

**INSTITUTO POLITÉCNICO NACIONAL**



**UNIDAD PROFESIONAL INTERDISCIPLINARIA DE INGENIERÍA Y CIENCIAS SOCIALES Y ADMINISTRATIVAS**

**“Repositorio de Datos de Conocimiento”**

**Ingeniería del Conocimiento**

**Nombres:**

Cordero Hernández Pablo Cesar 2015600437

Osorio cuello Edgar Hiram 2016602330

# ÍNDICE

[**ÍNDICE**](#_t4ib2tsdt7vs) **2**

[**Introducción**](#_fwr21x7a2tf2) **3**

[**Marco Teórico**](#_bhtj3m4p0v1m) **4**

[**Metodología y/o Herramientas Utilizadas**](#_ns4ugcmjt7o7) **5**

[Herramientas Utilizadas](#_ae3goz8af0z7) 5

[Diseño de la Aplicación](#_5j7cuifki6uh) 7

[Proceso de Negocios](#_7tfa8qy8bdyh) 9

[¿Por qué es importante la minería de datos?](#_pcuftlinzq0w) 9

[¿Quien lo utiliza?](#_pgfynqe9sk51) 10

[Ecuaciones y Fórmulas](#_k3ndebcmf0if) 12

[Teorema de bayes](#_omp8mskigusu) 12

[Fragmentos de Código](#_lfu9odvorma6) 13

[Citas y Referencias](#_rtvllu4c3svu) 16

[**Conclusiones**](#_n4jr3phuf9t1) **18**

# Introducción

La industria 4.0 está en pleno desarrollo, por ello el automatizar procesos es importante para llevar a cabo mayores tareas en menos tiempo. La minería de datos es actualmente importante en todo ámbito, los datos representan en conjunto información y la información es valiosa para cualquier entidad que necesite explotarla.

En el desarrollo de este proyecto, detallaremos y pondremos en práctica los conocimientos obtenidos durante el semestre en la materia de Ingeniería del Conocimiento para llevar a cabo una arquitectura extensible para minería de datos para aplicaciones móviles usando diferentes tecnologías en diferentes lenguajes en un servidor disponibles de manera pública en GitHub.

# Marco Teórico

**Prolog**

ProLog es un lenguaje de programación simple, pero poderoso. Se basa en nociones matemáticas de relaciones de inferencia. Es un lenguaje declarativo e interpretado, esto quiere decir que el lenguaje se usa para representar conocimientos sobre un determinado dominio y las relaciones entre objetos de ese dominio.

Un programa en ProLog consiste de una base de hechos de relaciones lógicas y detalles que se cumplen para la aplicación. Dicha base de datos no tiene una estructura impuesta, ni un procedimiento o clase principal. Escribir un programa en ProLog consiste en declarar el conocimiento disponible acerca de los objetivos, además de sus relaciones y sus reglas. En lugar de correr en un programa para obtener una solución, se hace una pregunta, el programa revisa la base de datos para encontrar la solución a la pregunta. Si existe más de una solución, ProLog hace backtracking para encontrar soluciones distintas. El propio sistema es el que deduce las respuestas a las preguntas que se le plantean, dichas respuestas las deduce del conocimiento obtenido por el conjunto de reglas dadas. La ejecución de ProLog consiste en una búsqueda en profundidad de un árbol conteniendo todas las posibles soluciones. Para cada una de ellas se evaluará su validez. La estructura de un programa en ProLog es lógica y directa. Se explican cada una de sus partes y operadores disponibles a continuación.[[1]](#footnote-0)

Historia

* Coulmerauer (1970) desarrolló un lenguaje para hacer deducciones de texto.
* Se aplica un mecanismo de resolución sobre predicados especiales, cláusulas de Horn, llamado unificación.
* La difusión del lenguaje se produce en los 80, pero de forma muy limitada debido a la falta de aplicaciones en dicho lenguaje.

