

**INSTITUTO POLITÉCNICO NACIONAL**

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

**“Diseño e implementación colaborativa de una arquitectura extensible de servicios de minería de datos “**

**Ingeniería del Conocimiento**

**Nombres:**

Cordero Hernández Pablo Cesar 2015600437

Osorio cuello Edgar Hiram 2016602330

Pastrana Jiménez Brenda Sulem 2016602726

Diseño e implementación colaborativa de una arquitectura extensible de servicios de minería de datos

MCE. Gilberto Antonio Ramírez Flores

Academia de Inteligencia Artificial, Unidad Profesional Interdisciplinaria de Ingeniería y Ciencias Sociales y Administrativas, Instituto Politécnico Nacional,  
Av. Té #950 Col Granjas México, Iztacalco, 08400, CDMX

gramirezfl@ipn.mx

Resumen. Somos 3 estudiantes de UPIICSA de la carrera ingeniería en informática que actualmente cursamos la materia de ingeniería del conocimiento y hoy día tenemos la misión de implementar un servicio de minería de datos. La intención es que nuestra aplicación sea fácil de usar para cualquier persona y que le sea de gran apoyo al realizar sus consultas.

Las operaciones de minería de datos consideradas son la clasificación y la agrupación. Gracias a la automatización del proceso de la comunicación con la plataforma podremos realizarlo de manera eficaz y sencilla. Además de que el servicio debe implementarse en una arquitectura extensible.

Palabras Clave:

* Automatización.
* Inteligencia artificial.
* Aplicaciones móviles.
* Prolog.
* Organizaciones inteligentes.
* Base de datos.
* Tecnología.
* Minería de datos.

1. Introducción

Automatizar procesos que hasta ese momento se han realizado de forma manual nos asegura ahorrar costes, tiempo y, además, ser mucho más eficaces a la hora de realizar nuestro trabajo diario.

Hoy en día el mercado es muy competitivo, son cada vez muchas más organizaciones las que optan por automatizar procesos, con el fin de ser mucho más rentables y eficaces. La automatización de las gestiones administrativas se trata de la automatización de sistemas con el fin de hacer las cosas mucho más fáciles, efectivas y eficientes, además del funcionamiento de la organización.

En cuanto a la minería de datos o exploración de datos es un campo de la estadística y las ciencias de la computación referida al proceso que intenta descubrir patrones en grandes volúmenes de conjuntos de datos.​ Utiliza los métodos de la inteligencia artificial, aprendizaje automático, estadística y sistemas de 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.

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.

* 1. Marco teórico

**Automatizar**

Es, en informática, el conjunto de métodos que sirven para realizar tareas repetitivas en un ordenador. Algunos métodos para la automatización de tareas son la programación simple, los macros, los intérpretes y las bombas lógicas. También hay algunos programas específicos que automatizan tareas. Incluso los virus informáticos utilizados de forma benéfica podrían considerarse otro método para la automatización de tareas para el usuario.



La mayoría de las apps web tienen una API (Application Program Interface) la cual pueden usar desarrolladores terceros para interactuar con ellas. Por lo tanto estas aplicaciones de automatización toman dos servicios y actúan como intermediarios para sus APIs. Gracias a ello podemos enlazar aplicaciones diferentes para conseguir una potencialidad exponencial de ellas.

**Inteligencia artificial**

La Inteligencia artificial es el campo científico de la informática que se centra en la creación de programas y mecanismos que pueden mostrar comportamientos considerados inteligentes. En otras palabras, la IA es el concepto según el cual “las máquinas piensan como seres humanos”.

Normalmente, un sistema de IA es capaz de analizar datos en grandes cantidades (big data), identificar patrones y tendencias y, por lo tanto, formular predicciones de forma automática, con rapidez y precisión. Para nosotros, lo importante es que la IA al integrar análisis predictivos, permite que nuestras experiencias cotidianas sean más inteligentes.

La siguiente tabla muestra algunos acontecimientos importantes en la historia de la Inteligencia Artificial.

|  |  |
| --- | --- |
| **Año** | **Acontecimiento** |
| 1842 | Lovelace: máquina analítica programable. |
| 1950 | Turing: la prueba de Turing. |
| 1956 | McCarthy, Minsky, Rochester y Shannon celebran la primera conferencia sobre IA. |
| 1965 | Weizenbaum: "ELIZA", el primer sistema especialista. |
| 1993 | Horswill: "Polly" (robótica basada en el comportamiento). |
| 2005 | TiVo: tecnología de recomendaciones. |
| 2011 | Apple, Google y Microsoft: aplicaciones móviles de recomendaciones. |
| 2013 | Varios: avances tecnológicos en aprendizaje automático y profundo. |
| 2016 | Google DeepMind: AlphaGo supera a Lee Sedol en el juego “Go”. |

**Técnicas principales de la Inteligencia artificial**

* **Aprendizaje automático**

Es en este campo en donde los avances más importantes de la IA se están llevando a cabo. En términos prácticos, “el Aprendizaje automático es la ciencia que se encarga de hacer que las computadoras realicen acciones sin necesidad de programación explícita”. La idea principal aquí es que se les puede proporcionar datos a los algoritmos de Aprendizaje automático y luego usarlos para saber cómo hacer predicciones o guiar decisiones.

