



DM-CRISP

KENIA FERNANDA FLORES ALVA
LEONARDO KENJI MINEMURA
ANGEL ESQUEDA OCHOA
CHRISTIAN MATOS

¿Qué es la minería de datos?

La minería de datos es el proceso de descubrir patrones y conocimientos útiles en grandes conjuntos de datos. Utiliza técnicas estadísticas e inteligencia artificial para extraer información valiosa y tomar decisiones informadas en campos como segmentación de clientes, detección de fraudes, optimización de procesos, análisis de mercado y medicina.



Ejemplos

Sector financiero

Detección de fraudes en transacciones bancarias.



Marketing y publicidad

Segmentación de clientes y personalización de campañas.



Transporte y logística

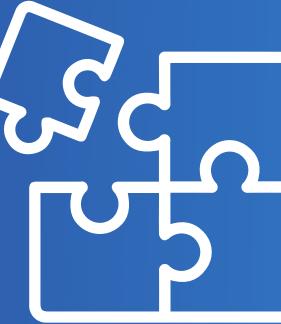
Optimización de rutas de entrega y gestión de la cadena de suministro



Metodología CRISP-DM

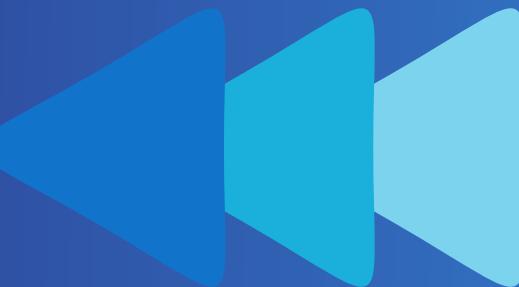
CRoss Industry Standard Process for Data Mining (CRISP-DM) es el arreglo de etapas por las que pasa una unidad específica de información desde su era de inicio hasta su posible documentación y/o cancelación al concluir su vida valiosa.





- Como **metodología**, incluye descripciones de las fases normales de un proyecto, las tareas necesarias en cada fase y una explicación de las relaciones entre las tareas.
- Como **modelo de proceso**, CRISP-DM ofrece un resumen del ciclo vital de minería de datos.





1. Comprensión del negocio

¿Qué necesita el negocio?

2. Comprensión de los datos

¿Qué información tenemos y necesitamos?

3. Preparación de los datos

¿Cómo transformamos nuestros datos para análisis?

