



PROCESO SELECTIVO POR EL SISTEMA DE PROMOCIÓN INTERNA PARA INGRESO EN LA ESCALA DE TÉCNICOS SUPERIORES ESPECIALIZADOS DE LOS ORGANISMOS PÚBLICOS DE INVESTIGACIÓN, CONVOCADO POR RESOLUCIÓN DE 24 DE MAYO DE 2024 (BOE N° 131 DE 30 DE MAYO)

Cuestionario del primer ejercicio

A6 - CIENCIA DE DATOS

Especialidad: D3 - Modelado y Análisis de Datos

- No abra el **CUESTIONARIO** ni empiece el examen hasta que se le indique.
- Solo se calificarán las respuestas marcadas en la **HOJA DE RESPUESTAS**
- El cuestionario consta de **80 preguntas**, cada una de ellas con **cuatro respuesta alternativas**, de las cuales **sólo una de ellas es correcta**.
- Una vez abierto el cuestionario, compruebe que consta de todas las páginas y preguntas y que sea legible. En caso contrario solicite uno nuevo al personal del aula.
- Las **contestaciones erróneas se PENALIZARÁN** con un 25 % de su valoración.
- Lea atentamente las **instrucciones** para contestar la **HOJA DE RESPUESTAS**, que figuran al dorso de la misma.
- Cumplimente los datos personales y firme la **HOJA DE RESPUESTAS**.
- El tiempo para la realización de este ejercicio será de **noventa (90) minutos**.
- **NO SEPRE** ninguna de las copias de la **HOJA DE RESPUESTAS**. Una vez finalizado el ejercicio, el personal del aula le indicará los pasos a seguir.
- El **CUESTIONARIO** se podrá utilizar como borrador y se podrá llevar por el opositor al finalizar el tiempo marcado para el ejercicio.

Barcelona 20 de septiembre de 2024

1. Indicar de las siguientes características cuál NO se corresponde con el modelo de computación en la nube:

- A) Acceso bajo demanda.
- B) Flexibilidad en la configuración de los recursos.
- C) Elasticidad y rapidez en el aprovisionamiento de recursos.
- D) Dependencia de recursos físicos locales.

2. La virtualización de recursos consiste en:

- A) El conjunto de técnicas software que permiten abstraer las características físicas de un computador, de forma que se puedan ejecutar diferentes máquinas virtuales.
- B) La utilización de máquinas virtuales hardware para proporcionar acceso al software existente en un sistema para un grupo virtual de usuarios.
- C) La explotación de organizaciones virtuales para el aprovisionamiento de recursos de computación a gran escala, especialmente en colaboraciones científicas.
- D) La sustitución de los recursos humanos de una organización por asistentes basados en inteligencia artificial.

3. Indicar cuál de los siguientes conceptos NO es una propiedad de seguridad de un activo de información en una organización.

- A) Confidencialidad.
- B) Integridad.
- C) Migralidad.
- D) Autenticidad.

4. Indicar cuál de los siguientes es un motor de bases de datos relacional:

- A) MariaDB.
- B) MongoDB.
- C) Apache Cassandra.
- D) Amazon DynamoDB.

5. ¿Cuál de las siguientes características es fundamental en el desarrollo de software abierto?

- A) Código cerrado y restrictivo.
- B) Licencias propietarias.
- C) Colaboración y transparencia.
- D) Propiedad exclusiva del autor.

6. ¿Cuál de las siguientes afirmaciones es correcta acerca del aprendizaje automático (machine learning)?

- A) El aprendizaje automático se refiere únicamente a la capacidad de las máquinas para procesar grandes volúmenes de datos.
- B) El aprendizaje automático es un enfoque que permite a las máquinas aprender y mejorar automáticamente a través de la experiencia sin ser programadas explícitamente.
- C) El aprendizaje automático solo se aplica en la investigación científica y no tiene aplicaciones en otros campos.
- D) El aprendizaje automático es una técnica exclusiva de la programación orientada a objetos.

7. ¿Qué es un Data Management Plan (Plan de Gestión de Datos)?

