

Ciencia de Datos

Metodologías



Rogelio Ferreira Escutia

Profesor / Investigador
Tecnológico Nacional de México
Campus Morelia

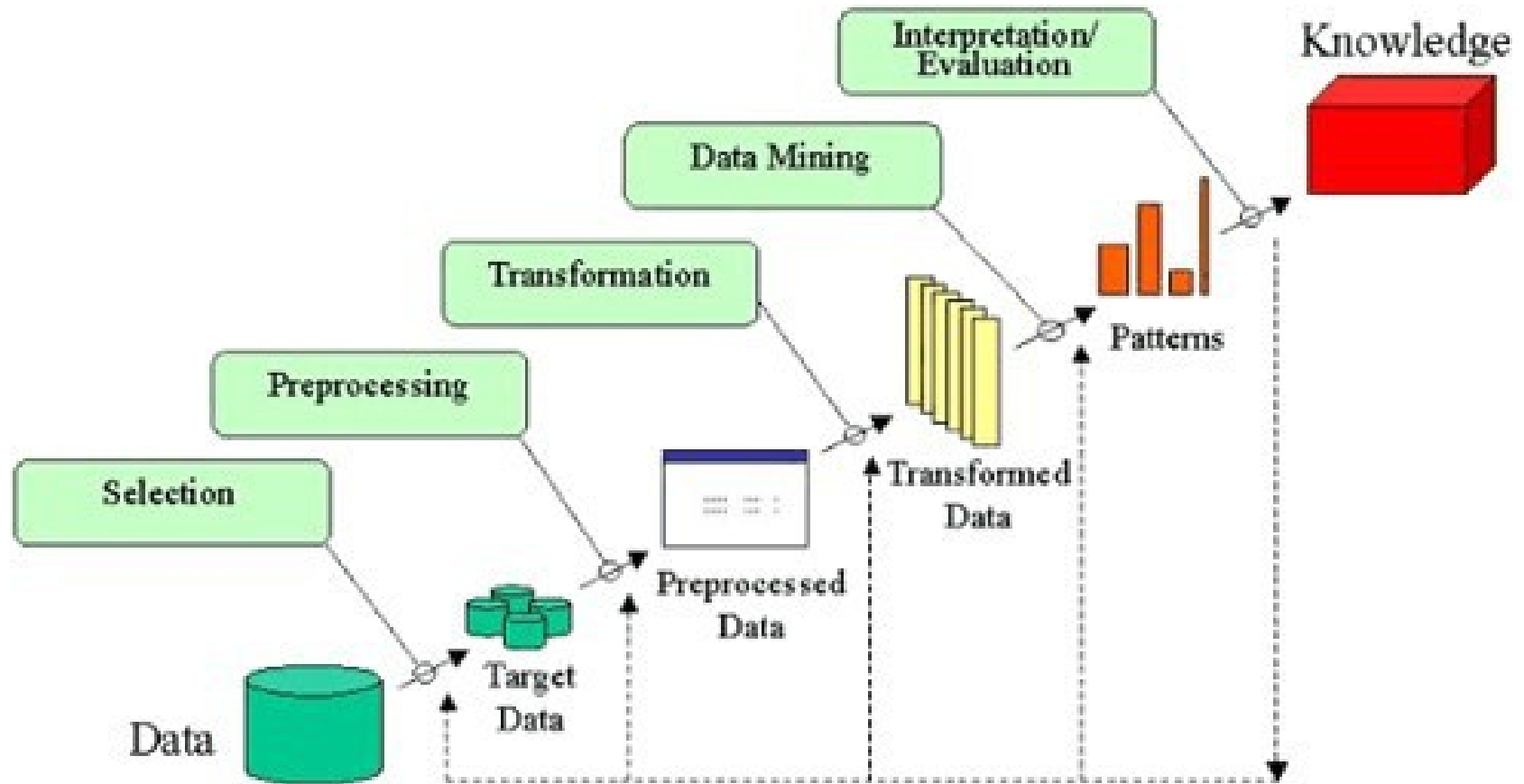


KDD

KDD

- **Es un método de 5 pasos que se utiliza para "Minería de Datos" (Data Mining), en donde lo que se busca patrones de interés o representativos.**

KDD



KDD (5 Pasos):

