



Introducción a Minería de datos

Rafael Contreras Agudelo
Edwyn Guzmán
Yarima Contreras Blanco

6213 - Minería de datos
Facultad de ciencias
Universidad Central de Venezuela

13 de febrero de 2026

Contenido

- 1 Objetivo de la Minería de datos
- 2 Definición del problema
- 3 Taxonomía de técnicas de MD

"Notoriamente no hay clasificación del universo que no sea arbitraria y conjetural."^a

^aEl idioma analítico de John Wilkins - Jorge Luis Borges

Objetivo de la Minería de datos

Conceptos básicos

Definición Es el proceso de **descubrir conocimiento** o **patrones** a partir de **datos**. [3](*)

Conceptos básicos

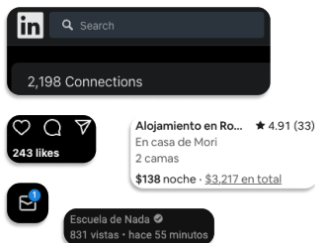
Definición Es el proceso de **descubrir conocimiento** o **patrones** a partir de **datos**. [3] (*)

Datos Abstracción representada estructuralmente de uno o varios hechos.

Conceptos básicos

Definición Es el proceso de **descubrir conocimiento** o **patrones** a partir de **datos**. [3](*)

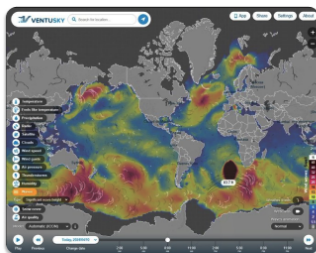
Datos Abstracción representada estructuradamente de uno o varios hechos.



Conceptos básicos

Definición Es el proceso de **descubrir conocimiento** o **patrones** a partir de **datos**. [3](*)

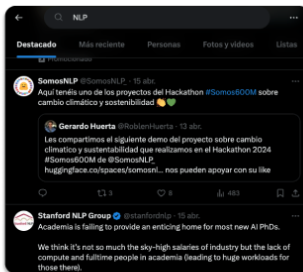
Datos Abstracción representada estructuralmente de uno o varios hechos.



Progress bar for loops

*

```
from tqdm import tqdm
for i in tqdm(range(10000)):
    pass
```



Definición del problema

Definición del problema

Definición Es el proceso de **descubrir conocimiento** o **patrones** a partir de **datos**. [3](*)

Datos Abstracción representada estructuralmente de uno o varios hechos.

Conocimiento Información implícita, desconocida y potencialmente útil observadas en datos. [1]

Ejemplos de datos

- ▶ # búsquedas de 'IA' en tiempo
- ▶ Tweets de muestra de interés
- ▶ Series vistas por usuarios
- ▶ datos_i

Ejemplos de conocimiento

- ▶ Interés académico en IA
- ▶ Opinión pública*
- ▶ "You may also like"
- ▶ conocimiento_{ij}

Definición del problema

Definición Es el proceso de **descubrir conocimiento** o **patrones** a partir de **datos**. [3](*)

Datos Abstracción representada estructuralmente de uno o varios hechos.

Conocimiento Información implícita, desconocida y potencialmente útil observadas en datos. [1]

Ejemplos de datos

- ▶ # búsquedas de 'IA' en tiempo
- ▶ Tweets de muestra de interés
- ▶ Series vistas por usuarios
- ▶ $datos_i$

Ejemplos de conocimiento

- ▶ Interés académico en IA
- ▶ Opinión pública*
- ▶ "You may also like"
- ▶ $conocimiento_{ij}$

Tipos de conjuntos de datos

- ▶ Datos multi-dimensionales
- ▶ Datos transaccionales
- ▶ Datos temporales
- ▶ Lenguaje natural
- ▶ Audio
- ▶ Imágenes
- ▶ Datos espaciales
- ▶ Datos orientados a gráficos

