



Objetivo general

Enseñar al público lector la importancia de la minería de datos y de los procesos de metodologías que esta cuenta sirviendo de ayuda a empresas, negocios, mercado y usuarios sobre todo al momento de buscar una toma de decisiones.

Alcanzar a comprender el uso de toma de decisiones de una empresa y aprender a aplicar una metodología para la resolución de un problema siguiendo paso a paso.

Introduccion

Business Intelligence (BI) se ha convertido en una herramienta esencial en el campo de la minería de datos, especialmente en la industria minera. BI ayuda en la extracción de información valiosa de grandes cantidades de datos y proporciona una ventaja competitiva a las empresas. Discutiremos el concepto de BI en la minería de datos, sus aplicaciones en la industria minera y los beneficios y desafíos de implementar BI en el análisis de datos de minería.

La minería de datos es un proceso de extracción de información útil de grandes conjuntos de datos. Implica el uso de varias técnicas y algoritmos para descubrir patrones, tendencias y relaciones en los datos. Uno de los métodos más populares utilizados para la minería de datos es Warehouse KDD (Descubrimiento de conocimiento en bases de datos).

Warehouse KDD es un método poderoso que permite a las organizaciones descubrir patrones ocultos y correlaciones en sus datos. Les ayuda a identificar nuevas oportunidades de crecimiento y desarrollar estrategias para una mejor toma de decisiones. Este método se puede utilizar para analizar varios tipos de datos, como estructurados, semiestructurados o no estructurados.



El proceso Warehouse KDD

Empezó a utilizarse en 1989 para referirse al amplio proceso de búsqueda de conocimiento en bases de datos, y para enfatizar la aplicación a "alto nivel" de métodos específicos de minería de datos. En general, el descubrimiento es un tipo de inducción de conocimiento, no supervisado, que implica dos procesos:

- Búsqueda de regularidades interesantes entre los datos de partida
- Formulación de leyes que las describan.

Implica varios pasos, incluido el preprocesamiento, el descubrimiento de patrones, la representación del conocimiento, el posprocesamiento y la evaluación. Al utilizar este método, las organizaciones pueden obtener información valiosa a partir de sus datos, lo que les puede ayudar a tomar mejores decisiones y mejorar sus operaciones.

El descubrimiento Busca más definidamente recoger datos, formar hipótesis para explicar las observaciones, diseñar experimentos, comprobar la corrección de las hipótesis, comparar nuestros hallazgos con los de otros investigadores y repetir el ciclo.

Sin embargo, el descubrimiento no requiere realizar simultáneamente todas estas tareas. De igual modo que un investigador puede descubrir nuevo conocimiento a través del análisis de sus datos, una computadora puede examinar los datos disponibles o recogidos por otras computadoras y encontrar relaciones y explicaciones previamente desconocidas, realizando así descubrimiento en un sentido más restringido.

