

# **Data mining and Predictive Models**

**ECEL Research Group**

**Kevin Rojas - Enero 2021**

# Cómo se divide el análisis?

Esto va a depender de que queremos hacer...

- Análisis Descriptivo.- Hechos, realidades, certezas - Utiliza gráficos, tablas informes, clasificación y segmentación
- Análisis Predictivo.- Pronósticos, Clasificaciones - Modelos de aprendizaje supervisado y NO supervisado
- Análisis Prescriptivo.- Qué podemos hacer para obtener el resultado que queremos ?

**“The fundamental purpose of data preparation is to manipulate and transform raw data so that the information content enfolded in the data set can be exposed, or made more easily accessible.”**

# Importancia del procesamiento de datos

# Datos -casi- nunca son puros

## Reglas poco útiles

- Datos Incompletos
- Datos con Ruido
- Datos Inconsistentes

# Volumen de datos

## Mejora eficiencia

- Selección relevante de datos: Eliminación de duplicados, anomalías
- Reducción de datos: Selección de características relevantes, muestreos, selección de instancias, discretización

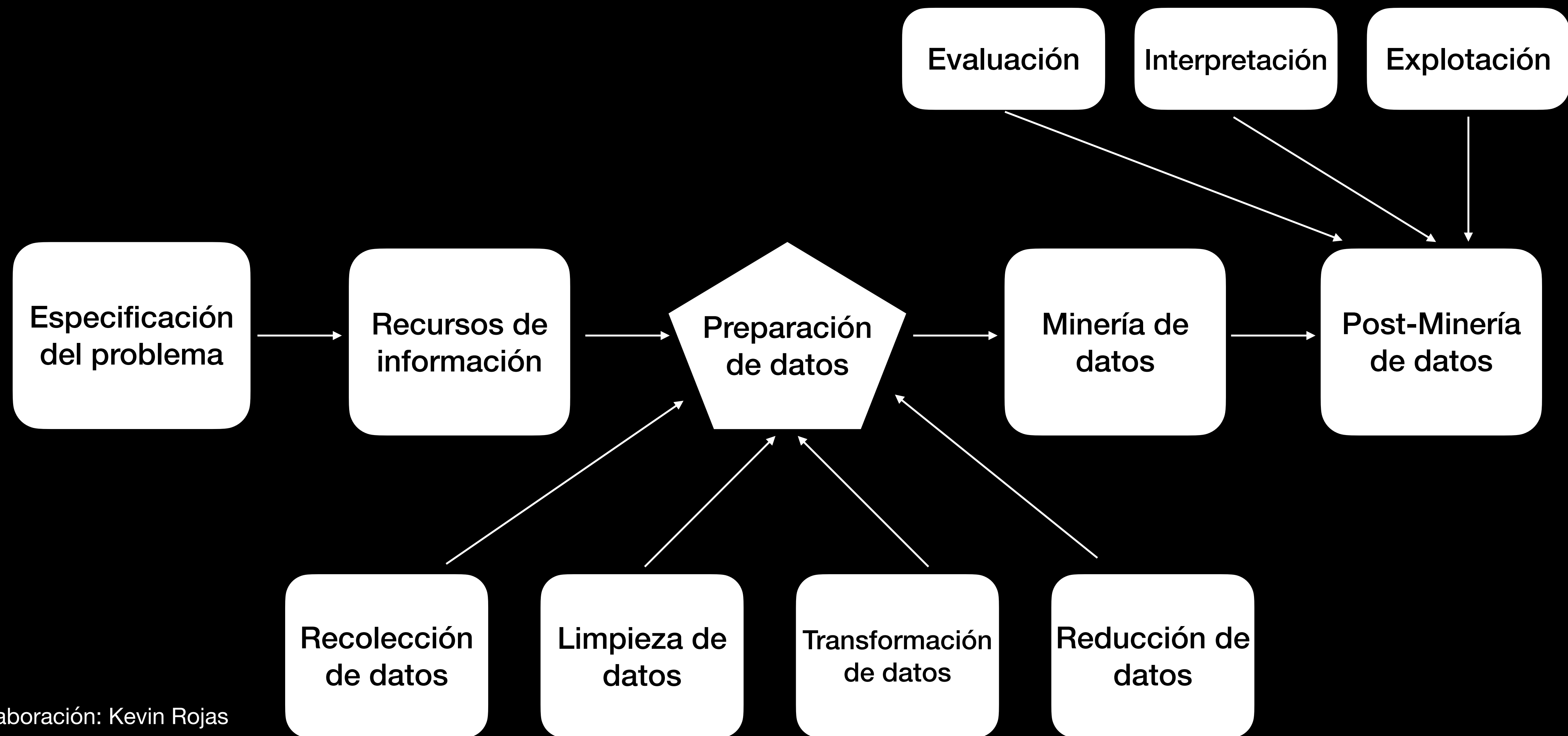
# Datos de calidad

## Reglas de calidad

- Recuperación de información incompleta
- Solución de conflictos
- Eliminación de datos atípicos

**"The Data Preprocessing" / "The Preparation of Data "encompasses all those data analysis techniques data that allows improving the quality of a set of data so that data extraction techniques knowledge / data mining can obtain greater and better information (better ranking percentage, rules with more completeness, etc.)**





Elaboración: Kevin Rojas

# Etapas de la preparación de datos

# Recolección de datos

También se considera la integración...

- Obtener datos de distintas fuentes de información
- Resolver problemas de representación y codificación
- Integración de datos de diferentes fuentes para la generación de información homogénea

# Limpieza de datos

**Importante para definir las reglas adecuadas**

- Resolver conflictos existentes entre los datos (problemas con bases)
- Eliminación de outliers y datos atípicos
- Resolución de casos de valores atípicos, ruidos, entre otros

# Transformación de datos

A formas que permitan el análisis

- Transformación y consolidación
- Resumen del conjunto de datos disponible
- Operaciones de agregación o desagregación
- Procesos de normalización

# Reducción de datos

## Mejora eficiencia de los algoritmos

- Elección de datos relevantes
- Consideramos procesos de selección de atributos y de instancias
- Procesos de discretización (especialmente en bases relacionales )