

Data Mining – KDD

Yael Emiliano Salinas Lozano
Guillermo Hernández Sosa
Cristopher Ditter Gochicoa Ángeles
Sebastián Castillo Femenías

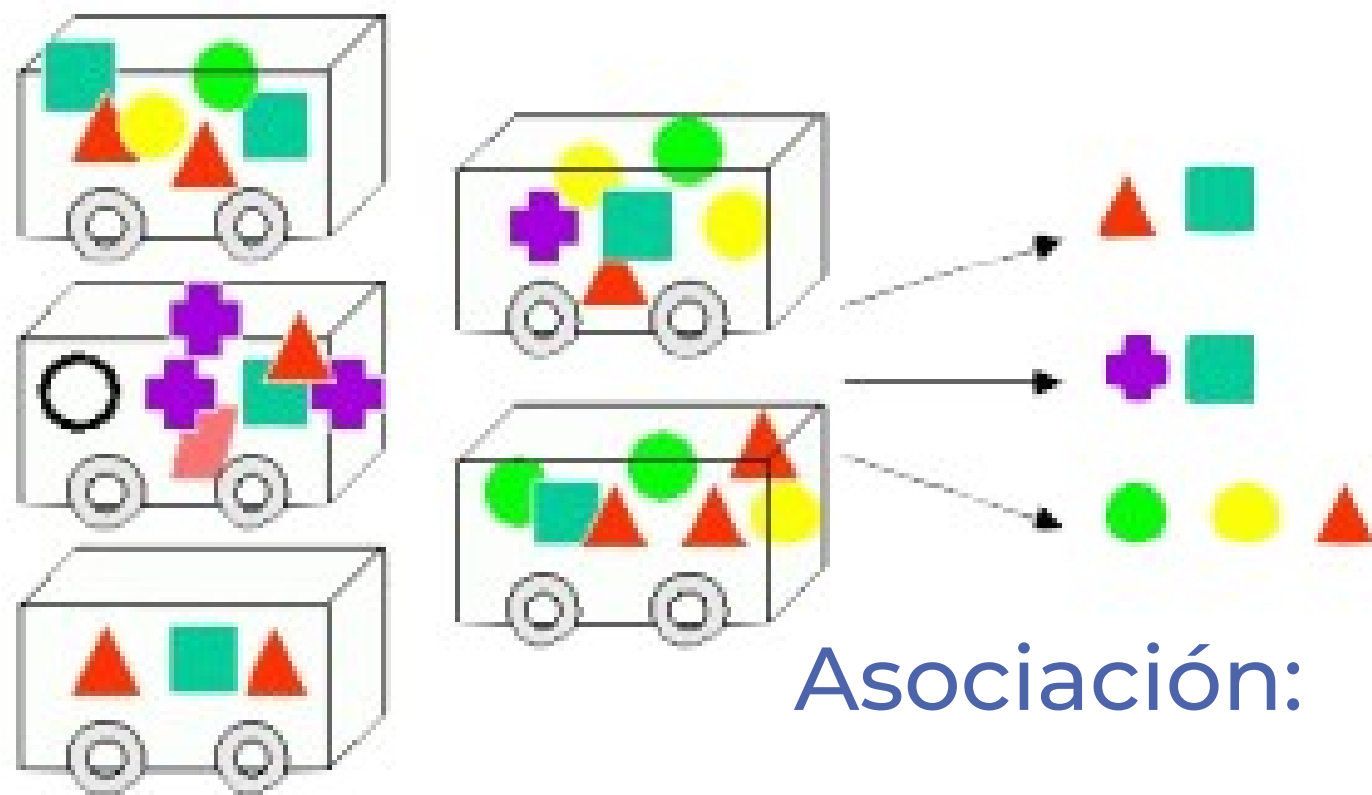


Data Mining

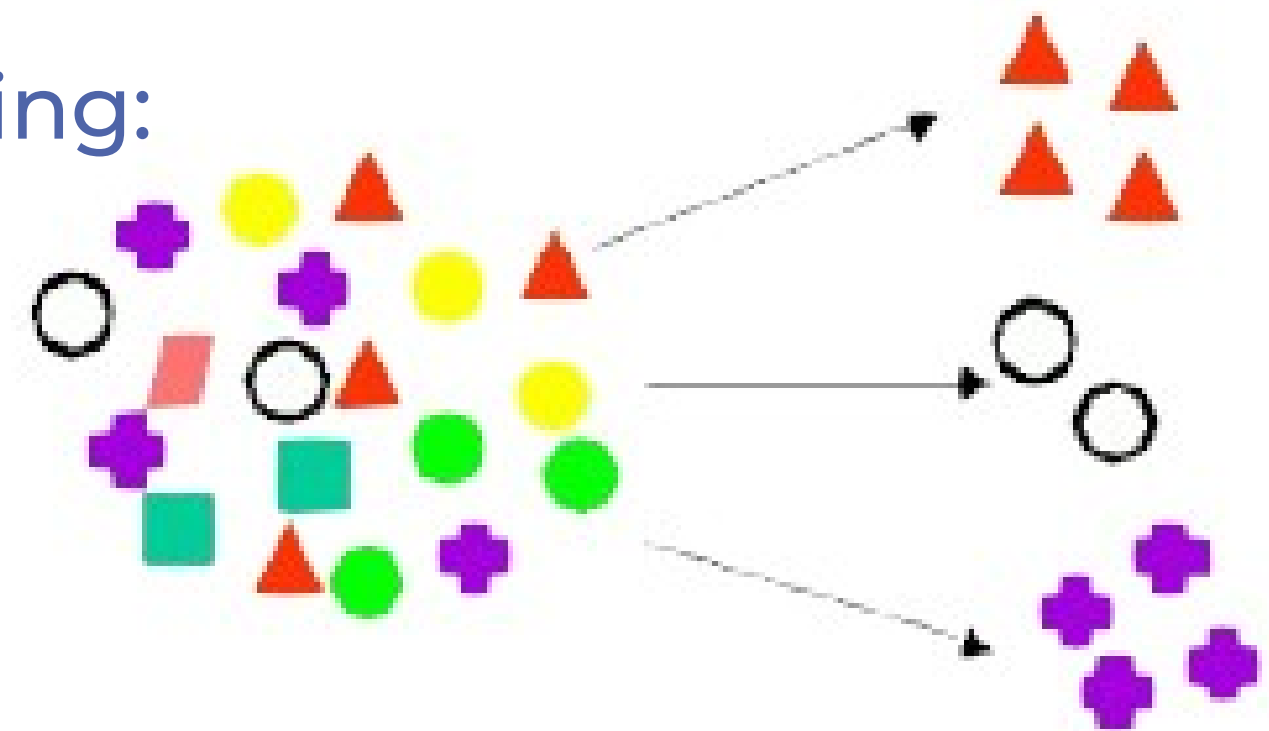
- ✓ Refiere al conjunto de técnicas y procedimientos para encontrar patrones en grandes volúmenes de datos. Esto con el fin de definir su comportamiento.
- ✓ Estos patrones son hallados a través de el uso de estadísticas o algoritmos de búsqueda próximos a la Inteligencia Artificial
- ✓ Esta información proporciona a las empresas una forma de monitorear ciertas tendencias de sus clientes, y así generar nuevas oportunidades en el mercado.

Técnicas en el DM

- ✓ **Asociación:** se usa una transacción y relación entre los elementos para identificar un patrón. Se utiliza, para conocer todos aquellos productos que los clientes compran juntos habitualmente, por ejemplo.
- ✓ **Agrupación o clustering:** Crea agrupaciones de objetos significativos que comparten las mismas características. A diferencia de una clasificación común, los clusters colocan los objetos en clases definidas por nosotros, y no predefinidas



Clustering:

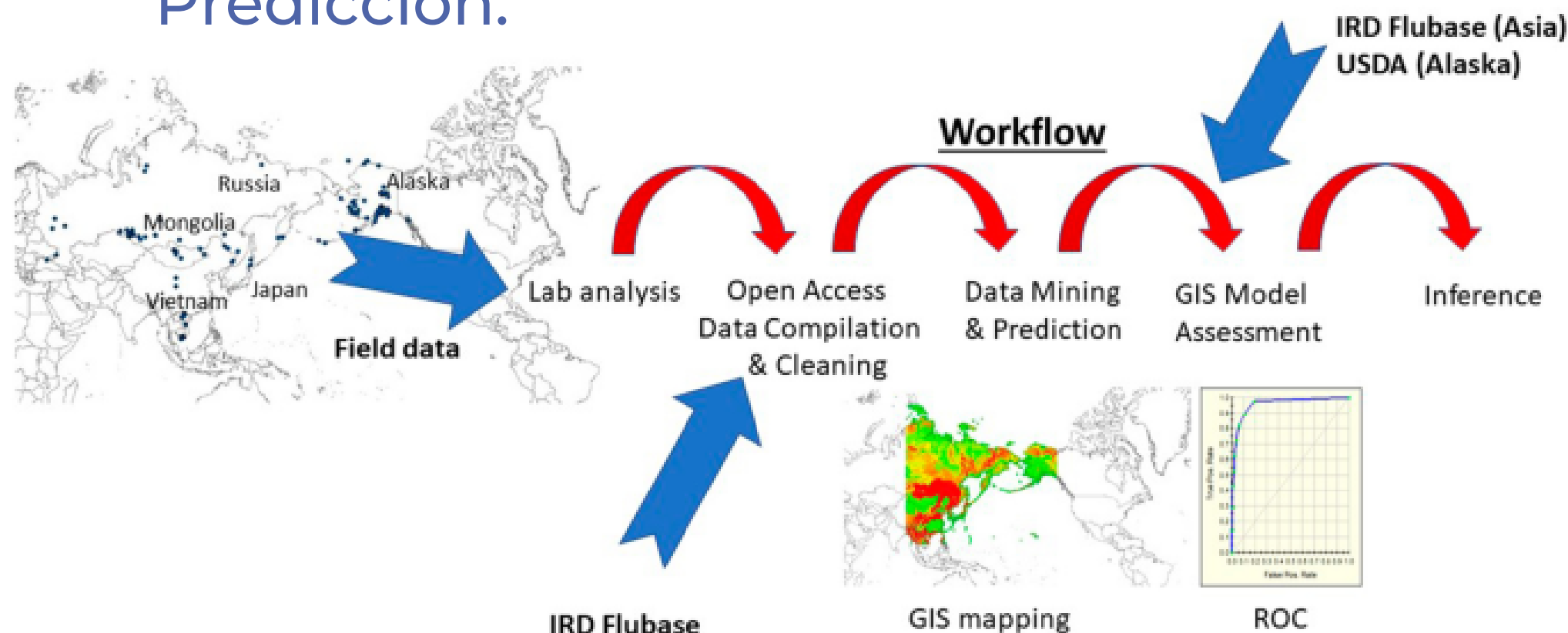


Técnicas en el DM

✓ **Predicción:** predice la relación que existe entre las variables independientes y dependientes, así como las variables independientes por sí solas. Puede usarse para predecir ganancias futuras de una venta.

✓ **Patrones secuenciales:** tiene como objetivo utilizar datos de transacciones y luego identificar tendencias, patrones y eventos similares en ellos durante un período de tiempo.

Predicción:



Patrones secuenciales:

SID	Sequence
1	$\langle \{a, b\}, \{c\}, \{f, g\}, \{g\}, \{e\} \rangle$
2	$\langle \{a, d\}, \{c\}, \{b\}, \{a, b, e, f\} \rangle$
3	$\langle \{a\}, \{b\}, \{f, g\}, \{e\} \rangle$
4	$\langle \{b\}, \{f, g\} \rangle$

Pasos para el análisis en el DM



1.



Investigación Comercial

2.



Análisis de Calidad

3.