* **Aprendizaje profundo**

Es una técnica de Aprendizaje automático que utiliza redes neuronales (el concepto de que las neuronas se pueden simular mediante unidades computacionales) para realizar tareas de clasificación (pensemos en clasificar una imagen de un gato, de un perro o personas, por ejemplo).

* **Descubrimiento de datos inteligentes**

La idea consiste en permitir la automatización total del ciclo de la IE (inteligencia empresarial): la incorporación y preparación de datos, el análisis predictivo y los patrones y la identificación de hipótesis. Este es un ejemplo interesante de la recuperación de datos inteligentes en acción. La información que ninguna herramienta de IE había descubierto.

**Prolog**

Proveniente del francés PROgrammation en LOGique,​ es un lenguaje de programación lógico e interpretado usado habitualmente en el campo de la Inteligencia artificial.

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.

**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 o exploración de datos (es la etapa de análisis de "Knowledge Discovery in Databases" o KDD) es un campo de la estadística y las ciencias de la computación referido al proceso que intenta descubrir patrones en grandes volúmenes de conjuntos de datos.​ Utiliza los métodos de la inteligencia artificial, aprendizaje automático, estadística y sistemas de 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 y de bases de datos, de procesamiento de datos, del modelo y de las consideraciones de inferencia, de métricas de Intereses, de consideraciones de la teoría de la complejidad computacional, de post-procesamiento de las estructuras descubiertas, de la visualización y de la actualización en línea.

**Historia**

La minería de datos es algo que empieza por los alrededores de los años sesenta, ya los estadísticos manejaban términos como Data Fishing, Data Mining (DM) o Data Archaeology con la idea de encontrar correlaciones sin una Definición previa de bases de datos.

En los principios de los años ochenta, Rakesh Agrawal, GioWiederhold, Robert Blum y Gregory Piatetsky-Shapiro y otros, empezaron a consolidar los términos de Minería de Datos y KDD.

Estas tecnologías han sido una buena herramienta para personas que se desenvuelven en el ámbito de los negocios y académico.

La evolución de sus herramientas en el transcurso del tiempo puede dividirse en cuatro etapas principales:

|  |  |
| --- | --- |
| **Año** | **Acontecimiento** |
| **1960** | Colección de Datos |
| **1980** | Acceso de Datos |
| **1990 (principios)** | Almacén de Datos y Apoyo a las Decisiones |
| **1990 (finales)** | Minería de Datos Inteligente. |

**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.

**Un proceso típico de minería de datos consta de los siguientes pasos generales:**

1. Selección del conjunto de datos, tanto en lo que se refiere a las variables objetivo (aquellas que se quiere predecir, calcular o inferir), como a las variables independientes (las que sirven para hacer el cálculo o proceso), como posiblemente al muestreo de los registros disponibles.
2. Análisis de las propiedades de los datos, en especial los histogramas, diagramas de dispersión, presencia de valores atípicos y ausencia de datos (valores nulos).
3. Transformación del conjunto de datos de entrada, se realizará de diversas formas en función del análisis previo, con el objetivo de prepararlo para aplicar la técnica de minería de datos que mejor se adapte a los datos y al problema, a este paso también se le conoce como preprocesamiento de los datos. Un problema sustancial asociado al desarrollo de este tipo de sistemas cuando contienen texto en Inglés es el tamaño de su vocabulario; que es más grande que el de cualquier otra Lengua en el mundo. Un método que se está ahora utilizando en estos casos es el de simplificación antes de proceder con el proceso, por medio de convertir dicho texto a Ingles básico; mismo que contiene solo 1,000 palabras que también se utilizan para describir en notas al pie el sentido de las más de 30,000 palabras definidas en el “Diccionario Básico de Ciencias”.
4. Selección y aplicación de la técnica de minería de datos, se construye el modelo predictivo, de clasificación o segmentación.
5. Extracción de conocimiento, mediante una técnica de minería de datos, se obtiene un modelo de conocimiento, que representa patrones de comportamiento observados en los valores de las variables del problema o relaciones de asociación entre dichas variables. También pueden usarse varias técnicas a la vez para generar distintos modelos, aunque generalmente cada técnica obliga a un preprocesado diferente de los datos.
6. Interpretación y evaluación de datos, una vez obtenido el modelo, se debe proceder a su validación comprobando que las conclusiones que arroja son válidas y suficientemente satisfactorias. En el caso de haber obtenido varios modelos mediante el uso de distintas técnicas, se deben comparar los modelos en busca de aquel que se ajuste mejor al problema. Si ninguno de los modelos alcanza los resultados esperados, debe alterarse alguno de los pasos anteriores para generar nuevos modelos.

