

Minería de Datos



Minería de Datos



Datos



Conocimiento
Información
Conclusiones

Varias perspectivas sobre el proceso

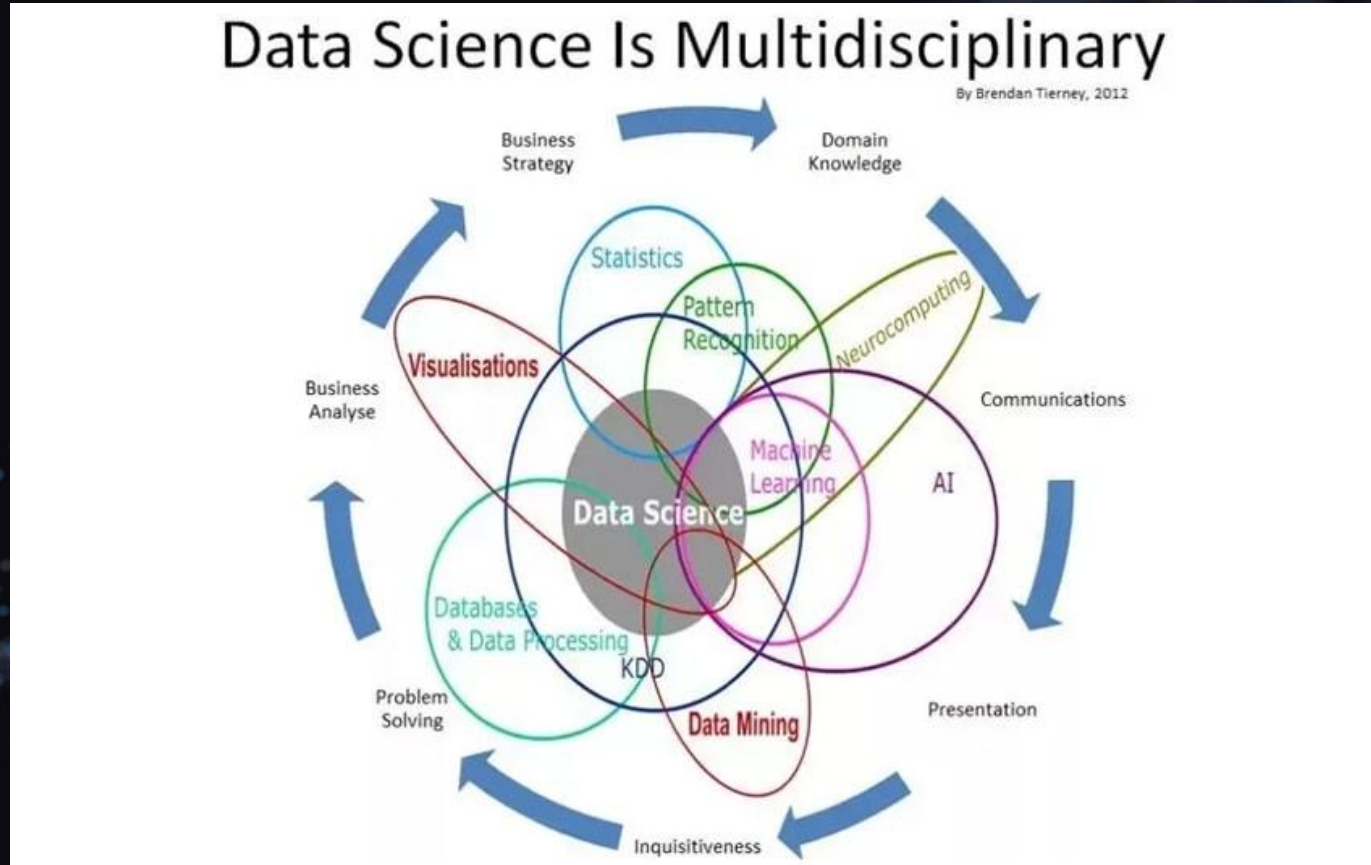
Minería de datos vs Ciencia de datos

- Marketing de nombres
- Diferencias mínimas
- Otros:
 - Data Engineering
 - Data Analytics
 - Information Retrieval



