

PEP N°2 2022-2

1. En el dominio de las reglas de asociación, explique para que son las medidas de calidad. Determine la monotonicidad de **Medida** e identifique a qué medida de las vistas en clases corresponde. **(1.2)**

$$Medida(A \Rightarrow B) = n(\hat{P}(A \cap B) - \hat{P}(A) \hat{P}(B))$$

2. Para determinar el potencial de germinación de una semilla de girasol, se consideran dos variables. Si germinó en tiempo adecuado su progenitor y el área frontal en $[mm^2]$. En la tabla adjunta, se muestran los valores de 12 diferentes semillas, que fue probada su germinación, cuyo resultado corresponde a la columna "Germinó". Estime las probabilidades para diseñar un clasificador Bayesiano Ingenuo que permitan determinar si una nueva semilla germinará o no. Determine si una semilla de 25 de área frontal y su ancestro NO germinó, ¿podrá germinar? **(1.8)**

Germinó Progenitor	Área Semilla	Germinó
Si	40	Si
Si	20	No
No	28	No
No	47	Si
Si	37	Si
No	32	Si
Si	41	Si
No	38	Si
Si	15	No
Si	27	No
No	38	Si
No	42	Si

3. Para el problema anterior, considere que se reconocen tres tipos de áreas de las semillas, menores que 26 $[mm^2]$, entre 27 y 36 $[mm^2]$ y mayores que 37 $[mm^2]$. Utilice el método de árboles de decisión para determinar el atributo (**germinó el progenitor o área**) es más eficiente para clasificar accidentes. **(2.0)**
4. Explique el concepto de poda pesimista en el algoritmo C4.5 propuesto por Quinlan. **(1.0)**

- Ingenuo: $p(c_i / a_i) = p(c_i) \prod_j p(a_j / c_i)$; para $a_i = x$ continua $p(x / c_i) = e^{-\frac{(x-\bar{x})^2}{2\sigma^2}} / \sigma\sqrt{2\pi}$

- Árboles de decisión: $Ganancia(V) = -\sum_i p(c_i) \log p(c_i) + \sum_i p(v_i) \sum_j p(c_j / v_i) \log p(c_j / v_i)$;

$RazonGanancia(V) = -Ganancia(V) / \sum_i p(v_i) \log p(v_i)$.