

Minería de Datos

Prof. Dr. Jorge Zavaleta
zavaleta.jorge@gmail.com

Sobre – Jorge Zavaleta

- Postdoctorado en Análisis de datos médicos (UERJ)
- Postdoctorado en Métodos de machine Learning para fake-news (UFRJ)
- Doctor en Ingeniería de Sistemas y Computación (UFRJ)
- Maestría en Ciencias de la Computación (UFRGS)
- Licenciado em Matemáticas (UNT)

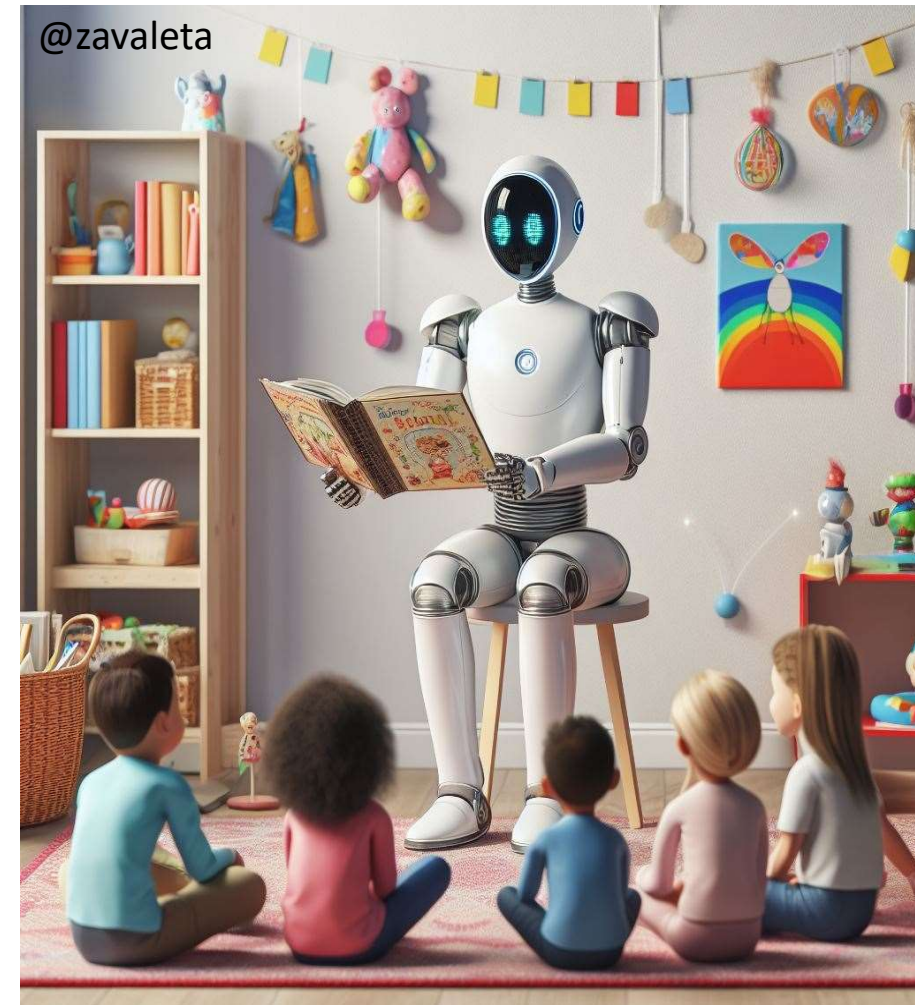


- Áreas de Interés e investigación
 - Inteligencia Artificial
 - Ciencia de Datos
 - *Machine Learning*
 - *Deep Learning*
 - Ciencia de Datos Aplicados a: Educación, Salud e Meteorología.
 - Lenguaje Java, R e Python.



Sumilla del curso

1. Introducción a los fundamentos de minería de datos,
2. Entendimiento y preparación de datos,
3. Modelos de clasificación
4. Árboles de decisión, modelos de regresión
5. Redes neuronales, modelos de agrupamiento y algoritmo, modelos de asociación
6. Algoritmo a priori, análisis de casos.



Introducción a Minería de Datos

- La **minería de datos**, también conocida como **KDD** (*Knowledge Discovery in Databases*), es un **proceso interdisciplinario** que busca ***extraer conocimiento útil a partir de grandes conjuntos de datos***.
- Este proceso implica la aplicación de técnicas **matemáticas**, **estadísticas** y de **aprendizaje automático** para **identificar patrones, tendencias y anomalías en los datos**.
- El **objetivo final** es convertir estos patrones en información útil que pueda ser utilizada para tomar mejores decisiones en diversos campos.

