

Estudio comparativo de software para la aplicación de minería de datos en inteligencia de negocios

Patricio Celi¹ 

¹ Instituto Tecnológico Quito

patricio.celi@itq.edu.ec

Corresponsal: patricio.celi@itq.edu.ec; Telf.: +593 999749541

Resumen: El manejo de grandes volúmenes de información en empresas ha sido crucial, impulsando la adopción de la inteligencia de negocios (BI) para decisiones basadas en datos, el avance tecnológico ha generado cantidades masivas de datos, desafiando el análisis convencional y dando lugar al Big Data y la Ciencia de Datos, las herramientas como redes sociales, movilidad, computación en la nube y el Internet de las Cosas han sentado las bases para estos avances, además comprender las herramientas de software para minería de datos se vuelve vital para optimizar objetivos estratégicos y mejorar la toma de decisiones empresariales, en este artículo se abordó la recolección de información, problemas, comparativas de herramientas y llegó a conclusiones relevantes para la gestión eficiente de datos en las organizaciones.

Palabras claves: Inteligencia de negocios, Extracción de información, Big Data, Ciencia de Datos, Minería de datos, Toma de decisiones basada en datos.

Abstract: The management of large volumes of information in companies has been crucial, driving the adoption of business intelligence (BI) for data-driven decisions, technological advancement has generated massive amounts of data, challenging conventional analysis and giving rise to Big Data and Data Science, tools such as social networks, mobility, cloud computing and the Internet of Things have laid the foundation for these advances, cloud computing and the Internet of Things have laid the foundation for these advances, also understanding data mining software tools becomes vital to optimize strategic objectives and improve business decision making, in this article we addressed the collection of information, problems, comparative tools and reached relevant conclusions for efficient data management in organizations.

Keywords: Business Intelligence, Information Extraction, Big Data, Data Science, Data Mining, Data-driven Decision Making.



Cita: Celi. Estudio comparativo de software para la aplicación de minería de datos en inteligencia de negocios. Revista DOXA ITQ, 2(1), 002.

<https://itq.edu.ec/estudio-comparativo-de-software-para-la-aplicacion-de-mineria-de-datos-en-inteligencia-de-negocios-2-1-2/>

Recibido: 26/02/2024

Aceptado: 22/04/2024

Publicado: 15/05/2024

Santiago del Castillo G., M.Sc.
Editor en jefe, Revista DOXA ITQ
Quito, Ecuador

Nota del editor: La Revista DOXA ITQ mantiene una posición neutral con respecto a cualquier reclamo legal que pueda surgir del contenido publicado. La responsabilidad de la información recae enteramente en los autores.

1. Introducción

En la actualidad, el manejo de grandes volúmenes de información representa un desafío significativo para las empresas, ya que esta información se ha convertido en uno de los activos intangibles más valiosos de su patrimonio. Para abordar esta complejidad, resulta imperativo el uso de la inteligencia de negocios (BI). Esta disciplina combina análisis de negocios, minería de datos, visualización de datos, herramientas e infraestructura de datos, junto con las mejores prácticas, para asistir a las organizaciones en la toma de decisiones fundamentadas en datos concretos.

A finales del siglo XX surgió el término KDD (Knowledge Discovery in Databases), que abarca el proceso completo de extracción e interpretación de información. Este proceso es esencial para preparar los datos y extraer resultados significativos. (Martínez et al., n.d.)

El desarrollo tecnológico ha sido el catalizador para la sistematización de procesos empresariales, generando cantidades masivas de información que desafían el análisis con software convencional. Es aquí donde entran en juego algoritmos especializados para el estudio de estos datos a gran escala.

