



Introducción a la Ingeniería de Datos. Metodologías para la Ciencia de Datos.

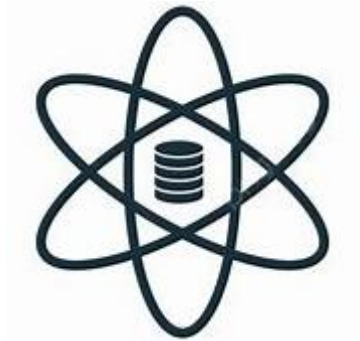
Detalles del curso



Conferencias.
Clases practices.
Seminarios.
Proyectos.
Evaluaciones sistemáticas.

Definiendo la ciencia de datos

Definiendo la ciencia de datos



Campo multidisciplinario

Estudio de los datos

Extraer conocimiento y valor

Definiendo qué es una metodología

Definiendo qué es una metodología

Estrategia, guía o conjunto de pautas

Proporcionan un marco de trabajo

Ayudan a desarrollar un proceso o actividad de manera estructurada

Desligado de un software o hardware particular

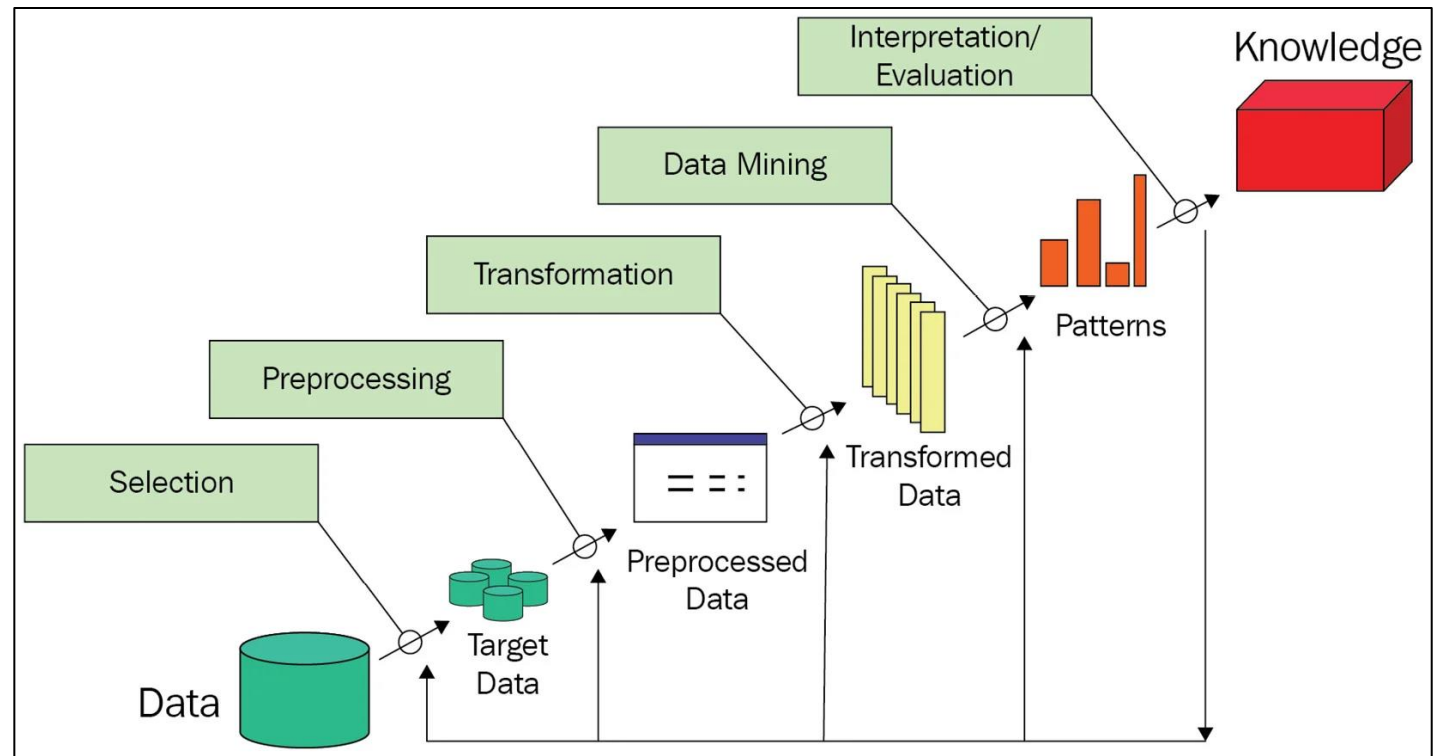


KDD (Knowledge Discovery in Databases)

