GUÍA DE PREPARACIÓN

Prueba Mixta Minería de Datos

Unidad I · 12 preguntas de práctica

A B C D

_____ completar _____

CASO

Interpretar → justificar → decidir

Cada pregunta tiene su respuesta en la lámina siguiente.

Estructura de la guía

La práctica mantiene 12 preguntas, siguiendo la hoja “Relación con la prueba”.

5 **Alternativas**

Reconocer y distinguir

3 **Completación**

Relacionar conceptos

4 **Casos aplicados**

Interpretar y decidir

Distribución de práctica

1–2	Fundamentos
3–4	Tipo de problema / aprendizaje
5 y 8	KDD / CRISP-DM
6 y 12	Entropía / ganancia
9	Estadística
10	Distancia / escalamiento
7 y 11	Aplicación, text/web mining y decisión

La guía no reproduce los casos del material anterior: usa situaciones nuevas con la misma lógica de interpretación y decisión.

🎯 Criterios evaluados

Los ejercicios se alinean con los cinco criterios indicados para la prueba.

1.1.1	**Minería de datos y aprendizaje automático**	Características, propósito y diferencias generales.
1.1.2	**Descubrimiento de conocimiento**	KDD y organización del proceso.
1.1.3	**Modelos, técnicas y áreas de aplicación**	Clasificación, regresión, clustering, anomalías, distancias, entropía.
1.1.4	**Text mining y web mining**	Técnicas aplicables en empresas y elección según el problema.
1.1.5	**Información y variables**	Identificar qué datos se necesitan, variable objetivo y variables explicativas.

🧠 Pregunta 1 · Fundamentos de Minería de Datos

Alternativas · Criterio 1.1.1

ALTERNATIVAS Criterio 1.1.1

Una cadena de farmacias posee años de ventas y quiere descubrir combinaciones de productos que suelen comprarse juntas para diseñar promociones. ¿Cuál descripción representa mejor el propósito de la Minería de Datos?

- **A** Mostrar un reporte mensual con las ventas totales.
- **B** Descubrir patrones útiles no evidentes en los datos.
- **C** Guardar las ventas en una base de datos.
- **D** Diseñar una interfaz para registrar las ventas.

Elige una alternativa y explica en una frase qué diferencia hay entre “mostrar datos” y “descubrir patrones”.

✅ Respuesta 1 · Fundamentos de Minería de Datos

Pauta de corrección y clave de interpretación

RESPUESTA

Alternativa correcta

B. Descubrir patrones útiles no evidentes en los datos.

La Minería de Datos busca extraer regularidades, relaciones o comportamientos que puedan transformarse en conocimiento útil para una decisión.

CLAVE DE INTERPRETACIÓN

Un reporte puede pertenecer a BI. Aquí el objetivo no es solo describir lo ocurrido, sino encontrar relaciones entre productos que no estaban explícitas.

Pregunta 2 · BI, Minería de Datos y Machine Learning

Completación · Criterio 1.1.1

COMPLETACIÓN Criterio 1.1.1

Completa las tres frases usando: BI · MINERÍA DE DATOS · MACHINE LEARNING

1 Un tablero que muestra ventas históricas por sucursal corresponde principalmente a: _____

2 Buscar patrones de abandono de clientes en una base histórica corresponde principalmente a: _____

3 Entrenar un modelo que prediga si un cliente abandonará el servicio corresponde principalmente a: _____

✅ Respuesta 2 · BI, Minería de Datos y Machine Learning

Pauta de corrección y clave de interpretación

RESPUESTA

Completación esperada

1. BI
2. Minería de Datos
3. Machine Learning

La frontera puede superponerse en proyectos reales, pero la intención principal de cada frase permite distinguirlas.

IDEA CENTRAL

BI: describir/monitorear.
Minería de Datos: descubrir patrones.
ML: aprender un modelo para predecir o decidir.