Es importante destacar que, a medida que el tiempo avanza, han surgido diversas innovaciones tecnológicas a nivel mundial que han sentado las bases para el Big Data y la Ciencia de Datos. Entre estas innovaciones se encuentran los medios sociales, la movilidad a través de dispositivos como teléfonos inteligentes y tabletas, las aplicaciones móviles, la Computación en la Nube y el Internet de las Cosas. (Big Data, Análisis de grandes volúmenes de datos en organizaciones - Luis Joyanes Aguilar - Google Libros, s/f)

En este contexto, comprender las características, ventajas y desventajas de las herramientas de software utilizadas para el análisis en minería de datos, que están en tendencia, se vuelve fundamental para las empresas, esto les permite iniciar un camino hacia la optimización de sus objetivos

estratégicos, mejorar la toma de decisiones tácticas y estratégicas, así como enfocar sus esfuerzos en la realización de su misión y visión empresarial. El objetivo principal radica en interpretar estos datos en busca de patrones o relaciones, utilizando técnicas de aprendizaje automático, estadísticas, bases de datos, técnicas de representación del conocimiento, razonamiento basado en casos, entre otras. (Martínez et al., s/f)

Dada la diversidad en la forma en que las organizaciones gestionan su información, es crucial reconocer que no existe una solución tecnológica única que garantice el éxito. Es por eso por lo que es vital explorar las alternativas más adecuadas para cada contexto específico en términos de minería de datos.

Este análisis se encuentra establecido de la siguiente manera en la sección posterior abarcaremos el análisis de la información recolectada, la problemática que abarca esta corriente, la comparativa y la discusión de los resultados, para finalizar con las conclusiones acerca de esta comparativa.

2. Materiales y métodos

El autor (Castorena Peña et al., 2018) menciona: "Hoy en día existen diferentes herramientas tecnológicas de minería de datos que sirven a las organizaciones como soporte para la toma de decisiones, pues permiten identificar comportamientos o patrones sobre un determinado bien o servicio."

(Quesada López et al., s/f) afirma: "En los últimos años, una gran cantidad de técnicas de minería de datos y de aprendizaje máquina han sido utilizadas para la construcción de modelos de estimación del esfuerzo de desarrollo del software"

(Camana, 2016) sostiene que: "La acumulación masiva de datos por años, en empresas u organizaciones ha dado lugar al uso del análisis de datos, por ello; se han requerido de herramientas de alto rendimiento capaces de procesar ingentes cantidades de datos. Para llevar a cabo tal efecto se usan técnicas especializadas, identificadas como algoritmos que son quienes realizan el trabajo sobre el conjunto de datos."

En el ámbito de Inteligencia de negocios (Erazo et al., 2020) indican que: “Desde el punto de vista empresarial, los sistemas de información facilitan la recopilación, administración, procesamiento, análisis y creación de estrategias con la finalidad de cumplir con los objetivos y metas organizacionales, cabe puntualizar que los SI, comprenden tanto equipos informáticos, tecnológicos, recursos materiales, económicos y humanos que se utilizan para el cumplimiento operacional, por otra parte, un factor importante de resaltar es el conocimiento en el manejo de los sistemas tecnológicos de información, la identificación del sistema adecuado para la empresa y el uso correcto del sistema facilitarán obtener resultados informativos eficientes, a tiempo y confiables”.

Además (Ismael, 2020) en relación con la toma de decisiones indica lo siguiente: “La importancia de las decisiones correctas, efectivamente, previo a una toma de decisiones se requiere la suficiente información, misma que debe ser basada en datos certeros para que los administradores puedan optar por un correcto criterio de decidir. Cuando la información que disponga el gerente es escasa o se encuentre basada en datos poco confiables, la toma de decisiones se convierte en limitada. Otras causas que limitan una decisión oportuna tienen que ver con el entorno externo del negocio como por ejemplo la situación política, económica, el medio cultural y social, el mercado local, criterios de moda y preferencias del comprador, entre otras variables.”