**Las técnicas más representativas son:**

* **Redes neuronales:**

Son un paradigma de aprendizaje y procesamiento automático inspirado en la forma en que funciona el sistema nervioso de los animales. Se trata de un sistema de interconexión de neuronas en una red que colabora para producir un estímulo de salida. Algunos ejemplos de red neuronal son:

* El perceptrón.
* El perceptrón multicapa.
* Los mapas autoorganizados, también conocidos como redes de Kohonen.
* **Regresión lineal:**

Es la más utilizada para formar relaciones entre datos. Rápida y eficaz pero insuficiente en espacios multidimensionales donde puedan relacionarse más de 2 variables.

* **Árboles de decisión:**

Un árbol de decisión es un modelo de predicción utilizado en el ámbito de la inteligencia artificial y el análisis predictivo, dada una base de datos se construyen estos diagramas de construcciones lógicas, muy similares a los sistemas de predicción basados en reglas, que sirven para representar y categorizar una serie de condiciones que suceden de forma sucesiva, para la resolución de un problema. Ejemplos:

* Algoritmo ID3.
* Algoritmo C4.5
* **Modelos estadísticos:**

Es una expresión simbólica en forma de igualdad o ecuación que se emplea en todos los diseños experimentales y en la regresión para indicar los diferentes factores que modifican la variable de respuesta.

* **Agrupamiento o Clustering:**

Es un procedimiento de agrupación de una serie de vectores según criterios habitualmente de distancia; se tratará de disponer los vectores de entrada de forma que estén más cercanos aquellos que tengan características comunes. Ejemplos:

* Algoritmo K-means
* Algoritmo K-medoids
* **Reglas de asociación:**

Se utilizan para descubrir hechos que ocurren en común dentro de un determinado conjunto de datos.

Según el objetivo del análisis de los datos, los algoritmos utilizados se clasifican en supervisados y no supervisados (Weiss y Indurkhya, 1998):

* Algoritmos supervisados (o predictivos): predicen un dato (o un conjunto de ellos) desconocido a priori, a partir de otros conocidos.
* Algoritmos no supervisados (o del descubrimiento del conocimiento): se descubren patrones y tendencias en los datos.

**¿Cómo funciona la minería de datos?**

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.

**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.

* 1. Metodología y/o 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**

Es un sistema de gestión de bases de datos relacional desarrollado bajo licencia dual: Licencia pública general/Licencia comercial por Oracle Corporation y está considerada como la base de datos de código abierto más popular del mundo, y una de las más populares en general junto a Oracle y Microsoft SQL Server, sobre todo para entornos de desarrollo web.

**Características**

Inicialmente, MySQL carecía de elementos considerados esenciales en las bases de datos relacionales, tales como integridad referencial y transacciones. A pesar de ello, atrajo a los desarrolladores de páginas web con contenido dinámico, justamente por su simplicidad.

Poco a poco los elementos de los que carecía MySQL están siendo incorporados tanto por desarrollos internos, como por desarrolladores de software libre. Entre las características disponibles en las últimas versiones se puede destacar:

-**Amplio subconjunto del lenguaje SQL. Algunas extensiones son incluidas igualmente.**