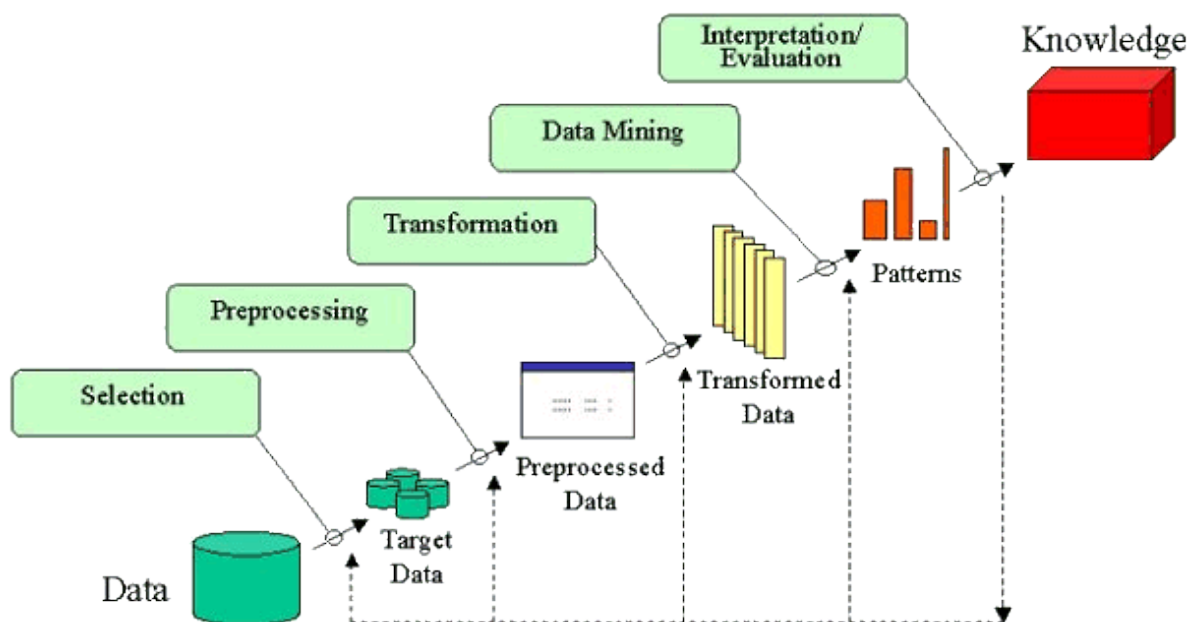


Proceso KDD

La minería de datos es el proceso de extraer información útil de grandes conjuntos de datos. Utiliza algoritmos para analizar e identificar patrones en los datos, que luego se pueden usar para hacer predicciones o tomar decisiones.

KDD (Descubrimiento de conocimiento en bases de datos) es un tipo de minería de datos que se enfoca en descubrir conocimiento de grandes conjuntos de datos. Implica técnicas como la agrupación en clústeres, la clasificación, la minería de reglas de asociación y la detección de anomalías.

KDD puede ayudar a las organizaciones a descubrir relaciones ocultas entre diferentes variables en sus datos de las que quizás no hayan sido conscientes antes. Esto se puede utilizar para tomar mejores decisiones o desarrollar nuevos productos y servicios.

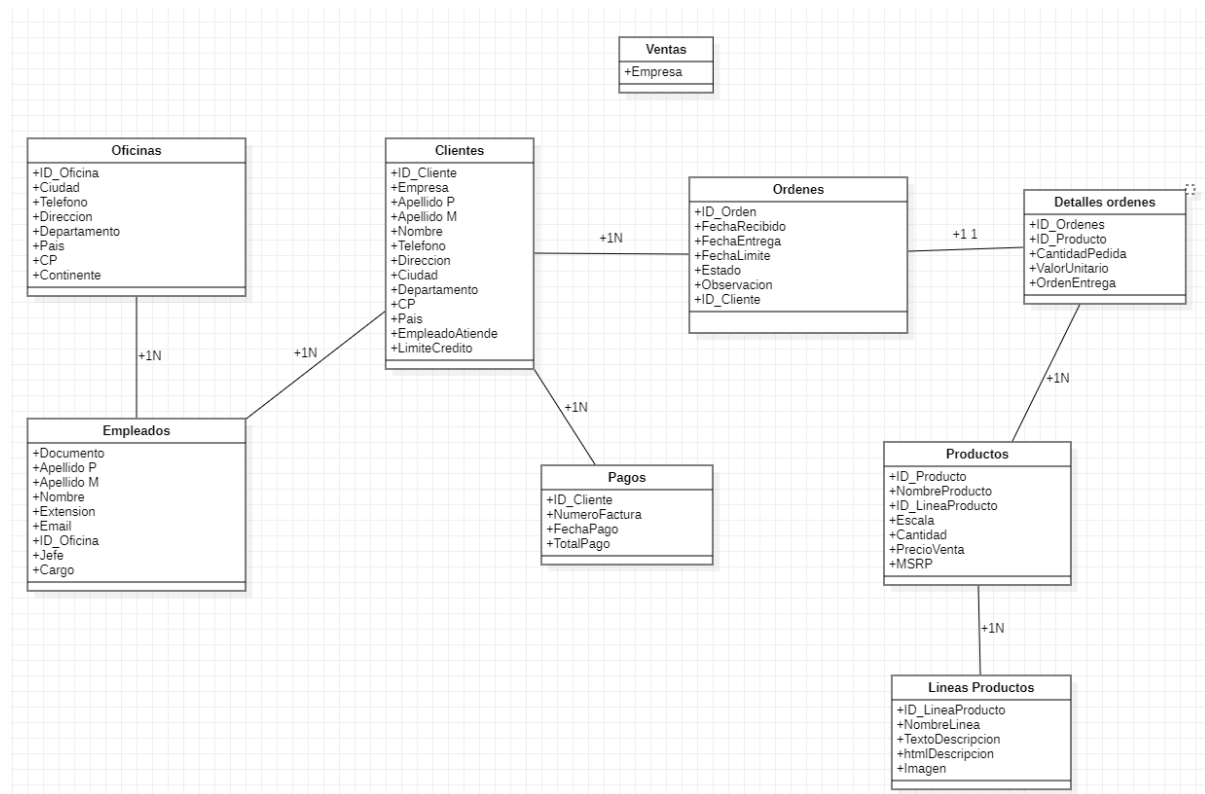




(img: ejemplo de metodología KDD)