Definición de Minería de Datos

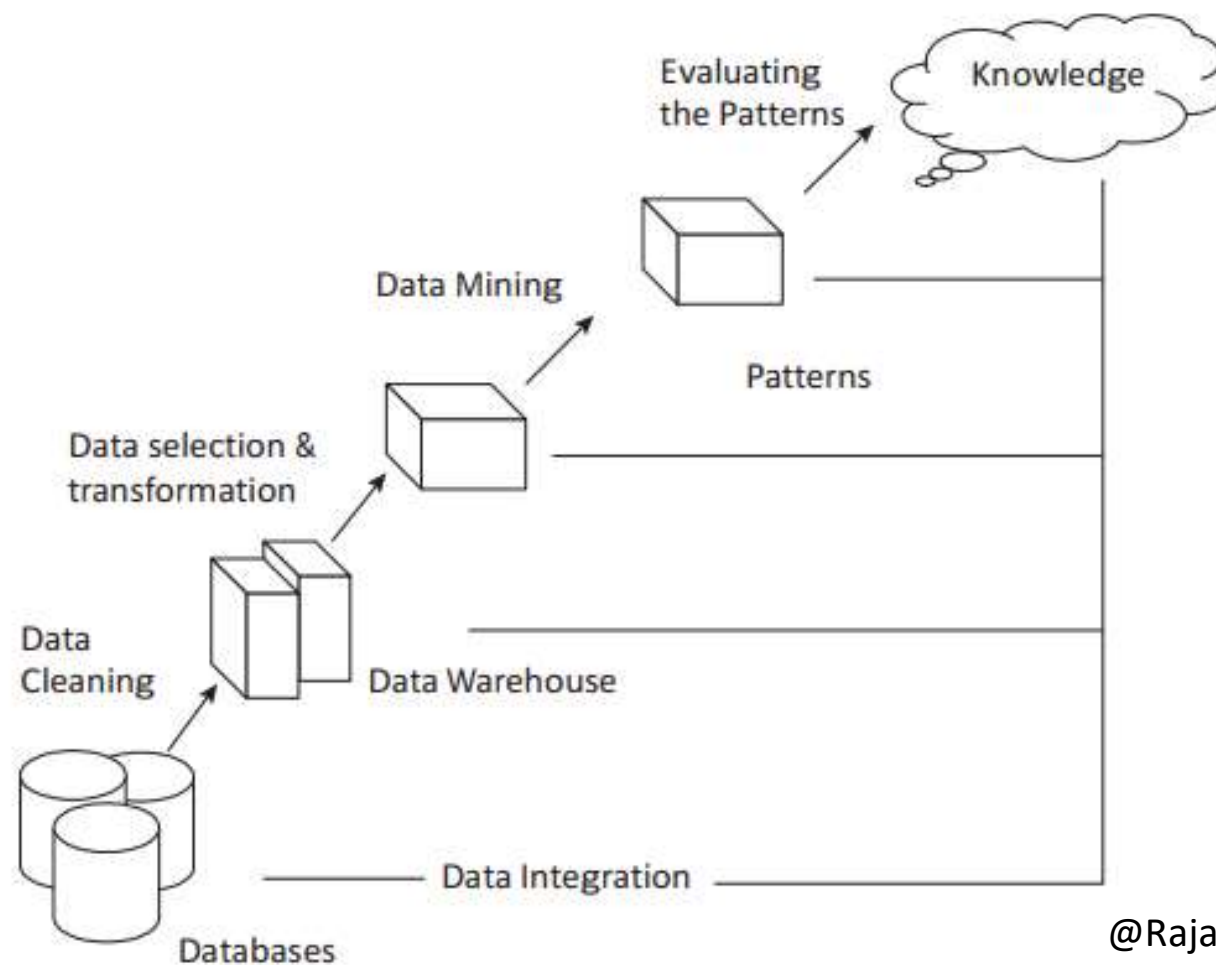
- **‘Minería’:** *extrae la información significativa de las bases de datos.*
- Este método ayuda a los investigadores, estudiantes y otros profesionales de TI a eliminar los detalles significativos exactos y desarrollar las aplicaciones deseadas.
- También se conoce como Descubrimiento de Conocimiento a partir de Bases de Datos (**KDD**).

Es el proceso de analizar grandes conjuntos de datos para descubrir patrones y tendencias que no son evidentes a simple vista.

Minería de datos

¿?

Ayuda a detectar los nuevos patrones de datos previamente desconocidos.



Limpieza de datos

- Este paso se puede definir como la **eliminación de datos irrelevantes**.
- Eliminar datos irrelevantes no es más que datos no deseados.
- Los registros se pueden eliminar.
- La recopilación de datos puede consistir en valores faltantes que deben eliminarse o deben imputar la información faltante.