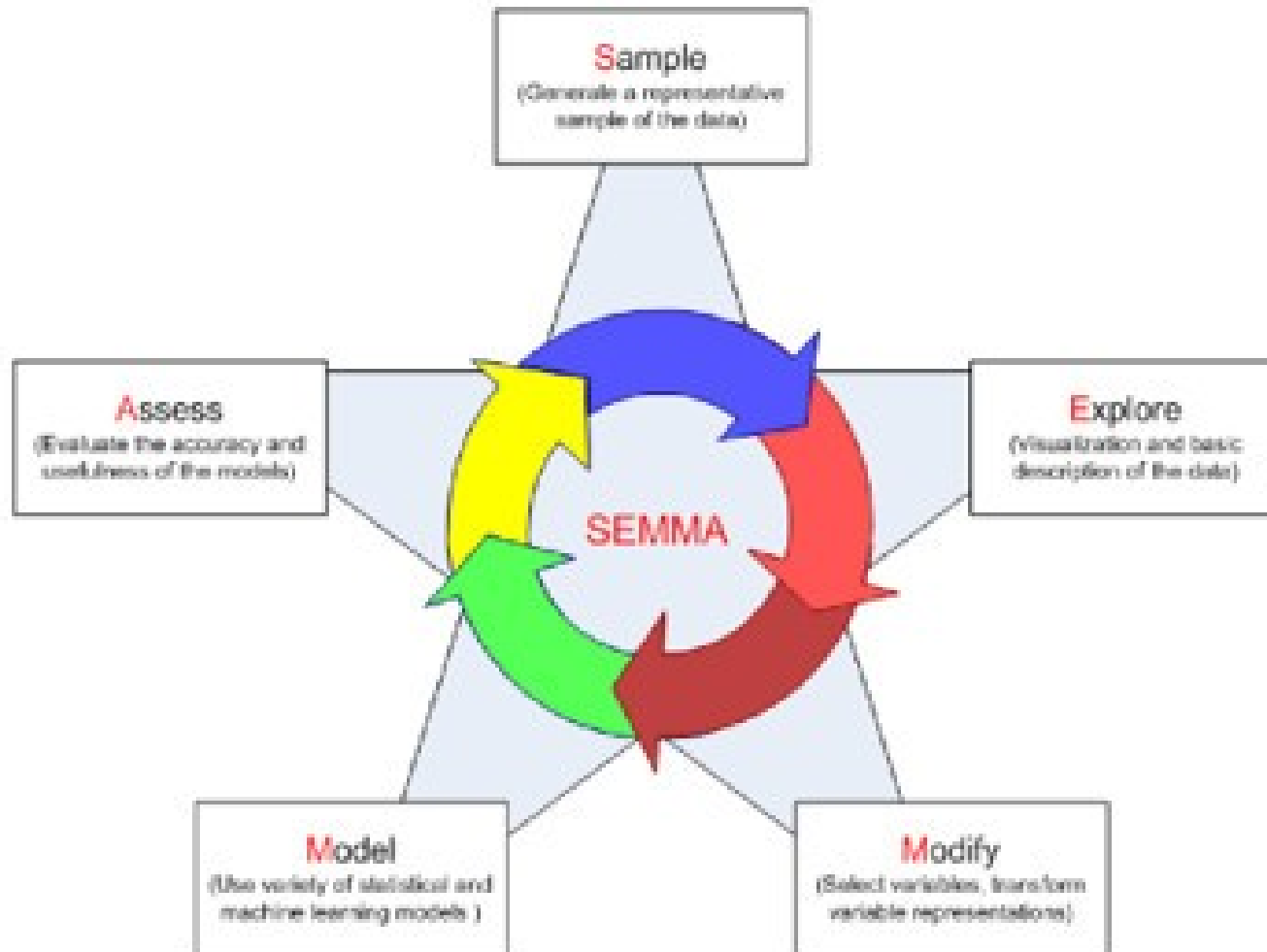
- **Selection:**
- **Preprocessing:**
- **Transformation:**
- **Data Mining:**
- **Interpretation/Evaluation:**

SEMMA

SEMMA

- **SEMMA** es el acrónimo de "Sample, Explore, Modify, Model, and Assess" (por sus siglas en inglés), y es un modelo que se compone de una lista de pasos secuenciales.

SEMMA



SEMMA (5 etapas):