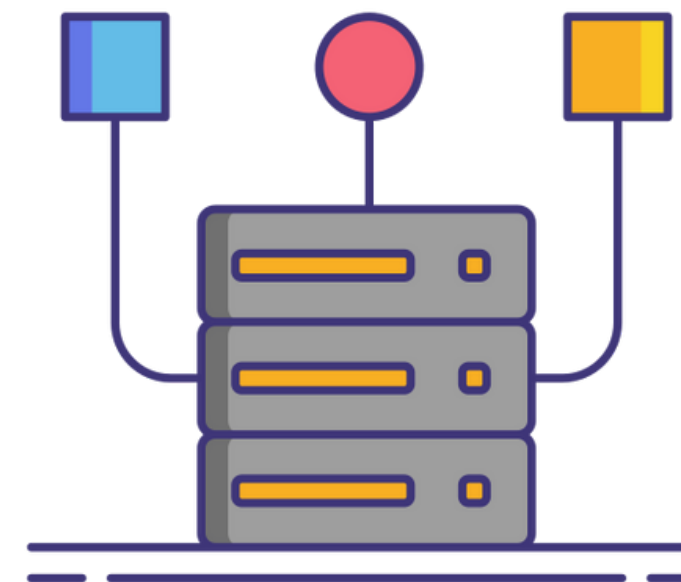
Limpieza de Datos

4.



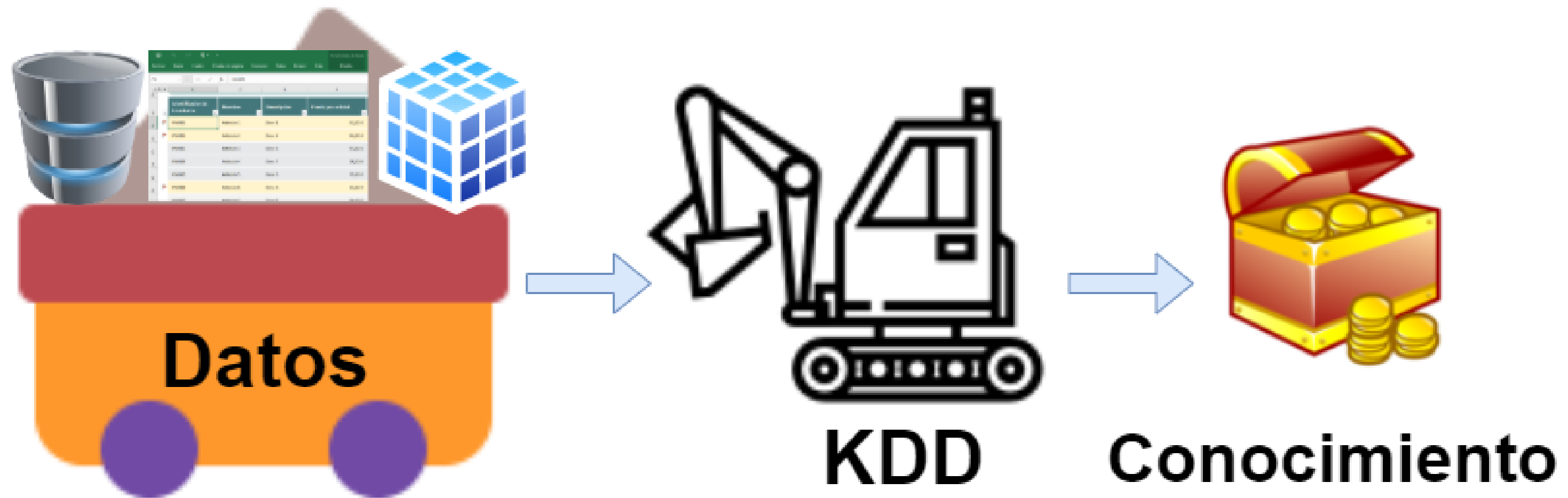
Transformación de Datos

5.