- A) Un plan para la eliminación de datos obsoletos y sin valor.
- B) Un plan para asegurar que los datos sean inaccesibles y no puedan ser utilizados.
- C) Un plan que establece cómo se gestionarán, organizarán, almacenarán y compartirán los datos durante un proyecto de investigación.
- D) Un plan que prioriza la recopilación de datos sin tener en cuenta su calidad o relevancia.

8. ¿Cuál de los siguientes métodos es un ejemplo de accounting en los Sistemas de Autorización, Autenticación y Accounting?

- A) Control de acceso físico.
- B) Registro de tiempos de sesión.
- C) Cifrado de datos.
- D) Detección de intrusiones.

9. ¿Qué es GitHub?

- A) Una plataforma de desarrollo colaborativo.
- B) Un sistema operativo .
- C) Un lenguaje de programación.
- D) Un motor de búsqueda.

10. En un CPD de un centro de investigación se va a adquirir un Sistema de Alimentación Ininterrumpida (SAI) y tenemos las siguientes especificaciones: 500 VA (Voltiamperios) y 300 W (vatios), que indican:

- A) La potencial real y la potencia aparente respectivamente del SAI.
- B) La energía eléctrica y la potencia que puede suministrar el SAI.
- C) La potencia aparente y la potencia real del SAI.
- D) La potencia del SAI solo que dados en dos unidades diferentes de potencia.

11. En el contexto de seguridad de sistemas informáticos, la autenticación:

- A) Otorga certificados digitales a los recursos que están disponibles.
- B) Confirma la disponibilidad de los recursos.
- C) Confirma que los usuarios son quienes dicen ser validando su identidad.
- D) Proporciona permisos de acceso a los datos almacenados en los recursos.

12. ¿Cuál de estos lenguajes proporciona una plataforma unificada y un modelo de programación que permite la programación paralela de GPUs, CPUs, DSPs y FPGAs?

- A) OpenCL.
- B) OpenMP.
- C) CUDA.
- D) SYCL.

13. ¿Cuál de las siguientes afirmaciones sobre XML es correcta?

- A) XML es un modelo de metadatos utilizado para describir la presentación de una página web
- B) XML es un lenguaje de programación utilizado para crear páginas web interactivas.
- C) XML es un lenguaje de marcado que permite definir etiquetas personalizadas para estructurar y almacenar datos.
- D) XML es un formato de imagen utilizado para gráficos y multimedia en la web

14. Indica cuál de estos sistemas operativos es de tiempo real

- A) QNX
- B) Fedora Linux
- C) Windows RT
- D) Minix

15. ¿Cuál de las siguientes afirmaciones describe mejor el concepto de escalabilidad en la computación científica en la nube?

- A) Se refiere a la capacidad de almacenar grandes cantidades de datos científicos en la nube, en escalas mayores del Petabyte.
- B) Implica la posibilidad de acceder a recursos de cómputo adicionales según las necesidades del usuario.
- C) Significa utilizar software especializado en la nube para realizar cálculos científicos complejos.
- D) Se refiere a la colaboración y el intercambio de datos científicos entre diferentes investigadores en las diferentes escalas o etapas de la investigación.

16. De entre las siguientes funcionalidades, indica cuál NO está relacionada con la gestión ágil de proyectos:

- A) Diagramas de Gantt interactivos.
- B) Tableros Kanban.
- C) Hojas de ruta de commits automatizados.
- D) Planificación de minitrabajos Sprints.

17. ¿Qué función principal tiene un sistema de control de versiones en un repositorio digital de software?

- A) Almacenar y organizar los archivos de código fuente.
- B) Permitir a los desarrolladores colaborar y trabajar en equipo en el mismo proyecto.
- C) Gestionar y rastrear los cambios realizados en los archivos a lo largo del tiempo.
- D) Compilar y ejecutar el software en diferentes entornos de desarrollo.

18. En la programación orientada a objetos, una clase es:

- A) Una plantilla en la que se definen los atributos y métodos predeterminados de un tipo de objeto.
- B) Una propiedad que permite subdividir una aplicación en partes más pequeñas e independientes de la aplicación en sí.
- C) Una técnica por la cual el entorno de objetos se encarga de destruir automáticamente los objetos no referenciados.
- D) Una ubicación de almacenamiento abstracta asociada a un nombre simbólico, que contiene una cantidad de información conocida.

