



### **Convocatoria referencia 03/2024**

CONVOCATORIA DE PRUEBAS SELECTIVAS PARA CUBRIR 10 PLAZAS DE PERSONAL LABORAL EN LA COMISIÓN NACIONAL DEL MERCADO DE VALORES. TÉCNICOS (ANALISTAS DE DATOS) PARA VARIAS DIRECCIONES Y DEPARTAMENTOS.

## **Primera prueba de la fase oposición: ejercicio escrito**

### **Cuestionario tipo test**

No debe firmar, ni escribir su nombre o poner marcas o señales en las hojas que pudieran permitir su identificación. En caso contrario, será descalificado. No se podrá sacar ningún tipo de documentación de la sala, a excepción de la "Plantilla de Respuesta para autocomprobación".

**PEGATINA**



## PARTE 1. CUESTIONES TIPO TEST

---

Rodee con un círculo la respuesta correcta. Para los errores, tache con una "X" y vuela a rodear la respuesta correcta con un círculo. Solo se darán por válidas aquellas respuestas rodeadas con un círculo sin ninguna otra señal.

---

1. ¿Cuál de las siguientes afirmaciones sobre la temperatura de un LLM es correcta?
  - a. Las temperaturas más altas son adecuadas si se espera una respuesta más determinista.
  - b. A mayor temperatura, las probabilidades de los tokens se igualan, haciendo más aleatoria la generación.
  - c. Temperaturas cercanas a 0 generan una salida más creativa e impredecible.
2. Para obtener salidas con una estructura clara sin necesidad de reentrenar, ¿qué enfoque es el más útil?
  - a. Few-shot prompting
  - b. Fine-tuning
  - c. Data augmentation
3. ¿Qué es prompting en modelos de lenguaje?
  - a. Un proceso para entrenar un modelo desde cero.
  - b. Una forma de proporcionar instrucciones o ejemplos al modelo.
  - c. Un método de optimización de hiperparámetros.
4. ¿Cuál de las siguientes afirmaciones sobre sistemas RAG (Retrieval-Augmented Generation) es FALSA?
  - a. Un índice vectorial mal construido puede hacer que el modelo genere respuestas incorrectas o irrelevantes.
  - b. Un prompt mal formulado puede afectar negativamente a la calidad de la respuesta generada.
  - c. Usar un modelo de lenguaje más grande siempre garantiza respuestas más precisas, aunque los documentos recuperados sean irrelevantes.
5. ¿Cómo se suele caracterizar a un agente en el contexto de modelos de lenguaje?
  - a. Es una entidad con un rol específico, capaz de tomar decisiones para cumplir una tarea
  - b. Es una herramienta externa usada para traducir texto
  - c. Es una API que permite consultar datos en tiempo real

6. ¿Qué característica haría a un agente más robusto frente a errores de una herramienta?
  - a. Almacenar logs en base de datos.
  - b. Implementar estrategias de fallback.
  - c. Limitar el número de tokens.
7. ¿Cuál de las siguientes afirmaciones sobre los agentes de IA generativa es correcta?
  - a. Permite almacenar y recuperar información de interacciones anteriores para mantener contexto.
  - b. Aumenta la capacidad de procesamiento paralelo del modelo.
  - c. No pueden tomar decisiones autónomas sin supervisión constante.
8. ¿Cuál de las siguientes acciones no puede realizar un modelo base sin estar integrado en un agente con herramientas?
  - a. Responder una pregunta general de conocimiento.
  - b. Generar una historia corta de ficción.
  - c. Realizar una llamada a una API externa para obtener información actualizada.
9. ¿Qué tipo de Redes de Neuronas se suelen aplicar a problemas de visión artificial?
  - a. Convolucionales.
  - b. Recurrentes.
  - c. Capas somatotópicos.
10. ¿Cuánto tiempo y esfuerzo se suele dedicar a limpiar y preparar los datos, en contraste con el tiempo que lleva entrenar el modelo predictivo?
  - a. 80%
  - b. 40%
  - c. 20%
11. ¿Cuáles son los ejes principales de una estrategia de Datos para que una compañía logre ser Data Driven?
  - a. Un buen Gobierno del dato, elegir bien las apuestas para crear valor con el apoyo de la dirección, rodearse de talento y acertar con la tecnología.
  - b. Contar con un gran presupuesto, rodearse de un buen equipo de Data Science y generar los primeros 4 algoritmos de Machine Learning.
  - c. Contar con una buena herramienta de visualización de datos, usar Hadoop y Python y montar un Data Lake con todos los datos de tu empresa.