Pregunta 3 · Tipo de problema

Alternativas · Criterio 1.1.3

ALTERNATIVAS Criterio 1.1.3

Una empresa desea estimar el monto en pesos que gastará un cliente durante el próximo mes. ¿Qué tipo de problema corresponde?

A Clasificación

B Regresión

C Clustering

D Asociación

Pista: mira el tipo de salida que se quiere obtener.

✅ Respuesta 3 · Tipo de problema

Pauta de corrección y clave de interpretación

RESPUESTA

Alternativa correcta

B. Regresión.

La salida esperada es un valor numérico continuo: el monto futuro de compra. Si la pregunta fuera “¿comprará o no?”, sería clasificación.

REGLA RÁPIDA

Categoría → clasificación
Número continuo → regresión
Grupos sin etiqueta → clustering

Pregunta 4 · Supervisado y no supervisado

Completación · Criterios 1.1.3 y 1.1.5

COMPLETACIÓN

Criterios 1.1.3 y 1.1.5

Completa el razonamiento:

Si existe una columna histórica “fraude = sí/no”, entonces hay una _____; por ello el aprendizaje puede ser _____.

Si solo queremos agrupar clientes parecidos sin una etiqueta previa, el enfoque será _____.

VARIABLE OBJETIVO / ETIQUETA · SUPERVISADO · NO SUPERVISADO

✅ Respuesta 4 · Supervisado y no supervisado

Pauta de corrección y clave de interpretación

RESPUESTA

Completación esperada

VARIABLE OBJETIVO / ETIQUETA → SUPERVISADO → NO SUPERVISADO.

El punto decisivo es si existe o no una respuesta histórica conocida que el modelo debe aprender a reproducir.

CRITERIO 1.1.5

Antes de elegir la técnica, identifica qué variable representa la respuesta que deseas predecir.

Pregunta 5 · Proceso KDD

Alternativas · Criterio 1.1.2

ALTERNATIVAS Criterio 1.1.2

Un equipo ya seleccionó los datos relevantes y detectó valores faltantes, duplicados y formatos inconsistentes. ¿Qué etapa de KDD está realizando principalmente?

A Preprocesamiento

B Minería de datos

C Interpretación

D Despliegue

✅ Respuesta 5 · Proceso KDD

Pauta de corrección y clave de interpretación

RESPUESTA

Alternativa correcta

A. Preprocesamiento.

En esta etapa se mejora la calidad de los datos antes de transformarlos o aplicar técnicas de minería.

SECUENCIA KDD

Selección → Preprocesamiento → Transformación → Minería de datos → Interpretación/Evaluación

Pregunta 6 · Entropía

Alternativas · Criterio 1.1.3

50	50
COMPRAN	NO COMPRAN

ALTERNATIVAS Criterio 1.1.3

Un nodo de un árbol de decisión contiene 50 clientes que compran y 50 que no compran. ¿Cómo describirías su entropía en términos conceptuales?

A Baja: el grupo es muy puro.

B Alta: existe mucha incertidumbre o mezcla.

C Cero: ambas clases aparecen por igual.

D No puede interpretarse sin calcular la media.

✅ Respuesta 6 · Entropía

Pauta de corrección y clave de interpretación

RESPUESTA

Alternativa correcta

B. Alta: existe mucha incertidumbre o mezcla.

Cuando las clases están muy mezcladas, la incertidumbre es alta. Cuando un grupo queda dominado por una clase, la entropía disminuye.

IMAGEN MENTAL

50 / 50 → máxima mezcla
100 / 0 → grupo puro

La entropía mide incertidumbre, no “cantidad de datos”.

Pregunta 7 · Text Mining: opiniones de clientes

Caso aplicado · Criterios 1.1.3, 1.1.4 y 1.1.5

CASO APLICADO

Criterios 1.1.3, 1.1.4 y 1.1.5

CASO

Una empresa de delivery recibe 40.000 comentarios escritos al mes. La gerencia quiere saber cuáles reclamos se relacionan con retrasos, cobros y mala atención, y priorizar acciones.