1. Selección
2. Procesamiento/Limpieza
3. Transformación/Reducción
4. Minería de Datos
5. Interpretación/Evaluación

KDD (Knowledge Discovery in Databases)

1. Selección
2. Procesamiento/Limpieza
3. Transformación/Reducción
4. Minería de Datos
5. Interpretación/Evaluación



CRISP-DM

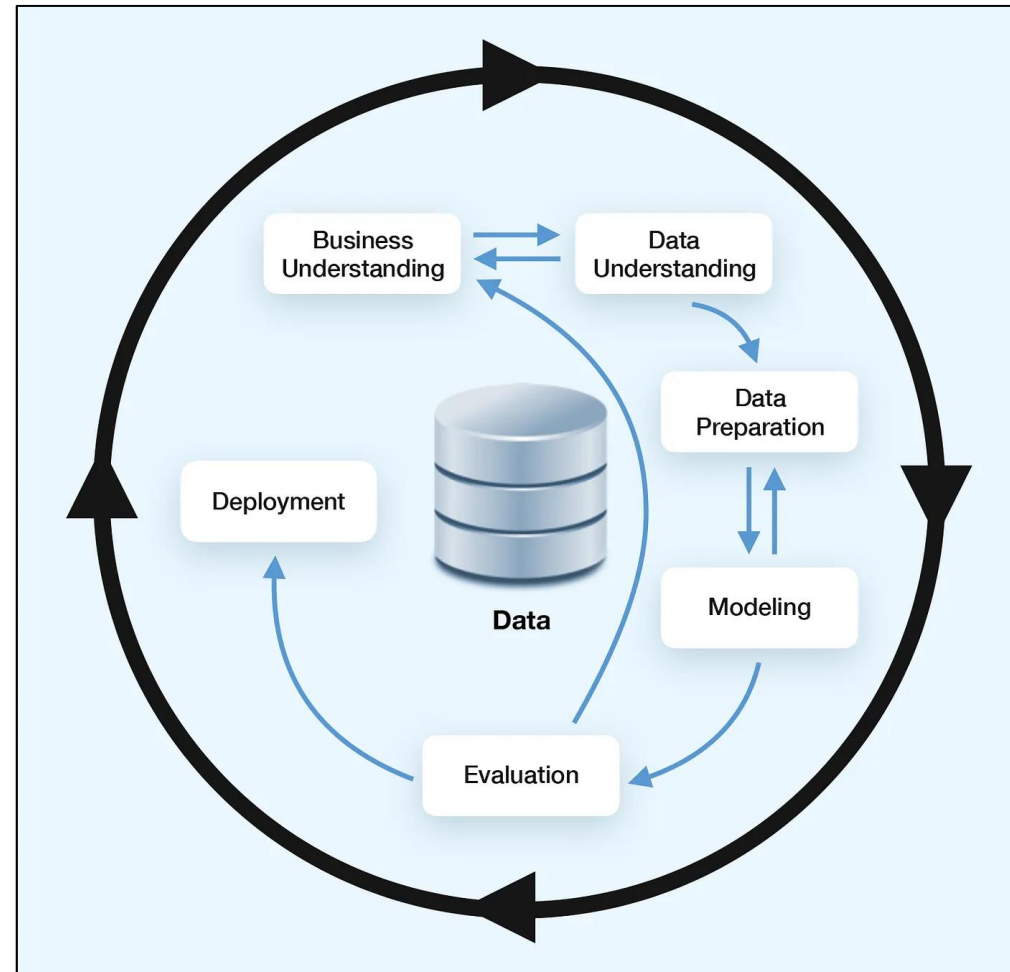
Cross-Industry Standard Process for Data Mining