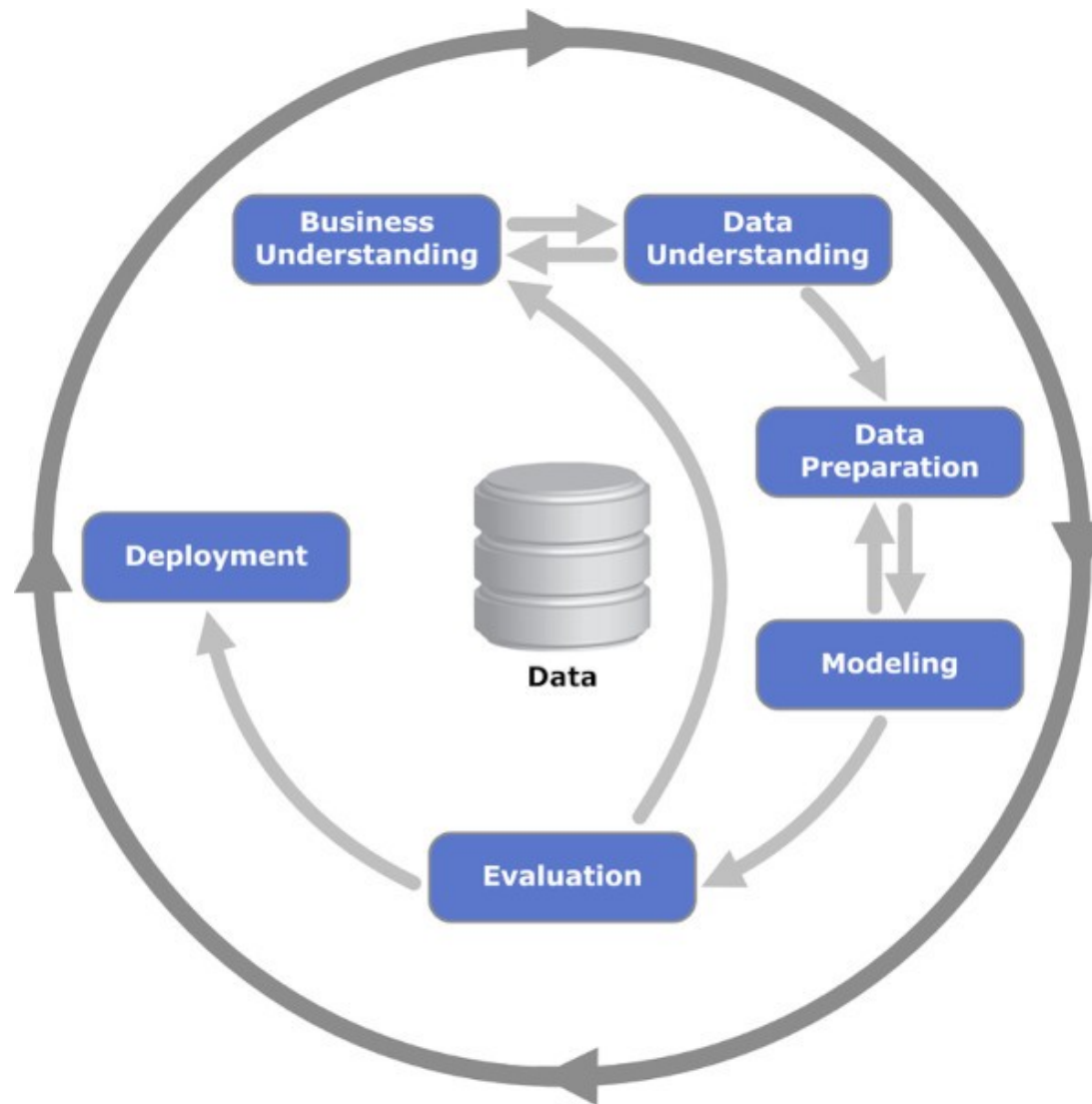
- **Sample:** Se toma una muestra.
- **Explore:** Entender los datos y sus relaciones.
- **Modify:** Pre-Procesar los datos antes del modelado.
- **Model:** Aplicar varias técnicas para crear un modelo.
- **Asses:** Evaluar los resultados del modelo empleado.

CRISP-DM

CRISP-DM

- **Es un estándar abierto que se utiliza para Ciencia de Datos y es el mas usado actualmente.**
- **Inició en 1996 como un proyecto de la Unión Europea y la primera versión se presentó en 1999 en Bruselas Bélgica durante el "4th CRISP-DM SIG Workshop".**
- **IBM es la compañía que lo adoptó y lo usa como su estándar, principalmente para su producto "SPSS Modeler". Este impulso que IBM le dió, permitió que CRISP-MD se convirtiera en el estándar "De Facto" del área de Ciencia de Datos.**

CRISP-DM



CRISP-DM (6 Etapas):

- **1) Business Understanding**
- **2) Data Understanding**
- **3) Data Preparation**
- **4) Modeling**
- **5) Evaluation**
- **6) Deployment**

