0.7219$$

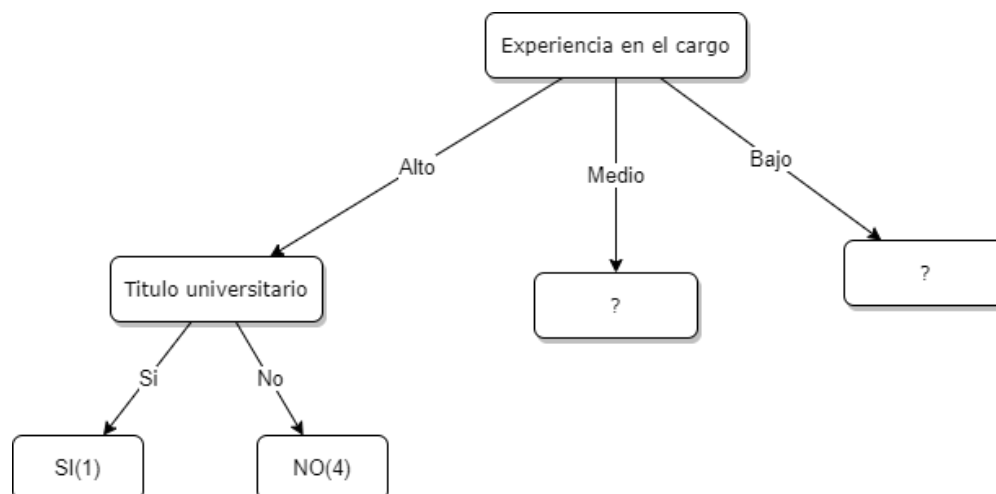
$$\text{Ganancia (E, Experiencia en el cargo)} = 0.7219 - 0.6490 = 0.0729$$

$$\text{Ganancia (E, Trabaja actualmente)} = 0.7219 - 0.5509 = 0.1709$$

El atributo con más ganancia es el título universitario, el cual subdividirá la rama alta

	Entropía(E,A)	Ganancia(E,A)
Título universitario	0	0.7219
Cantidad de trabajos anteriores	0.6490	0.0729
Trabaja actualmente	0.5509	0.1709

Resultado parcial:



RAMA MEDIA

$$P(SI) = 5/6$$

$$P(NO) = 1/6$$

$$\text{Entropía}(E) = - (5/6) * \log_2(5/6) - (1/6) * \log_2(1/6) = 0.6500$$

Cálculos para el título universitario

	P(SI)	P(NO)
SI	0	0
NO	5/6	1/6

Este atributo se descarta debido a que no hay ejemplos si el aspirante tiene título

Cálculos para trabajo actualmente

	P(SI)	P(NO)
SI	0/1	1/1
NO	5/5	0/5

$$\text{Entropía}(E_{SI}) = -(1/1) * \log_2(1/1) = 0$$

$$\text{Entropía}(E_{NO}) = -(5/5) * \log_2(5/5) = 0$$

$$P(SI) = 1/6$$

$$P(NO) = 5/6$$

$$\text{Entropía}(E, \text{Trabajo actualmente}) = 1/6 * 0 + 5/6 * 0 = 0$$

Cálculos para cantidad de trabajos anteriores

Corte = 7

	P(SI)	P(NO)
<=7	1/1	0/1
>7	4/5	1/5

$$\text{Entropía}(E>7) = -(4/5) * \log_2(4/5) - (1/5) * \log_2(1/5) = 0.7219$$

$$\text{Entropía}(E<=7) = -(1/1) * \log_2(1/1) = 0$$

$$P(>7) = 1/6$$

$$P(<=7) = 5/6$$

$$\text{Entropía}(E, \text{Cantidad de trabajos anteriores}) = 1/6 * 0 + 5/6 * 0.7219 = 0.6016$$

Corte = 8.5

	P(SI)	P(NO)
≤ 8.5	4/4	0/4
> 8.5	1/2	1/2

Entropía($E > 8.5$) = $-(1/2) * \log_2(1/2) - (1/2) * \log_2(1/2) = 1$

Entropía($E \leq 8.5$) = $-(4/4) * \log_2(4/4) = 0$

$P(>7) = 4/6$

$P(\leq 7) = 2/6$

Entropía (E, Cantidad de trabajos anteriores) = $4/6 * 0 + 2/6 * 1 = 0.3333$

Corte	Valor entropía
7	0.6016
8.5	0.3333

Como el valor del corte 8.5 es menor utilizaremos ese punto para el corte.