1. Comprensión Empresarial
2. Comprensión de Datos
3. Preparación de Datos
4. Modelado
5. Evaluación
6. Implementación

CRISP-DM

Cross-Industry Standard Process for Data Mining

1. Comprensión Empresarial
2. Comprensión de Datos
3. Preparación de Datos
4. Modelado
5. Evaluación
6. Implementación



SEMMA

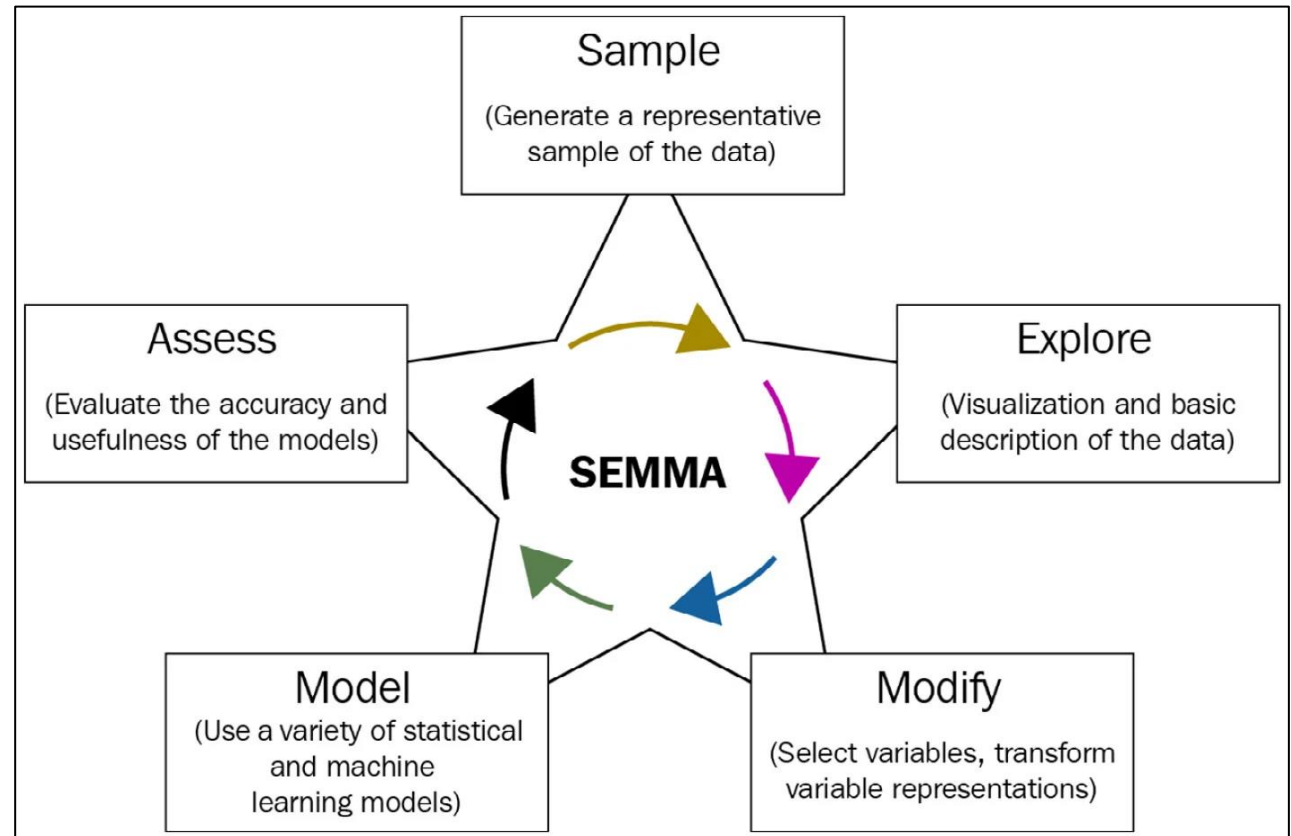
Sample, Explore, Modify, Model, Assess

1. Muestreo
2. Exploración
3. Modificación
4. Modelado
5. Evaluación

SEMMA