Esquema (Almacén de datos)

Business Intelligence (BI) es un conjunto de herramientas, técnicas y tecnologías utilizadas para transformar datos sin procesar en información significativa y útil. BI ayuda a analizar datos y proporcionar información sobre las operaciones comerciales, que se pueden utilizar para tomar decisiones informadas.



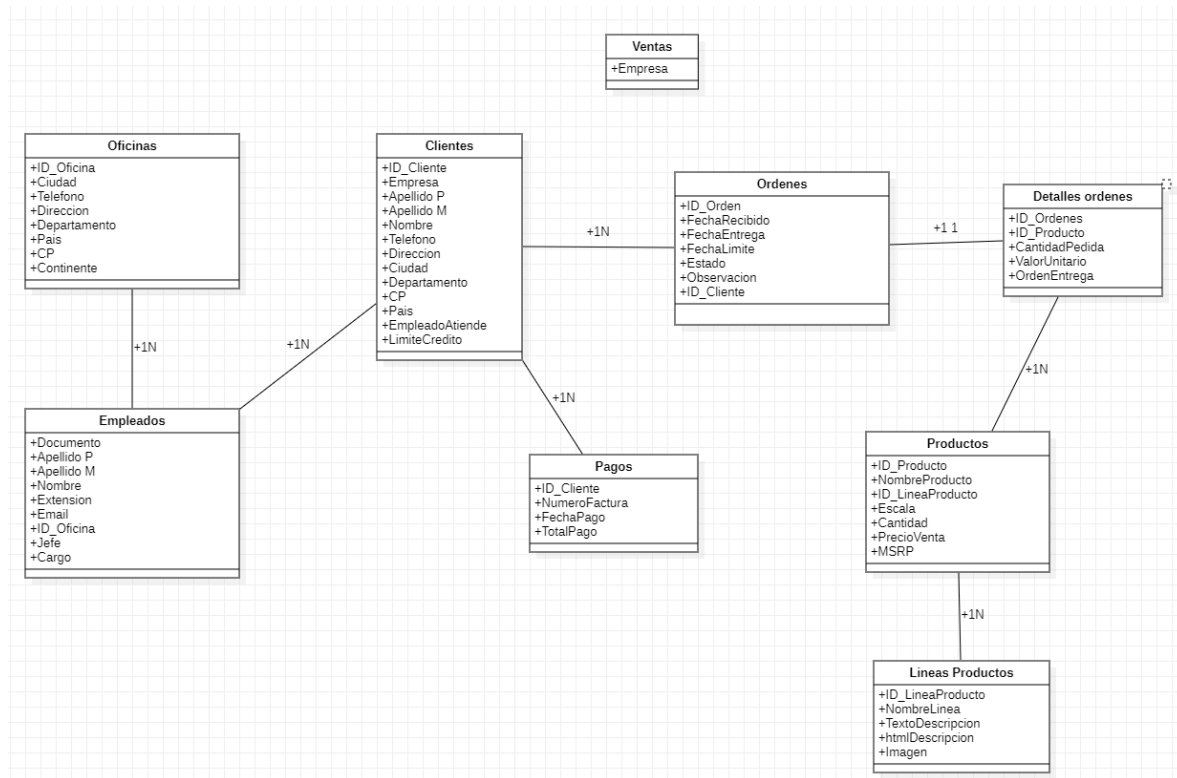


En este esquema se aplicará la metodología KDD con el fin de explicar cómo funciona la minería de datos y la toma de decisiones. Realizaremos un reporte que explique los análisis, resultados y conclusiones.