Cálculos de entropía y ganancia con esta fórmula:

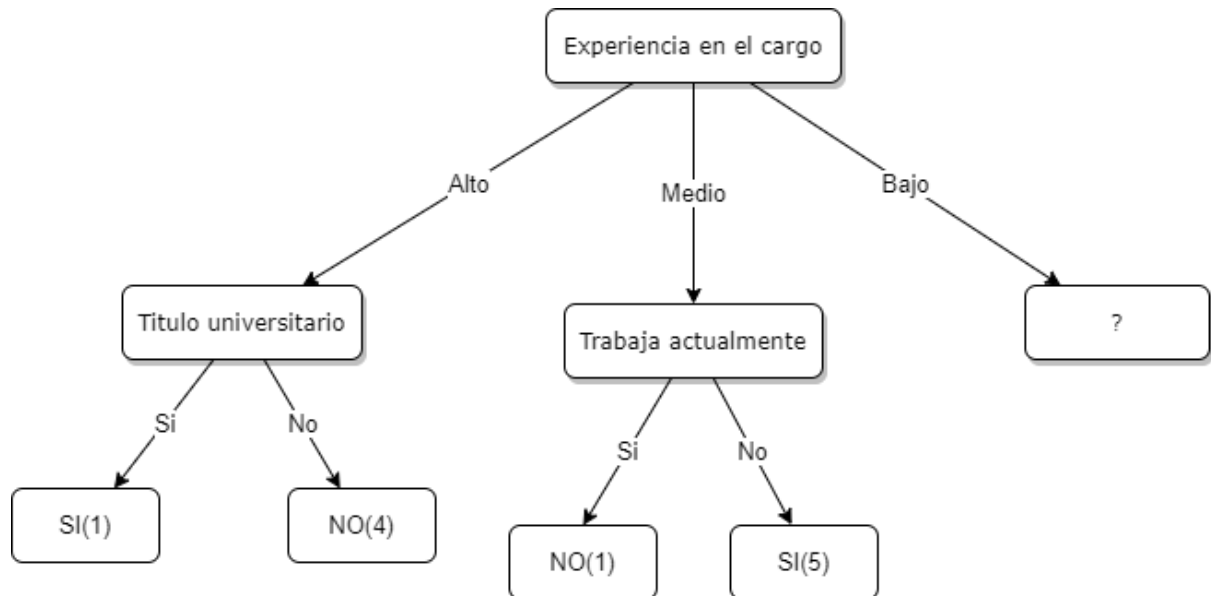
Ganancia (E, Cantidad de trabajos anteriores) = $0.6500 - 0.3333 = 0.3166$

Ganancia (E, Trabaja actualmente) = $0.6500 - 0 = 0.6500$

El atributo con más ganancia es el título universitario, el cual subdividirá la rama media

	Entropía(E,A)	Ganancia(E,A)
Cantidad de trabajos anteriores	0.3333	0.3166
Trabaja actualmente	0	0.6500

Resultado parcial:



RAMA BAJA

$$P(SI) = 3/5$$

$$P(NO) = 2/5$$

$$\text{Entropía}(E) = - (3/5) * \log_2(3/5) - (2/5) * \log_2(2/5) = 0.9709$$

Cálculos para el título universitario

	P(SI)	P(NO)
SI	2/2	0/2
NO	1/3	2/3

$$\text{Entropía}(E_{SI}) = -(2/2) * \log_2(2/2) = 0$$

$$\text{Entropía}(E_{NO}) = -(1/3) * \log_2(1/3) - (2/3) * \log_2(2/3) = 0.9182$$

$$P(SI) = 2/5$$

$$P(NO) = 3/5$$

$$\text{Entropía}(E, \text{Título universitario}) = 2/5 * 0 + 3/5 * 0.9182 = 0.5509$$

Cálculos para trabaja actualmente

	P(SI)	P(NO)
SI	1/2	1/2
NO	2/3	1/3

$$\text{Entropía(Esi)} = -(1/2) * \log_2(1/2) - (1/2) * \log_2(1/2) = 1$$

$$\text{Entropía(Eno)} = -(2/3) * \log_2(2/3) - (1/3) * \log_2(1/3) = 0.9182$$

$$P(SI) = 2/5$$

$$P(NO) = 3/5$$

$$\text{Entropía (E, Trabaja actualmente)} = 2/5 * 1 + 3/5 * 0.9182 = 0.9509$$

Cálculos para cantidad de trabajos anteriores

Corte = 7

	P(SI)	P(NO)
<=7	2/4	2/4
>7	1/1	0/1

$$\text{Entropía(E>7)} = -(1/1) * \log_2(1/1) = 0$$

$$\text{Entropía(E<=7)} = -(2/4) * \log_2(2/4) - (2/4) * \log_2(2/4) = 1$$

$$P(>7) = 1/5$$

$$P(<=7) = 4/5$$

$$\text{Entropía (E, Cantidad de trabajos anteriores)} = 4/5 * 1 + 1/5 * 0 = 0.8$$