Sample, Explore, Modify, Model, Assess

1. Muestreo
2. Exploración
3. Modificación
4. Modelado
5. Evaluación

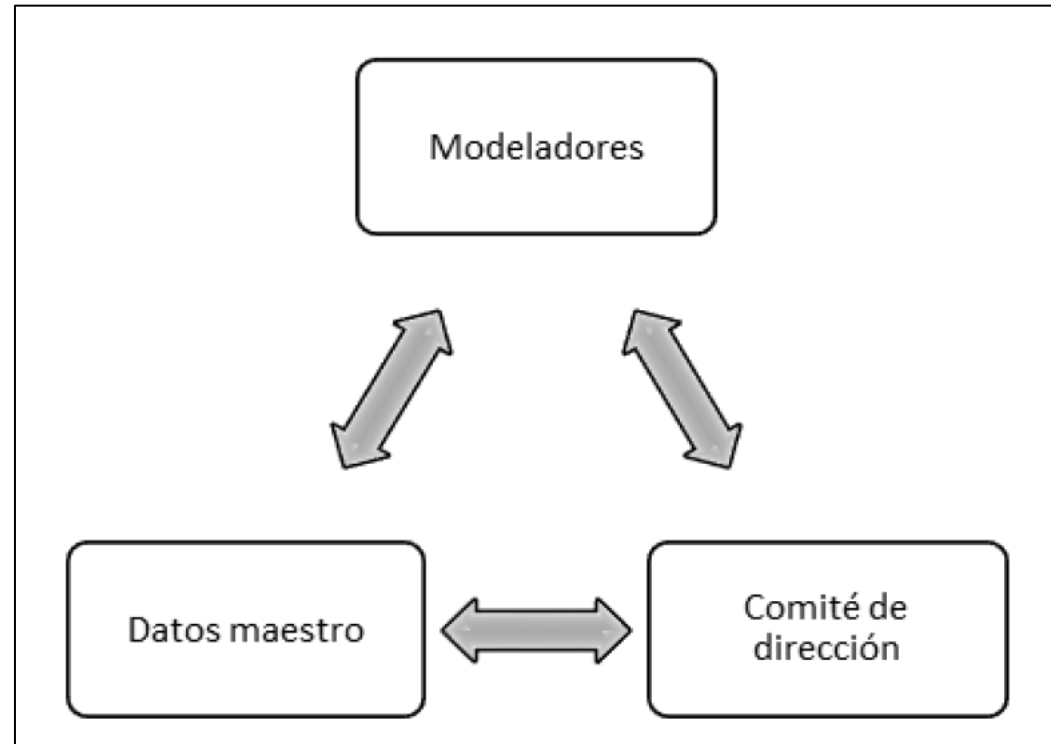


RAMSYS

Rapid collaborative data Mining System

Roles:

1. Modeladores
2. Data Master
3. Comité de Dirección



TDSP (Team Data Science Process)

Etapas:

1. Comprensión empresarial
2. Adquisición y comprensión de datos
3. Modelado
4. Implementación
5. Aceptación del cliente

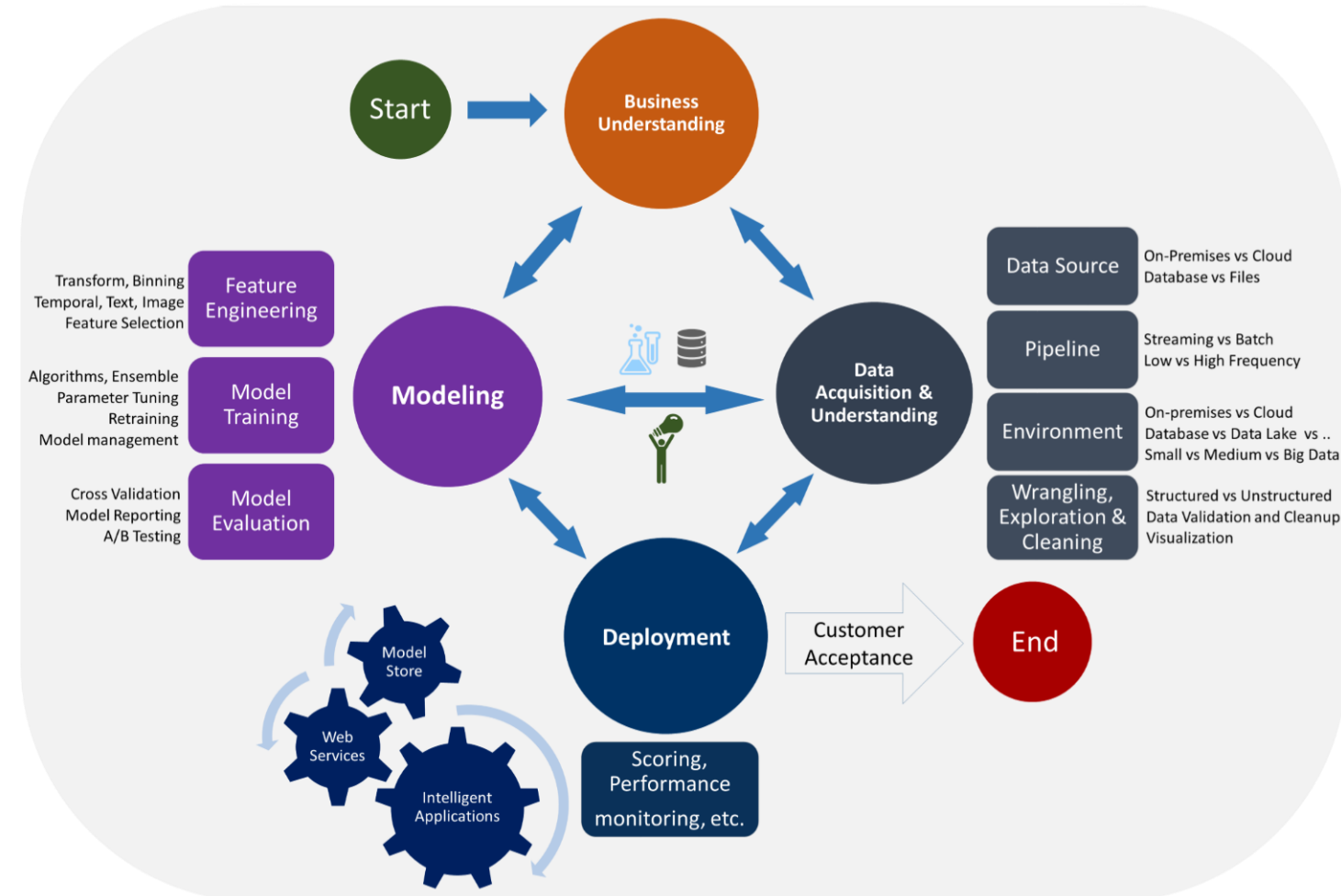
Roles:

- Arquitecto de soluciones
- Project Manager
- Ingeniero de datos
- Científico de datos
- Desarrollador de aplicaciones
- Líder de proyecto (Project lead)

TDSP (Team Data Science Process)

Etapas:

1. Comprensión empresarial
2. Adquisición y comprensión de datos
3. Modelado
4. Implementación
5. Aceptación del cliente



Elementos comunes en las metodologías

Bussiness Understanding



Comprensión del Negocio: Definir el problema y los objetivos.

Elementos comunes en las metodologías

Data Understanding

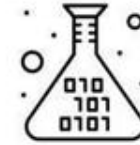


Comprensión del Negocio: Definir el problema y los objetivos.

Comprensión de los Datos: Identificar, recopilar, explorar, comprender.

Elementos comunes en las metodologías

Data Preparation



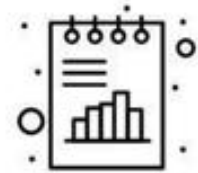
Comprensión del Negocio: Definir el problema y los objetivos.

Comprensión de los Datos: Identificar, recopilar, explorar, comprender.

Preparación de los Datos: Limpiar, transformar y preparar los datos.

Elementos comunes en las metodologías

Data Modeling - Evaluation



Comprensión del Negocio: Definir el problema y los objetivos.

Comprensión de los Datos: Identificar, recopilar, explorar, comprender.

Preparación de los Datos: Limpiar, transformar y preparar los datos.

Modelado de Datos: Selección de algoritmos, ajuste de parámetros y validación del modelo.

Evaluación: Evaluar el rendimiento del modelo.

Elementos comunes en las metodologías

Deployment



Comprensión del Negocio: Definir el problema y los objetivos.

Comprensión de los Datos: Identificar, recopilar, explorar, comprender.

Preparación de los Datos: Limpiar, transformar y preparar los datos.

Modelado de Datos: Selección de algoritmos, ajuste de parámetros y validación del modelo.

Evaluación: Evaluar el rendimiento del modelo.

Despliegue: Desplegar en un entorno de producción.

Elementos comunes en las metodologías

Monitoring & Maintenance

Comprensión del Negocio: Definir el problema y los objetivos.

Comprensión de los Datos: Identificar, recopilar, explorar, comprender.

Preparación de los Datos: Limpiar, transformar y preparar los datos.

Modelado de Datos: Selección de algoritmos, ajuste de parámetros y validación del modelo.

Evaluación: Evaluar el rendimiento del modelo.

Despliegue: Desplegar en un entorno de producción.

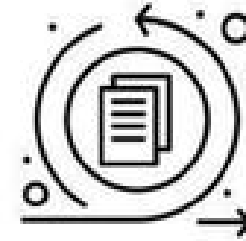
Monitoreo y Mantenimiento: Supervisar el rendimiento en producción y realizar ajustes.

Conclusiones

Proceso Iterativo y Cíclico

Aplicar metodologías con Flexibilidad

Comunicación y Colaboración





**Data Creation
& Data Capture**

**Analytics &
Data Scientists**