

Examen Teórico 2 de Inteligencia Artificial

Sección 1: Preprocesado (15 preguntas)

1. ¿Cuál es el objetivo principal del preprocesado de datos en Machine Learning?

- a) Mejorar la estética de los datos.
- b) **Preparar los datos para que sean utilizados por los algoritmos de aprendizaje.**
- c) Reducir el tamaño del código del modelo.
- d) Aumentar la complejidad del modelo.

2. ¿Por qué el preprocesado de datos es considerado una etapa fundamental en los proyectos de ML?

- a) Porque es la etapa que más tiempo consume.
- b) **Porque impacta directamente en la calidad, eficiencia y precisión del modelo.**
- c) Porque es la etapa más sencilla de automatizar.
- d) Porque es la etapa que requiere menos conocimiento técnico.

3. ¿Qué componente de MLOps se encarga del preprocesado de datos?

- a) La gestión del modelo.
- c) La monitorización del modelo.
- b) El entrenamiento del modelo.
- d) **El pipeline de datos.**

4. ¿Cuál de las siguientes NO es una tarea común del preprocesado de datos?

- a) Limpieza de datos.
- c) **Desarrollo del algoritmo de aprendizaje.**
- b) Transformación de datos.
- d) Manejo de valores faltantes.

5. ¿Qué se entiende por "limpieza de datos" en el contexto del preprocesado?

- a) Ordenar los datos alfabéticamente.
- b) **Eliminar o corregir datos erróneos, incompletos o inconsistentes.**
- c) Convertir todos los datos a minúsculas.
- d) Eliminar todas las filas con valores faltantes.

6. ¿Qué tipo de transformación de datos se utiliza para evitar que algunas variables dominen el entrenamiento del modelo?

- a) Codificación de variables categóricas.
- b) Reducción de dimensionalidad.
- c) **Escalado de características.**
- d) Manejo de valores faltantes.

7. ¿Cuál es el propósito de la codificación de variables categóricas?

- a) Eliminar las variables categóricas del conjunto de datos.
- b) **Transformar variables no numéricas en representaciones numéricas.**
- c) Convertir todas las variables a un mismo tipo de dato.
- d) Reemplazar los valores faltantes en las variables categóricas.

8. ¿Qué técnica de preprocesado se utiliza para seleccionar las características más relevantes para el modelo?

- a) Escalado de características.
- b) Codificación de variables categóricas.
- c) **Reducción de dimensionalidad.**
- d) Manejo de valores faltantes.

9. ¿Cómo se relaciona el Big Data con el preprocesado de datos en el contexto de la IA?

- a) El Big Data no requiere preprocesado.
- b) El Big Data solo se utiliza en modelos de IA simples.
- c) **El Big Data suele ser heterogéneo y desorganizado, por lo que el preprocesado es crucial.**
- d) El preprocesado se realiza después de entrenar el modelo de IA con Big Data.

10. ¿Qué tipo de operaciones de manipulación de cadenas de texto en Python podrían ser útiles en el preprocesado de datos?

- a) Operaciones matemáticas como la suma o la multiplicación.
- b) **Funciones para eliminar espacios en blanco, convertir a mayúsculas o minúsculas, reemplazar subcadenas o segmentar cadenas.**
- c) Creación de gráficos y visualizaciones de datos.
- d) Importación de librerías externas para el análisis de datos.

11. ¿Cómo pueden las estructuras de datos como listas, tuplas y diccionarios ser útiles en el preprocesado de datos?

- a) Para almacenar el código del modelo de Machine Learning.
- b) Para visualizar los datos antes del preprocesado.
- c) **Para organizar y manipular datos para su posterior análisis.**
- d) Para entrenar el modelo de Machine Learning.

12. ¿Qué se debe considerar al manejar valores faltantes en un conjunto de datos?

- a) Siempre se deben eliminar las filas con valores faltantes.
- b) Se debe decidir la mejor estrategia según el contexto, como eliminar filas, imputar valores o utilizar algoritmos que manejen datos faltantes.**
- c) Los valores faltantes no afectan el rendimiento del modelo.
- d) Se deben reemplazar todos los valores faltantes por cero.

