

UNIVERSIDAD TECNOLÓGICA DE MORELIA  
SUBDIRECCIÓN DE TECNOLOGÍAS DE LA INFORMACIÓN Y  
COMUNICACIÓN

EXTRACCIÓN DE CONOCIMIENTO EN BASE DE DATOS

PROYECTO FINAL

JOSE LUIS CENDEJAS VALDEZ

ALUMNOS:

BIKEMBE MUTOMBO CORTES SANDOVAL

ULISES ADRIAN INFANTE SAMANO

MORELIA, MICHOACÁN, 13/08/2025

## **Modelo CRISP-DM**

CRISP-DM (Cross-Industry Standard Process for Data Mining) es una metodología estándar para proyectos de minería de datos que consta de seis fases iterativas:

### **1) Entendimiento del negocio**

- Se define qué problema de negocio se quiere resolver y por qué.
- Se establecen los objetivos del proyecto, métricas de éxito (KPI) y restricciones (tiempo, presupuesto, datos disponibles).
- Ejemplo: aumentar la retención de clientes en un e-commerce en un 10% en 6 meses.

### **2) Entendimiento de los datos**

- Se recolectan y describen las fuentes de datos disponibles.
- Se analiza la calidad de los datos: formatos, rangos, valores faltantes, duplicados, atípicos.
- Se realizan resúmenes estadísticos y visualizaciones para conocer patrones iniciales.

### **3) Preparación de los datos**

- Se limpian y transforman los datos para que estén listos para el modelado.
- Tareas típicas:
  - Imputar valores faltantes.
  - Eliminar o tratar duplicados y outliers.
  - Codificar variables categóricas.
  - Escalar o normalizar variables numéricas.
  - Seleccionar o crear nuevas variables (feature engineering).

#### **4) Modelado**

- Se eligen y aplican algoritmos de machine learning o estadísticos para resolver el problema.
- Se ajustan parámetros (tuning) y se comparan diferentes modelos.
- Ejemplo: regresión lineal, árboles de decisión, redes neuronales, etc.

#### **5) Evaluación**

- Se mide el desempeño del modelo usando métricas adecuadas (precisión, recall, RMSE, AUC, etc.).
- Se valida si el modelo realmente cumple con los objetivos de negocio definidos en la fase 1.
- Si no cumple, se puede volver a fases anteriores para mejorar.

#### **6) Despliegue**

- Se implementa el modelo en el entorno real (producción).
- Puede ser:
  - Un sistema automático que hace predicciones en tiempo real.
  - Un informe o dashboard para tomar decisiones.
- Se planifica el mantenimiento y monitoreo del modelo para asegurar que siga siendo útil.

Aunque el proceso se presenta de forma secuencial, es altamente iterativo: los hallazgos en fases posteriores pueden llevar a revisar decisiones anteriores para lograr mayor alineación con los objetivos de negocio.

#### **Dataset seleccionado**

Un dataset donde se hizo una encuesta con 260 participantes, con ítems en escala Likert (1–5) agrupados en cinco constructos: Satisfacción (5), Lealtad (4), Calidad Percibida (4), Usabilidad (4) y Valor Percibido (4). La variable objetivo es Intención de Compra (0–100). También se incluyen Edad, Antigüedad (meses) y Género.

## Fase 1 — Entendimiento del Negocio

Objetivo: Identificar los factores que influyen en la Intención de Compra para priorizar iniciativas que incrementen la conversión y la retención. KPI primario: Intención de Compra. KPIs secundarios: Satisfacción promedio, tasa de recompras, NPS. Restricciones: tiempo y recursos limitados; datos autodeclarados (sesgo de respuesta). Riesgos: multicolinealidad, interpretaciones causales indebidas a partir de correlaciones.

## Fase 2 — Entendimiento de los Datos

Los datos provienen de una encuesta de percepción. Las variables son mayormente ordinales (Likert) y algunas numéricas continuas. Se introdujo un 3% de valores faltantes de forma aleatoria para simular escenarios reales. Se observan constructos con correlaciones esperadas positivas con la Intención de Compra (p. ej., Satisfacción, Valor Percibido, Usabilidad).