Pregunta A

¿Qué enfoque es más apropiado: text mining o web mining? Justifica.

Pregunta B

¿Qué información o variables extra pedirías para convertir el hallazgo en una decisión concreta?

No basta nombrar una técnica: debes relacionarla con el tipo de dato y la decisión que se quiere tomar.

✅ Respuesta 7 · Text Mining: opiniones de clientes

Pauta de corrección y clave de interpretación

RESPUESTA

Respuesta esperada

A) Text mining, porque la fuente principal son comentarios de texto no estructurado. Se podrían extraer temas, palabras frecuentes, categorías o sentimiento.

B) Conviene asociar el texto con variables como fecha, local, tiempo real de entrega, tipo de pedido, monto, canal y evaluación del cliente. Así el análisis permite detectar dónde y cuándo ocurren los problemas y priorizar acciones.

DECISIÓN POSIBLE

Si “retraso” aparece especialmente en determinadas zonas y franjas horarias, la empresa puede ajustar dotación, rutas o tiempos prometidos.

Pregunta 8 · KDD y CRISP-DM

Completación · Criterio 1.1.2

COMPLETACIÓN Criterio 1.1.2

Completa con KDD o CRISP-DM:

A _____ pone especial énfasis en comprender el problema de negocio y definir qué significa éxito.

B _____ describe el proceso de descubrir conocimiento desde la selección hasta la interpretación de patrones.

C Ambos procesos son _____: al encontrar problemas, puede ser necesario volver a etapas anteriores.

PALABRAS: CRISP-DM · KDD · ITERATIVOS

✅ Respuesta 8 · KDD y CRISP-DM

Pauta de corrección y clave de interpretación

RESPUESTA

Completación esperada

A) CRISP-DM
B) KDD
C) ITERATIVOS

La diferencia central está en el énfasis: CRISP-DM parte explícitamente desde el negocio; KDD organiza el descubrimiento de conocimiento en los datos.

NO CONFUNDIR

“Iterativo” significa que el proyecto puede retroceder y ajustar decisiones anteriores. No es una secuencia rígida que se ejecuta una sola vez.

Pregunta 9 · Estadística e interpretación

Caso aplicado · Criterios 1.1.3 y 1.1.5

CASO APLICADO

Criterios 1.1.3 y 1.1.5

TIEMPO DE ATENCIÓN (min)

7 8 8 9 28

La jefatura quiere definir un tiempo “típico” de atención.

Pregunta A

Calcula media, mediana y moda. ¿Cuál representa mejor el tiempo típico?

Pregunta B

¿Qué debería revisar la empresa antes de cambiar su estándar de atención?

✅ Respuesta 9 · Estadística e interpretación

Pauta de corrección y clave de interpretación

RESPUESTA

Respuesta esperada

Media = 12 min; mediana = 8 min; moda = 8 min.

La mediana representa mejor el comportamiento típico porque el valor 28 eleva fuertemente la media. Antes de cambiar un estándar, conviene revisar si 28 es un caso real, un evento excepcional o un error de registro.

DECISIÓN

Un estadístico no debe tomarse aisladamente. La medida elegida debe representar el fenómeno y considerar posibles outliers.

Pregunta 10 · Distancia y escalamiento

Caso aplicado · Criterios 1.1.3 y 1.1.5

CASO APLICADO

Criterios 1.1.3 y 1.1.5

CLIENTES

A: edad 25 · ingreso $600.000
B: edad 27 · ingreso $650.000
C: edad 55 · ingreso $620.000

Se quiere usar KNN para recomendar productos.

Pregunta A

¿Por qué sería problemático usar directamente edad e ingreso en la distancia euclidiana?

Pregunta B

¿Qué acción de preparación de datos realizarías antes de calcular similitud?