-**Disponibilidad en gran cantidad de plataformas y sistemas.**

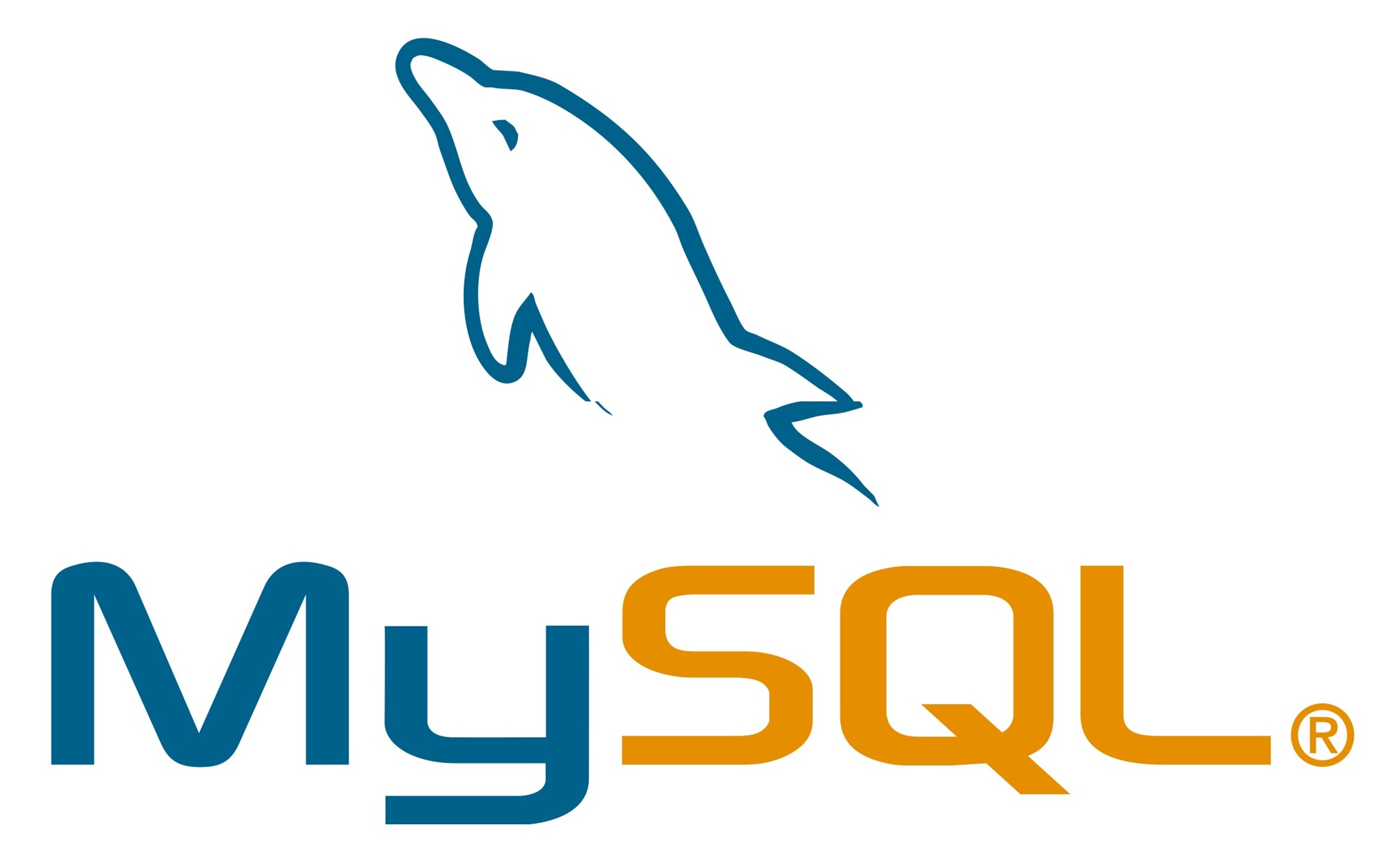
**-Posibilidad de selección de mecanismos de almacenamiento que ofrecen diferentes velocidades de operación, soporte físico, capacidad, distribución geográfica, transacciones...**