19. ¿En qué consiste el aprendizaje no supervisado?

- A) Un tipo de aprendizaje en el que no se utilizan etiquetas para entrenar el modelo.
- B) Un tipo de aprendizaje que solo se aplica a problemas de regresión.
- C) Un tipo de aprendizaje que utiliza redes neuronales recurrentes (RNN).
- D) Un tipo de aprendizaje que utiliza regresión logística.

20. ¿Cuál de las siguientes opciones describe qué es Big Data?

- A) Una cantidad de datos que es demasiado grande para ser procesada por sistemas tradicionales.
- B) Un conjunto de herramientas utilizadas para el procesamiento de datos en tiempo real.
- C) Una base de datos de gran tamaño.
- D) Una técnica para el análisis de datos utilizando algoritmos de inteligencia artificial.

21. ¿Qué es una TPU?

- A) Es una unidad de procesamiento gráfico.
- B) Es una unidad de procesamiento central.
- C) Es una unidad de procesamiento tensorial.
- D) Es una unidad de procesamiento vectorial.

22. El aprendizaje automático se ha visto revolucionado en los últimos años por la evolución de distintas formas de computación basada en redes neuronales. El factor clave para el uso extendido de estos nuevos algoritmos ha sido

- A) El rápido acceso a disco.
- B) La eficiente implementación en Unidades Centrales de Procesamiento.
- C) La eficiente implementación en Unidades Gráficas de Procesamiento.
- D) La reducción de la latencia en la transmisión de información.

23. La principal característica de un sistema de tiempo real es:

- A) Disponer de un planificador basado en prioridades.
- B) Respuesta a interrupciones con niveles de prioridad.
- C) Respuesta a los eventos dentro de un plazo determinado.
- D) Ausencia de gestor de memoria virtual.

24. El soporte y coordinación para el tratamiento de vulnerabilidades y la resolución de incidentes de seguridad que tengan las Administraciones de las Comunidades Autónomas corresponde a:

- A) INCIBE-CERT.
- B) CCN-CERT.
- C) Al CERT correspondiente de la Comunidad Autónoma.
- D) Al CERT del Ministerio de Defensa.

25. ¿Cuál es el protocolo más común para iniciar una sesión remota segura mediante línea de comandos a una máquina Linux?

- A) SSH.
- B) Telnet.
- C) RDP.
- D) VPN.

26. ¿Cuál de las siguientes técnicas se utiliza para reducir la dimensionalidad de los datos en aprendizaje profundo?

- A) Data augmentation (aumento de datos).
- B) Reducción de ruido.
- C) Redes neuronales convolucionales (CNN).
- D) Análisis de componentes principales (PCA).

27. ¿Cuál de las siguientes bibliotecas NO es ampliamente utilizada para implementar deep learning en Python?

- A) TensorFlow.
- B) Scikit-learn.
- C) PyTorch.
- D) Wamba.

28. ¿Qué tipo de estructura tienen los documentos XML?

- A) Jerárquica, organizados en una estructura de árbol con elementos anidados.
- B) Lineal, con una sola línea de etiquetas que definen el contenido del documento.
- C) Matricial, con una matriz de celdas que contienen los datos y metadatos.
- D) Circular, donde cada elemento hace referencia a todos los demás elementos.

29. ¿Cuál de las siguientes afirmaciones describe, define o delimita correctamente al Método de Monte-Carlo:

- A) Se trata de un método de resolución numérica de integración
- B) Se aplicó por primera vez al resolver las ecuaciones del movimiento uniformemente acelerado
- C) Es un método no determinista o estadístico numérico usado para aproximar expresiones matemáticas complejas y costosas de evaluar con exactitud
- D) Su ámbito de aplicación es solo en Física Cuántica

30. ¿Cuál de las siguientes opciones describe correctamente qué es el aprendizaje por refuerzo en el contexto del aprendizaje profundo?