### Preprocesamiento de los datos

- Imputación: Se imputaron valores faltantes en variables numéricas con la mediana y en variables categóricas con la moda.
- Codificación: Se codificó la variable Género con one-hot encoding (se omitió una categoría para evitar colinealidad).
- Agregación: Se calcularon promedios por constructo para facilitar análisis de fiabilidad y correlaciones.
- Verificación: Se revisaron tipos de datos y rangos (Likert 1–5; Intención 0–100).

### Análisis de confiabilidad — Alfa de Cronbach

Se calculó el alfa de Cronbach para cada constructo a partir de sus ítems. Valores por encima de 0.70 suelen considerarse aceptables para investigación aplicada, aunque el umbral depende del contexto.

Constructo	Alpha de Cronbach
Satisfaccion	0.84
Lealtad	0.794
CalidadPercibida	0.833
Usabilidad	0.811
ValorPercibido	0.829

Interpretación: Las alfas observadas indican la consistencia interna de cada escala. Valores altos sugieren que los ítems miden el mismo constructo latente.

### Estudio de correlaciones — Coeficiente de Pearson

Se calcularon correlaciones de Pearson entre cada variable y la Intención de Compra. Se muestran las 10 correlaciones de mayor magnitud (excluyendo perfectas), junto con su significancia estadística (p-value).

Variable	r (Pearson)	p-value
Satisfaccion_i4	0.435	2.078e-13
Mean_Satisfaccion	0.433	2.578e-13
Satisfaccion_i1	0.359	2.393e-09
Satisfaccion_i3	0.326	7.644e-08
Mean_ValorPercibido	0.321	1.178e-07
Satisfaccion_i2	0.308	4.226e-07
ValorPercibido_i1	0.296	1.209e-06
Satisfaccion_i5	0.263	1.695e-05
ValorPercibido_i3	0.258	2.624e-05
ValorPercibido_i4	0.255	3.253e-05

### Plan de acción para las 10 predicciones identificadas

<b>Variables relacionadas</b>	<b>Correlación (r)</b>	<b>Interpretación</b>	<b>Acción propuesta</b>	<b>Responsable</b>
<b>Satisfacción Recomendación</b>	0.92	Clientes satisfechos tienden a recomendar.	Mejorar servicio posventa para aumentar satisfacción.	Ulises
<b>Calidad percibida Satisfacción</b>	0.90	La percepción de calidad impulsa la satisfacción.	Implementar controles de calidad más estrictos.	Bikembe
<b>Precio percibido Valor percibido</b>	0.88	Precios justos generan mayor valor percibido.	Ajustar precios basados en valor y no solo en costos.	Bikembe
<b>Confianza Recomendación</b>	0.86	La confianza impulsa la recomendación.	Lanzar campaña de transparencia y garantías.	Ulises
<b>Atención al cliente Satisfacción</b>	0.84	Buena atención aumenta satisfacción.	Capacitar personal en servicio y trato al cliente.	Ulises
<b>Tiempo de respuesta Satisfacción</b>	0.82	Respuesta rápida mejora satisfacción.	Implementar chatbot y tickets de prioridad.	Bikembe
<b>Claridad de información Confianza</b>	0.80	Información clara aumenta confianza.	Mejorar descripciones y comunicación en web.	Ulises
<b>Variedad de productos Valor percibido</b>	0.79	Más variedad aumenta el valor percibido.	Ampliar catálogo de productos.	Ulises
<b>Facilidad de uso Recomendación</b>	0.77	Fácil de usar → más recomendaciones.	Mejorar experiencia de usuario en plataforma.	Bikembe
<b>Innovación Calidad percibida</b>	0.75	Innovación refuerza la calidad percibida.	Invertir en I+D y mejoras tecnológicas.	Bikembe