**-Transacciones y claves foráneas.**

**-Conectividad segura.**

**-Replicación.**

**-Búsqueda e indexación de campos de texto.**



**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.

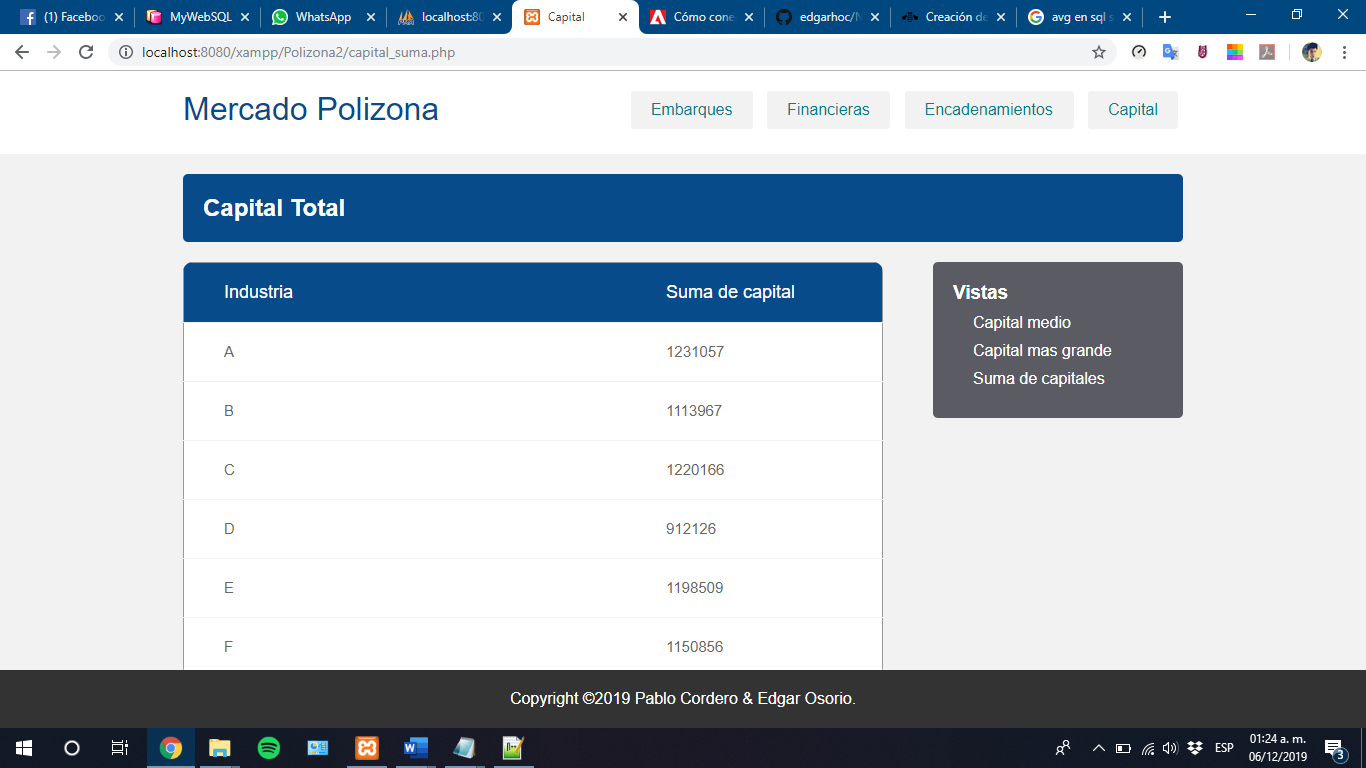
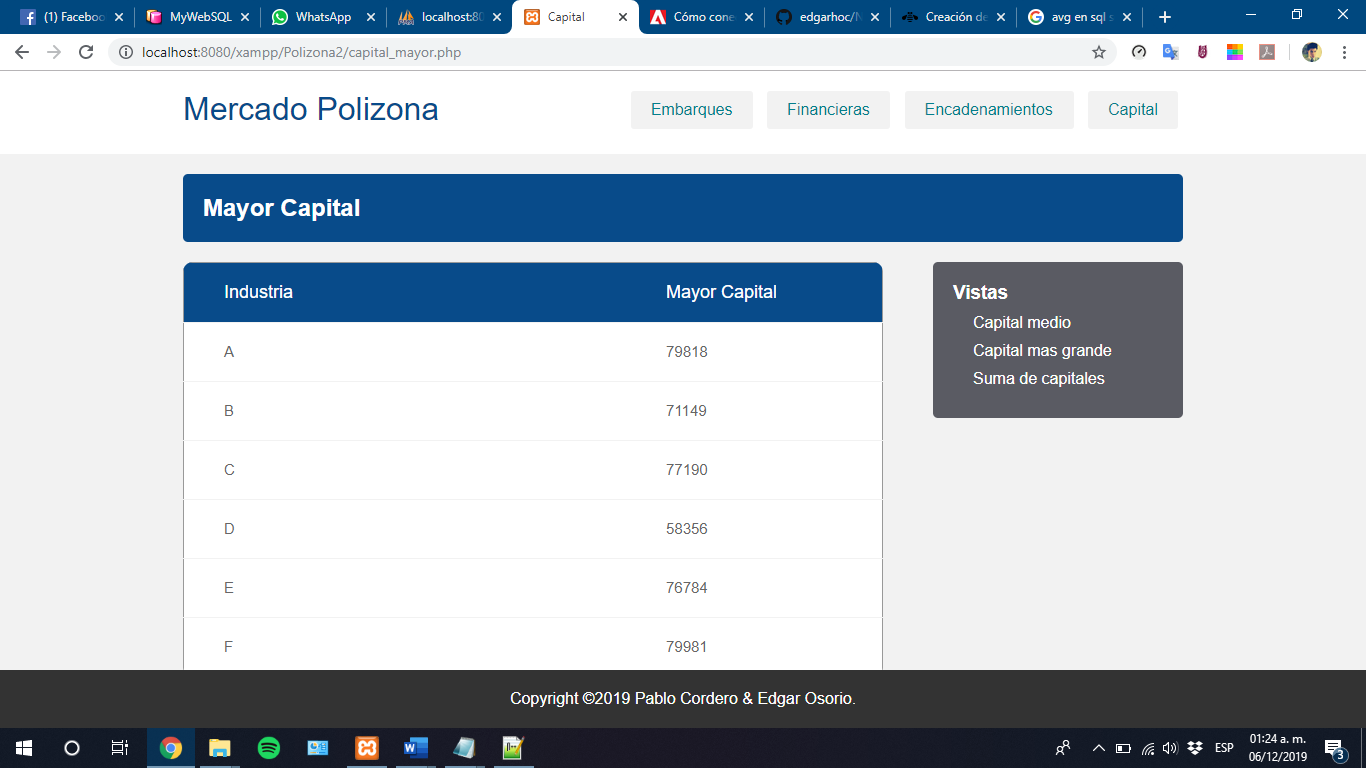
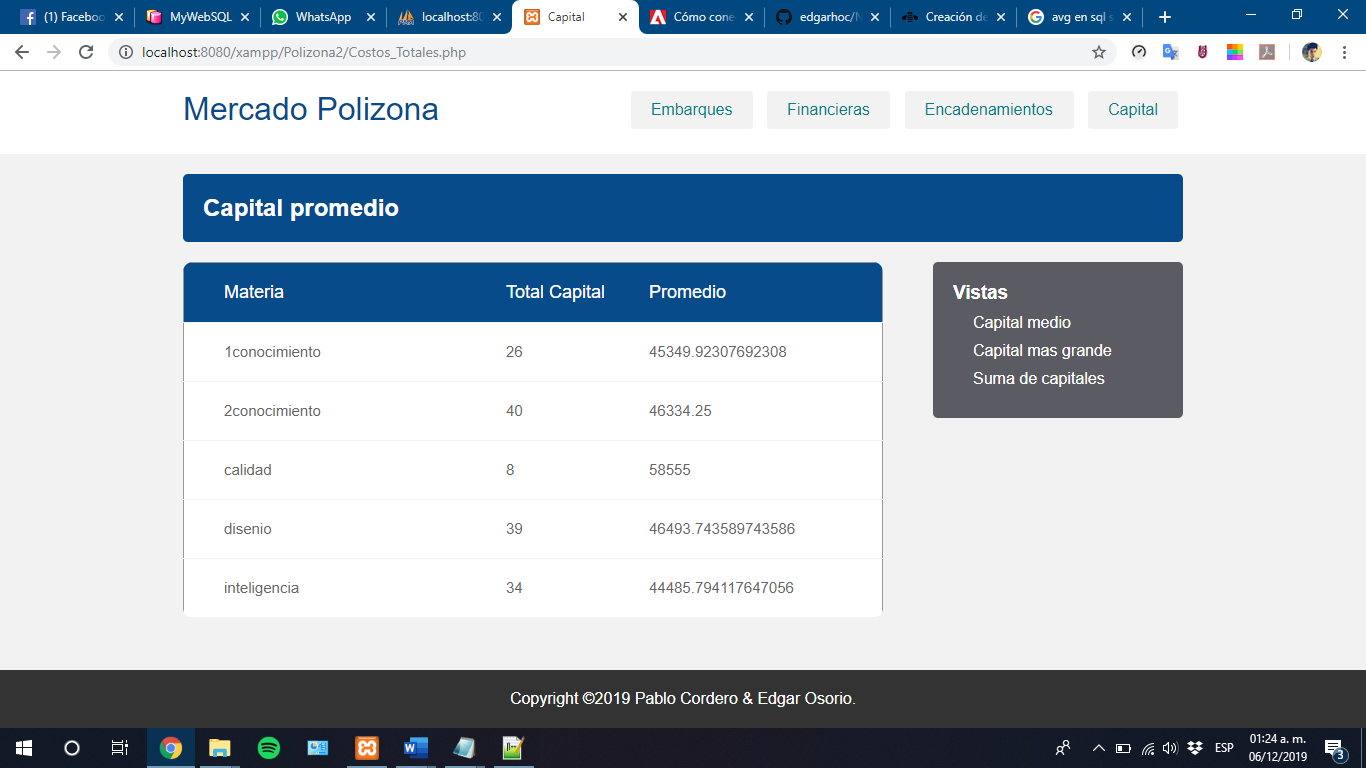
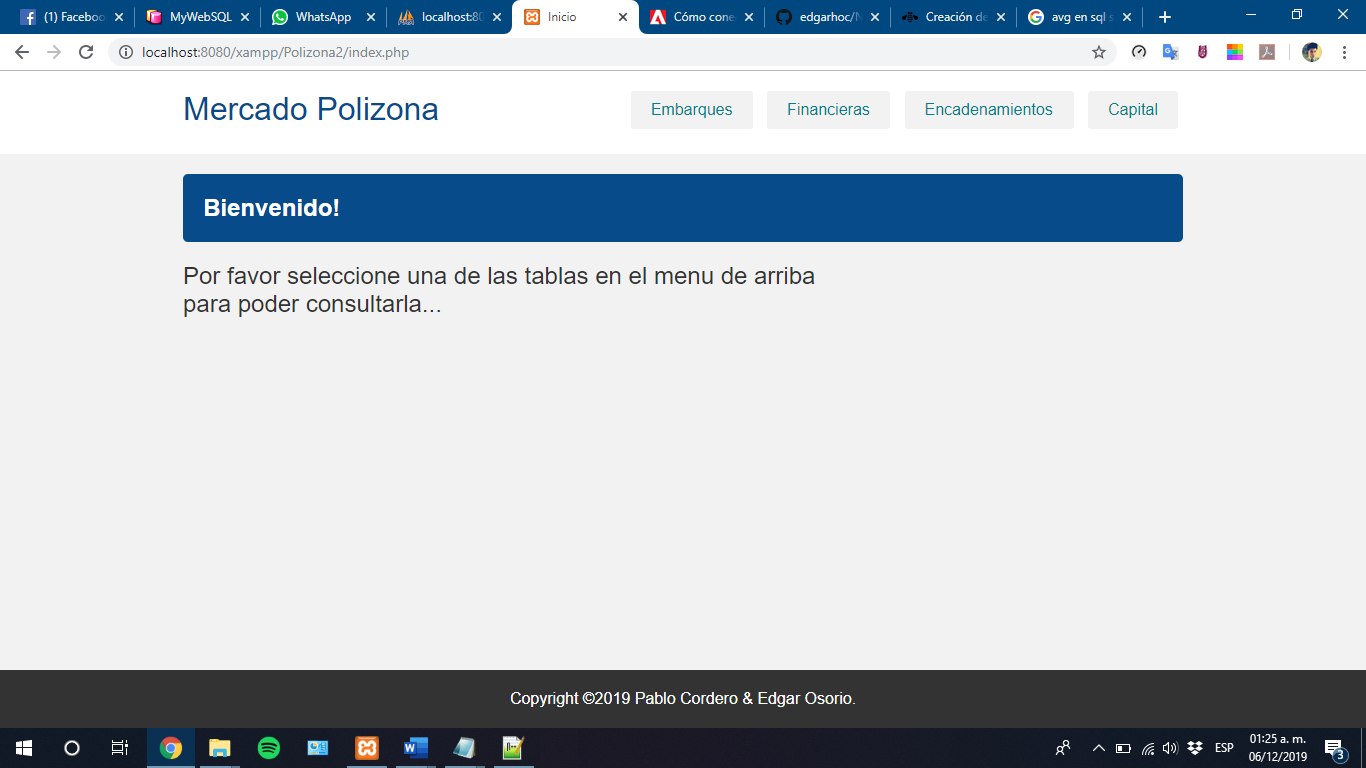
Las ventajas JSON sobre XML como formato de intercambio de datos es que resulta mucho más sencillo escribir un analizador sintáctico. En la práctica, JSON se emplea habitualmente en entornos donde el tamaño del flujo de datos entre cliente y servidor es de vital importancia (de aquí su uso por Yahoo!, 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.