12. En el contexto actual de Narrow Artificial Intelligence, podemos decir que un algoritmo de Machine Learning hace:
- Una tarea específica.
  - Diferentes tareas de manera simultánea.
  - Las mismas que en el contexto de Artificial General Intelligence.
13. ¿Cuál de los siguientes no es un ejemplo de diseño de investigación?
- Test A/B.
  - Máquina de soporte vectorial.
  - Análisis multivariado.
14. A la hora de presentar los resultados de un proyecto de Analítica Avanzada, simplemente con decir qué tal salió el modelo es suficiente:
- Depende de quién presenta los resultados, si es el Jefe de Proyecto, el enfoque será lo que decida.
  - A la hora de presentar los resultados, con indicar que el modelo tuvo una ROC buena, es suficiente para Negocio.
  - No, los resultados del modelo son importantes, pero más aún es comunicar el valor que se espera generar con la implantación/explotación del proyecto.
15. El incumplimiento de la GDPR en un proyecto de Big Data puede conllevar:
- La GDPR no tiene nada que ver con los Datos
  - Un daño grave para tu empresa tanto en términos económicos como reputacionales.
  - Una probabilidad muy alta de que el científico de datos sea despedido.
16. "Data as a Service" es un concepto indispensable en una estrategia de Datos por...:
- Hay que gestionar los Datos con APIs que permitan proveer soluciones escalables basadas en Machine Learning y Visualización de datos a todas las áreas de la compañía.
  - Crear una unidad de Datos que pueda proveer de buenos informes e insights a todas las unidades de negocio es crítico para el éxito de la empresa.
  - Que los analistas y científicos de datos estén al servicio de la ciencia es indispensable para el éxito de cualquier proyecto.

17. ¿A qué necesidades podemos dar respuesta con una adecuada estrategia de analítica Big data?
- Procesos de análisis para la optimización del ciclo de vida del cliente (CLTV) sabiendo que podemos modelar si va a dejar de consumir mi producto o servicio (Propensión a la fuga) o si es proclive a hacerlo (propensión a la compra).
  - Identificar el talento y gestionar el equipo en mi organización (People analytics), desarrollar una estrategia de comunicación adecuada (Right moment, Right person, right offer) y vigilar lo que hace mi competencia directa (Inteligencia competitiva)
  - Ambas
18. ¿Cómo se clasifican los datos en función de su estructura?
- Estructurados, no estructurados e híbridos
  - Estructurados y no estructurados
  - Estructurados, polivalentes y mixtos
19. Afirmación: Una función se puede asignar a una variable, pasarse como parámetro o bien devolverse en una función
- Solo se pueden pasar como parámetro las funciones lambda
  - Verdadero
  - Falso
20. Afirmación: Una lista puede almacenar cualquier tipo de dato
- Verdadero
  - Falso.
  - Las listas solo pueden tener valor del mismo tipo
21. Una lista se diferencia de una tupla en:
- Una tupla es de solo lectura y no se puede modificar.
  - Una tupla se define con () y sus valores separados por comas.
  - Ambas son correctas.
22. En Python dadas estas dos listas a=[1,2,3] y b=[1,2,3]
- a is b retorna true.
  - a==b retorna true.
  - Ambas son correctas.
23. ¿Cómo cuento valores diferentes de la columna "id\_pedido" de la tabla "pedidos"?
- SELECT count(distinct id\_pedido) FROM pedidos
  - SELECT count\_distinct(id\_pedido) FROM pedidos
  - SELECT DISTINCT id\_pedido FROM pedidos