Integración de datos

- Los datos se recopilan de fuentes heterogéneas y se integran en una fuente común como el almacén de datos (DW).
- Una técnica muy común, la Extracción-Transformación-Carga (ETL), es beneficiosa en este sentido.
- La integración de los datos de múltiples fuentes requiere una sincronización adecuada entre los sistemas

Selección y transformación de datos

- Una vez seleccionados los datos requeridos, la siguiente tarea es la transformación de datos.
- Como su nombre indica, no es más que transformarlo en el procedimiento de transformación deseado.

Evaluación de patrones

- La evaluación se basa en algunas medidas.
- Una vez que se aplican estas medidas, los resultados recuperados se comparan/evalúan estrictamente en función de los patrones almacenados.

Representación del conocimiento

- No es otra cosa que representar los datos procesados en los formatos requeridos, como tablas e informes.
- Se puede decir que la representación del conocimiento genera las **reglas**, y es posible utilizar la visualización exacta.

Importancia de la Minería de Datos

- Útil en análisis predictivo.
- Almacenan y gestionan datos en sistemas multidimensionales .
- Están identificando los patrones ocultos.
- Representación del conocimiento en los formatos deseados, etc.

Aplicaciones de Minería de Datos

- Detección de fraudes:
 - La minería de datos identifica patrones, i.e., patrones específicos del usuario, y construye un modelo basado en estados válidos e inválidos. Utilizando técnicas de gestión de datos, se pueden clasificar los registros en función de patrones fraudulentos y no fraudulentos.
- Análisis de Marketing
 - Se basa en la minería asociada, i.e., identificar las preferencias del usuario. Con estas técnicas, se pueden identificar los hábitos de compra de los usuarios. Con esta técnica, se pueden comparar diferentes artículos, precios de los artículos, etc.

Aplicaciones de Minería de Datos ... cont.

- Gestión de las relaciones con los clientes
 - Todas las organizaciones observan y mantienen atentamente este segmento que se conoce popularmente como CRM. En este segmento, se puede distinguir a los usuarios/clientes en función de la lealtad hacia la organización. Los datos del usuario/cliente pueden recopilarse y analizarse para obtener los resultados deseados.
- Banca y Finanzas
 - El sector bancario y financiero dispone de numerosos datos relacionados con los clientes. Los sistemas de software bancario y financiero ayudan a los diferentes gestores a identificar el segmento de clientes correctos, es decir, los clientes.

Aplicaciones de Minería de Datos ... cont.

- Banca y Finanzas ... cont.
 - Estos sistemas de software procesan y transaccionan que una persona no puede manejar manualmente. Estos almacenes de sistemas de software procesan un gran volumen de datos y producen los resultados deseados en menos tiempo.
- Industrias de la salud
 - Todo el mundo se preocupa por la salud. Diferentes parámetros y valores ayudan a los profesionales de la salud a diagnosticar la enfermedad. El número de pacientes, enfermedades y síntomas se puede procesar para obtener una predicción precisa.

Aplicaciones de Minería de Datos ... cont.

- Industrias de la salud ... cont.
 - Los sistemas de software utilizados en la industria de la atención médica procesan una gran parte de los valores almacenados y los comparan con los patrones almacenados para obtener una conclusión de la calidad
- Finalidad educativa
 - A partir de la recopilación de datos, se pueden identificar los intereses de los estudiantes en diferentes campos. También ayuda a mejorar la metodología de enseñanza con las nuevas tendencias.

Aplicaciones de Minería de Datos ... cont.

- Investigación de delitos

- La minería de datos ayuda a identificar diferentes patrones aplicados en otros delitos.
- Los crímenes, los delincuentes y sus características criminales se clasifican en esta categoría.
- Se puede procesar un gran volumen de (datos almacenados) para identificar diferentes relaciones con los delincuentes.
- En esta categoría, el reconocimiento facial, el reconocimiento de huellas dactilares, etc., se consideran y se utilizan en la investigación

iii Trabajos de casa!!!

- Instalar/actualizar los ambientes de programación

Bibliografía

1. Raja R. et al. Data Mining and Machine Learning Applications. John Wiley & Sons, Inc. 2022. ISBN 978-1-119-79178-2.
2. Kantardzic, M. Data Mining: Concepts, Models, Methods, and Algorithms. 3ed. John Wiley & Sons, Inc. 2020. ISBN: 9781119516040
3. Fortino, A. Data Mining and Predictive Analytics for Business Decisions: A Case Study Approach. Mercury Learning and information LLC, 2023. ISBN: 978-1-68392675-7.
4. Spector, AZ et al. Data Science in context: Foundations, Challenges, Opportunities. Cambridge University Press, 2023. SBN 978-1-009-27220-9.
5. Tyagi, AK. Data Science and Data Analytics: Opportunities and challenges. CRC Press, 2022.
6. Mariani, MC. et al. Data Science in theory and practice: Techniques for big data analytics and complex data sets. John Wiley and Sons, Inc., 2022.
7. Wes McKinny, Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython, 2nd edition O'Reilly Media, 2017.