Características del lenguaje

* El programa prolog se compone de unos hechos (datos) y un conjunto de reglas, es decir, relaciones entre objetos de la base de datos.
* La ejecución del programa cargado en memoria consiste en realizar una pregunta de forma interactiva: el intérprete generará por inferencia los resultados que se deducen a partir del contenido de la base de datos.
* PROLOG tiene una sintaxis y semántica simples. Sólo busca relaciones entre los objetos creados, las variables y las listas, que son sus estructuras básicas.
* Comentarios entre /\* \*/

**Minería de datos**

La minería de datos es el proceso de hallar anomalías, patrones y correlaciones en grandes conjuntos de datos para predecir resultados. Empleando una amplia variedad de técnicas, puede utilizar esta información para incrementar sus ingresos, recortar costos, mejorar sus relaciones con clientes, reducir riesgos y más.[[2]](#footnote-1)

Historia

El proceso de hurgar en los datos para descubrir conexiones ocultas y predecir tendencias futuras tiene una larga historia. Conocido algunas veces como "descubrimiento de conocimientos en bases de datos", el término "minería de datos" no se acuño sino hasta la década de 1990. Pero su base comprende tres disciplinas científicas entrelazadas: estadística (el estudio numérico de relaciones de datos), inteligencia artificial (inteligencia similar a la humana exhibida por software y/o máquinas) y machine learning (algoritmos que pueden aprender de datos para hacer predicciones). Lo que era antiguo es nuevo otra vez, ya que la minería de datos continúa evolucionando para igualar el ritmo del potencial sin límites del big data y poder de cómputo asequible.

En la última década, los avances en el poder y la velocidad de procesamiento nos han permitido llegar más allá de las prácticas manuales, tediosas y que toman mucho tiempo al análisis de datos rápido, fácil y automatizado. Cuanto más complejos son los conjuntos de datos recopilados, mayor es el potencial que hay para descubrir insights relevantes. Los comerciantes detallistas, bancos, fabricantes, proveedores de telecomunicaciones y aseguradoras, entre otros, utilizan la minería de datos para descubrir relaciones entre todas las cosas, desde precios, promociones y demografía hasta la forma en que la economía, el riesgo, la competencia y los medios sociales afectan sus modelos de negocios, ingresos, operaciones y relaciones con clientes.

¿Cómo funciona la minería de datos?

Si bien la tecnología de información a gran escala ha ido evolucionando por separado las transacciones y sistemas de análisis, la minería de datos proporciona un enlace entre los dos. El software de minería de datos analiza las relaciones y patrones en los datos de transacción almacenados sobre la base de consultas de los usuarios de composición abierta.

Existen varios tipos de software de análisis que están disponibles: estadísticos, de aprendizaje automático, y redes neuronales. En general, se buscan cualquiera de estos cuatro tipos de relaciones:

• Clases:

Los datos almacenados se utilizan para localizar datos en grupos predeterminados. Por ejemplo, una cadena de restaurantes podría minar datos de compra del cliente para determinar cuando los clientes los visitan y lo que normalmente ordenan. Esta información podría utilizarse para aumentar el tráfico al tener especiales del día.

• Grupos:

Los elementos de datos se agrupan de acuerdo a las relaciones lógicas o preferencias de los consumidores. Por ejemplo, los datos pueden ser extraídos para identificar segmentos de mercado o afinidades de los consumidores.

• Asociaciones:

Los datos pueden ser minados para identificar asociaciones. El ejemplo de cerveza-pañal es un ejemplo de minería asociativa.

• Patrones secuenciales:

Los datos se minan para anticipar patrones de comportamiento y tendencias. Por ejemplo, un distribuidor especializado en sistemas al aire libre puede predecir la probabilidad de compra de una mochila basado en la compra de un consumidor de sacos de dormir y calzado de senderismo.

La minería de datos se compone de cinco elementos principales:

1. Extraer, transformar y cargar datos de transacciones en el sistema de almacenamiento de datos.
2. Almacenar y manejar los datos en un sistema de base de datos multidimensional.
3. Proporcionar acceso a datos para los analistas de negocios y profesionales de tecnologías de la información.
4. Analizar los datos por un software de aplicación.
5. Presentar los datos en un formato útil, como un gráfico o tabla.

**Teorema de bayes**

El teorema de Bayes es utilizado para calcular la probabilidad de un suceso, teniendo información de antemano sobre ese suceso.

Podemos calcular la probabilidad de un suceso A, sabiendo además que ese A cumple cierta característica que condiciona su probabilidad. El teorema de Bayes entiende la probabilidad de forma inversa al teorema de la probabilidad total. El teorema de la probabilidad total hace inferencia sobre un suceso B, a partir de los resultados de los sucesos A. Por su parte, Bayes calcula la probabilidad de A condicionado a B.

El teorema de Bayes ha sido muy cuestionado. Lo cual se ha debido, principalmente, a su mala aplicación. Ya que, mientras se cumplan los supuestos de sucesos disjuntos y exhaustivos, el teorema es totalmente válido.[[3]](#footnote-2)

\*Este teorema es detallado más adelante en la sección de “Ecuaciones y Fórmulas” del documento.

