Logotipo, nombre de la empresa

Descripción generada automáticamente

Universidad de Montevideo – Introducción a la Ciencia de Datos

Examen 2022

03 de Febrero, 2022

1. Por qué cree que no debe utilizarse el método de validación K-Fold Cross-Validation cuando se tiene una serie de tiempo? Cómo cree que deberían dividirse los conjuntos de entrenamiento y validación?
2. Por qué es necesario el pre-procesamiento o limpieza de datos?
3. Cual es la desventaja principal del modelo Bag of Words?
4. Cómo determinaría el número correcto de clusters en un problema no supervisado?
5. Qué cree que sucederá con la performance de un modelo que fue entrenado mediante validación de resustitución a la hora de enfrentarse a datos reales? Por qué sucede este fenómeno?
6. Cuál es la diferencia entre aprendizaje supervisado y no supervisado?
7. Qué es el p-value y cuál es su utilidad?
8. Qué significa que una serie de tiempo sea estacionaria?
9. Cómo se calcula el precision y recall de un modelo y qué significado tiene?
10. Cuál es la diferencia entre un box plot y un histograma? Explique sus usos.