Modelado de Datos

KDD (Knowledge Discovery in Databases)





Metodología KDD

01 Identificación del problema.

02 Selección de datos

03 Limpieza de datos.

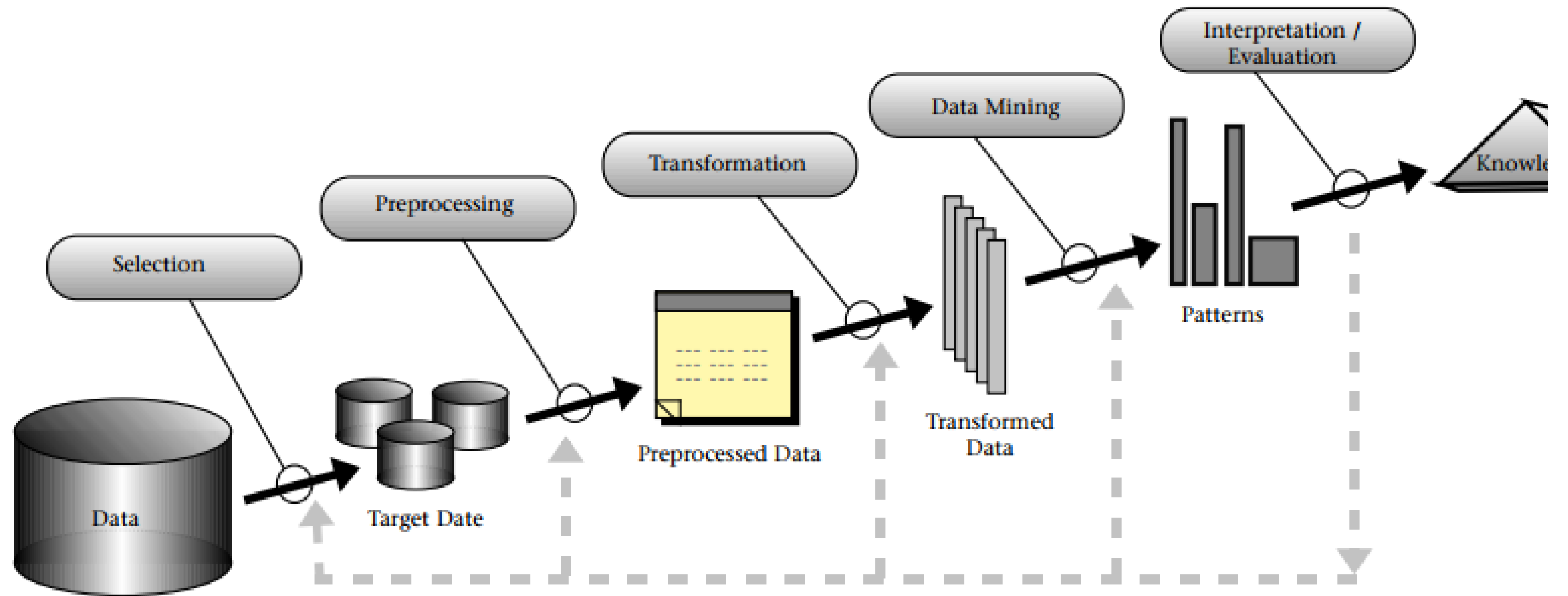
04 Integración de datos

05 Transformación de datos

06 Minado de datos

07 Evaluación

08 Descubrimiento de nuevo conocimiento



Identificar Problema

Se trata de entender la problemática a la que nos vamos a enfrentar y tener contexto para proponer soluciones viables y poder definir nuestras metas a alcanzar.