# Metodología y/o Herramientas Utilizadas

## Herramientas Utilizadas

**PHPMyAdmin**

Es una herramienta de software libre escrito en PHP pensado para gestionar la administración de MySQL. Es compatible con una amplia gama de operaciones con MySQL, las más utilizadas son compatibles con la interfaz de usuario, además tiene la capacidad de ejecutar cualquier sentencia SQL directamente.

Cuenta con soporte para la mayoría de características de MySQL: buscar y eliminar bases de datos, tablas, vistas, campos e índices crear, copiar, eliminar y cambiar el nombre de alterar las bases de datos, tablas, campos e índices de mantenimiento del servidor, bases de datos y tablas, con propuestas sobre la configuración del servidor ejecutar, editar y marcar cualquier sentencia SQL, incluso consultas por lotes administrar usuarios MySQL y privilegios gestión de procedimientos almacenados y disparadores. Importar datos de CSV y SQL. Exportar datos a varios formatos: CSV, SQL, XML, PDF, ISO / IEC 26300 - OpenDocument Text y hoja de cálculo, Word, LATEX y otros.

**MySQL**

MySQL es un sistema de gestión de bases de datos relacionales de código abierto con un modelo cliente-servidor. MySQL es un manejador de bases de datos principalmente utilizado en las aplicaciones WEB en conjunto con PHP.

**PHP**

PHP (Hypertext Preprocessor) es un lenguaje de código abierto muy popular especialmente adecuado para el desarrollo web y que puede ser incrustado en HTML. Lo que distingue a PHP de algo del lado del cliente como Javascript es que el código es ejecutado en el servidor, generando HTML y enviándolo al cliente. El cliente recibirá el resultado de ejecutar el script, aunque no se sabrá el código subyacente que era. El servidor web puede ser configurado incluso para que procese todos los ficheros HTML con PHP, por lo que no hay manera de que los usuarios puedan saber qué se tiene debajo de la manga.

**JSON**

JSON (JavaScript Object Notation) es un formato ligero de intercambio de datos. JSON es un formato de texto que es completamente independiente del lenguaje pero utiliza convenciones que son ampliamente conocidos por los programadores de la familia de lenguajes C, incluyendo C, C++, C#, Java, JavaScript, Perl, Python, y muchos otros. Estas propiedades hacen que JSON sea un lenguaje ideal para el intercambio de datos.

Las ventajas JSON sobre XML como formato de intercambio de datos es que resulta mucho más sencillo escribir un analizador sintáctico. En la práctica, JSON se emplea habitualmente en entornos donde el tamaño del flujo de datos entre cliente y servidor es de vital importancia (de aquí su uso por Yahoo!, Google, Mozilla, etc, que atienden a millones de usuarios) cuando la fuente de datos es explícitamente de fiar y donde no es importante el hecho de no disponer de procesamiento XSLT para manipular los datos en el cliente.