Para complementar la toma de decisiones (Rozo, 2013) indaga acerca de las etapas en el proceso de toma de decisiones en la que menciona: “La toma de decisiones es un proceso en lugar de un simple acto de escoger entre diferentes alternativas. La figura siguiente muestra el proceso de toma de decisiones como una serie de ocho pasos que comienza con la identificación del problema, los pasos para seleccionar una alternativa que pueda resolver el problema, y finalizan con la evaluación de la eficacia de la decisión.”

Después de haber entendido los aspectos importantes de la investigación vamos a

detallar las características principales de cada herramienta de trabajo, al final realizaremos una comparativa entre ellas para concluir cual es la mas adecuada para tipos de trabajo en escenarios.

3. Herramientas de Estudio

Pentaho es una plataforma de inteligencia de negocios de código abierto que ofrece un conjunto completo de herramientas para la integración de datos, el análisis de negocios, la generación de informes y la minería de datos. Incluye componentes como Pentaho Data Integration (PDI) para la extracción, transformación y carga (ETL) de datos, Pentaho Business Analytics para el análisis y visualización de datos, y Pentaho Data Mining para técnicas avanzadas de minería de datos.

Dentro de las ventajas importantes de este software se encuentran:

- **Código Abierto:** Pentaho es una plataforma de inteligencia de negocios de código abierto, lo que significa que ofrece flexibilidad y la posibilidad de personalización según las necesidades específicas de cada organización.
- **Integración Completa de Datos:** Pentaho ofrece una integración completa de datos, incluyendo la extracción, transformación y carga (ETL), lo que facilita la preparación de datos para el análisis.
- **Escalabilidad:** Es escalable y puede manejar grandes volúmenes de datos, lo que lo hace adecuado para empresas con necesidades de análisis de datos a gran escala.

Tableau es una herramienta líder en visualización de datos que permite a los usuarios crear gráficos interactivos, dashboards y paneles de control de manera intuitiva. Ofrece una amplia gama de capacidades para el análisis exploratorio de datos, la creación de informes personalizados y la colaboración en equipos además Tableau se destaca por su facilidad de uso y su capacidad para conectar y visualizar datos de múltiples fuentes.

Esta herramienta se destaca principalmente por:

- **Interfaz Intuitiva:** Tableau se destaca por su interfaz intuitiva y fácil de usar, lo que permite a los usuarios crear visualizaciones complejas de datos de manera sencilla.
- **Análisis Visual Avanzado:** Ofrece capacidades avanzadas de análisis visual que permiten identificar patrones y tendencias en los datos de manera efectiva.
- **Colaboración en Tiempo Real:** Facilita la colaboración en tiempo real, lo que permite a los equipos trabajar juntos en la exploración y análisis de datos de manera colaborativa.

Power BI es una plataforma de análisis de negocios desarrollada por Microsoft que permite a las organizaciones transformar datos en información significativa y visualmente atractiva. Ofrece capacidades de creación de informes interactivos, visualizaciones avanzadas, análisis predictivo y colaboración en tiempo real. Power BI se integra fácilmente con otras herramientas y servicios de Microsoft, como Excel y Azure, para ofrecer una solución completa de inteligencia empresarial.

La herramienta estrella de Microsoft se destaca principalmente por:

- **Integración con Microsoft:** Al ser desarrollado por Microsoft, Power BI ofrece una integración perfecta con otras herramientas y servicios de Microsoft, como Excel y Azure, lo que simplifica el flujo de trabajo para los usuarios que ya utilizan productos de Microsoft.
- **Análisis Predictivo:** Power BI ofrece capacidades avanzadas de análisis predictivo que permiten a las empresas predecir tendencias futuras y tomar decisiones informadas.
- **Visualizaciones Interactivas:** Ofrece visualizaciones interactivas y personalizables que permiten a los usuarios explorar y presentar datos de manera dinámica.

El análisis comparativo es un proceso fundamental en toda actividad cognitiva, busca establecer conexiones causales entre diferentes fenómenos o causales e identificar elementos únicos que expliquen la existencia de un fenómeno considerado único e irrepetible.

