

Examen SAA

Nombre: _____

1. En la práctica se eliminan los atributos "Name", "Investigator or Source", "pdf", "href formula", "href", "original order". ¿Por qué crees que no se tienen en cuenta?

¿Crees que se podría sacar algún tipo de información de alguno de estos atributos?
¿Por qué?
2. ¿Por qué sacamos el mes y el día de "Case Number" y no de "Date"?
3. En el caso de ds_USA['Day'] y ds_USA['Month'], el valor 0 se corresponde con datos que no se conocen, ¿cómo utilizarías un algoritmo para inferir estos datos?
4. La variable ds_USA['Area'] contiene un valor nulo. En la práctica se ha optado por eliminar el registro. ¿Se podría haber rellenado con la media, la mediana o la moda? Razona la respuesta y escribe el comando correspondiente.
5. ¿En la práctica aparecen outliers?, explícalo.
6. ¿Crees que el dataset está balanceado? Explícalo.
7. En la práctica, ¿qué normalizaciones de datos aparecen?
8. Al aplicar el método de la Varianza aparecen magnitudes muy diferentes, ¿a qué crees que es debido?
9. ¿Con qué atributos nos quedaremos si aplicamos Correlación? Explica la respuesta
10. En Backward Elimination, ¿hay algún error en el código? Explícalo
11. ¿Qué es la discretización y el 1 hot encoding? Pon un ejemplo para ambos casos aplicado a la práctica.
12. ¿Qué diferencias hay entre PCA1 y PCA2? ¿Con qué variables nos quedaríamos?

NOTA: Todas las preguntas puntúan lo mismo.