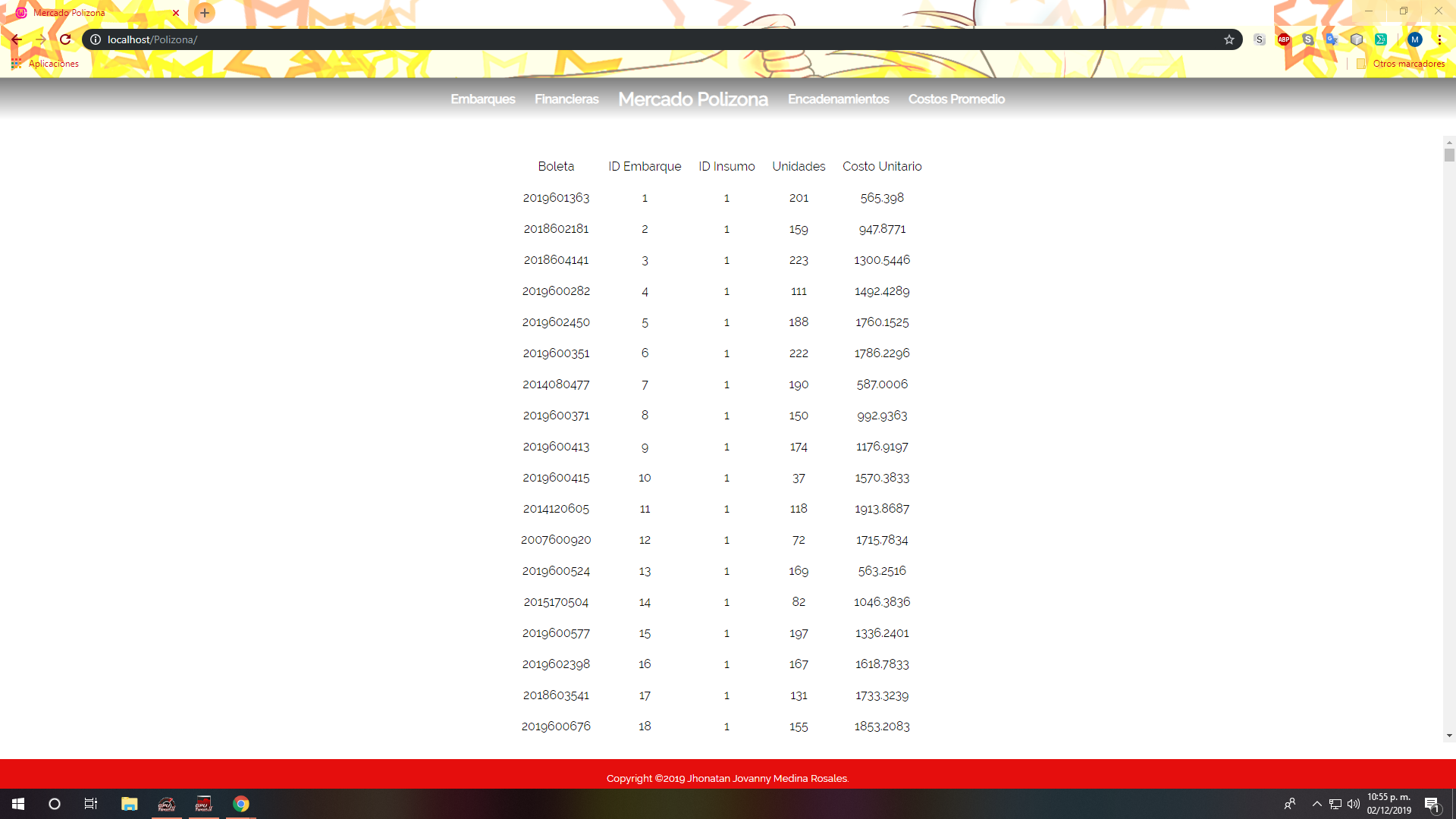
JSON está constituído por dos estructuras:

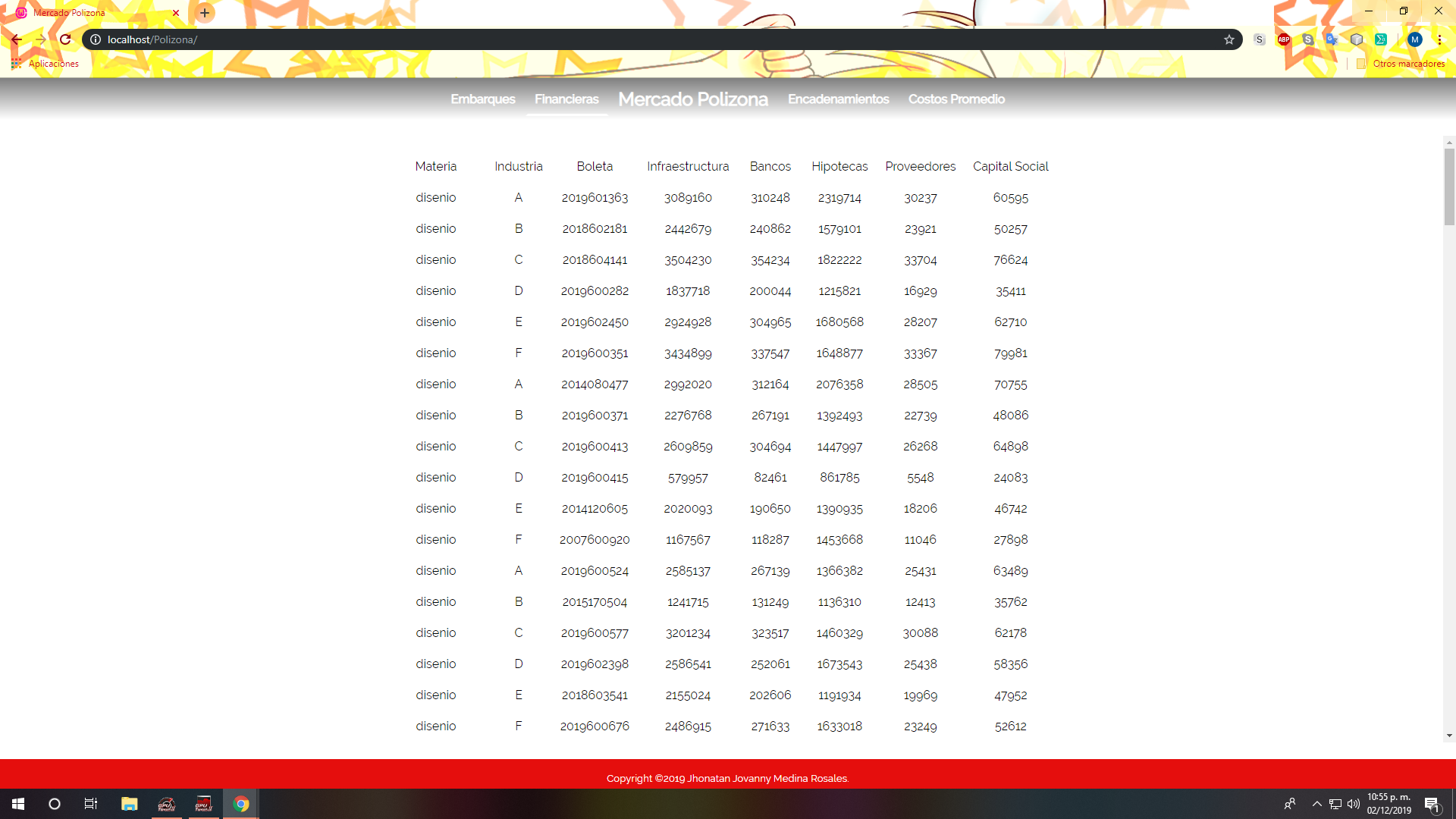
* Una colección de pares de nombre/valor. En varios lenguajes esto es conocido como un objeto, registro, estructura, diccionario, tabla hash, lista de claves o un arreglo asociativo.
* Una lista ordenada de valores. En la mayoría de los lenguajes, esto se implementa como arreglos, vectores, listas o secuencias.

Estas son estructuras universales; virtualmente todos los lenguajes de programación las soportan de una forma u otra. Es razonable que un formato de intercambio de datos que es independiente del lenguaje de programación se base en estas estructuras.