4. Modelado

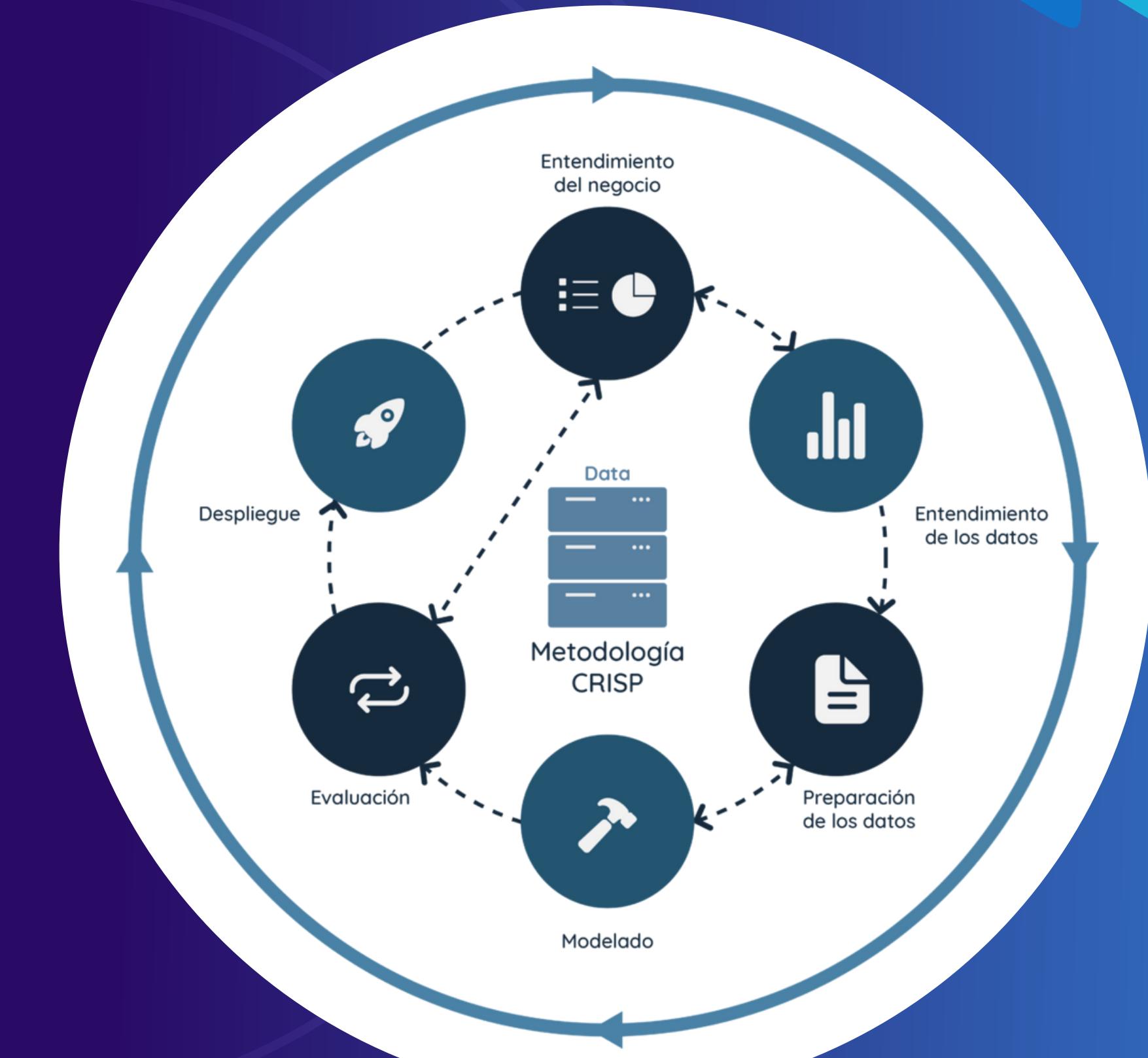
¿Qué técnicas de minería deberíamos aplicar?

5. Evaluación

¿Qué modelo coincide con los criterios de éxito?

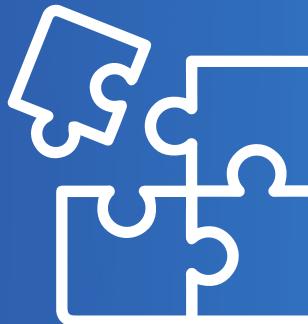
6. Despliegue

¿Cómo comunicamos nuestros descubrimientos?





Publicado en 1999 para estandarizar los procesos de minería de datos en todas las industrias, desde entonces se ha convertido en la metodología más común para proyectos de minería de datos, análisis y ciencia de datos.



Es probable que los equipos de ciencia de datos que combinen una implementación flexible de CRISP-DM con enfoques generales de gestión de proyectos ágiles basados en equipos obtengan los mejores resultados.

El ciclo vital del modelo contiene seis fases con flechas que indican las dependencias más importantes y frecuentes entre fases. La secuencia de las fases no es estricta. De hecho, la mayoría de los proyectos avanzan y retroceden entre fases si es necesario.



El modelo de CRISP-DM es flexible y se pueden personalizar fácilmente.

Por ejemplo, si su organización intenta detectar actividades de blanqueo de dinero, es probable que necesite realizar una criba de grandes cantidades de datos sin un objetivo de modelado específico. En lugar de realizar el modelado, su trabajo se centrará en explorar y visualizar datos para descubrir patrones sospechosos en datos financieros. CRISP-DM permite crear un modelo de minería de datos que se adapte a sus necesidades concretas.

