

- Preguntas Guía
- Tareas Clave
- Tipos de problemas
- Preguntas SMART

Preguntas Guía

Preguntas que te ayudan a definir y entender el problema desde diferentes perspectivas.

Ejemplos:

- ¿Qué problema estás tratando de resolver?
Define el problema y cómo una solución basada en datos puede abordarlo.
Ejemplo: ¿Cómo predecir la demanda de productos para optimizar inventarios?
- ¿Qué datos están disponibles?
Evalúa qué datos necesitas, su calidad y disponibilidad.
Ejemplo: Datos históricos de ventas, inventario y eventos promocionales.
- ¿Qué modelo será más adecuado?
Considera el tipo de problema (clasificación, regresión, agrupamiento) y selecciona algoritmos apropiados.
Ejemplo: Usar un modelo de regresión lineal para predecir la demanda.
- ¿Cómo validarás los resultados?
Define métricas y estrategias de validación para evaluar la calidad del modelo.
Ejemplo: Usar RMSE para medir la precisión del modelo de predicción.
- ¿Cómo implementarán los interesados este modelo?
Considera cómo el modelo será usado en la práctica y quién lo usará.

Tareas Clave

Pasos esenciales para abordar un problema desde su definición hasta la implementación de la solución.

Pasos clave:

1. Identificar el problema: Define el contexto y sus implicaciones.
2. Recolectar datos relevantes: Asegúrate de que la información es suficiente y de calidad.
3. Seleccionar y entrenar modelos: Probar varios algoritmos y Ajustar hiperparámetros.
4. Evaluar el modelo: validación cruzada o comparar métricas como precisión, recall, F1-Score.
5. Implementar: Crear una API o tablero para que los interesados usen el modelo.

Tipos de problemas

1. Predicción (Regresión): Pronosticar valores futuros basándose en datos históricos (por ejemplo, predecir el precio de una vivienda).
2. Clasificación: Asignar etiquetas o categorías a los datos (por ejemplo, determinar si un correo es spam o no spam).
3. Detección de Anomalías: Identificar valores atípicos o comportamientos inusuales en los datos (por ejemplo, detectar transacciones fraudulentas en una tarjeta de crédito).
4. Agrupamiento (Clustering): Dividir los datos en grupos basados en similitudes, sin usar etiquetas predefinidas (por ejemplo, segmentar clientes según su comportamiento de compra).
5. Descubrimiento de Temas: Usar técnicas como el modelado de tópicos para identificar temas comunes en un conjunto de datos textuales (por ejemplo, descubrir temas en reseñas de productos).
6. Descubrimiento de Relaciones: Aplicar análisis de redes o métodos estadísticos para encontrar conexiones entre diferentes variables (por ejemplo, determinar si existe una correlación entre la edad y el ingreso).
7. Detección de Patrones: Usar algoritmos como series temporales o reglas de asociación para identificar tendencias recurrentes (por ejemplo, descubrir patrones de ventas estacionales).

Preguntas SMART

- (S)pecifico: definir claramente el problema.
- (M)edible: cuantificar el problema o el objetivo.
- (A)ction-oriented) Orientado a la acción: centrarse en los pasos para resolver el problema.
- (R)elavent) Relevante: relacionarse con los objetivos generales del proyecto.
- (T)ime-bound) Limitado en el tiempo: establecer un cronograma para alcanzar el objetivo.

Preguntas SMART aplicadas al conjunto de datos Iris

- Específicas: Define claramente el problema que deseas resolver.
Ejemplo: ¿Cómo se puede clasificar correctamente las especies de Iris (setosa, versicolor, virginica) basándose en las características de los pétalos y sépalos?

- **Medibles:** Cuantifica los objetivos o resultados deseados.
Ejemplo: ¿Qué nivel de precisión logra un modelo de clasificación al predecir las especies utilizando todas las variables?
- **Orientado a la acción:** Identifica pasos concretos para resolver el problema.
Ejemplo: ¿Qué características (largo y ancho del pétalo o del sépalo) tienen mayor peso en la clasificación de especies?
- **Relevantes:** Relaciona tus preguntas con el objetivo principal del análisis.
Ejemplo: ¿Cómo se pueden usar estos hallazgos para mejorar la clasificación de plantas en aplicaciones botánicas?
- **Limitadas en tiempo:** Establece plazos para completar las tareas.
Ejemplo: ¿Qué tan rápido puede un modelo basado en árboles de decisión entrenarse y validarse para este conjunto de datos?

Cosas que evitar al formular preguntas en el conjunto de datos Iris

- **Preguntas tendenciosas:** No supongas una respuesta sin análisis.
Ejemplo a evitar: ¿El ancho del pétalo siempre determina la especie de Iris?
Mejor: ¿Qué características tienen mayor impacto en la predicción de especies?
- **Preguntas cerradas:** Permite una exploración más profunda.
Ejemplo a evitar: ¿El largo del pétalo afecta la clasificación?
Mejor: ¿Cómo se correlaciona el largo del pétalo con las especies de Iris?
- **Preguntas vagas:** Sé específico y claro.
Ejemplo a evitar: ¿Qué variables son importantes?
Mejor: ¿Qué tan útiles son las medidas del sépalo en comparación con las del pétalo para clasificar las especies?