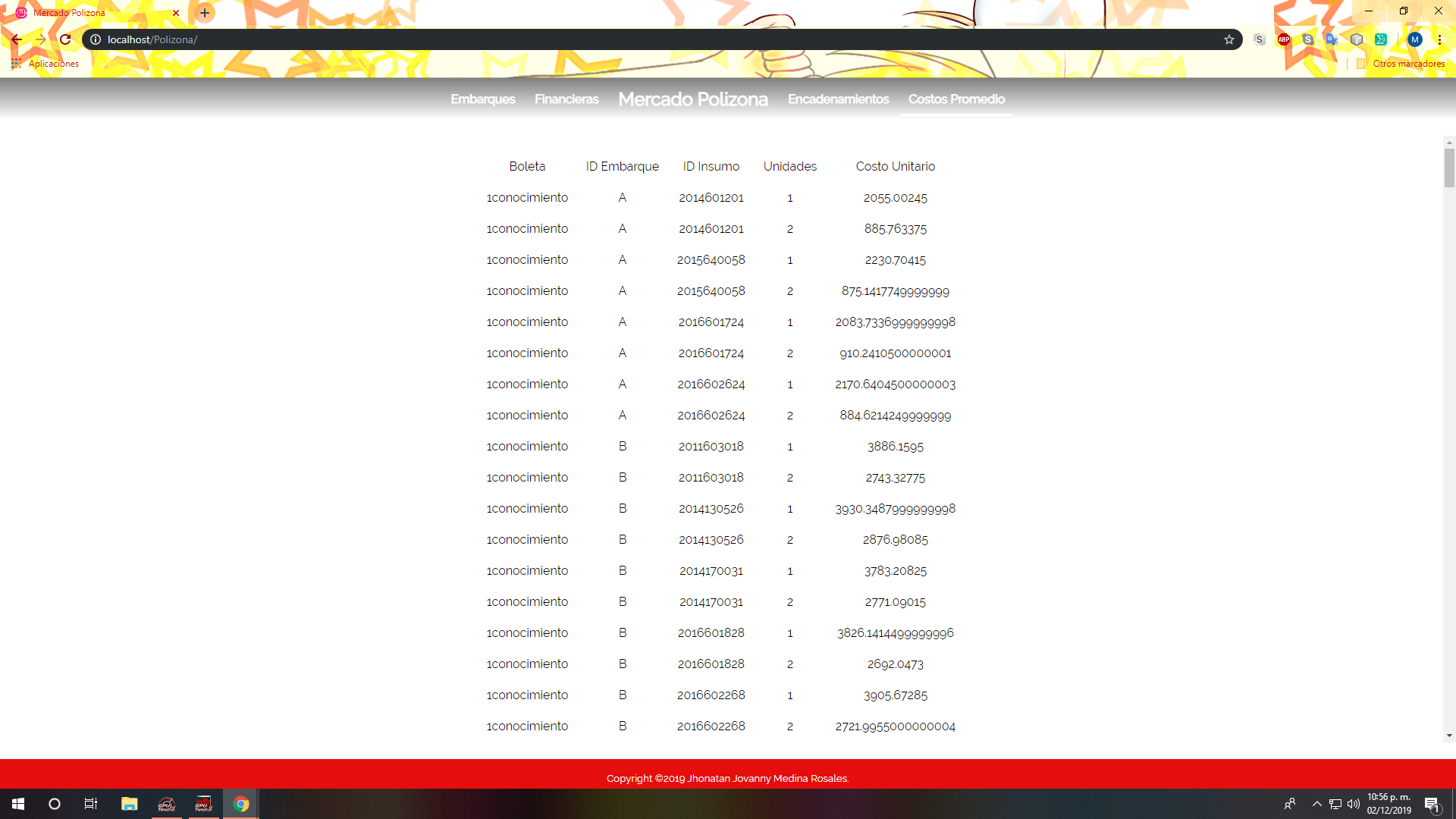
# 

## Diseño de la Aplicación









## Proceso de Negocios

la minería de datos o la exploración de datos es un campo de la estadística y las [ciencias de la computación](https://es.wikipedia.org/wiki/Ciencias_de_la_computaci%C3%B3n) referido al proceso que intenta descubrir patrones en grandes volúmenes de conjuntos de datos-​ Utiliza los métodos de la [inteligencia artificial](https://es.wikipedia.org/wiki/Inteligencia_artificial), [aprendizaje automático](https://es.wikipedia.org/wiki/Aprendizaje_autom%C3%A1tico), [estadística](https://es.wikipedia.org/wiki/Estad%C3%ADstica) y sistemas de [bases de datos](https://es.wikipedia.org/wiki/Bases_de_datos). El objetivo general del proceso de minería de datos consiste en extraer información de un conjunto de datos y transformarla en una estructura comprensible para su uso posterior. Además de la etapa de análisis en bruto, supone aspectos de [gestión de datos](https://es.wikipedia.org/wiki/Gesti%C3%B3n_de_datos) y de bases de datos, de [procesamiento de datos](https://es.wikipedia.org/wiki/Procesamiento_de_datos), del modelo y de las consideraciones de inferencia, de métricas de Intereses, de consideraciones de la [teoría de la complejidad computacional](https://es.wikipedia.org/wiki/Teor%C3%ADa_de_la_complejidad_computacional), de post-procesamiento de las estructuras descubiertas, de la visualización y de la actualización en línea.

La tarea de minería de datos real es el análisis automático o semi-automático de grandes cantidades de datos para extraer patrones interesantes hasta ahora desconocidos, como los grupos de registros de datos (análisis clúster), registros poco usuales (la detección de anomalías) y dependencias (minería por reglas de asociación). Esto generalmente implica el uso de técnicas de bases de datos como los índices espaciales. Estos patrones pueden entonces ser vistos como una especie de resumen de los datos de entrada, y pueden ser utilizados en el análisis adicional o, por ejemplo, en el aprendizaje automático y [análisis predictivo](https://es.wikipedia.org/wiki/An%C3%A1lisis_predictivo).

### ¿Por qué es importante la minería de datos?

Ha podido apreciar los números asombrosos – el volumen de datos producidos se duplica cada dos años. Los datos no estructurados por sí solos conforman el 90% del universo digital. Pero más información no significa necesariamente más conocimientos.