13. ¿Las fuentes proporcionan una descripción exhaustiva de todas las técnicas de preprocesado de datos?

- a) Sí, las fuentes cubren todas las técnicas de preprocesado en detalle.
- b) No, las fuentes solo mencionan la relevancia del preprocesado, pero no profundizan en las técnicas específicas.**
- c) Las fuentes solo se enfocan en el preprocesado de datos de texto.
- d) Las fuentes solo describen las técnicas de preprocesado para imágenes.

14. ¿Dónde se recomienda buscar información más completa sobre técnicas de preprocesado?

- a) En foros de programación general.
- b) En tutoriales básicos de Python.
- c) En recursos específicos sobre Machine Learning.**
- d) En blogs de análisis de datos generales.

15. ¿Cuál es la relación entre el código Python y el preprocesado de datos?

- a) El código Python no se utiliza en el preprocesado de datos.
- b) Python solo se usa para visualizar datos, no para preprocesarlos.
- c) Python ofrece herramientas y librerías que facilitan el preprocesado de datos.**
- d) El preprocesado solo se puede realizar con lenguajes de bajo nivel.

Sección 2: Librerías de Python (10 preguntas)

16. ¿Cuál de las siguientes opciones NO es una tarea principal que se puede realizar con Scikit-learn?

- a) Clasificación
- c) Agrupación
- b) Visualización de datos**
- d) Reducción de dimensionalidad

Explicación: Scikit-learn se enfoca en algoritmos de aprendizaje automático, mientras que la visualización de datos se realiza con otras bibliotecas como Matplotlib.

17. ¿Cuál de las siguientes afirmaciones describe mejor la API de Scikit-learn?

- a) Compleja y difícil de aprender c) Limitada en cuanto a la variedad de algoritmos
- b) Consistente y fácil de usar** d) Desarrollada principalmente para el análisis de datos

Explicación: La API consistente de Scikit-learn facilita el uso e intercambio de diferentes algoritmos.

18. ¿Qué tipo de datos se utiliza principalmente en NumPy?

- a) DataFrames c) Series de tiempo
- b) Arrays multidimensionales** d) Listas enlazadas

Explicación: NumPy se destaca por su capacidad para trabajar eficientemente con arrays multidimensionales.

19. ¿Qué biblioteca se utiliza en Python para la manipulación y análisis de datos?

- a) NumPy **c) Pandas**
- b) Scikit-learn d) Matplotlib

Explicación: Pandas es la biblioteca de Python especializada en manipulación y análisis de datos.

20. ¿Qué funcionalidad NO se menciona en las fuentes como parte de NumPy?

- a) Operaciones matemáticas en arrays c) Transformada de Fourier
- b) Creación de gráficos interactivos** d) Generación de números aleatorios

Explicación: La creación de gráficos se realiza con otras bibliotecas como Matplotlib. NumPy se centra en operaciones numéricas y matemáticas.

21. ¿Qué tipo de archivos se pueden leer y escribir con Pandas, según las fuentes?

- a) Solo archivos CSV c) Archivos de imagen
- b) Archivos CSV y Excel** d) Archivos de audio

Explicación: Las fuentes mencionan ejemplos de uso de Pandas para leer y escribir archivos CSV y Excel.

22. ¿Cuál de las siguientes opciones NO es una característica clave de Scikit-learn?

- a) Documentación completa **c) Visualización de datos avanzada**
- b) Comunidad activa d) Integración con otras bibliotecas

Explicación: Si bien Scikit-learn se integra con bibliotecas de visualización, su enfoque principal no es la visualización en sí misma.

23. ¿Para qué tipo de tareas es más adecuado NumPy?

- a) Procesamiento de lenguaje natural
- b) **Cálculos científicos y numéricos**
- c) Desarrollo web
- d) Análisis de redes sociales

Explicación: NumPy es la base para la computación científica en Python debido a su eficiencia en el manejo de arrays multidimensionales.

24. ¿Qué biblioteca permite realizar operaciones de álgebra lineal en Python?

- a) Pandas
- b) Scikit-learn
- c) **NumPy**
- d) Todas las anteriores

Explicación: NumPy incluye funciones específicas para realizar operaciones de álgebra lineal como la multiplicación de matrices.

