

El proceso de minería de datos, siguiendo el enfoque CRISP-DM (Cross-Industry Standard Process for Data Mining). Este proceso consta de seis fases principales:

1. **Business Understanding (Comprensión del Negocio):**

- **Descripción:** En esta fase se establece el conocimiento del dominio de negocio, se determinan los objetivos y requisitos del proyecto desde la perspectiva del negocio y se convierte este conocimiento en un problema de minería de datos.
- **Perfiles Profesionales:** Expertos en el negocio, analistas de negocio, consultores de negocio.
- **Nivel de Trabajo:** Alto, ya que requiere una comprensión profunda del contexto del negocio y una alineación clara entre los objetivos de negocio y los objetivos de minería de datos.

2. **Data Understanding (Comprensión de los Datos):**

- **Descripción:** Aquí se recopilan los datos iniciales y se procede a familiarizarse con ellos, identificando problemas de calidad de datos, detectando patrones y entendiendo las propiedades de los datos.
- **Perfiles Profesionales:** Científicos de datos, analistas de datos, ingenieros de datos.
- **Nivel de Trabajo:** Moderado a alto, debido a la necesidad de explorar, limpiar y validar los datos disponibles.

3. **Data Preparation (Preparación de los Datos):**

- **Descripción:** En esta fase se seleccionan los datos que se utilizarán en el análisis y se preparan mediante limpieza, transformación y formateo, asegurando que los datos están listos para el modelado.
- **Perfiles Profesionales:** Científicos de datos, ingenieros de datos, analistas de datos.
- **Nivel de Trabajo:** Alto, ya que puede implicar una cantidad significativa de manipulación de datos y procesamiento para asegurar que los datos sean adecuados para los modelos.

4. **Modeling (Modelado):**

- **Descripción:** Se seleccionan y aplican las técnicas de modelado, calibrando los parámetros del modelo para optimizar su desempeño. Se construyen los modelos de datos para resolver el problema planteado.
- **Perfiles Profesionales:** Científicos de datos, ingenieros de machine learning, estadísticos.
- **Nivel de Trabajo:** Moderado, dependiendo de la complejidad de los modelos y la cantidad de pruebas necesarias para optimizarlos.

## 5. Evaluation (Evaluación):

- **Descripción:** Se evalúan los modelos construidos para verificar que cumplen con los objetivos del negocio y de minería de datos, asegurando que no se ha pasado por alto ningún aspecto importante del problema.
- **Perfiles Profesionales:** Científicos de datos, analistas de negocio, consultores de datos.
- **Nivel de Trabajo:** Moderado, ya que implica análisis críticos y aseguramiento de la calidad del modelo antes de su implementación.

## 6. Deployment (Despliegue):

- **Descripción:** En esta fase, los modelos se implementan en el entorno productivo y se realizan las acciones necesarias para monitorear y mantener los modelos en operación.
- **Perfiles Profesionales:** Ingenieros de datos, desarrolladores de software, administradores de sistemas.
- **Nivel de Trabajo:** Moderado a bajo, dependiendo de la complejidad del sistema en el que se despliega y la necesidad de monitoreo continuo.

### Fases con Más y Menos Trabajo:

- **Más Trabajo:**
  - **Preparación de los Datos:** Por lo general, esta fase es la más laboriosa debido a la necesidad de limpiar y transformar grandes volúmenes de datos para que sean aptos para el modelado.
  - **Comprensión del Negocio:** Requiere un entendimiento profundo del contexto del negocio y cómo los objetivos de negocio se traducen en objetivos de minería de datos.
- **Menos Trabajo:**
  - **Despliegue:** Una vez que los modelos están listos y evaluados, el despliegue puede ser relativamente menos laborioso, aunque depende de la complejidad del entorno de producción.
  - **Evaluación:** Aunque crucial, tiende a ser menos laboriosa comparada con la preparación de datos y la comprensión del negocio.

Este diagrama ilustra un ciclo iterativo donde cada fase puede retroalimentar a la anterior, lo que asegura un proceso continuo de mejora y ajuste.