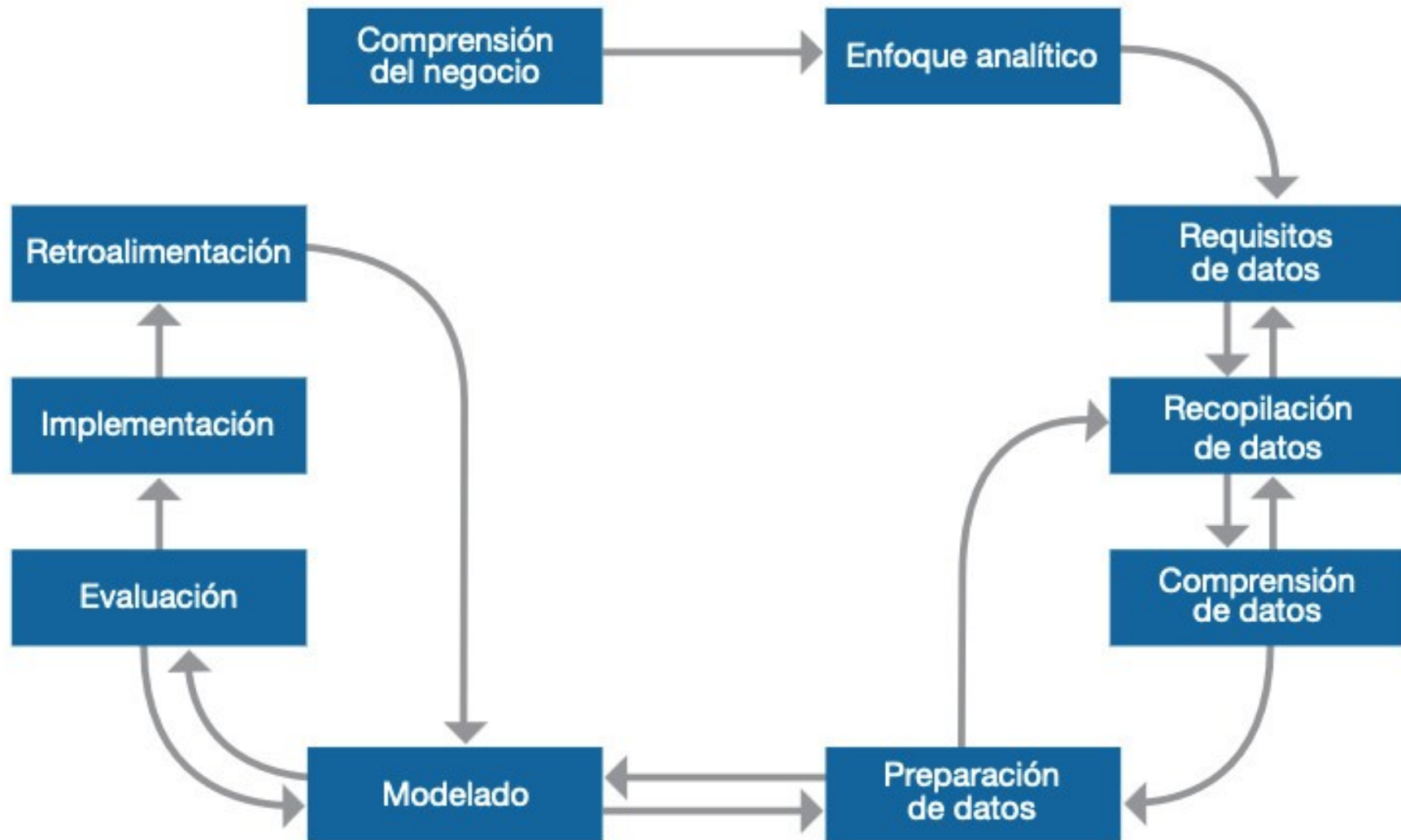
Comparación

KDD vs SEMMA vs CRISP-DM

KDD	SEMMA	CRISP-DM
---	---	Business Understanding
Selection	Sample	Data Understanding
Preprocessing	Explore	
Transformation	Modify	Data Preparation
Data Mining	Model	Modeling
Interpretation/Evaluation	Assess	Evaluation
---	---	Deployment