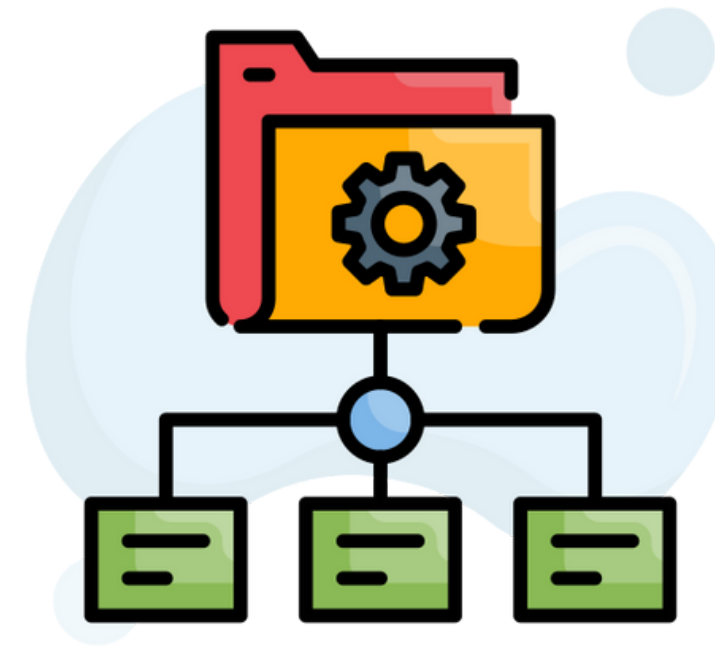
Selección

Se decide que tipo de informacion encontramos util para nuestra investigacion y la que no encontremos util se queda fuera del warehouse.



Limpieza

En esta parte del procedimiento, lo que se hace es que se remueven cualquier tipo de errores que podamos encontrar



Integración

Si se busca información proveniente de distintas bases de datos, se crea una base de datos llamada warehouse y ahí se almacena toda la información



Transformación

Se transforma nuestra información para que este en un formato adecuado como la de reducción de dimensiones (se disminuye la cantidad de variables) para el siguiente proceso que es el de minar información