- A) El aprendizaje por refuerzo es un enfoque de entrenamiento en el que el modelo se entrena utilizando técnicas de agrupamiento para clasificar los datos en diferentes categorías.
- B) El aprendizaje por refuerzo es un enfoque de entrenamiento en el que el modelo se entrena sin tener información sobre los resultados deseados.
- C) El aprendizaje por refuerzo es un enfoque de entrenamiento en el que el modelo se entrena utilizando ejemplos de entrada y salida etiquetados, para aprender a predecir la salida correcta.
- D) El aprendizaje por refuerzo es un enfoque de entrenamiento en el que el modelo se entrena mediante interacción con un entorno, recibiendo recompensas o castigos según su desempeño.

31. ¿Cómo funciona el método bootstrap?

- A) Utiliza el muestreo estratificado para seleccionar una muestra representativa.
- B) Realiza múltiples ajustes de modelos a los datos y promedia los resultados.
- C) Genera muestras de re-muestreo a partir de los datos originales con reemplazo.
- D) Aplica transformaciones a los datos para mejorar su ajuste a una distribución teórica.

32. ¿Cuál de las siguientes afirmaciones describe correctamente un árbol de decisión?

- A) Un árbol de decisión es un algoritmo de aprendizaje no supervisado.
- B) Un árbol de decisión es un método de clasificación que solo puede manejar atributos numéricos.
- C) Un árbol de decisión es una estructura jerárquica de nodos que representan decisiones y resultados.
- D) Un árbol de decisión es un método utilizado exclusivamente para la regresión de datos.

33. ¿Cuál de las siguientes técnicas se utiliza comúnmente para optimizar los pesos de una red neuronal durante el entrenamiento?

- A) Gradiente descendente estocástico (SGD).
- B) Algoritmo de vecinos más cercanos (KNN).
- C) Algoritmo de agrupamiento aglomerativo.
- D) Algoritmo de árboles de decisión.

34. ¿Para qué se utiliza el método de backpropagation (retropropagación) en aprendizaje profundo?

- A) Para propagar el error de salida hacia las capas anteriores de una red neuronal y estimar el gradiente
- B) Para aumentar la velocidad de entrenamiento de una red neuronal.
- C) Para inicializar los pesos de una red neuronal.
- D) Para agrupar y dividir los datos de entrenamiento en clases.

35. Para evaluar el rendimiento de un modelo de aprendizaje automático o algoritmo de clasificación/regresión se suelen utilizar la técnica de validación cruzada k-fold. ¿Cuál es la ventaja de utilizar esta técnica en lugar de una única división del conjunto de datos en entrenamiento y prueba?

- A) Proporciona una estimación más precisa del rendimiento del modelo.
- B) Reduce el tiempo de entrenamiento del modelo.
- C) Evita el sesgo introducido por la división arbitraria de los datos.
- D) No hay ventajas, ambos métodos son equivalentes.

36. Dado un algoritmo de clasificación binaria entrenado con un conjunto de datos que presenta un alto desequilibrio entre clases, ¿qué métrica NO es la más adecuada para evaluar su rendimiento?

- A) Puntuación F1 (F1-score).
- B) Sensibilidad
- C) Especificidad
- D) Accuracy

37. ¿Cuál de las siguientes actividades NO es un tema central en el aprendizaje automático?

- A) Aprendizaje por refuerzo
- B) Aprendizaje semisupervisado
- C) Aprendizaje bayesiano
- D) Aprendizaje global

38. ¿Cuál de las siguientes no es una metodología de regresión regularizada?

- A) Regresión riscal (o ridge)
- B) Regresión LASSO
- C) Regresión por mínimas desviaciones absolutas
- D) Elasticnet

39. En el contexto de la estimación de parámetros, ¿cuál de las siguientes afirmaciones describe mejor el métodos de los mínimos cuadrados?

- A) Es un método que utiliza la probabilidad para encontrar los valores óptimos de los parámetros
- B) Es un método que minimiza la suma de los residuos al cuadrado entre los valores observados y los predichos por el modelo
- C) Es un método que busca maximizar la función de verosimilitud de los parámetros
- D) Es un método que utiliza la regresión lineal para estimar los parámetros del modelo

40. Si un aparato de medida está mal calibrado, las medidas obtenidas de él son un ejemplo de ...

- A) Error sistemático.
- B) Error aleatorio.
- C) Error estadístico.
- D) Error de cuantificación.