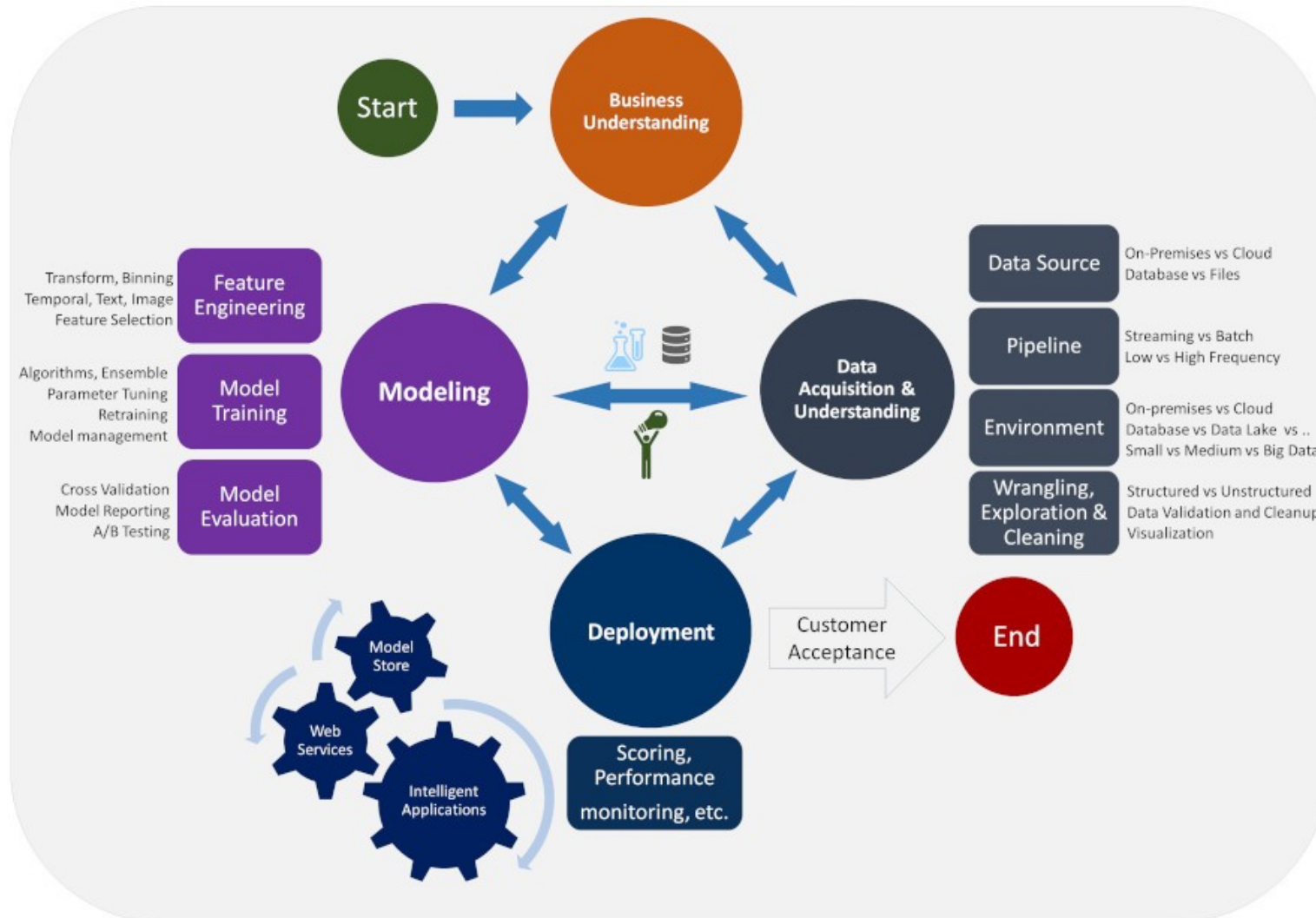
Otros Modelos

Modelo de IBM

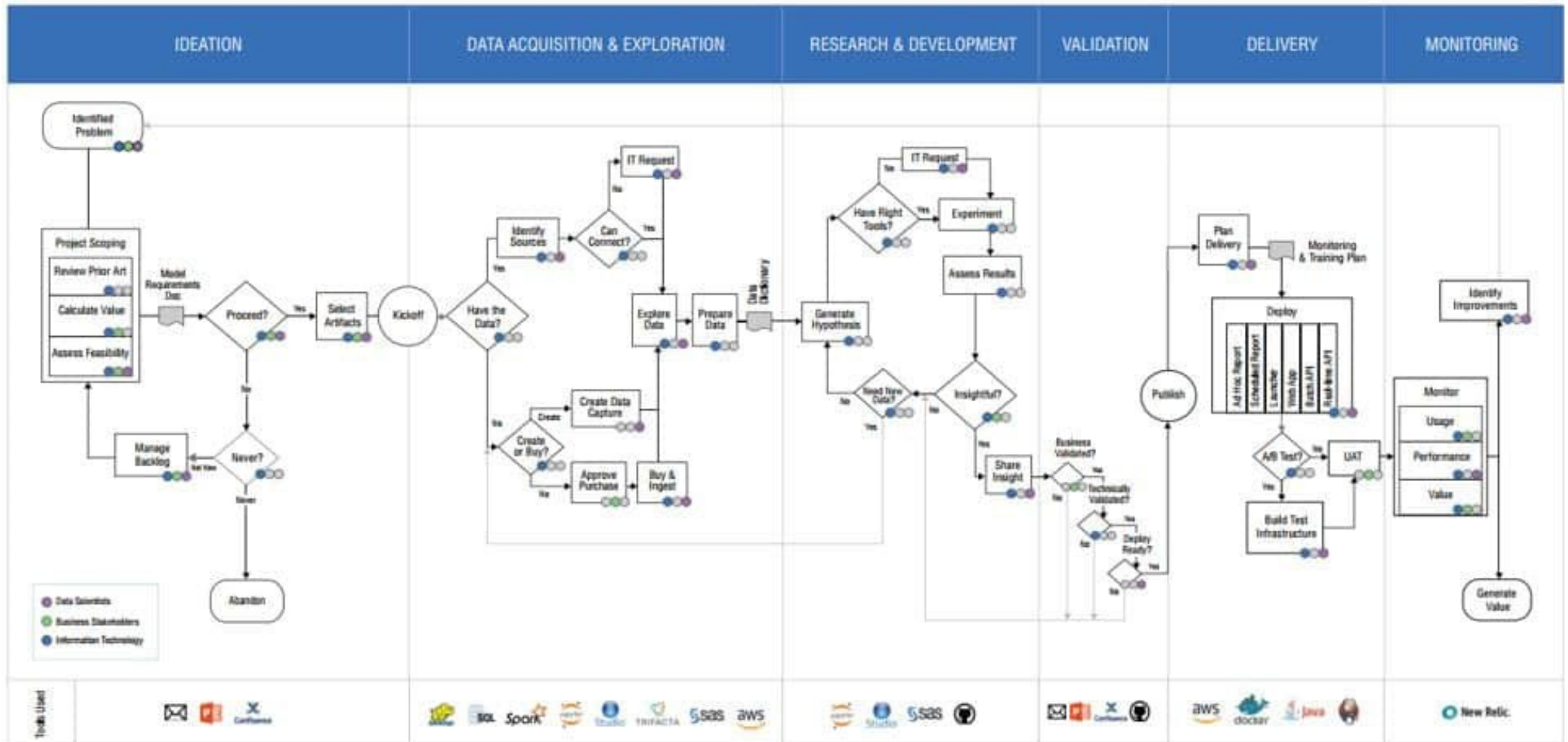


TDSP (Microsoft)

Data Science Lifecycle



Domino Data Science Life Cycle



OSEMN

Data Science **OSEMN** Model

Obtain

- from other location
- Query from database or API
- Extract from another file
- Generate data (e.g. Sensors)



Scrub

- Filtering lines
- Extracting columns or words
- Replacing values
- Handling missing values
- Converting formats