Paso 1

Desarrollo y entendimiento del dominio de la aplicación, el conocimiento relevante y los objetivos del usuario final. Este paso requiere cierta dependencia usuario/analista, pues intervienen factores como: conocer los cuellos de botella del dominio, saber qué partes son susceptibles de un procesamiento automático y cuáles no, cuáles son los objetivos, los criterios de rendimiento exigibles, para qué se usarán los resultados que se obtengan, compromisos entre simplicidad y precisión del conocimiento extraído, etc

Se identificarán los diversos campos y opciones que ofrece el diagrama con el cual ayudara a buscar y identificar el punto de partida para la minería de datos.



Paso 2

Una vez identificado y analizado bien el diagrama de la empresa es hora de empezar a aplicar la metodología KDD

Empezaremos por recopilar todos los datos que se encuentran en nuestra base de datos (Diagrama). En dicho diagrama contamos con las siguientes tablas de una empresa

Oficinas, Empleados, Clientes, Ordenes, Pagos, detalles de órdenes, productos y líneas de producto.



En cada una de estas encontraremos los siguientes datos.

Oficinas
+ID_Oficina
+Ciudad
+Telefono
+Direccion
+Departamento
+Pais
+CP
+Continente

Empleados
+Documento
+Apellido P
+Apellido M
+Nombre
+Extension
+Email
+ID_Oficina
+Jefe
+Cargo

Lineas Productos
+ID_LineaProducto
+NombreLinea
+TextoDescripcion
+htmlDescripcion
+Imagen

Clientes
+ID_Cliente
+Empresa
+Apellido P
+Apellido M
+Nombre
+Telefono
+Direccion
+Ciudad
+Departamento
+CP
+Pais
+EmpleadoAtiende
+LimiteCredito

Ordenes
+ID_Orden
+FechaRecibido
+FechaEntrega
+FechaLimite
+Estado
+Observacion
+ID_Cliente

Detalles ordenes
+ID_Ordenes
+ID_Producto
+CantidadPedida
+ValorUnitario
+OrdenEntrega

Pagos
+ID_Cliente
+NumeroFactura
+FechaPago
+TotalPago

Productos
+ID_Producto
+NombreProducto
+ID_LineaProducto
+Escala
+Cantidad
+PrecioVenta
+MSRP

La recopilación de datos es un proceso muy importante puesto que es necesario a la hora de aplicar la metodología ya establecida.

Esto implica consideraciones sobre la homogeneidad de los datos, su variación a lo largo del tiempo, estrategia de muestreo, grados de libertad, etc.

Paso 3



El procesamiento de los datos

Esta parte definirá que datos limpiar para una mejor selección y extracción de la información, para poder aplicar esto es necesario tener los datos ya recopilados previamente solicitados en el paso número 2

Preprocesado de los datos: eliminación de ruido, estrategias para manejar valores ausentes, normalización de los datos, etc.

Este paso se aplicará a nuestro diagrama de la siguiente forma:

*Limpiar valores (ATRIBUTOS) No necesarios para la empresa. Visto de otra forma servirá para manejar valores ausentes y normalizar más los datos.

Una vez aplicado quedará de la siguiente forma

Oficinas	Empleados	Cientes	Lineas Productos
+ID_Oficina +Ciudad +Telefono +Direccion +Departamento +Pais +CP	+Nombre +Extension +Email +ID_Oficina +Cargo	+ID_Cliente +Nombre +Telefono +Direccion +Ciudad +Departamento +CP +Pais	+ID_LineaProducto +NombreLinea +TextoDescripcion +Imagen

Ordenes	Detalles ordenes	Productos
+ID_Orden +FechaRecibido +FechaEntrega +Estado +Observacion +ID_Cliente	+ID_Ordenes +ID_Producto +CantidadPedida +ValorUnitario +OrdenEntrega	+ID_Producto +ID_LineaProducto +Cantidad +PrecioVenta

Paso 4

Transformación y reducción



Incluye la búsqueda de características útiles de los datos según sea el objetivo final, la reducción del número de variables y la proyección de los datos sobre espacios de búsqueda en los que sea más fácil encontrar una solución. Este es un paso crítico dentro del proceso global, que requiere un buen conocimiento del problema y una buena intuición, y que, con frecuencia, marca la diferencia entre el éxito o fracaso de la minería de datos.

Junto con la normalización de datos ya utilizada se aplica la reducción que fue la eliminación de campos innecesarios para el mejor manejo de la información por eso tanto el preprocesamiento junto con la transformación y reducción de los datos son fundamentales para poder lograr un mejoramiento diferencial de los datos y que en este se pueda observar como ayuda a un mejor control y facilidad para la empresa.