Minería de datos vs otros campos

- En la práctica, no hay diferencias claras
- Enfocarse en técnicas y no en campos

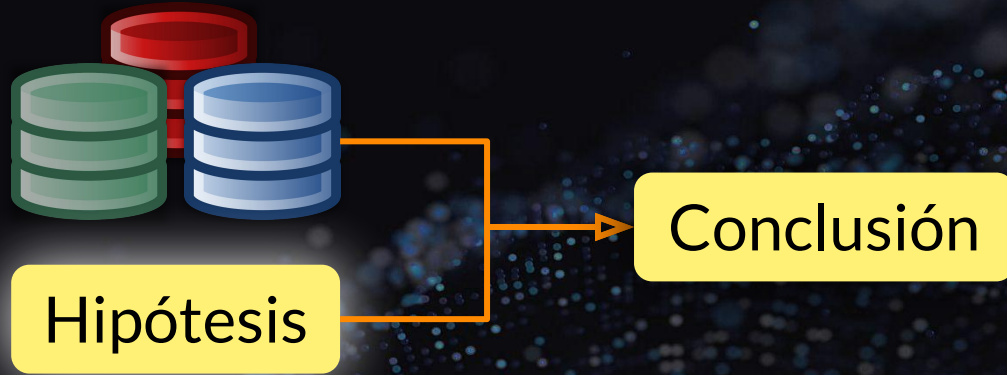


Minería de Datos vs Estadística

Minería de Datos	Estadística
Exploratoria	Confirmatoria
Inductiva	Deductiva
Muchas variables	Pocas variables
Muchos datos	Pocos datos
Datos numéricos y no numéricos	Datos numéricos
Pocas asunciones	Muchas asunciones
Pocas garantías	Muchas garantías

Enfoque Confirmativo vs Exploratorio

Confirmativo

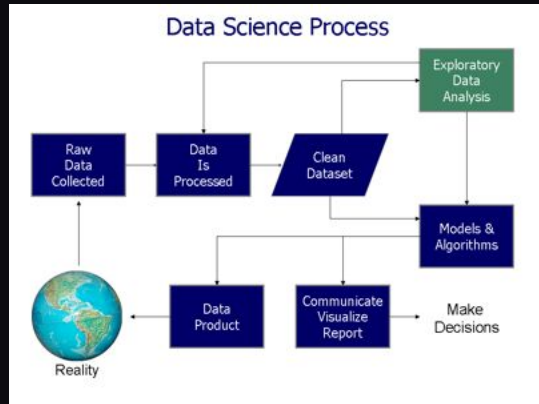


Exploratorio

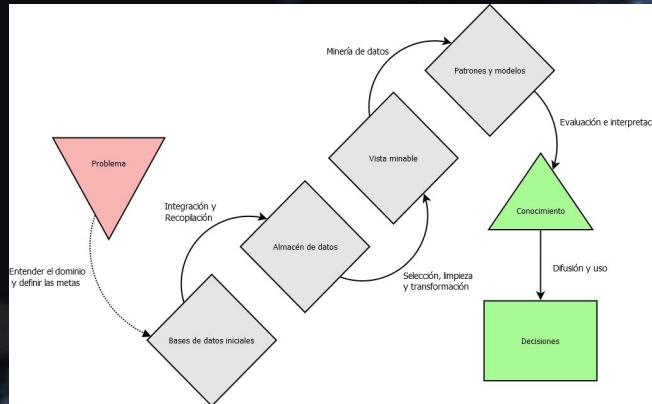


Metodologías en Minería de Datos

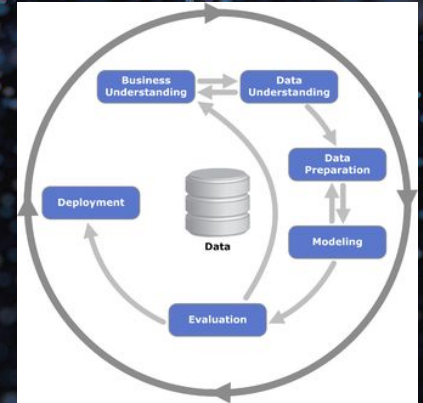
- Visiones complementarias



EDA: Exploratory Data Analysis (1977, John Tukey)
Estadística con énfasis en Datos