Explore

- Understanding data
- Deriving statistics
- Creating visualization



Model

- Clustering
- Classification
- Regression
- Dimensionality reduction



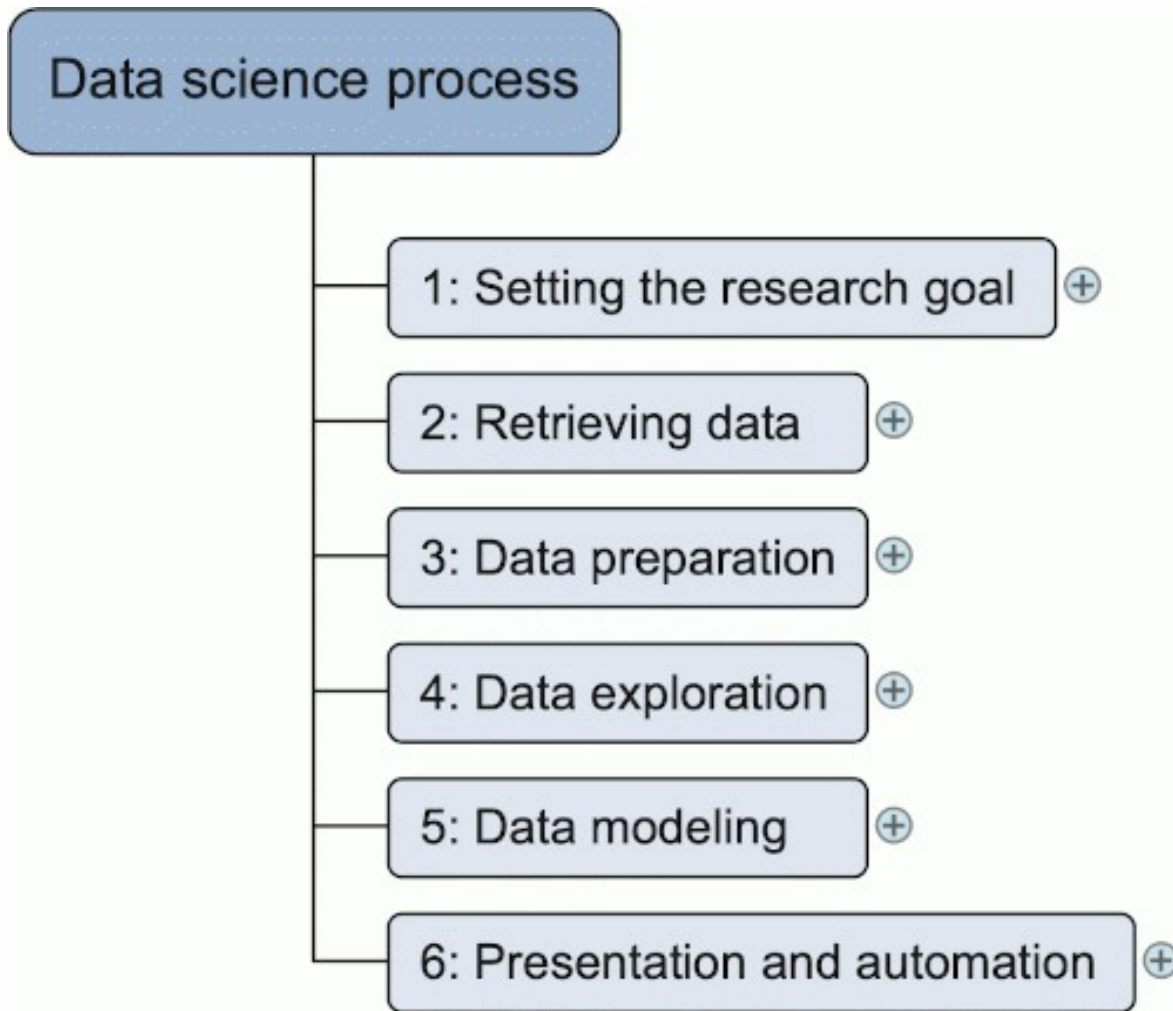
INterpret

- Drawing conclusion from data
- Evaluating meaning of results
- Communicating result



Etapas de un Proyecto de Ciencia de Datos

Pasos que sigue un Científico de Datos



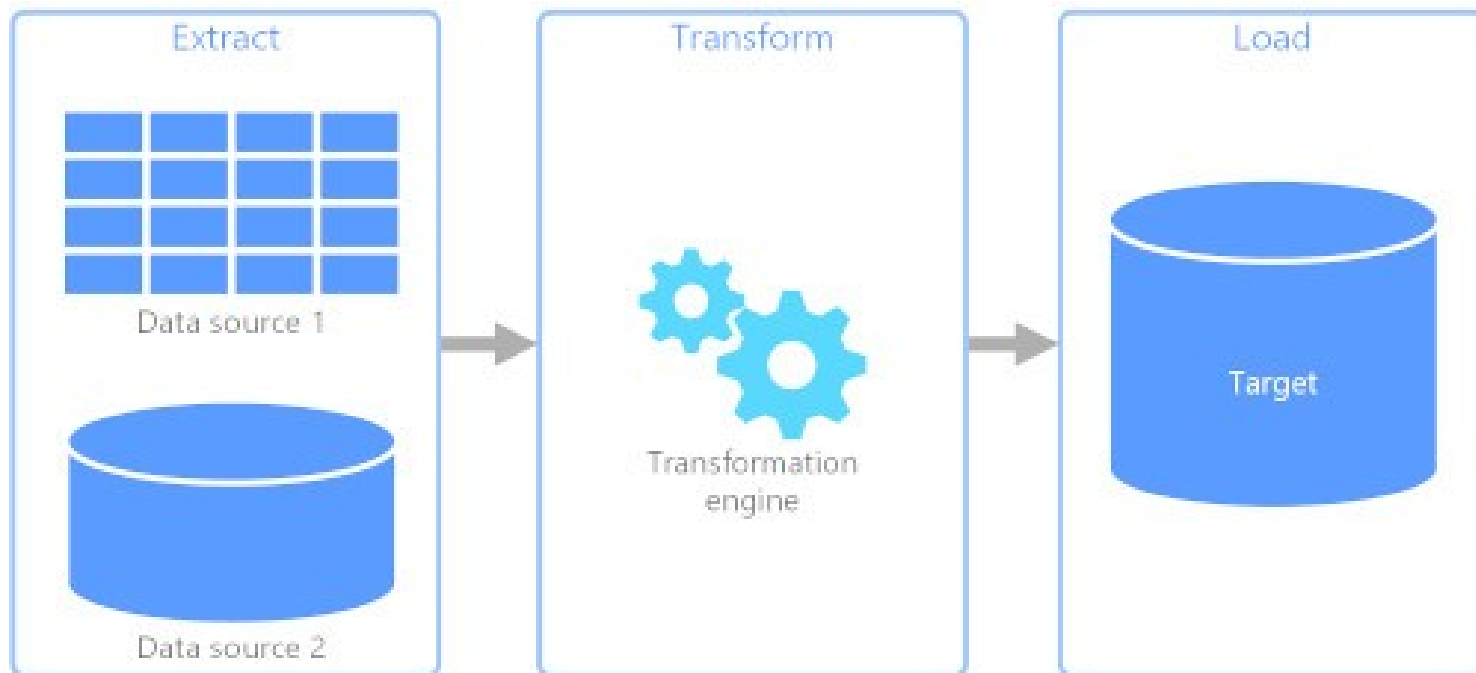
Paso 1) Definir objetivo

- **Especificar a donde se quiere llegar:**
 - **Predecir resultados.**
 - **Descubrir tendencias.**
 - **Encontrar patrones.**



Paso 2) Extracción de datos

- **Localizar la ubicación de los datos para poder leerlos y en su caso almacenarlos en nuestro sistema de base de datos.**