41. ¿Cuál es la medida de tendencia central que se define como el valor que divide a la muestra en dos partes iguales?

- A) A. Media aritmética
- B) Mediana
- C) Moda
- D) Media armónica

42. ¿Cuál de los siguientes métodos se utiliza comúnmente en el análisis de series temporales para predecir valores futuros en función de los datos pasados?

- A) Media móvil autorregresiva integrada (ARIMA).
- B) Análisis de Varianza (ANOVA)
- C) Análisis de componentes principales (PCA).
- D) Suavizado Gaussiano (GSM)

43. Existen varios modelos y métodos utilizados para analizar y predecir series temporales. La elección del modelo depende de la naturaleza de los datos y los objetivos del análisis. ¿Cuál de estos nombres NO se corresponde con un modelo de series temporales?

- A) Modelo Autorregresivo (AR)
- B) Modelo de media móvil (AM)
- C) Modelo integrador sobre kernel gaussiano (IG)
- D) Modelo autorregresivo integrador con media móvil (ARIMA)

44. Una red semántica es:

- A) Una forma de representación del conocimiento lingüístico en la que los conceptos y sus interrelaciones se representan mediante un grafo
- B) La forma en la que una página web se estructura para expresar el significado del contenido
- C) Una forma de representación de la estructura sujeto, verbo, predicado de un documento
- D) Una estructura de elementos de comunicación de red que se organizan dinámicamente en función de los requerimientos de las aplicaciones

45. Respecto al análisis de componentes principales (ACP), que frase NO es cierta

- A) El ACP asume que los datos observados son combinación lineal de una cierta base.
- B) Se utiliza comúnmente para reducción de la dimensionalidad del conjunto de datos
- C) Tiene como limitación que es necesario partir de una propuesta del número de dimensiones a las que se quiere ajustar el conjunto de datos.
- D) Cuando los datos no son dimensionalmente homogéneos, la ACP se resuelve por un método basado en la matriz de correlación.

46. ¿Qué tarea de visión por computador implica localizar y categorizar objetos dentro de una imagen?

- A) Clasificación de objetos.
- B) Detección de objetos.
- C) Segmentación semántica
- D) Segmentación de instancias.

47. ¿Cuál de los siguientes estadísticos se ve más afectado por los valores extremos (outliers) en un conjunto de datos?

- A) Media
- B) Mediana
- C) Moda
- D) Moda y Mediana

48. ¿Cuál de las siguientes acciones sobre los datos NO debería ser una etapa típica del flujo de tratamiento de datos?

- A) Limpiar
- B) Importar
- C) Comunicar
- D) Tunear

49. ¿Cuál de las siguientes medidas NO es propia en la evaluación de modelos de clasificación?

- A) F1
- B) Especificidad
- C) G2
- D) Precisión

50. ¿Cuál de los siguientes criterios NO se emplea como criterio de calidad de particiones en árboles de clasificación?

- A) Índice de Gini
- B) Entropía cruzada
- C) Tasa de clasificación de error
- D) Suma de cuadrados de residuos

51. ¿Cuál de las siguientes afirmaciones sobre las máquinas de vector soporte (SVMs) es cierta?

- A) No pueden emplearse para regresión
- B) Admiten interpretación bayesiana
- C) En clasificación dan siempre peores resultados que los modelos de redes neuronales
- D) No admiten interpretación probabilística

52. ¿Cuál de las siguientes técnicas NO es de reducción de la dimensionalidad?

- A) Análisis de componentes principales
- B) Factorización de matrices no negativas
- C) UMAP
- D) Proceso de Dirichlet

53. ¿Cuál de los siguientes NO es un concepto típico del aprendizaje por refuerzo?

- A) Q-aprendizaje
- B) Proceso de decisión markoviano
- C) Aprendizaje semisupervisado
- D) Programación dinámica

54. ¿Cuál de las siguientes afirmaciones es correcta sobre Scipy?