En este estudio, el análisis comparativo se centra en las características clave de tres herramientas de inteligencia de negocios: Power BI, Pentaho y Tableau. El objetivo es comprender cómo cada una de estas herramientas aborda y resuelve los desafíos en el análisis de datos y la toma de decisiones empresariales.

4. Discusión

Ahora nos adentramos en el análisis y la interpretación de los resultados obtenidos a partir de la comparativa entre las herramientas de inteligencia de negocios evaluadas, esta sección es crucial para comprender la relevancia práctica de las diferencias y similitudes identificadas entre las herramientas, así como para destacar las implicaciones de estos hallazgos en el contexto empresarial.

A lo largo de esta discusión, exploraremos las fortalezas y limitaciones de cada herramienta, identificaremos áreas de mejora y reflexionaremos sobre su aplicabilidad y utilidad en entornos reales de toma de decisiones basadas en datos.

Tabla 1
Comparativa de Herramientas

criterio	Power BI	Pentaho	Tableau				
Fuentes de entrada	Orígenes de datos incluye formatos como Excel, SQL Server, Access, CSV, entre otros.	Las herramientas tienen una conectividad amplia con diversas fuentes de datos o plataformas, con una compatibilidad cercana al 95%.	Múltiples orígenes de datos como Excel, SQL Server, Access, CSV, y archivos planos, es necesario destacar que Tableau requiere datos que estén más limpios y estructurados para un rendimiento óptimo.	Exportación de la base de datos	Permite exportar datos en diferentes extensiones. (Excel, XML, CSV, JSON, Archivo Plano, etc.)	Permite exportar datos en diferentes extensiones. (Excel, XML, CSV, Archivo Plano, etc.)	Permite exportar datos en diferentes extensiones. (Excel, XML, CSV, JSON, Archivo Plano, etc.)
	Fuentes de salida	Exporta información a diversas plataformas, y es relevante destacar que no solo envía datos analizados del pasado y presente, sino que también permite la predicción de datos futuros.	Pentaho organiza las salidas en diferentes etapas, que incluyen informes, tableros de KPI, integración de procesos y flujos de datos, así como ajustes de la solución. Después de pasar por todo este proceso, la información se envía a diversas plataformas de manera lineal o visual sin restricciones.		Facilita la exportación de datos en varias extensiones y también ofrece colaboración en tiempo real entre equipos para análisis de datos, compitiendo con otras plataformas. Estos datos pueden enviarse a servidores reales o publicarse para el consumo del equipo.	Creación de KPI	Facilita la creación de Indicadores Clave de Rendimiento.
Creación de dashboards		Creación de paneles de control fáciles de entender para el usuario final. Además, permite la implementación de diversos tipos de visualizaciones y gráficos predefinidos, así como la creación de nuevas métricas para mejorar las	Visualización de datos en línea gradual, analizarlos y desplegarlos de forma estructurada.	Facilita la creación rápida de paneles y elementos visuales mediante arrastrar y soltar, aunque su flexibilidad y dinamismo en la creación de informes es algo inferior en comparación con otros sistemas.	Gráficos estadísticos		Creación de diferentes tipos de visualizaciones, tales como gráficos de columnas, gráficos lineales, gráficos combinados de línea y columna, gráficos de puntos de referencia, gráficos de anillos y circulares, medidores, tarjetas, entre otros.
	Costo			Varios costes desde gratuito hasta planes empresariales		Gratuito es código Abierto	Varios planes hasta el premium, no existe gratuidad.