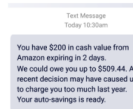
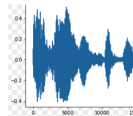
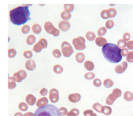
ID	$atributo_1$...	$atributo_i$
1	$val_{1,1}$...	$val_{1,i}$
2	$val_{2,1}$...	$val_{2,i}$

Tabla: Estructura ejemplo

$$type(atributo_i) \in [bool, enum, number, string, date]$$

Tipos de conjuntos de datos

- ▶ Datos multi-dimensionales
- ▶ Datos transaccionales
- ▶ Datos temporales
- ▶ **Lenguaje natural**
- ▶ **Audio**
- ▶ **Imágenes**
- ▶ Datos espaciales
- ▶ Datos orientados a gráficos



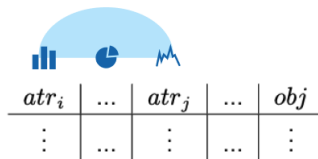
Preprocesamiento

Algoritmos
especializados

Taxonomía de técnicas

Tipos de conocimiento a extraer[2]

- ▶ Tareas descriptivas
 - ▶ Caracterización de datos
 - ▶ Discriminación de datos
- ▶ Tareas predictivas
 - ▶ Clasificación
 - ▶ Regresión
- ▶ Agrupación
- ▶ Detección de anomalías
- ▶ Asociaciones



Tipos de conocimiento a extraer

► Tareas descriptivas^a

- Caracterización de datos
- Discriminación de datos

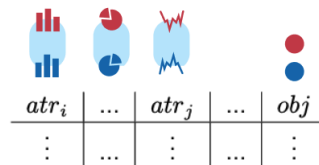
► Tareas predictivas

- Clasificación
- Regresión

► Agrupación

► Detección de anomalías

► Asociaciones



^a[kaggle.com/search?q=eda](https://www.kaggle.com/search?q=eda)

Tipos de conocimiento a extraer

- ▶ Tareas descriptivas
 - ▶ Caracterización de datos
 - ▶ Discriminación de datos
- ▶ Tareas predictivas
 - ▶ Clasificación
 - ▶ Regresión
- ▶ Agrupación
- ▶ Detección de anomalías
- ▶ Asociaciones

ID	...	$atributo_i$	$objetivo$
1	...	$val_{1,i}$	obj_1
2	...	$val_{2,i}$	obj_2

Tabla: Estructura ejemplo

$type(obj_i) = enum \rightarrow$ clasificación

$type(obj_i) = number \rightarrow$ regresión

Tipos de conocimiento a extraer

- ▶ Tareas descriptivas
 - ▶ Caracterización de datos
 - ▶ Discriminación de datos
- ▶ Tareas predictivas
 - ▶ Clasificación
 - ▶ Regresión
- ▶ Agrupación
- ▶ Detección de anomalías
- ▶ Asociaciones

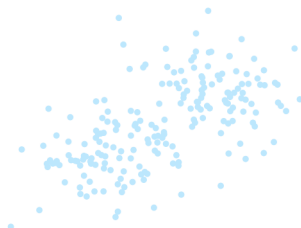
Ejemplo de aplicaciones

- ▶ Determinar la categoría de una página web
- ▶ Identificar el sentimiento predominante de un tweet
- ▶ Categorizar candidatos a créditos bancarios
- ▶ Estimar el costo de un inmueble
- ▶ Estimar el crecimiento de una compañía
- ▶ Estimar la popularidad de una canción en spotify^a

^ahf.co/datasets/maharshipandya/spotify-tracks-dataset

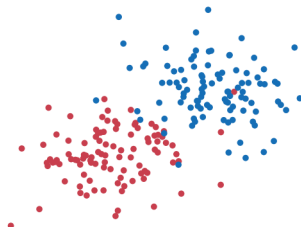
Tipos de conocimiento a extraer

- ▶ Tareas descriptivas
 - ▶ Caracterización de datos
 - ▶ Discriminación de datos
- ▶ Tareas predictivas
 - ▶ Clasificación
 - ▶ Regresión
- ▶ **Agrupación**
- ▶ Detección de anomalías
- ▶ Asociaciones