1. Diseño de la aplicación



* 1. Proceso de negocio

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).

### 

1. **Ventajas del uso de minería de datos**

1. La minería de datos descubre información que no se esperaba obtener. Como muchos modelos diferentes son usados, algunos resultados inesperados tienden a aparecer. Las combinaciones de distintas técnicas otorgan efectos inesperados que se transforma en un valor añadido a la empresa.

2. Enormes bases de datos pueden ser analizadas mediante la tecnología de data mining.

3. Los resultados son fáciles de entender: personas sin un conocimiento previo en ingeniería informática pueden interpretar los resultados con sus propias ideas.

4. Contribuye a la toma de decisiones tácticas y estratégicas para detectar la información clave.

5. Te permite encontrar, atraer y retener a los clientes. Reduce el riesgo de perder clientes: ofrecer promociones específicas o productos especiales para retenerlos.

6. Mejora la relación con el cliente: la empresa puede mejorar la atención al cliente a partir de la información obtenida.

7. Permite ofrecer a tus clientes los productos o servicios que necesitan.

8. Los modelos son confiables. Los modelos son probados y comprobados usando técnicas estadísticas antes de ser usado, para que las predicciones que se obtienen sean confiables y válidas.

9. En su mayoría, los modelos se generan y construyen de manera rápida. El modelado a veces se torna más fácil puesto que muchos algoritmos han sido probados previamente.

10. Abre nuevas oportunidades de negocios y ahorra costes a la empresa.

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.