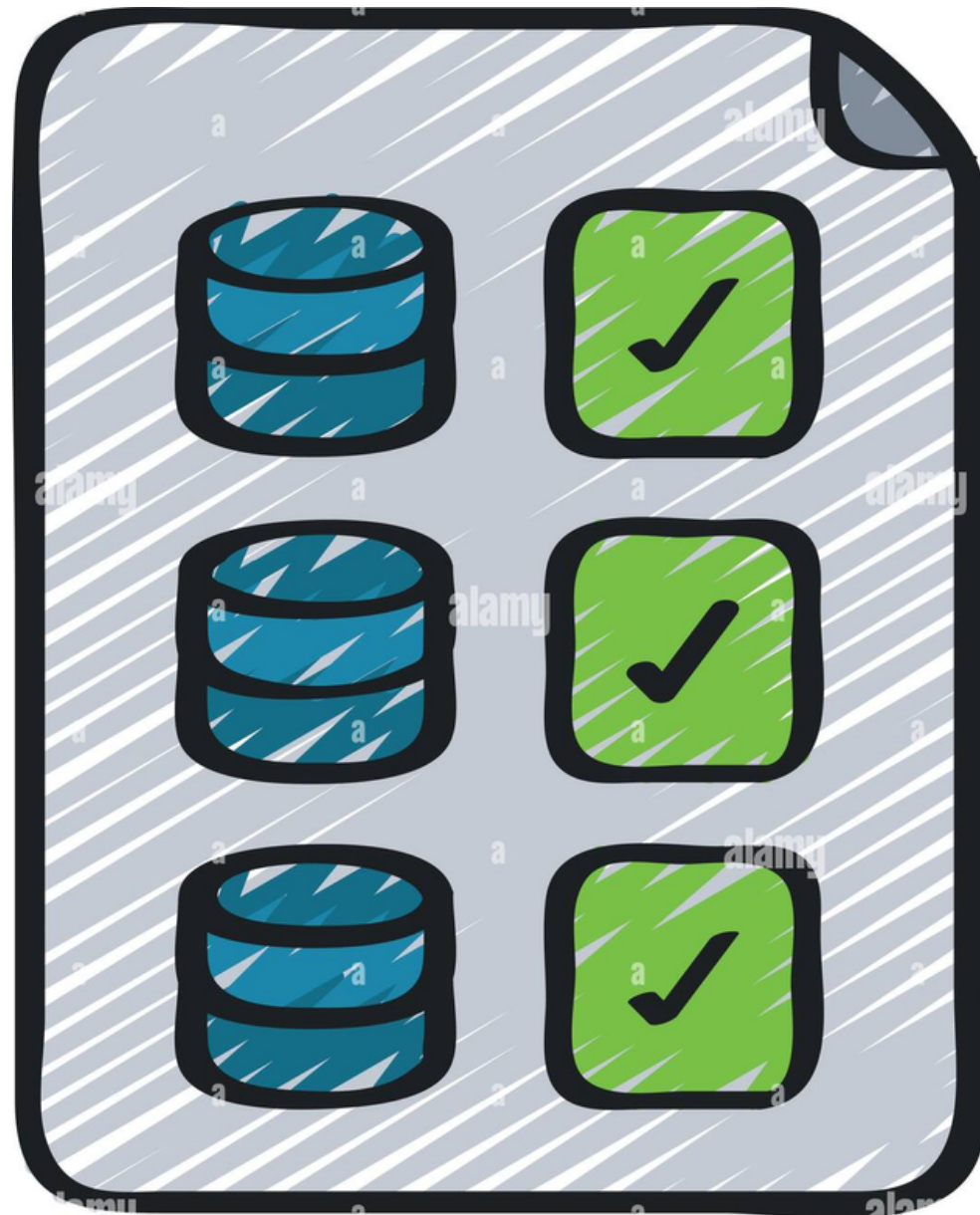
Minería de Datos

Se aplica este proceso para extraer los patrones significativos que veamos y nos sean útiles para llegar al conocimiento.



Evaluación

Una vez aplicado el minado de datos, se evalúan los patrones que se generaron. Esto para verificar que que cumpla las metas planteadas



Conocimiento

Se representa el conocimiento encontrado ayudándonos de gráficas o tablas y comenzar a resolver la problemática



El KDD es iterativo

Esto quiere decir que el KDD cuenta con ciclos con los que se va puliendo y adquiriendo nuevos conocimientos.

¿Se puede minar cualquier tipo de información?

Casi toda, siempre y cuando esté estructurada y se pueda representar en forma de datos



Ventajas del KDD

- Mejora la toma de decisiones (proporciona información valiosa)
- Aumento de la eficiencia (automatiza tareas repetitivas y que consumen tiempo)
- Mejor servicio al cliente (comprender mejor necesidades y preferencias)
- Detección de fraudes (detecta actividades fraudulentas mediante identificación de patrones y anomalías)
- Modelado predictivo (predice tendencias y patrones futuros)



Desventajas del KDD

- **Preocupaciones de privacidad:** (Implica recolección y análisis de grandes cantidades de datos)
- **Complejidad:** (puede ser un proceso complejo que requiere habilidad y conocimiento especializado)
- **Consecuencias no deseadas:** (sesgos o discriminación, si los datos no son comprensibles o utilizados adecuadamente)
- **Calidad de datos:** (depende en gran medida de la calidad de los datos)
- **Alto costo:** (requiere inversión de software, hardware y personal)
- **Sobreajuste:** (debido a el aprendizaje automatico, donde puede afectar de manera negativa el rendimiento del modelo)



Diferencias entre KDD Y DATA MINING



Técnica usada

- KDD: Limpieza de datos, integración de datos, selección de datos, transformación de datos, minería de datos, evaluación de patrones y representación y visualización del conocimiento.
- Data mining: Reglas de asociación, clasificación, clustering, regresión, árboles de decisión, redes neuronales y reducción de dimensionalidad.



Output o Salida

- KDD: Resultado Información estructurada, como reglas y modelos, que pueden utilizarse para tomar decisiones o hacer predicciones.
- Data mining: Patrones, asociaciones o ideas que se pueden utilizar para mejorar la toma de decisiones o la comprensión.



Enfoque

- KDD: El enfoque se centra en el descubrimiento de conocimiento útil, es más amplio que encontrar patrones en los datos.
- Data mining El enfoque se centra en el descubrimiento de patrones o relaciones en los datos.



Preguntas

- ¿Toda la información se puede minar?
- ¿En qué consiste el paso de limpieza de datos?
- 1 Ventaja del KDD
- 1 Desventaja del KDD
- Menciona dos técnicas para la minería de datos
- ¿Qué significa la transformación de datos?