25. Según las fuentes, ¿dónde se puede encontrar más información sobre Pandas?

- a) En la documentación oficial de Pandas
- b) En foros de desarrolladores
- c) En tutoriales específicos
- d) **Todas las anteriores**

Explicación: Las fuentes sugieren consultar la documentación oficial, foros o tutoriales para profundizar en el conocimiento de Pandas.

26. ¿Cuál es la función principal de la biblioteca Matplotlib en Python?

- a) Manipulación de datos
- b) Aprendizaje automático
- c) **Visualización de datos**
- d) Procesamiento de imágenes

Explicación: Matplotlib es la biblioteca principal en Python para crear gráficos y visualizaciones estáticas, interactivas y animadas.

Sección 3: Algoritmos ML (10 preguntas)

27. ¿Cuál es el objetivo principal de la inteligencia artificial?

- a) Desarrollar robots que puedan realizar tareas domésticas.
- b) **Crear máquinas que puedan imitar comportamientos inteligentes.**
- c) Reemplazar completamente la mano de obra humana en todas las industrias.
- d) Construir supercomputadoras capaces de resolver cualquier problema matemático.

28. ¿Cuál de las siguientes opciones NO es un ejemplo de comportamiento inteligente que la IA busca imitar?

- a) Conducir un vehículo.
- b) Analizar patrones en datos.
- c) Reconocer la voz humana.
- d) **Realizar la fotosíntesis.**

29. ¿Qué tipo de IA se considera que son todos los sistemas existentes en la actualidad?

- a) **IA Débil**
- b) IA Fuerte
- c) IA Superficial
- d) IA Profunda

30. ¿Cuál es la principal diferencia entre el aprendizaje supervisado y el no supervisado?

- a) **El aprendizaje supervisado utiliza datos etiquetados, mientras que el no supervisado no.**
- b) El aprendizaje supervisado se utiliza para la clasificación, mientras que el no supervisado se utiliza para la regresión.
- c) El aprendizaje supervisado es más complejo que el no supervisado.
- d) El aprendizaje supervisado se basa en redes neuronales, mientras que el no supervisado no.

31. ¿Cuál de los siguientes es un ejemplo de aplicación del aprendizaje supervisado?

- a) **Detección de spam**
- b) Clusterización de clientes
- c) Reducción de dimensionalidad
- d) Aprendizaje de espacios latentes

32. ¿Cuál de los siguientes es un ejemplo de técnica de aprendizaje no supervisado?

- a) Árboles de decisión
- b) Regresión lineal
- c) **Clusterización**
- d) Clasificación de imágenes

33. ¿Qué tipo de modelo predictivo utiliza una estructura de árbol para tomar decisiones?

- a) **Árboles de decisión**
- b) Modelos de regresión
- c) Modelos de clasificación
- d) Redes neuronales

34. ¿Qué tipo de modelo se utiliza para predecir una variable continua, como el precio de la vivienda?

- a) **Modelos de regresión**
- b) Modelos de clasificación
- c) Árboles de decisión
- d) Redes neuronales

35. ¿En qué se inspiran las redes neuronales?

- a) En la estructura del ADN
- b) En el funcionamiento de los ordenadores cuánticos
- c) **En la estructura del cerebro humano**
- d) En el comportamiento de las colonias de hormigas

36. ¿Qué término se utiliza para describir las redes neuronales con muchas capas que permiten el aprendizaje de patrones complejos?

- a) Machine Learning
- b) Inteligencia Artificial
- c) **Deep Learning**
- d) Big Data

Sección 4: Redes Neuronales (7 preguntas)

37. ¿Qué son las redes neuronales?

- a) **Un tipo de algoritmo de aprendizaje automático inspirado en la estructura y función del cerebro humano.**
- b) Un conjunto de reglas predefinidas para resolver problemas específicos.
- c) Un lenguaje de programación utilizado para desarrollar aplicaciones de inteligencia artificial.
- d) Un tipo de hardware especializado en el procesamiento de grandes cantidades de datos.

38. ¿Cuál de los siguientes NO es un componente de una red neuronal?