✅ Respuesta 10 · Distancia y escalamiento

Pauta de corrección y clave de interpretación

RESPUESTA

Respuesta esperada

El ingreso tiene una escala numérica muchísimo mayor que la edad, por lo que dominaría la distancia.

Antes de usar KNN conviene escalar o normalizar las variables para que ambas contribuyan de forma comparable. Luego la cercanía puede interpretarse como similitud en el espacio de características.

IDEA VISUAL

Sin escalar: “pesos” aplastan a “años”.
Con escalamiento: cada variable aporta según su posición relativa, no por el tamaño de su unidad.

Pregunta 11 · Web Mining y decisión comercial

Caso aplicado · Criterios 1.1.3, 1.1.4 y 1.1.5

CASO APLICADO

Criterios 1.1.3, 1.1.4 y 1.1.5

CASO

Un e-commerce detecta que muchas visitas llegan desde redes sociales, revisan varios productos y abandonan antes de pagar. La empresa quiere entender el recorrido digital y anticipar qué sesiones tienen mayor probabilidad de compra.

Pregunta A

¿Qué tipo de información de web mining sería útil para describir el recorrido?

Pregunta B

Si se quiere predecir compra/no compra, identifica variable objetivo, dos variables explicativas y el tipo de problema.

✅ Respuesta 11 · Web Mining y decisión comercial

Pauta de corrección y clave de interpretación

RESPUESTA

Respuesta esperada

A) Datos de navegación: páginas vistas, secuencia de clics, fuente de tráfico, tiempo por página, productos visitados, carrito y punto de abandono.

B) Objetivo: compra/no compra. Variables explicativas posibles: número de productos vistos, tiempo de sesión, origen del tráfico, uso de cupón, monto del carrito. Tipo: clasificación supervisada.

DECISIÓN POSIBLE

La empresa podría activar una intervención distinta según el comportamiento: recordatorio, cupón, recomendación o mejora del flujo de compra.

Pregunta 12 · Ganancia de información y selección de variables

Caso aplicado · Criterios 1.1.3 y 1.1.5

CASO APLICADO

Criterios 1.1.3 y 1.1.5

Un modelo quiere predecir si un estudiante aprobará. Se prueban dos variables para dividir los datos:

VARIABLE A

“Entregó actividades”

Después de dividir, los grupos quedan bastante puros.

VARIABLE B

“Color favorito”

Después de dividir, los grupos continúan muy mezclados.

¿Cuál variable tendría mayor ganancia de información? ¿Qué implica eso para la selección de variables del modelo?

✅ Respuesta 12 · Ganancia de información y selección de variables

Pauta de corrección y clave de interpretación

RESPUESTA

Respuesta esperada

“Entregó actividades” tendría mayor ganancia de información porque reduce más la incertidumbre sobre la aprobación.

La ganancia de información ayuda a identificar variables que separan mejor las clases. Una variable con baja ganancia puede aportar poco poder discriminativo, aunque su utilidad final debe validarse en el contexto del modelo.

DECISIÓN ANALÍTICA

No se selecciona una variable porque “suene interesante”, sino porque demuestra aportar información útil para distinguir la variable objetivo.

🏁 Cómo responder bien una prueba mixta

Una estrategia corta para alternativas, completación y casos.

1. **Identifica qué te preguntan** — Concepto, proceso, técnica, cálculo, interpretación o decisión.
2. **Busca la pista principal** — ¿Hay etiqueta? ¿La salida es categoría o número? ¿El dato es texto/web?
3. **Descarta incompatibles** — Una técnica debe coincidir con el objetivo y el tipo de datos.
4. **Justifica con evidencia** — En casos, cita una característica concreta del enunciado.
5. **Conecta con una decisión** — Un buen análisis termina explicando qué haría la organización.

Meta: demostrar comprensión + aplicación + interpretación, no solo memoria.