- A) Scipy es una biblioteca de Python para programación de bases de datos.
- B) Scipy es una biblioteca de Python para programación web
- C) Scipy es una biblioteca de Python para programación de videojuegos.
- D) Scipy es una biblioteca de Python para matemáticas, ciencia e ingeniería

55. La programación paralela es aquella en la que:

- A) un problema se divide en varias partes y las instrucciones de cada parte se ejecutan simultáneamente en diferentes procesadores
- B) un problema se divide en varias partes y las instrucciones de cada parte se ejecutan en secuencia
- C) varios programas independientes se ejecutan en varios procesadores distintos
- D) varios programadores trabajan en paralelo

56. En los lenguajes orientados a objetos, declarar y definir dos métodos con el mismo nombre pero distintos argumentos se denomina:

- A) Polimorfismo
- B) Sobrecarga
- C) Sobrescritura
- D) Herencia

57. ¿Cuál de las siguientes opciones es una característica de la metodología de desarrollo “ágil”?

- A) El equipo trabaja en un plan detallado desde el principio.
- B) El equipo trabaja en iteraciones cortas y regulares.
- C) El equipo no tiene contacto con el usuario final.
- D) El equipo no realiza pruebas hasta el final del proyecto.

58. ¿Qué es la herencia en programación orientada a objetos?

- A) Es la capacidad de una clase de heredar bucles *for*.
- B) Es la capacidad de recibir una herencia y dejar de programar.
- C) Es la capacidad de una clase de heredar sólo los métodos de otra clase.
- D) Es la capacidad de una clase de heredar los atributos y métodos de otra clase y agregar nuevos atributos y métodos.

59. En programación, ¿cuál es la estructura de control que permite repetir un conjunto de instrucciones mientras se cumpla una condición?

- A) if-else
- B) case
- C) while
- D) switch

60. ¿Cuál de los siguientes lenguajes de programación NO es orientado a objetos?

- A) Python
- B) C
- C) C++
- D) Java

61. En programación orientada a objetos (OOP), ¿qué elemento se utilizaría para representar el “color de un coche”?

- A) un atributo
- B) una clase
- C) un objeto
- D) una instancia

62. ¿Cuáles son los elementos que caracterizan una neurona artificial?

- A) Dendritas, cuerpo, axón.
- B) Entradas, núcleo, salidas.
- C) Entradas, pesos sinápticos, regla de propagación, función de activación.
- D) Entradas, función de activación, salidas.

63. ¿Cuándo decimos que hay *overfitting* en el contexto de modelado?

- A) Cuando el modelo se ajusta muy bien a los datos de entrenamiento y no generaliza bien a otros.
- B) Cuando el modelo no logra reproducir los datos de entrenamiento.
- C) Cuando el modelo se ajusta muy bien a los datos de entrenamiento y generaliza bien.
- D) Ninguna de las opciones anteriores.

64. La calidad del código mide la capacidad de

- A) Mantenibilidad del código.
- B) Mantenibilidad y legibilidad del código.
- C) Mantenibilidad, legibilidad y eficiencia del código.
- D) Complejidad, mantenibilidad, legibilidad y eficiencia del código.

65. ¿Qué redes de las utilizadas en procesamiento de lenguaje natural permiten procesar las secuencias de texto en paralelo?

- A) Redes neuronales convolucionales (CNN).
- B) Redes de memoria a corto plazo (LSTM).
- C) Redes transformadoras (Transformers).
- D) Redes neuronales recurrentes (RNN).

66. En una distribución normal...

- A) la media, la moda y la mediana coinciden.
- B) sólo la media y la mediana coinciden.
- C) sólo la moda y la media coinciden.
- D) la media, la moda y la mediana son distintas.

67. ¿Qué estructura de datos se utiliza para gestionar una colección de elementos en la que el último elemento añadido es el primero en ser retirado?

- A) Pila
- B) Cola
- C) Conjunto
- D) Árbol

68. Una red neuronal debe de tener...

- A) menos de 40 capas.
- B) al menos dos capas.
- C) exactamente cuatro capas.
- D) exactamente tres capas.

69. La normalización y centrado son técnicas usadas para...

- A) poner en la misma escala variables de diferente magnitud.
- B) aumentar las diferencias entre variables.
- C) obtener valores más extremos.
- D) uniformizar las varianzas.