Tipos de conocimiento a extraer

- ▶ Tareas descriptivas
 - ▶ Caracterización de datos
 - ▶ Discriminación de datos
- ▶ Tareas predictivas
 - ▶ Clasificación
 - ▶ Regresión
- ▶ **Agrupación**
- ▶ Detección de anomalías
- ▶ Asociaciones



Tipos de conocimiento a extraer

- ▶ Tareas descriptivas
 - ▶ Caracterización de datos
 - ▶ Discriminación de datos
- ▶ Tareas predictivas
 - ▶ Clasificación
 - ▶ Regresión
- ▶ Agrupación
- ▶ Detección de anomalías
- ▶ Asociaciones

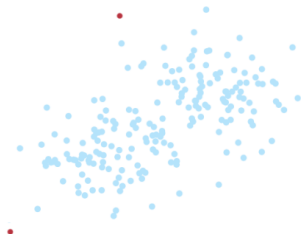
Ejemplo de aplicaciones

- ▶ Segmentación de clientes
- ▶ Agrupación de tweets por semejanza semántica^a
- ▶ Agrupar estudiantes con desempeños similares

^a[sbert.net](https://www.sbert.net)

Tipos de conocimiento a extraer

- ▶ Tareas descriptivas
 - ▶ Caracterización de datos
 - ▶ Discriminación de datos
- ▶ Tareas predictivas
 - ▶ Clasificación
 - ▶ Regresión
- ▶ Agrupación
- ▶ Detección de anomalías
- ▶ Asociaciones



Tipos de conocimiento a extraer

- ▶ Tareas descriptivas
 - ▶ Caracterización de datos
 - ▶ Discriminación de datos
- ▶ Tareas predictivas
 - ▶ Clasificación
 - ▶ Regresión
- ▶ Agrupación
- ▶ Detección de anomalías
- ▶ Asociaciones

Ejemplo de aplicaciones

- ▶ Detección de ataques a servidores
- ▶ Detección de fraudes bancarios
- ▶ Identificar las fallas de un servicio de telefonía móvil

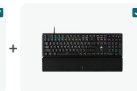
Tipos de conocimiento a extraer

- ▶ Tareas descriptivas
 - ▶ Caracterización de datos
 - ▶ Discriminación de datos
- ▶ Tareas predictivas
 - ▶ Clasificación
 - ▶ Regresión
- ▶ Agrupación
- ▶ Detección de anomalías
- ▶ Asociaciones

Comprados juntos habitualmente



Este producto: CORSAIR
HARPOON - Mouse inalámbrico
RGB para juegos, 10,000 DPI, 6...
US\$49⁹⁹



CORSAIR Teclado mecánico K70
CORE RGB para juegos con
reposamanos - Interruptores...
US\$104⁹⁹



CORSAIR MM700 RGB -
Alfombrilla de mouse para juegos
de tela extendida, 36.6 x 15.8...
US\$52⁹⁸

Tipos de conocimiento a extraer

- ▶ Tareas descriptivas
 - ▶ Caracterización de datos
 - ▶ Discriminación de datos
- ▶ Tareas predictivas
 - ▶ Clasificación
 - ▶ Regresión
- ▶ Agrupación
- ▶ Detección de anomalías
- ▶ Asociaciones

Ejemplo de aplicaciones

- ▶ *"También te puede gustar"*
- ▶ Detección de ocurrencias probables/improbables

Resumen

Minería de datos

Descubrir conocimiento y/o patrones a partir de datos.*

Referencias

- [1] William J. Frawley, Gregory Piatetsky-Shapiro y Christopher J. Matheus. "Knowledge Discovery in Databases: An Overview". En: (1992).
- [2] Jiawei Han, Micheline Kamber y Jian Pei. *Data mining concepts and techniques, third edition*. 2012.
- [3] Ling Liu y M. Tamer Özsu, eds. *Encyclopedia of Database Systems, Second Edition*. 2018.

¡Gracias!

dm-25.ucv.ai