Paso 5

Elección del tipo de sistema para minería de datos. Esto depende de si el objetivo del proceso de KDD es la clasificación, regresión, agrupamiento de conceptos (clustering), detección de desviaciones, etc.

Se aplicará un sistema con el objetivo de llevar los datos a un mejor agrupamiento tanto de conceptos como de información en general que sirva para el flujo correcto de estos mismos

El uso de Clustering o la cluterización es un proceso importante dentro del **Machine learning**. Este proceso desarrolla una acción fundamental que le permite a los algoritmos de aprendizaje automatizado entrenar y conocer de forma adecuada los datos con los que desarrollan sus actividades.

La agrupación en clústeres de un KDD (descubrimiento de conocimientos en bases de datos) es el proceso de agrupar puntos de datos similares en clústeres dentro de un conjunto de datos. Esta técnica es útil para identificar patrones subyacentes en grandes conjuntos de datos y facilita el análisis de datos complejos.

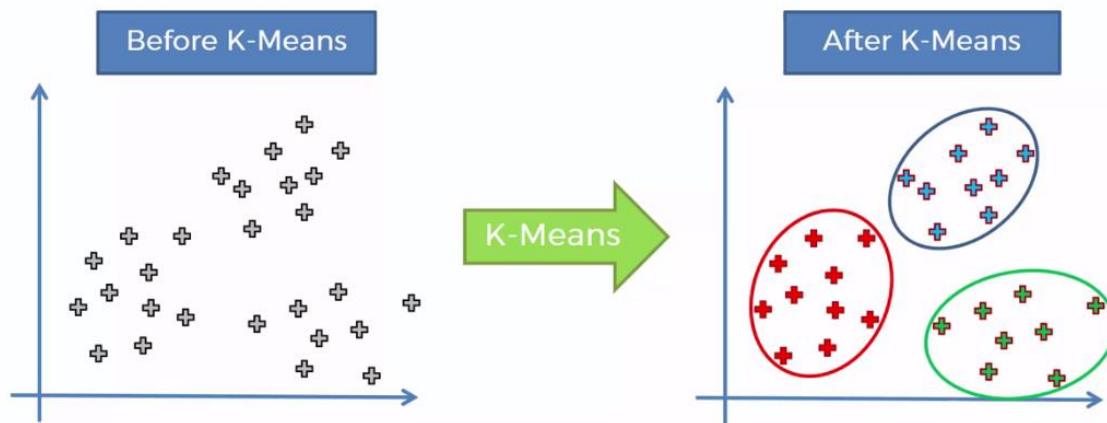
Existen diferentes tipos de algoritmos de agrupamiento que se pueden usar para este propósito, como k-means, agrupamiento jerárquico y modelos de mezcla gaussiana. También exploraremos los diferentes casos de uso para la agrupación, como la segmentación de clientes y el análisis de la cesta de la compra.

La agrupación en clústeres es una herramienta importante en el campo del descubrimiento de conocimiento y la minería de datos (KDD). Se utiliza para agrupar puntos de datos similares,



lo que permite una mejor comprensión de las relaciones subyacentes entre ellos. Los ejemplos de algoritmos de agrupamiento incluyen k-means, agrupamiento jerárquico y agrupamiento basado en densidad.

Cada uno de estos tipos tiene diferentes ventajas y se puede utilizar para lograr diferentes resultados. Al comprender estos ejemplos de agrupación, se pueden desarrollar modelos más precisos que capturen mejor la estructura de un conjunto de datos determinado.



Aquí dependerá de las necesidades del usuario. En este caso se plasma un ejemplo mas claro con el uso de clusters y la herramienta k-means para explicar de forma más detallada el agrupamiento de datos.

Paso 6

Minería de datos

En este paso se realiza la búsqueda de conocimiento con una determinada representación del mismo. El éxito de la minería de datos depende en gran parte de la correcta realización de los pasos previos: por parte del usuario.

Dependerá de la realización de los pasos anteriores dependiendo del trabajo que cada uno quiera aplicar, esto no es siempre igual ya que varía de cada tipo de dato.

En nuestro caso se uso k-means para una mejor agrupación y minado de datos.