Facilidad de manejo	La herramienta está diseñada para usuarios principiantes y sin experiencia en análisis de datos. Aunque su interfaz es amigable, la curva de aprendizaje puede ser rápida, pero varía según el usuario, no se necesitan conocimientos previos para utilizarla.	Pentaho ofrece soluciones que abarcan diversas áreas como Big Data, integración de datos, análisis empresarial, análisis de Internet de las cosas (IoT) y análisis empresarial en la nube, sin embargo, hay que tener bases sobre la comprensión de los datos y sus entradas y salidas.	Es una herramienta que se centra en su facilidad de uso, lo que la convierte en una de las más fáciles de utilizar en su categoría, según lo afirmado por expertos. Su enfoque principal está en mejorar la experiencia del usuario y hacer que la herramienta sea accesible incluso para aquellos sin experiencia técnica previa.
Confiabilidad	Completamente confiable, muchos expertos auguran una calificación de 10 en confiabilidad, ya que cuenta con el paraguas de Microsoft.	Completamente confiable, controla el acceso a los datos y análisis de esta con roles dinámicos por tal la seguridad que nos brinda es en un índice bastante alto.	Confiable, calificación 8 sobre 10, en relación con las plataformas de su misma especie.

Nota. en esta tabla se presenta una comparación de las herramientas disponibles en cada plataforma. Elaboración propia.

Al discutir sobre el uso de herramientas de inteligencia en el contexto del Big Data es crucial para entender cómo las empresas modernas están optimizando sus procesos y toma de decisiones, donde estas herramientas son fundamentales para el análisis y estudio de datos dentro del ámbito de la inteligencia empresarial, lo que lleva a la reducción de costos y la generación de mejores ingresos.

En el siglo XXI, las empresas reconocen la importancia de invertir en herramientas de inteligencia que les permitan analizar su propia data de manera efectiva y estructurada hoy en día se han convertido en un insumo vital para la toma de decisiones estratégicas, ya que proporcionan información clave proveniente de diferentes sistemas de

comunicación dentro de la organización, como contabilidad, finanzas, logística, educación, talento humano, entre otros.

La finalidad de ellas es presentar a la alta gerencia un análisis detallado de los indicadores del negocio que impactan directamente en la entidad, donde en un periodo de tiempo determinado, lo que permite establecer planes de acción y tomar decisiones empresariales que impulsen la mejora económica y la eficiencia del negocio.

Por tal que el estudio comparativo de estas herramientas es crucial para saber elegir cual es la mas adecuada para su modelo de negocio.

La discusión basada en el estudio comparativo de herramientas de inteligencia de negocio como Power BI, Pentaho y Tableau revela aspectos clave que las empresas deben considerar al elegir la plataforma más adecuada para sus necesidades.

En primer lugar, se observa que cada herramienta tiene su enfoque particular en términos de facilidad de uso, integración con otras plataformas, capacidades de análisis y escalabilidad. Power BI destaca por su integración con el ecosistema de Microsoft, lo que lo hace atractivo para usuarios familiarizados con herramientas de Microsoft. Por otro lado, Pentaho ofrece una amplia conectividad con diversas fuentes de datos y es altamente personalizable, siendo una opción robusta para entornos heterogéneos, Tableau se destaca por su facilidad de uso y capacidad para crear visualizaciones interactivas de manera rápida y sencilla, lo que lo convierte en una opción popular entre usuarios no técnicos.

En cuanto a la escalabilidad, Pentaho y Tableau son altamente escalables y pueden manejar grandes volúmenes de datos, siendo opciones viables para empresas de mediano a gran tamaño. Por su parte, Power BI ofrece diferentes planes que se adaptan a las necesidades de diferentes tipos de empresas, desde pequeñas hasta corporaciones.

En términos de precios, cada herramienta tiene sus propios modelos de precios y opciones de licencia. Power BI ofrece planes

gratuitos y de pago, Pentaho al ser de código abierto no tiene costos de licencia directos, pero puede requerir inversión en servicios de soporte, mientras que Tableau tiene precios estándar que varían según el tipo de usuario y la funcionalidad requerida.