- a) Neuronas artificiales
- b) Capas
- c) Pesos
- d) **Transistores**

39. ¿Cómo aprenden las redes neuronales a partir de los datos?

- a) **Ajustando los pesos de las conexiones entre neuronas.**
- b) Mediante la memorización de todos los datos de entrenamiento.
- c) A través de la intervención manual de un programador.
- d) Utilizando algoritmos de búsqueda exhaustiva.

40. ¿Qué es el error en el contexto del entrenamiento de una red neuronal?

- a) **La diferencia entre la salida de la red y la salida deseada.**
- b) Un fallo en el código del algoritmo de aprendizaje.
- c) La cantidad de datos de entrenamiento necesarios.
- d) El tiempo que tarda la red en procesar los datos.

41. ¿Cuál de los siguientes tipos de redes neuronales es adecuado para el procesamiento de imágenes?

- a) Redes neuronales convolucionales (CNN)
- b) Redes neuronales recurrentes (RNN)
- c) **Redes neuronales de transformadores**
- d) Redes neuronales de alimentación hacia adelante

42. ¿Cuál de las siguientes NO es una aplicación común de las redes neuronales?

- a) **Compilación de código fuente**
- b) Reconocimiento de imágenes
- c) Traducción automática
- d) Diagnóstico médico

43. ¿Cuál es la relación entre las redes neuronales, el aprendizaje automático y la inteligencia artificial?

- a) **Las redes neuronales son una herramienta del aprendizaje automático, que a su vez es una rama de la inteligencia artificial.**
- b) El aprendizaje automático es una herramienta de las redes neuronales, que a su vez es una rama de la inteligencia artificial.
- c) La inteligencia artificial es una herramienta de las redes neuronales, que a su vez es una rama del aprendizaje automático.
- d) Las redes neuronales, el aprendizaje automático y la inteligencia artificial son campos independientes sin relación entre sí.

Sección 5: Procesamiento del Lenguaje Natural (7 preguntas)

44. ¿Cuál es el principal desafío del Procesamiento del Lenguaje Natural (NLP)?

- a) La velocidad de procesamiento de las computadoras modernas.
- b) **La complejidad y ambigüedad del lenguaje humano.**
- c) La falta de datos de texto disponibles para entrenar modelos.
- d) La dificultad para traducir entre diferentes idiomas.

45. ¿Cuál de los siguientes NO es un ejemplo de una técnica de representación vectorial de palabras?

- a) One-Hot Encoding
- b) Embeddings
- c) Tokenización
- d) **Word2Vec**

46. ¿Cuál es la principal ventaja de los embeddings sobre el One-Hot Encoding?

- a) Son más fáciles de implementar.
- b) **Capturan las relaciones semánticas entre palabras.**
- c) Requieren menos datos para el entrenamiento.
- d) Son más rápidos de procesar.

47. ¿Qué tipo de aprendizaje se utiliza principalmente en los Modelos de Lenguaje a Gran Escala (LLMs) modernos?

- a) Aprendizaje Supervisado c) Aprendizaje por Refuerzo
- b) **Aprendizaje Autosupervisado** d) Aprendizaje No Supervisado

48. ¿Cuál es una de las principales ventajas de la arquitectura de Transformers en el NLP?

- a) Son más fáciles de interpretar que las redes neuronales recurrentes (RNNs).
- b) **Superan las limitaciones de las RNNs en cuanto a la memoria a largo plazo.**
- c) Son menos costosas de entrenar que otros modelos de lenguaje.
- d) Son más precisas en tareas de clasificación de texto.

49. ¿Cuál de las siguientes NO es una aplicación común del NLP?

- a) **Análisis de Redes Sociales** c) Generación de Texto
- b) Traducción Automática d) Análisis de Sentimientos

50. ¿Cuál de las siguientes opciones describe mejor la principal ventaja de usar embeddings sobre one-hot encoding para representar palabras en NLP?

- a) Los embeddings son más fáciles de implementar que one-hot encoding.
- b) Los embeddings pueden manejar vocabularios más grandes que one-hot encoding.
- c) **Los embeddings capturan las relaciones semánticas entre palabras, mientras que one-hot encoding no.**
- d) Los embeddings requieren menos recursos computacionales para procesar que one-hot encoding.