



Asesorías y Tutorías para la Investigación Científica en la Educación Puig-Salabarría S.C.
José María Pino Suárez 400-2 esq a Lerdo de Tejada. Toluca, Estado de México. 7223898475
 RFC: ATI120618V12

Revista Dilemas Contemporáneos: Educación, Política y Valores.
<https://www.dilemascontemporaneoseduccionpoliticaayvalores.com/>

Año: XII Número: 3 Artículo no.: 14 Período: 1 de mayo al 31 de agosto del 2025

TÍTULO: Análisis comparativo de herramientas para explorar grandes volúmenes de datos: propuesta metodológica.

AUTORES:

1. Máster. Zully Kristel Guzmán Caraveo.
2. Dra. Martha Patricia Silva Payró.
3. Dr. Guillermo de los Santos Torres.

RESUMEN: El trabajo, aquí presentado, resulta de un análisis comparativo realizado entre distintas herramientas para explorar grandes volúmenes de datos y cuáles son las herramientas más adecuadas para ser utilizadas en la realización de análisis de los datos. La comparación se realizó con diversas herramientas o *software* que son utilizados para efectuar exploraciones en un gran conjunto de datos almacenados en base de datos, las cuales son de *software* libre, comerciales y servicios en nube; éstas últimas proporcionan módulos o servicios para dicho análisis. El análisis comparativo se efectuó utilizando parámetros de calidad a través de la metodología de calidad de *software* bajo el acrónimo de FURPS.

PALABRAS CLAVES: análisis, herramientas, comparativa, exploración, volúmenes de datos.

TITLE: Comparative analysis of tools for exploring large volumes of data: methodological proposal.

AUTHORS:

1. Master. Zully Kristel Guzmán Caraveo.
2. PhD. Martha Patricia Silva Payró.
3. PhD. Guillermo de los Santos Torres.

ABSTRACT: The work presented here results from a comparative analysis carried out between different tools to explore large volumes of data and determine the most outstanding tools for data analysis. The comparison was made with various tools or software that are used to explore a large set of data stored in databases, which are free software, commercial, and cloud services; the latter provide modules or services for such analysis. The comparative analysis was performed using quality parameters through the software quality methodology under the acronym FURPS.

KEY WORDS: analysis, tools, comparative, exploration, data volumes.

INTRODUCCIÓN.

Con el transcurso de los años, las actividades y el desarrollo de nuevas tecnologías han aumentado de forma considerable el almacenamiento de información, donde todo ese flujo de información recolectado ha permitido satisfacer las necesidades diarias de las organizaciones, pero ha presentado un problema inherente en las capacidades humanas para analizar y transformar la información en conocimiento útil y relevante que apoye a la toma de decisiones (Ballesteros *et al.*, 2013).

Este tipo de problemas comúnmente se han presentado en los grandes sistemas conocidos como bases de datos, en los cuales se almacena una gran cantidad de información; estos datos con frecuencia están integrados por contenido valioso y pueden ser vistos como una recopilación masiva de información para uso de la mayoría de las organizaciones, aun con el uso de herramientas estadísticas clásicas para la manipulación, el análisis y la extracción de información; esta tarea es casi imposible, y por esta razón, se ha motivado el empleo de técnicas y herramientas para el análisis de los datos (Ballesteros *et al.*, 2013).

Ballesteros *et al.* (2013) y Parada (2024) describen a la minería de datos como el análisis de datos, que es posible por la extracción de conocimiento e información, gracias a los diferentes técnicas y tecnologías que permiten analizar los datos, en búsqueda de comportamientos, tendencias o variaciones que expliquen el comportamiento de dichos datos almacenados.

En la actualidad, se han implementado las técnicas y herramientas para minería de datos o análisis de datos, las cuales posibilitan explorar, descubrir y analizar nueva información con base en reglas y patrones a partir de un gran volumen de datos almacenados; es por ello, que existen diferentes herramientas, con las cuales se pueden aplicar la minería de datos; estas herramientas pueden ser de licencias libres o comerciales, servicios en nube, lenguajes de programación, entre otras.

Para el desarrollo del presente trabajo, fue necesario realizar una búsqueda y documentación de las herramientas existentes para la exploración de los datos; es decir, como se conoce actualmente, la minería de datos. En esta búsqueda se encontraron diferentes herramientas, las cuales se clasificaron en cuatro categorías:

- 1) Herramientas *framework* para análisis de datos.
- 2) Lenguajes de programación para análisis de datos.
- 3) Herramientas de Microsoft para análisis de datos.
- 4) Servicios en nube para análisis de datos.

En la tabla 1 se muestran las herramientas encontradas al momento de realizar la investigación sobre las herramientas y *software* que son utilizados para el análisis de datos. Se encontraron un total de 16 herramientas, las cuales se clasificaron de acuerdo con las categorías anteriormente mencionadas (ver tabla 1).