Cálculos de entropía y ganancia con esta fórmula:

$$\text{Ganancia (E, Título universitario)} = 0.9709 - 0.5509 = 0.4199$$

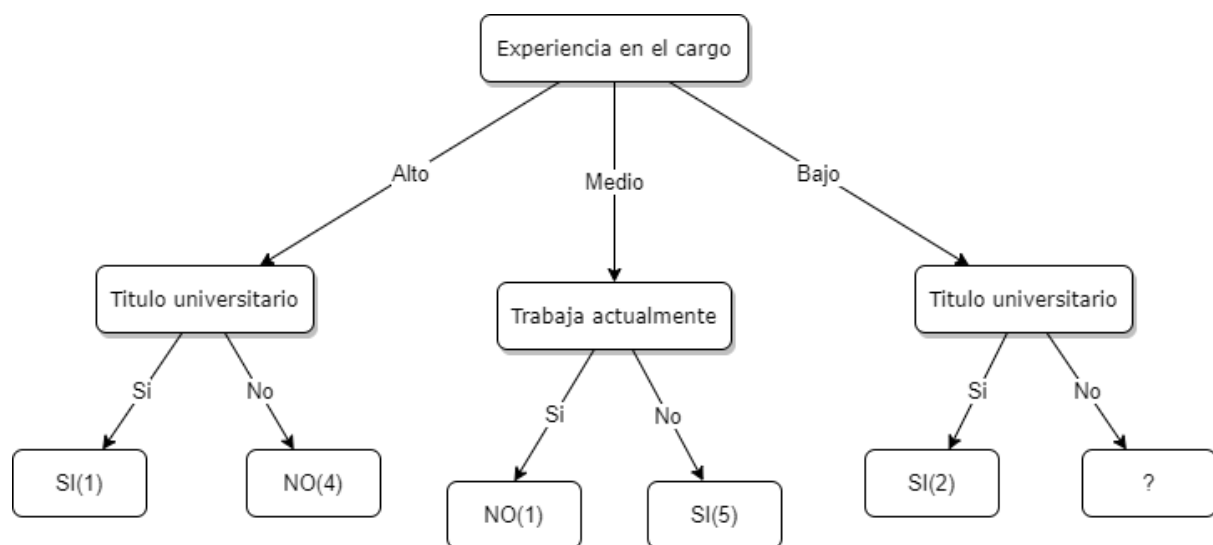
$$\text{Ganancia (E, Cantidad de trabajos anteriores)} = 0.9709 - 0.8 = 0.1709$$

$$\text{Ganancia (E, Trabaja actualmente)} = 0.9709 - 0.9509 = 0.0199$$

El atributo con más ganancia es el título universitario, el cual subdividirá la rama baja

	Entropía(E,A)	Ganancia(E,A)
Título universitario	0.5509	0.4199
Cantidad de trabajos anteriores	0.8	0.1709
Trabaja actualmente	0.9509	0.0199

Resultado parcial:



RAMA BAJA - SIN TÍTULO UNIVERSITARIO

$$P(SI) = 1/3$$

$$P(NO) = 2/3$$

$$\text{Entropía}(E) = - (1/3) * \log_2(1/3) - (2/3) * \log_2(2/3) = 0.9182$$

Cálculos para trabaja actualmente

	P(SI)	P(NO)
SI	0/1	1/1
NO	1/2	1/2

$$\text{Entropía}(E_{Si}) = -(1/1) * \log_2(1/1) = 0$$

$$\text{Entropía}(E_{No}) = -(1/2) * \log_2(1/2) - (1/2) * \log_2(1/2) = 1$$

$$P(SI) = 1/3$$

$$P(NO) = 2/3$$

$$\text{Entropía (E, Trabaja actualmente)} = 1/3 * 0 + 2/3 * 1 = 0.6666$$

Cálculos para cantidad de trabajos anteriores

Corte = 7

	P(SI)	P(NO)
≤ 7	0/2	2/2
> 7	1/1	0/1

$$\text{Entropía}(E > 7) = -(1/1) * \log_2(1/1) = 0$$

$$\text{Entropía}(E \leq 7) = -(2/2) * \log_2(2/2) = 0$$

$$P(>7) = 2/3$$

$$P(\leq 7) = 1/3$$

$$\text{Entropía (E, Cantidad de trabajos anteriores)} = 2/3 * 0 + 1/3 * 0 = 0$$

Cálculos de entropía y ganancia con esta fórmula:

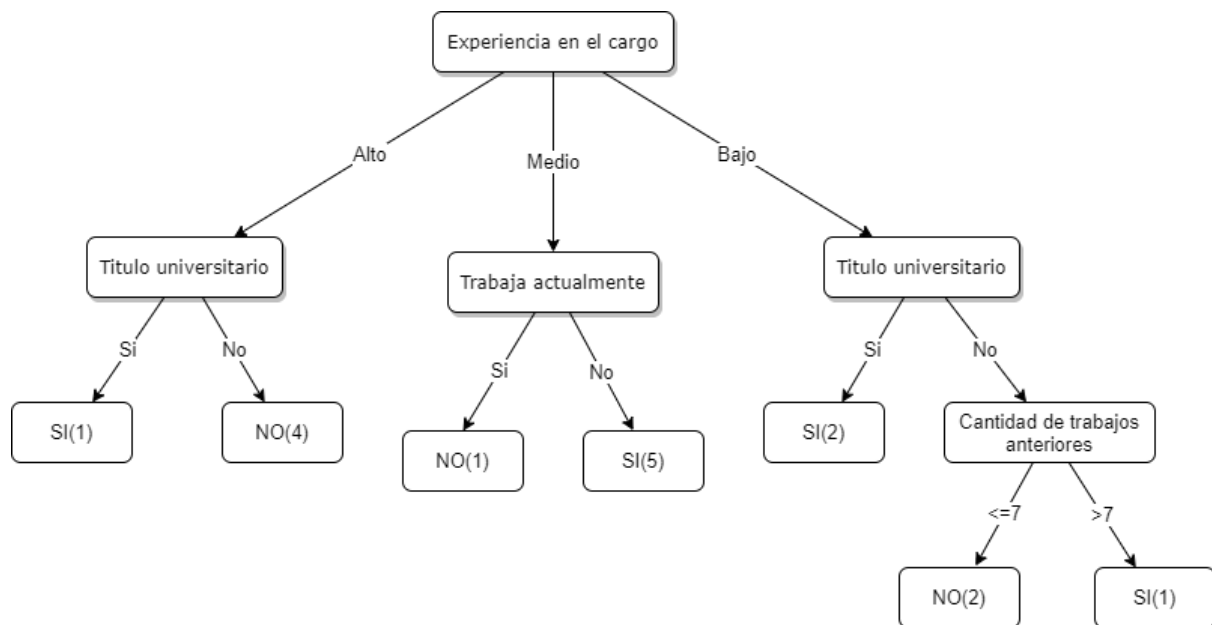
$$\text{Ganancia (E, Cantidad de trabajos anteriores)} = 0.9182 - 0 = 0.9182$$

$$\text{Ganancia (E, Trabaja actualmente)} = 0.9182 - 0.6666 = 0.2516$$

El atributo con más ganancia es el título universitario, el cual subdividirá la rama baja

	Entropía(E,A)	Ganancia(E,A)
Cantidad de trabajos anteriores	0	0.9182
Trabaja actualmente	0.6666	0.2516

Arbol completo:



4) c)

Título universitario	Experiencia en el cargo	Cantidad de Trabajos Anteriores	Trabaja actualmente	Obtiene trabajo
SI	ALTA	8	SI	SI
NO	MEDIA	6	SI	NO
NO	BAJA	2	NO	NO
NO	MEDIA	5	NO	SI

6-

Inciso		a)	b)	c)
minimal size for split		40	100	300
minimal leaf size		20	50	200
Tamaño del árbol		84	32	4
Accuracy	Train	79.59%	73.08%	71.06%
	Test	53.25%	57.14%	54.55%

d) El tamaño del árbol va decreciendo a medida que se incrementan los parámetros.

El conjunto de datos de entrenamiento tiene una mejor precisión que el test, esto es lo más lógico ya que los datos de entrenamiento es lo que se utiliza para hacer las predicciones.

A medida que los parámetros crecen la precisión podría decrementarse ya que se crea un árbol un poco más general, aunque esto siempre va a depender de los datos (con los datos de entrenamiento sucede).