Paso 3) Preparar los datos

- **Antes de hacer algún proceso, se hace un preprocesamiento o limpiado de los datos, con el objetivo de poder tener solamente los datos requeridos y en el formato deseado.**



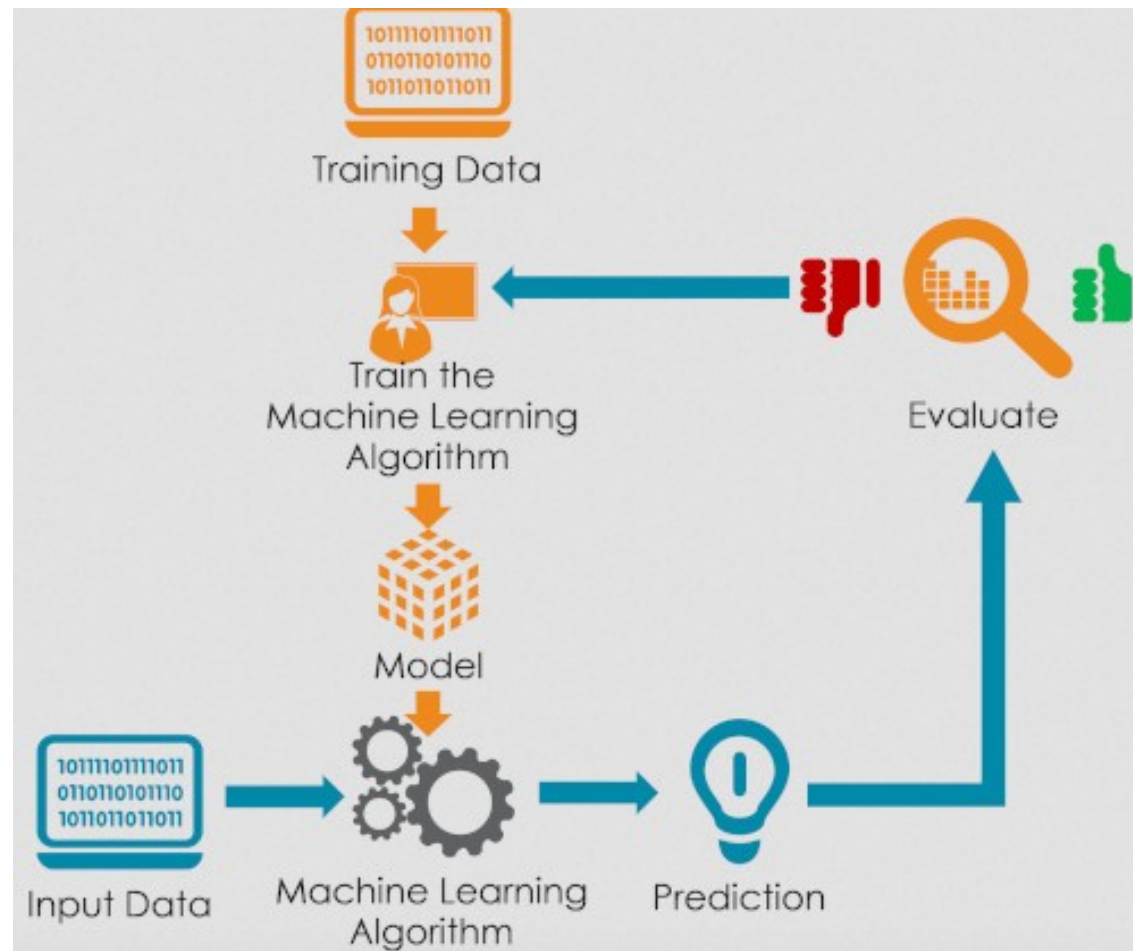
Paso 4) Exploración de los datos

- En este paso se analizan y procesan los datos



Paso 5) Modelado

- A partir de los datos se crea un modelo para lograr llegar a nuestro objetivo y encontrar resultados.



Paso 6) Visualización

- Finalmente hay que visualizar (graficar) los datos obtenidos para poder entender más fácilmente los resultados.



Metodología Propuesta (ITM)

Metodología Propuesta (ITM)

- **1) Objetivo**
- **2) Extracción**
- **3) Análisis**
- **4) Pre-Procesamiento**
- **5) Modelado**
- **6) Resultados**



Rogelio Ferreira Escutia

Profesor / Investigador
Tecnológico Nacional de México
Campus Morelia



rogelio.fe@morelia.tecnm.mx



rogeplus@gmail.com



xumarhu.net



[@rogeplus](https://twitter.com/rogeplus)



[https://www.youtube.com/
channel/UC0on88n3LwTKxJb8T09sGjg](https://www.youtube.com/channel/UC0on88n3LwTKxJb8T09sGjg)



[rogelioferreiraescutia](https://www.linkedin.com/in/rogelioferreiraescutia)