5. Conclusiones

En conclusión, la elección entre Power BI, Pentaho y Tableau dependerá de las necesidades específicas de cada empresa en términos de integración, análisis, escalabilidad y presupuesto. Es crucial evaluar detenidamente cada opción para seleccionar la herramienta que mejor se alinee con los objetivos y requerimientos de negocio de la organización.

La elección entre estas herramientas debe considerar aspectos como la facilidad de uso, la capacidad de integración con sistemas existentes, las necesidades de análisis de datos, la escalabilidad y los costos a largo plazo y es fundamental encontrar un equilibrio entre funcionalidad, accesibilidad y presupuesto para maximizar el valor obtenido de la herramienta seleccionada.

Aunque Power BI puede ser más adecuado para empresas que ya utilizan herramientas de Microsoft, Pentaho ofrece una mayor personalización y adaptabilidad a entornos heterogéneos, mientras que Tableau se destaca por su capacidad de crear visualizaciones interactivas de manera rápida y sencilla.

Contribución del autor: El autor ha contribuido en todos los apartados de la investigación.

Financiamiento: El autor financió totalmente el estudio.

Conflictos de intereses: El autor declara no tener ningún conflicto de intereses.

Referencias

Big Data, Análisis de grandes volúmenes de datos en organizaciones - Luis Joyanes Aguilar - Google Libros. (s/f). Recuperado el 8 de mayo de 2022, de <https://books.google.es/books?hl=es>

<https://doi.org/10.15503/revista.tecnologica.v29n1.a05>
&lr=&id=1GywDAAAQBAJ&oi=fnd&pg=PT6&dq=big+data+articulo+cientifico&ots=_XS6N-2d1S&sig=AzNXe7X2Um-BixQ73AWeoa8Flhg#v=onepage&q=big%20data%20articulo%20cientifico&f=false

Camana, R. (2016). Potenciales Aplicaciones de la Minería de Datos en Ecuador | Camana | Revista Tecnológica - ESPOL. *Revista Tecnológica ESPOL*, 29, 5–6.

<http://200.10.150.204/index.php/tecnologica/article/view/464>

Castorena Peña, J. A., Silva Avila, A. E., Domínguez Lugo, Alma Jovita, Rodríguez Montelongo, & Diana Laura. (2018). El uso de herramientas tecnológicas de minería de datos en el análisis de datos climatológicos. *RECI*, 7(Ene-Jun), 5–5.

<https://biblat.unam.mx/es/revista/reci/articulo/el-uso-de-herramientas-tecnologicas-de-mineria-de-datos-en-el-analisis-de-datos-climatologicos>

Erazo, J., Narváez, C., & Cordero, D. (2020, diciembre). Soluciones corporativas de inteligencia de negocios en las pequeñas y medianas empresas - Dialnet. *Revista Arbitrada Interdisciplinaria KOINONIA*, 6–6. <https://dialnet.unirioja.es/servlet/articulo?codigo=7439114>

Ismael, E., Carlos, J., Ivonne, C., & Marcelo, D. (2020). *Soluciones corporativas de inteligencia de negocios en las pequeñas y medianas empresas*. <https://doi.org/10.35381/r.k.v5i10.703>

Martinez, C., Mart Inez Alarcón, E., & Rivero, R. (s/f). *Ciencia de datos. Técnicas analíticas y aprendizaje estadístico*. Recuperado el 8 de mayo de 2022, de www.alfaomega.com.co

Quesada López, C., Murillo-Morera, J., & Jenkins, M. (s/f). *Un estudio comparativo de técnicas de minería de datos y aprendizaje máquina para la estimación del esfuerzo utilizando puntos de función*.

Rozo, F. E. C. (2013). INDICADORES DE GESTIÓN: TOMA DE DECISIONES BASADA EN INTELIGENCIA DE

NEGOCIOS. *Tecnología Investigación y
Academia*, 1(2).
[https://revistas.udistrital.edu.co/index
.php/tia/article/view/4639](https://revistas.udistrital.edu.co/index.php/tia/article/view/4639)