24. ¿Qué operador faltaría usar para obtener los campos que se pretenden obtener?  
SELECT nombre\_cliente, sum(revenue) FROM orders
- La sentencia WHERE.
  - La sentencia LIMIT.
  - La sentencia GROUP BY.
25. Si quiero juntar una tabla A con una tabla B, y que el resultado contenga todas las filas de la tabla A, con las columnas de la tabla B llenas o vacías utilizaré:
- RIGHT JOIN.
  - JOIN.
  - LEFT JOIN.
26. Dado un dataframe "df", con la sentencia df[0:8]...
- Obtenemos de la fila 0 a la 8
  - Obtenemos de la fila 0 a la 7
  - Obtenemos de la columna 0 a la 8.
27. ¿Con qué método rellenamos los valores nulos de un dataframe?
- fillna.
  - isna.
  - dropduplicates.
28. Los valores atípicos (también llamados outliers) en el análisis y predicción de una variable aleatoria pueden ser un inconveniente. Selecciona la respuesta que es FALSA.
- El recorrido intercuartílico es poco sensible a los outliers ya que elimina tanto los valores extremos máximos como los mínimos.
  - Como los valores atípicos son un gran problema siempre se eliminan directamente al principio del estudio.
  - Efectivamente, los outliers distorsionan algunos resultados, como es el caso del cálculo de la desviación estándar y la varianza.
29. Las visualizaciones se deben evaluar...
- Con contexto de negocio puesto que sin conocer el uso las gráficas no se pueden evaluar.
  - Sin contexto de negocio puesto que lo relevante son las propiedades visuales.
  - Sólo teniendo en cuenta a los power users de la visualización que serán los que más lo conozcan.

30. El principio de discriminación es...
- Un dashboard ante dos situaciones relevantes con accionamientos distintos discrimina ambas situaciones y no se ve la diferencia entre ellas.
  - Un dashboard ante dos situaciones relevantes con accionamientos distintos distingue bien ambas situaciones (se nota la diferencia).
  - Un dashboard discrimina entre dos grupos de usuarios de manera que cada uno de ellos ve algo distinto en el dashboard.
31. ¿Cuál es el propósito de la visualización de datos?
- El propósito de la visualización no es solamente la visualización en sí, si no ayudar a las personas a dar sentido a los datos.
  - Una representación gráfica de datos o conceptos que es una construcción gráfica internamente a nivel mental o bien un artefacto externo que respalda la toma de decisiones.
  - Ambas son correctas.
32. Es una buena práctica utilizar colores para datos categóricos para diferenciar entre categorías. Una buena práctica es utilizar...
- No más de 15 colores diferentes.
  - Siempre menos de 3 colores.
  - De 5 a 9 colores.
33. ¿Se puede leer datos desde una URL?
- Si.
  - No, solo de base de datos.
  - No solo de ficheros de texto.
34. ¿Qué nos permite realizar Power BI Desktop?
- Compartir los informes y ademas usar el modulo de machine learning de Azure.
  - Compartirlos con toda la compañía sin restricciones.
  - Solo crear informes.
35. ¿Qué cardinalidad tiene la relación de un maestro y una tabla de hechos?
- 1:\*
  - 1:1
  - :
36. ¿Para qué sirve el crear un maestro Calendario?
- Para relacionar varias tablas por un campo fecha y así evitar que se dupliquen los datos.
  - Para calcular los años bisiestos.
  - Ninguna de las anteriores.

37. Para decidir la posición de los elementos visuales existen dos posicionamientos importantes:
- F-Shaped y Left-To-Right.
  - Z-Shaped y Left-To-Right.
  - F-Shaped y Z-Shaped.
38. La consistencia de un dashboard...
- Es usar la misma paleta de colores en todas las gráficas.
  - Es usar los mismos mapeos de variables a canales visuales manteniendo homogeneidad en la forma de interpretar los gráficos.
  - Es usar el mismo tipo de gráfica en todo el dashboard.
39. ¿Qué función aplicamos a un modelo logístico para acotar las predicciones entre 0 y 1?
- Coseno.
  - Sigmoide.
  - Arcotangente.
40. ¿Cuál de las siguientes afirmaciones comparando árboles de decisión de clasificación y regresión es FALSA?
- En los de clasificación, la partición óptima se encuentra con el MSE o RMSE, en los de regresión, con el Gini.
  - En los de clasificación, la partición óptima se encuentra con el Gini; en los de regresión, con el MSE o RMSE.
  - En los de clasificación, se comprueba la calidad predictiva del modelo con la precisión; en los de regresión, con el MSE o RMSE.
41. Nuestro dataset tiene nulos y queremos entrenar un árbol de decisión con él:
- Debemos obtener un dataset sin nulos, ya sea eliminando las filas con nulos o rellenándolos.
  - Es preferible llenar los nulos, pero podemos admitirlos si representan una cantidad residual del total de las filas.
  - El método de entrenamiento del árbol de decisión permite considerar los nulos como otra posible partición.
42. El algoritmo de los K vecinos más próximos Knn ...
- Solo se puede aplicar a problemas de clasificación.
  - Requiere realizar un entrenamiento previo a la inferencia.
  - Se puede aplicar para resolver tanto problemas de regresión como de clasificación.