En tal situación, las fases de modelado, evaluación y despliegue pueden ser menos relevantes que las fases de preparación y comprensión de datos. Sin embargo, es muy importante considerar algunas cuestiones que surgen durante fases posteriores para la planificación a largo plazo y objetivos futuros de minería de datos.

Diferencias con KPP



Enfoque

CRISP-DM se centra en el proceso de minería de datos, mientras que KDD abarca todo el descubrimiento de conocimientos.

Ámbito de aplicación

CRISP-DM se utiliza en proyectos de minería de datos en diversos sectores, mientras que KDD abarca disciplinas como la IA y el aprendizaje automático.

Fases y pasos

CRISP-DM consta de 6 fases principales, mientras que KDD incluye etapas adicionales como selección y transformación de datos.

Orientación

CRISP-DM es práctico y orientado a proyectos, mientras que KDD se enfoca en principios teóricos y técnicas de descubrimiento de conocimientos.



Caso de Uso

Tomaremos como ejemplo una empresa de comercio electrónico que desea mejorar su sistema de recomendación de productos para aumentar las ventas y la satisfacción del cliente.





Comprendión del negocio

- La empresa quiere aumentar las ventas y la retención de clientes mediante recomendaciones personalizadas de productos.
- Se establece como objetivo principal mejorar la precisión y relevancia de las recomendaciones.

Comprendión de los datos



- Se recopilan datos transaccionales de los clientes, incluyendo historiales de compras, productos vistos y acciones en el sitio web.
- Se evalúa la calidad de los datos y se identifican los campos relevantes para el análisis, como productos comprados, categorías, preferencias y valoraciones de los clientes.

Preparación de los datos



- Se limpian los datos, eliminando valores nulos, duplicados o inconsistentes.
- Se transforman los datos en un formato adecuado para el análisis, como codificar variables categóricas y normalizar valores numéricos.
- Se seleccionan las variables relevantes para construir el modelo, como historiales de compras, preferencias de productos y datos demográficos de los clientes.

Modelado



- Se eligen técnicas de minería de datos, como filtrado colaborativo y análisis de asociación, para generar recomendaciones personalizadas.
- Se construye un modelo de filtrado colaborativo basado en la similitud de preferencias de los clientes para generar recomendaciones de productos.
- Se implementa un modelo de análisis de asociación para descubrir patrones de compra frecuentes y sugerir productos complementarios.

Evaluación



- Se evalúa la precisión y relevancia de las recomendaciones utilizando métricas como la tasa de aciertos y la medida de relevancia.
- Se realizan pruebas A/B para comparar las recomendaciones generadas por el modelo con el sistema anterior y se recopilan los comentarios de los usuarios.

Despliegue



- Se implementa el nuevo sistema de recomendación en la plataforma de comercio electrónico.
- Se realiza un seguimiento del rendimiento del modelo y se realizan ajustes según sea necesario.
- Se monitorea continuamente la satisfacción del cliente y se recopilan datos adicionales para mejorar aún más las recomendaciones.

CUESTIONARIO

01

¿Qué significa CRISP-DM?

02

¿Cuántas fases tiene el modelo CRISP-DM?

03

¿Cuál es el objetivo de la fase "Comprensión del negocio" en CRISP-DM?

04

¿Qué implica la fase "Comprensión de los datos" en CRISP-DM?

05

¿Cómo se diferencia CRISP-DM de KPP-DM?



¿Preguntas?



A hand, shown from the side and wearing a teal sleeve, is giving a thumbs-up gesture against a dark purple background. The hand is positioned in the upper right quadrant of the frame. In the lower left foreground, there are large, semi-transparent circular shapes in light blue, purple, yellow, and orange.

¡Gracias!