La minería de datos le permite:

1. Filtrar todo el ruido caótico y repetitivo en sus datos.
2. Entender qué es relevante y luego hacer un buen uso de esa información para evaluar resultados probables.
3. Acelerar el ritmo de la toma de decisiones informadas.

**La minería de datos es una piedra angular de la analítica, la cual le ayuda a desarrollar los modelos que pueden descubrir conexiones dentro de millones o miles de millones de registros. Aprenda cómo la minería de datos está dando forma al mundo en que vivimos.**

### ¿Quien lo utiliza?

Comunicaciones  
En un mercado sobrecargado donde la competencia es cerrada, las respuestas se encuentran a menudo en los datos de sus consumidores. Las compañías de multimedia y telecomunicaciones pueden utilizar modelos analíticos para entender montañas de datos de clientes, ayudándoles así a predecir el comportamiento de sus clientes y ofrecer campañas altamente dirigidas y relevantes.

Seguros  
Con conocimientos analíticos, las compañías de seguros pueden resolver problemas complejos concernientes a fraude, cumplimiento, gestión de riesgo y separación de clientes. Las compañías han utilizado técnicas de minería de datos para asignar precios a productos con mayor eficacia en líneas de negocios y hallar nuevas formas de ofrecer productos competitivos a su base de clientes existente.

Educación  
Con vistas unificadas basadas en datos del progreso de los estudiantes, los educadores pueden predecir el desempeño de sus alumnos antes de que pongan un pie en el salón de clases – y desarrollar estrategias de intervención para mantenerlos en curso. La minería de datos ayuda a los educadores a acceder a datos de los estudiantes, a predecir niveles de logro y a detectar estudiantes o grupos de estudiantes que necesitan atención extra.

Manufactura  
la alineación de planes de suministro con pronósticos de demanda es esencial, como lo es también la detección temprana de problemas, garantía de calidad e inversión en equidad de marca. Los fabricantes pueden predecir el desgaste de activos de producción y anticipar su mantenimiento, lo cual puede maximizar el tiempo en operación y mantener la línea de producción acorde a lo programado.

Bancos  
Los algoritmos automatizados ayudan a los bancos a entender a su base de clientes y también los miles de millones de transacciones en el corazón del sistema financiero. La minería de datos ayuda a las compañías de servicios financieros a tener una mejor vista de los riesgos del mercado, a detectar el fraude en menos tiempo, a gestionar las obligaciones de cumplimiento de las regulaciones y a obtener retornos óptimos de sus inversiones en marketing.

Retail  
Grandes bases de datos de clientes contienen insights ocultos que le pueden ayudar a mejorar las relaciones con clientes, optimizar campañas de marketing y pronosticar ventas. A través de modelos de datos más precisos, las compañías detallistas pueden ofrecer campañas más enfocadas – y encontrar la oferta que tenga el mayor impacto en el cliente.

## Ecuaciones y Fórmulas

### Teorema de bayes

Lo que normalmente se quiere saber en aprendizaje es cual es la mejor hipótesis o mejor dicho la más probable dados los datos.

El teorema de bayes expresa la probabilidad condicional de un evento aleatorio A dado B en términos de la distribución de probabilidad condicional del evento B dado A y la distribución de probabilidad marginal de sólo A.

En términos más generales y menos matemáticos, el teorema de bayes es de enorme relevancia puesto que vincula la probabilidad de A dado B con la probabilidad de B dado A.

Sea un conjunto de sucesos mutuamente excluyentes y exhaustivos, y tales que la probabilidad de cada uno de ellos es distinta de cero, sea B un suceso cualquiera del que se conocen las probabilidades condicionales P(B|A1). Entonces, la probabilidad P(A1|B) viene dada por la expresión.

**P(A1|B) = ( P(B|A1) P(A1) ) / P(B)**

Donde:

* P(A1) - son las probabilidades a priori.
* P(B|A1) es la probabilidad de B en la hipótesis A1.
* P(A1|B) son las probabilidades a posterior.