43. El aprendizaje no supervisado...
- Tiene como objetivo predecir el valor de una variable continua a partir de datos históricos de la misma.
  - Tiene como objetivo aprender patrones de datos no etiquetados.
  - Es el aprendizaje que puede realizar una máquina sin tener que ser supervisada por un humano.
44. Donde empieza el overfitting (sobreajuste):
- Depende del problema que estemos resolviendo.
  - Cuando nuestra métrica de error sube para el conjunto de test y baja para el de train.
  - Cuando nuestra métrica de error sube para el conjunto de train y baja para el de test.
45. ¿Cuál es el objetivo principal de la Cross Validación?
- Tener métricas fiables de cómo funcionará tu modelo.
  - Preparar al modelo para la mejor representación de datos futuros.
  - Ambas son correctas.
46. Antes de llevar a cabo la realización de un modelo de clustering K-means en el problema planteado, se lleva a cabo la realización del gráfico de elbow, ¿Por qué?
- Para determinar el número de clusters para parametrizar el algoritmo.
  - Para determinar el tipo de distancias que debe de utilizar el algoritmo.
  - Ninguna de las anteriores.
47. La maldición de la dimensionalidad:
- Afecta únicamente a problemas de ML supervisados.
  - Hace que los datos sea densos y eso les viene bien a muchos algoritmos de ML.
  - Implica que a medida que se añaden más características necesitamos una cantidad exponencialmente creciente de observaciones.
48. ¿Qué parte de un proyecto de Data Science consume más tiempo?
- Entrenamiento de modelos.
  - Exploración y validación del dato.
  - Ingeniería de Variables.
49. ¿Qué afirmación es correcta sobre correlación y causalidad?
- Una correlación alta siempre implica causalidad.
  - Una correlación alta sugiere una relación, pero no prueba causalidad.
  - Correlación y causalidad son términos intercambiables.
50. ¿Cuál es el objetivo principal de la estadística inferencial?
- Describir los datos de una muestra.
  - Hacer generalizaciones sobre una población a partir de una muestra.
  - Calcular promedios.

## **PREGUNTAS RESERVA**

1. Cuando un prompt se vuelve demasiado extenso, ¿qué tipo de problema puede surgir?
  - a. El modelo puede priorizar secciones recientes del prompt y olvidar partes clave del inicio
  - b. El modelo aumenta automáticamente su número de capas
  - c. El modelo pasa a utilizar embeddings estáticos en lugar de contextuales
2. Cuando ajustamos un modelo, ¿que métrica estamos optimizando?
  - a. La métrica de evaluación.
  - b. La función pérdida / loss function
  - c. Los pesos de mi modelo.
3. En una regresión, ¿qué valor tenemos en consideración para saber si una variable es estadísticamente significativa?
  - a. La desviación estandar del coeficiente Beta
  - b. El Coeficiente Beta de esa variable
  - c. El contraste de Hipótesis Nula / p-value
4. La función apply()...
  - a. Sirve para aplicar una función sobre todas las columnas o filas de un dataframe, pudiendo seleccionar si la orientación es por filas o por columnas con un parámetro.
  - b. Sirve para aplicar una función sobre una fila
  - c. Ambas son correctas.
5. Dadas dos variables independientes X e Y se calcula la correlación entre ambas obteniéndose un valor. Señala cual es la opción correcta.
  - a. 1 -> Función lineal entre X e Y.
  - b. -0,1 -> Correlación negativa baja..
  - c. 1,5 -> Correlación positiva muy alta.