

1. Indique cuales son las etapas del proceso de descubrimiento de conocimiento. ¿Qué características tiene el proceso?
2. *El método de binning para eliminación de ruido es supervisado.* Indique si esa afirmación es V o F y justifique.
3. *El método de binning es un método multivariado para la eliminación de ruido.* Indique si esa afirmación es V o F y justifique.
4. Dado $X = \{5, 5, 5, 5, 5, 6, 6, 15, 15, 16, 16, 48\}$. Realice una discretización de igual ancho con $k = 3$. Indicar si la solución es óptima y justifique su respuesta.
5. *El método hot-deck es el método adecuado de imputación de faltantes de variables con dominio continuo en \mathbb{R} .* Indique si esa afirmación es V o F y justifique.
6. *La normalización con Z-Score reduce el sesgo de las variables escaladas.* Indique si esa afirmación es V o F y justifique.
7. En una estrategia de reducción de datos el filtro de varianzas es un método seguro para cualquier tipo de variables. Indique si esa afirmación es V o F y justifique.
8. Dada la siguiente tabla de datos:

	A	B	C	D
I1	2	0	0	0
I2	0	0	0	1
I3	0	4	3	0
I4	0	0	1	0

- A. ¿Qué tipo de almacenamiento NoSQL es aplicable?
- B. De un ejemplo de la implementación del tipo seleccionado en A)

9. Para el siguiente conjunto de transacciones responda:

TID	Items
1	Y, X
2	X, L, C, E
3	M, C, L, Y
4	X, L, C, Y
5	C, Y, X, M

- A. ¿Cuál debería ser el min_sup para que todos los 1-itemsets sean frecuentes?
- B. Dado un $min_sup = 3$, de un ejemplo de k-itemsets frecuentes para $k \in \{2, 3\}$.

10. Si un itemset no es frecuente el principio de Apriori garantiza que todos sus superset serán frecuentes. Indique si esa afirmación es V o F y justifique.
11. Si aplica el algoritmo de Apriori para la tabla de transacciones del punto 8 y considerando un $min_sup = 3$. A) ¿Cuál sería la lista de C_2 ? B) ¿Cuál sería la lista de L_2 ? C) ¿Cómo queda L_3 ?
12. El itemset $\{YC\}$ es un maximal frequent itemset. Indique si esa afirmación es V o F y justifique.
13. Indique si la regla $X \Rightarrow C$ es frecuente. Justifique.
14. Para un $min_sup = 3$ realice el FP_Tree de las transacciones del punto 8)
15. Para el siguiente conjunto de secuencias responda:

SID	Secuencias
10	$\langle a(abc)(ac)d(cf) \rangle$
20	$\langle (ad)c(bc)(ae) \rangle$
30	$\langle (ef)(ab)(df)cb \rangle$
40	$\langle eg(af)cbc \rangle$

- A. Encuentre una supersecuencia de $\langle f(ab)c \rangle$ en la base de secuencias.
- B. Para un $min_sup = 2$ encuentre dos patrones secuenciales.



Data Mining

Examen Comisión 1 - 17/07/2018

Apellido y Nombre: _____