Paso 7

Interpretación del conocimiento extraído, con posibilidad de iterar de nuevo desde el primer paso. La obtención de resultados aceptables dependerá de factores como: definición de medidas del interés del conocimiento (de tipo estadístico, en función de su sencillez, etc.) que permitan filtrarlo de forma automática, existencia de técnicas de visualización para facilitar la valoración de los resultados o búsqueda manual de conocimiento útil entre los resultados obtenidos.

Los resultados de cada tipo dependerán de cada trabajo y de un buen manejo de los datos, se aplicarán las medidas de interés para definir que tan complejo o sencillo es.

El filtrado de información hace referencia a la búsqueda de datos y la complejidad y sencillez de encontrarlos.





Paso 8

Consolidación del conocimiento descubierto, incorporándolo al sistema, o simplemente documentándolo y enviándolo a la parte interesada. Este paso incluye la revisión y resolución de posibles inconsistencias con otro conocimiento extraído previamente.

Si esta parte es incompleta o errónea el usuario deberá de realizar nueva mente todo el proceso hasta lograr llegar al resultado deseado, es importante seguir paso a paso ademas de tener lo que requiere a la mano.

Para finalizar la revisión de estos datos podremos hacerla en forma de analisis, resultados y conclusiones.

Analisis sobre la minería de datos

El análisis de minería de datos se está convirtiendo en una herramienta cada vez más importante para que las empresas obtengan información valiosa a partir de sus datos. Puede ayudarlos a identificar patrones, tendencias y correlaciones en sus datos que pueden usarse para mejorar la toma de decisiones y hacer mejores predicciones para el futuro. También se puede utilizar para descubrir relaciones ocultas entre diferentes conjuntos de datos y detectar anomalías en los datos.



Analisis

La metodología KDD funciona para un mejor ordenamiento de los datos, funciona para la agrupación y relaciones entre los mismos ayudando a depurar (limpiar) y sobre todo lo mas importante a tomar una mejor decisión puesto que nos basamos en buscar una alternativa para limpiar y agilizar los datos de una empresa, con esta metodología logramos acortar el cuello de botella que esta tenia con los datos antiguos a una mejor optimizada gracias al uso de K-means que sirvió para agrupar.

La toma de decisiones es algo importante para las empresas puesto que se requiere y es fundamental a la hora de buscar mejores rendimientos o para facilitar mas el registro y uso de datos.

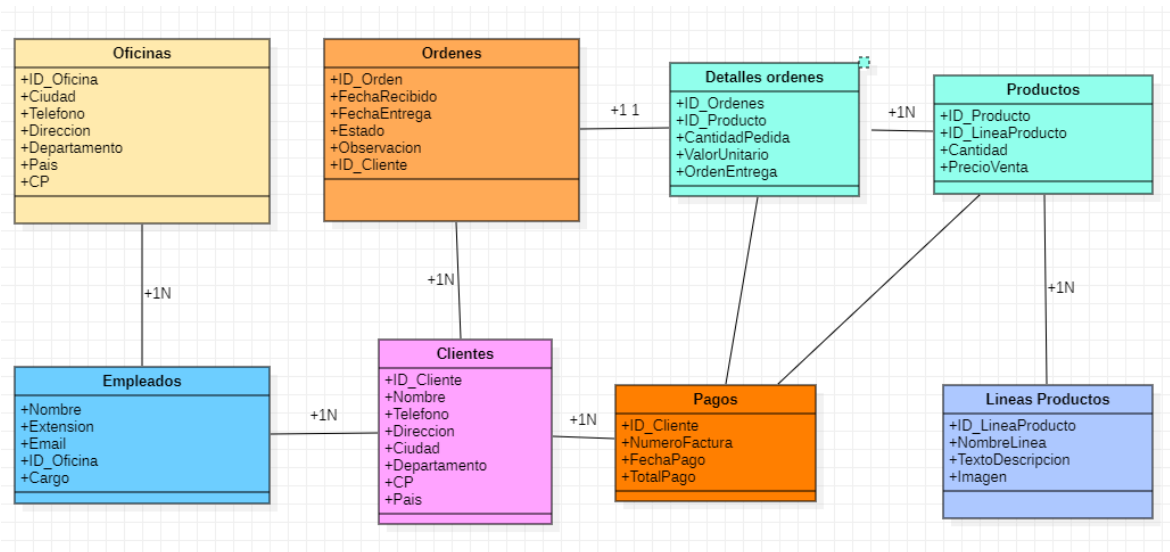
Resultados

- *Mejor agrupamiento de datos gracias a k-means
- *Depuración de datos (variables) innecesarias vistas en los ejemplos.
- *Buena interpretación y evaluación de nuestro diagrama