En este caso específico el teorema de bayes es fundamental para la obtención de las probabilidades. utilizando la fórmula que permite estimar la probabilidad de los elementos basados en otras dimensiones. utilizamos:

**( P(A|X) \* P(X) ) / ( P(A|X)\* P(X|N) ) + ( P(A|Y) \* P(Y|N) )**

## 

## Fragmentos de Código

**Embarques**

<?php

$sql="SELECT \* from embarques";

$result=mysqli\_query($conexion,$sql);

while($mostrar=mysqli\_fetch\_array($result)){

?>

<tr>

<td><?php echo $mostrar['boleta'] ?></td>

<td><?php echo $mostrar['idembarque'] ?></td>

<td><?php echo $mostrar['idinsumo'] ?></td>

<td><?php echo $mostrar['unidades'] ?></td>

<td><?php echo $mostrar['costoUnitario'] ?></td>

</tr>

<?php

}

?>

**Encadenamientos**

<?php

$sql="SELECT \* from encadenamientos";

$result=mysqli\_query($conexion,$sql);

while($mostrar=mysqli\_fetch\_array($result)){

?>

<tr>

<td><?php echo $mostrar['vendedora'] ?></td>

<td><?php echo $mostrar['compradora'] ?></td>

<td><?php echo $mostrar['coeficiente'] ?></td>

</tr>

<?php

}

?>

**Financieras**

<?php

$sql="SELECT \* from financieras";

$result=mysqli\_query($conexion,$sql);

while($mostrar=mysqli\_fetch\_array($result)){

?>

<tr>

<td><?php echo $mostrar['materia'] ?></td>

<td><?php echo $mostrar['industria'] ?></td>

<td><?php echo $mostrar['boleta'] ?></td>

<td><?php echo $mostrar['infraestructura'] ?></td>

<td><?php echo $mostrar['bancos'] ?></td>

<td><?php echo $mostrar['hipotecas'] ?></td>

<td><?php echo $mostrar['proveedores'] ?></td>

<td><?php echo $mostrar['capital\_social'] ?></td>

</tr>

<?php

}

?>

**SUPER QUERY**

CREATE VIEW costosPromedio as

(

SELECT materia, industria, financieras.boleta, idinsumo, sum(unidades\*costoUnitario)/sum(unidades)

as costoPromedio FROM financieras, embarques

where financieras.boleta=embarques.boleta

group by idinsumo,boleta

order by materia, industria, boleta

)

**Costos Promedios**

<?php

$sql="SELECT \* from costospromedio";

$result=mysqli\_query($conexion,$sql);

while($mostrar=mysqli\_fetch\_array($result)){

?>

<tr>

<td><?php echo $mostrar['materia'] ?></td>

<td><?php echo $mostrar['industria'] ?></td>

<td><?php echo $mostrar['boleta'] ?></td>

<td><?php echo $mostrar['idinsumo'] ?></td>

<td><?php echo $mostrar['costoPromedio'] ?></td>

</tr>

<?php

}

?>

## 

## 

## 

## 

## Citas y Referencias

**Minería de datos**

1° operación Laudon NYU (Business Intelligence).

Clasificador asignar a cada objeto una y solo una categoría (Turing).

[ttps://www.ecured.cu/ProLog](https://www.ecured.cu/ProLog)

<https://desarrolloweb.com/articulos/1573.php>

<https://www.tecnologias-informacion.com/mineria-de-datos.html>

<https://www.sas.com/es_mx/insights/analytics/data-mining.html>

<https://www.ecured.cu/PhpMyAdmin>

<https://www.json.org/json-es.html>

<https://www.aprenderaprogramar.com/index.php?option=com_content&view=article&id=956:ique-es-y-para-que-sirve-json-especificacion-oficial-javascript-object-notation-diferencia-de-xml-cu01213f&catid=83&Itemid=212>

# Conclusiones

Las aplicaciones móviles, en una sociedad que tiene la tecnología al alcance de la mano, son vitales para una mayor comodidad, eficiencia. El poner a la disposición nuestra mina de datos a todo aquel interesado podrá ayudarnos a ver en la práctica si el impacto que habíamos predicho está llevándose a cabo o no. El llevar a cabo este proyecto fue completamente satisfactorio puesto que hemos llevado a la práctica conocimientos nuevos y conocimientos ya adquiridos en otras materias. Esperamos que nuestro proyecto sea de gran ayuda a cualquier interesado en él y lo pueda usar con provecho.

1. "ProLog - EcuRed." <https://www.ecured.cu/ProLog>. Accessed 3 Dec. 2019. [↑](#footnote-ref-0)
2. "La minería de datos: Definición, usos y funciones." <https://www.tecnologias-informacion.com/mineria-de-datos.html>. Accessed 3 Dec. 2019. [↑](#footnote-ref-1)
3. "Teorema de Bayes - Definición, qué es y concepto ...." <https://economipedia.com/definiciones/teorema-de-bayes.html>. Accessed 3 Dec. 2019. [↑](#footnote-ref-2)