KDD: Knowledge Discovery in Databases (1989, KDDSIG)

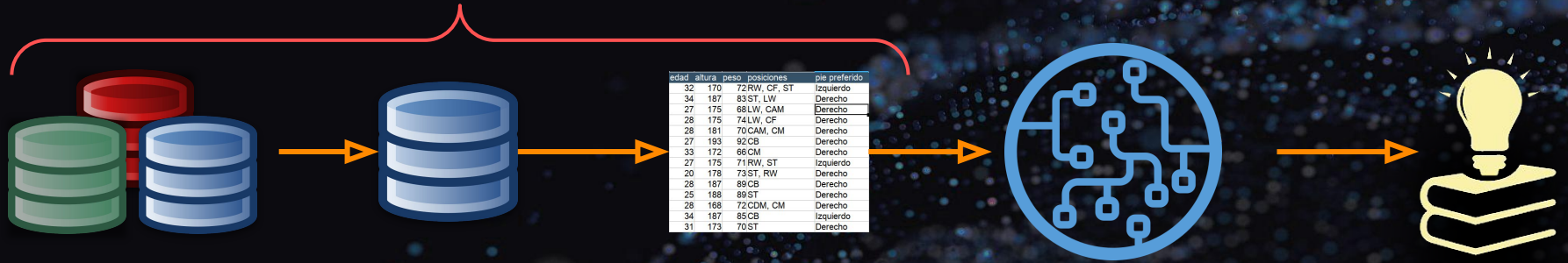


CRISP-DM: Cross Industry Standard Process for Data Mining (1999, EU)

Knowledge Discovery in Databases (KDD, 1989)

- Minería de datos = etapa de KDD

Datos: 80% del esfuerzo



Obtención

Selección y
Limpieza

Preprocesamiento
y transformación (Minería de datos)

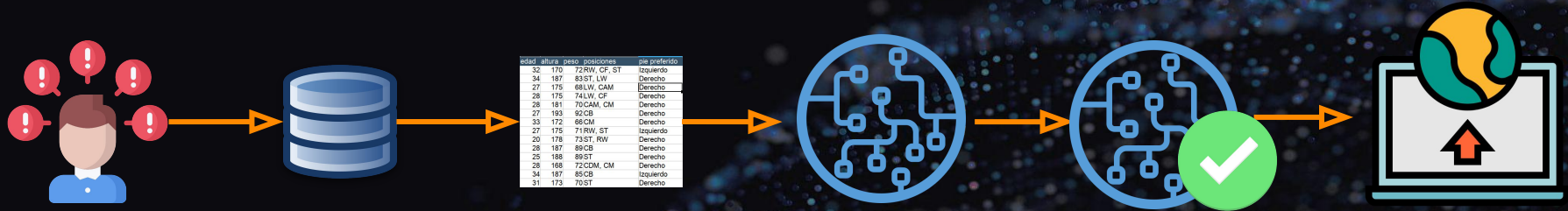
Modelado

Conocimiento

Proceso iterativo

CRISP-DM: Cross Industry Standard Process (1996)

- Minería de datos = todo el proceso



Proceso iterativo

Exploratory Data Analysis (1977, Tukey)

- Metodología sólo para fases de datos
- Aplicar en KDD o CRISP o estadística común



Definición del
Problema

Exploración
de datos

Preparación de
datos

Modelado

Evaluación

Despliegue



- Visualizar y transformar los datos sin hipótesis previas
- Inspiro a los lenguajes S y luego R (R for Data Science)