70. En el contexto de *machine learning*, los hiperparámetros son...

- A) variables con más de tres dimensiones.
- B) las variables que se estiman durante el proceso de aprendizaje.
- C) los parámetros que se usan para controlar el proceso de aprendizaje.
- D) Todas las anteriores son correctas.

71. El análisis de agrupamiento o *clustering* se refiere a...

- A) las técnicas más usadas actualmente en el tratamiento de imágenes digitales.
- B) un conjunto de técnicas estadísticas para clasificar datos individuales en grupos homogéneos.
- C) un método determinista basado en cálculo numérico.
- D) la aplicación de ecuaciones diferenciales a modelos científicos.

72. ¿Cuál de las siguientes opciones describe correctamente qué es el dropout en el contexto del aprendizaje profundo?

- A) El dropout es una técnica de regularización utilizada para reducir el sobreajuste
- B) El dropout es una técnica de preprocesamiento de datos utilizada para reducir su dimensionalidad antes de aplicar otros algoritmos de aprendizaje automático.
- C) El dropout es una técnica que permite transferir el conocimiento aprendido en un modelo a otro modelo sin tener que entrenarlo desde cero.
- D) El dropout es una técnica de optimización utilizada para encontrar el mínimo global de la función de pérdida en un modelo de aprendizaje profundo.

73. En una prueba de hipótesis, ¿qué valor se compara con el estadístico de prueba para tomar una decisión?

- A) Valor crítico
- B) Estadístico de estimación
- C) P-valor
- D) Error estándar

74. En una tabla de un estudio en el que se recoge la estatura de 50 estudiantes, la frecuencia acumulada representa:

- A) El número de individuos con una determinada estatura
- B) El número de individuos con una determinada estatura dividido por 50
- C) El número de individuos con una estatura menor o igual a una dada
- D) El porcentaje de individuos con una estatura mayor o igual a una dada

75. Indicar cuál de los siguientes parámetros NO es una medida de tendencia central de una variable cuantitativa:

- A) La moda
- B) La media geométrica
- C) La mediana
- D) El cuantil

75. Un coeficiente de correlación de Pearson de valor 1 entre dos variables cuantitativas siempre indica que:

- A) Las dos variables están correlacionadas y una variable es la causa de la otra
- B) Las dos variables son independientes
- C) Las dos variables están relacionadas de forma lineal
- D) La relación de las dos variables en una gráfica sigue una línea horizontal.

76. ¿Cuál de las siguientes afirmaciones describe mejor la "Ingestión de datos"?

- A) Es el proceso de manipulación y transformación de datos en diferentes formatos.
- B) Se refiere a la limpieza y preparación de datos para su análisis posterior.
- C) Es el proceso de carga y almacenamiento de datos en un sistema o plataforma.
- D) Se refiere a la exploración y visualización de datos para obtener perspectivas (insights) y patrones.

77. ¿Cuál de las siguientes no es una aplicación típica del análisis de componentes principales?

- A) Reducción de la dimensionalidad
- B) Equilibrado de muestras
- C) Visualización
- D) Selección de predictores en regresión (por componentes principales)

78. ¿Qué es un clúster de computación?

- A) Es una agrupación local de recursos computacionales autónomos, normalmente de prestaciones similares, interconectados y que trabajan de manera estrecha
- B) Un sistema paralelo distribuido por varios países (de ámbito global)
- C) Un sistema distribuido paralelo de ámbito global
- D) Un sistema virtualizado

79. ¿Cuál de los siguientes conceptos es central en la programación funcional?

- A) Uso de clases y objetos
- B) Funciones como elementos básicos
- C) Herencia múltiple
- D) Manejo explícito de memoria

80. Un portátil tiene 4 cores. ¿Es un computador paralelo?

- A) No. Es un único computador y no puede paralelizar.
- B) No. Puede ser utilizado para conectarse a un computador que sea realmente paralelo, pero no lo es.
- C) Sí. Podemos ejecutar varios programas simultáneamente o lanzar varios hilos de ejecución de un mismo programa.
- D) Sí. Pero solo si tienes permiso del administrador.