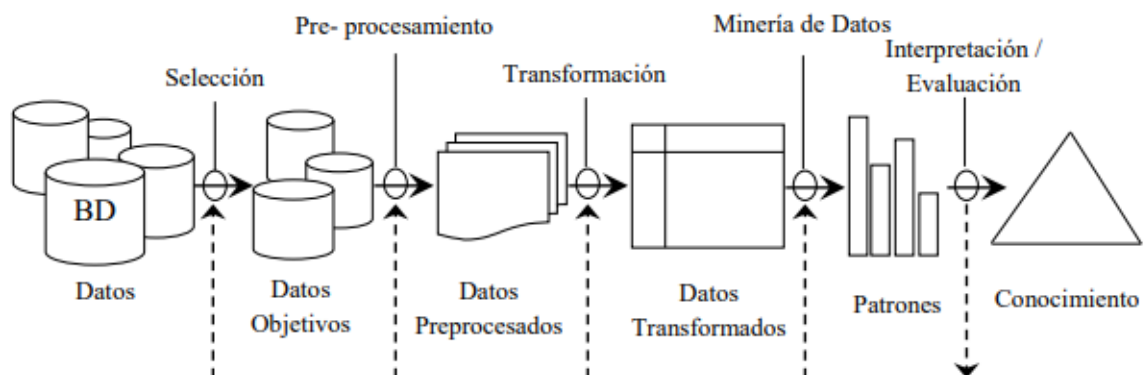


*Mejoramiento de datos con el uso de la transformación

Diagrama terminado con el uso de metodología KDD



Metodologia usada KDD



Material usado

Materia: Inteligencia de Negocios.

Carrera: Ingeniería Informática

Tema: Minería de Datos

Equipo: Hernández Flores Angel

Luna Frausto Roberto

Martínez Luna Aldair

Montoya Herrera Brian

Nava Bazán Brenda Vanessa



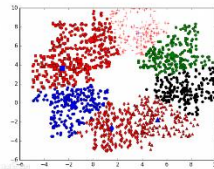
Software (StarUML)



Rytr (inteligencia artificial)



Algoritmo de K-means (adaptado para el diagrama)



Conclusión

En conclusión, la minería de datos es una herramienta poderosa que se ha vuelto cada vez más importante en el mundo actual basado en datos. Sus técnicas y herramientas se utilizan para analizar grandes conjuntos de datos y obtener información sobre patrones y relaciones complejas.

La minería de datos tiene numerosas aplicaciones del mundo real en diversas industrias, incluidas la salud, las finanzas y el marketing.

A medida que los datos continúan desempeñando un papel cada vez más importante en nuestras vidas, la minería de datos seguirá siendo una herramienta esencial para las empresas, los gobiernos y las personas.

Materia: Inteligencia de Negocios.

Carrera: Ingeniería Informática

Tema: Minería de Datos

Equipo: Hernández Flores Angel

Luna Frausto Roberto

Martínez Luna Aldair

Montoya Herrera Brian

Nava Bazán Brenda Vanessa



La toma de decisiones ayuda demasiado a la hora de decidir algo de suma importancia teniendo en cuenta que puede fallar o salir de forma exitosa

Bibliografías

R. (s. f.). Rytr - Best AI Writer, Content Generator & Writing Assistant.
Rytr. <https://rytr.me/?via=07101991>

Data Mining: Concepts and Techniques. By Han, Jiawei / Kamber, Micheline Morgan
Kaufmann Publishers; 08/2000

Predictive Data Mining: A Practical Guide By Weiss. Sholom M. /Indurkha, Nitin Morgan
Kaufmann Pub; 08/1997

Materia: Inteligencia de Negocios.

Carrera: Ingeniería Informática

Tema: Minería de Datos

Equipo: Hernández Flores Angel

Luna Frausto Roberto

Martínez Luna Aldair

Montoya Herrera Brian

Nava Bazán Brenda Vanessa



Data Mining: Technologies, Techniques, Tools, and Trends By Thuraisingham, Bhavani
CRC Pr; 12/1998

Methodologies for Knowledge Discovery and Data Mining By Zhong, SPRINGER
VERLAG INC; 07/1999