### Usos de la minería de datos

**La minería de datos es muy útil en los siguientes dominios:**

**Análisis y gestión del mercado:**

**Perfil del cliente:** la minería de datos ayuda a determinar qué tipo de personas compran qué tipo de productos.

**Identificación de los requisitos del cliente:** la extracción de datos ayuda a identificar los mejores productos para diferentes clientes. Utiliza la predicción para encontrar los factores que pueden atraer nuevos clientes.

**Análisis de mercado cruzado:** la minería de datos realiza asociaciones / correlaciones entre las ventas de productos.

**Objetivos Marketing:** la minería de datos ayuda a encontrar grupos de clientes modelo que comparten las mismas características, como intereses, hábitos de gasto, ingresos, etc.

**Patrón de compra del cliente:** la extracción de datos ayuda a determinar el patrón de compra del cliente.

**Análisis empresarial y gestión de riesgos**

**La minería de datos se utiliza en los siguientes campos del Sector Corporativo:**

**Planificación financiera y evaluación de activos:** involucra análisis y predicción de flujo de efectivo, análisis de reclamos contingentes para evaluar activos.

**Planificación de recursos:** implica resumir y comparar los recursos y los gastos.

**Competencia:** Implica el monitoreo de competidores y tendencias del mercado.

Detección de fraude

La minería de datos también se utiliza en los campos de servicios de tarjetas de crédito y telecomunicaciones para detectar fraudes.

En llamadas telefónicas fraudulentas, ayuda a encontrar patrones como el destino de la llamada, la duración de la llamada, la hora del día o la semana, etc. También analiza los patrones que se desvían de las normas esperadas.

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.

* 1. Ecuaciones y fórmulas

### Teorema de bayes

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) )**

1. Fragmentos de código

**Capital promedio**

<?php

$sql="SELECT \* from v\_sumatotal";

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

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

?>

<tr class="row">

<td class="cell"><?php echo $mostrar['materia'] ?></td>

<td class="cell"><?php echo $mostrar['total\_capital'] ?></td>

<td class="cell"><?php echo $mostrar['promedio'] ?></td>

</tr>

**Mayor capital**

<?php

$sql="SELECT \* from mayor\_capital";

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

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

?>

<tr class="row">

<td class="cell"><?php echo $mostrar['industria'] ?></td>

<td class="cell"><?php echo $mostrar['Mayor\_capital'] ?></td> </tr>

**Capital Total**

<?php

$sql="SELECT \* from sumacapital";

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

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

?>

<tr class="row">

<td class="cell"><?php echo $mostrar['industria'] ?></td>

<td class="cell"><?php echo $mostrar['Mayor\_capital'] ?></td>

</tr>

**SUPER QUERY**

VIEW SUMA DE CAPITAL

create view v\_sumatotal as(

select materia, count(capital\_social) as total\_capital, avg(capital\_social) as promedio

from financieras

group by materia

);

VIEW DE MAYOR CAPITAL

create view Mayor\_capital as(

select industria, max(capital\_social) as Mayor\_capital

from financieras

group by industria

);

VIEW SUMA TOTAL DE CAPITAL

Create view sumacapital as(

select industria, sum(capital\_social) as Mayor\_capital

from financieras

group by industria

);

1. Conclusiones y trabajos futuros

La tendencia del mundo entero se ha centrado alrededor de los Smartphones y sus funcionalidades. Estos pequeños dispositivos nos acompañan a todas horas y posiblemente sean los aparatos electrónicos que más utilizamos durante el día.

Las apps han cambiado nuestra forma de movernos por la ciudad, son vitales para una mayor comodidad, eficiencia y rendimiento.

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; ya que una gran ventaja de tener esto en una app nos proporciona datos fáciles de entender, porque, aunque sean personas sin el conocimiento previo a la informática, pueden interpretar los resultados con sus propias ideas.

En su mayoría, los modelos se generan y construyen de manera rápida, abriendo así nuevas oportunidades de negocios y ahorrar costes a las empresas o emprendedores, porque si de tecnología se trata, siempre hay que estar un paso adelante.

* 1. Citas y referencias

[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>

4.2 Referencias

1. Baldonado, M.; Chang, C.-C.K.; Gravano, L.; Paepcke, A.: The Stanford Digital Library Metadata Architecture. *Proceedings of the 3rd International Conference on Web Information System and Technologies* (WEBIST 2007), pp.108–121 (1998)

2. Bruce, K.B.; Cardelli, L.; Pierce, B.C.: Comparing Object Encodings. Abadi, M.; Ito, T. (Eds): Springer-Verlag, pp. 415–438 (1997)

3. van Leeuwen, J.: Plability in Actions Videogames. Gamasutra Game Developer. http://gamasutra.net/playability.html. Accedido el 13 de Febrero de 2008