Tabla 1. Clasificación de las herramientas para el análisis de datos.

Clasificación	Nombre	Características
Herramientas <i>framework</i> para análisis de datos	<i>RapidMiner</i>	Permite el desarrollo de procesos de análisis de datos mediante el encadenamiento de operadores, a través de una Interfaz Gráfica de Usuario (GUI). Permite realizar análisis estadísticos, minería de datos y análisis predictivos (Vargas <i>et al.</i> , 2016).
	<i>Orange</i>	Esta herramienta cuenta con un fácil, potente, rápido y versátil <i>front end</i> de programación visual para el análisis exploratorio de datos y visualización, para preprocesamiento de datos, características de

Clasificación	Nombre	Características
		puntuación y filtrado, modelado, evaluación del modelo y técnicas de exploración (García, 2013).
	<i>Weka</i>	Es de libre distribución (licencia GPL) y destaca por la cantidad de algoritmos que presenta, así como por la eficiencia de estos. Proporciona herramientas para la realización de tareas propias de minería de datos, la visualización y permite la programación en JAVA de algoritmos (Rodríguez y Díaz, 2009).
Lenguajes de programación para análisis de datos	<i>Lenguaje R</i>	Es un lenguaje y entorno de programación, cuya característica principal es que forma un entorno de análisis estadístico para la manipulación de datos, cálculos y la creación de gráficos; con la particularidad de que es un <i>software GNU</i> ; es decir, de uso libre (Contreras <i>et al.</i> , 2010).
	Python	Es un lenguaje de programación de alto nivel, interpretado y multipropósito; cuenta con facilidades para la programación orientada a objetos, imperativa y funcional, por lo que se considera un lenguaje multiparadigmas (López <i>et al.</i> , 2019).
	Matlab	<i>Matlab</i> facilita la ciencia de datos con herramientas para acceder, organizar, depurar y analizar conjuntos de datos complejos de diversos campos, además permite la visualización de datos interactivos, contiene extensiones para el análisis a <i>big data</i> (MathWorks, s.f.a) Además, Matlab contiene una <i>suite</i> para realizar análisis y modelar los datos, donde se utilizan métodos estadísticos descriptivos para la visualización y agrupación de los datos, también de proporcionar una variedad de algoritmos de <i>Machine Learning</i> supervisados, semisupervisados y no supervisados (MathWorks, s.f.b)
Herramientas de Microsoft para análisis de datos	Microsoft SQL Server	Es un sistema de administración y análisis de base de datos relacionales, para soluciones de comercio electrónico, línea de negocio y almacenamiento de datos (Microsoft, 2014).
	Microsoft Excel	Análisis de datos en Excel permite comprender los datos a través de consultas de lenguaje natural que le permiten formular preguntas sobre los datos sin tener que escribir fórmulas complicadas, proporciona resúmenes visuales de alto nivel, tendencias y patrones (Microsoft, s.f.).
	Microsoft Analysis Services	Es un motor de datos analíticos, utilizado en el soporte de decisiones y análisis de negocios. Proporciona modelos de datos semánticos de nivel empresarial para informes comerciales y aplicaciones de clientes como <i>Power BI</i> , <i>Excel</i> , informes de <i>Reporting Services</i> y otras herramientas de visualización de datos (Microsoft, 2023).
Servicios en nube para análisis de datos	Oracle: Oracle Business Intelligence Server	Optimiza la simultaneidad y el paralelismo para hacer que el valor de las aplicaciones de BI esté al alcance de tantas personas como sea posible. Proporciona acceso a los datos y cálculos centralizados, creando esencialmente un gran canal a través del cual todos los usuarios pueden

Clasificación	Nombre	Características
		usar toda la información, de cualquier forma, dentro de la empresa (Oracle, s.f.).
	IBM: SPSS Modeler	<i>SPSS Modeler</i> es una solución líder en ciencia de datos visual y <i>machine learning</i> . Ayuda a las empresas a acelerar el tiempo para obtener valor para la empresa y lograr los resultados deseados acelerando las tareas operativas para los científicos de datos (IBM, s.f.).
	Intel: DAAL	Intel® <i>Data Analytics Acceleration Library</i> (Intel DAAL) es una solución de <i>software</i> que ofrece componentes básicos que cubren todas las etapas del análisis de datos, desde el preprocesamiento hasta la toma de decisiones. La versión beta de Intel DAAL ofrece compatibilidad con el lenguaje <i>Python</i> (Fedorov, 2016).
	Amazon: EMR	Es la plataforma de <i>big data</i> nativa en la nube líder del sector que permite que los equipos procesen grandes cantidades de datos de forma rápida y rentable a escala (Amazon Web Services [AWS], s.f.)
	Google: Google Cloud Datalab	Es una herramienta interactiva fácil de usar para la exploración, transformación, análisis y visualización de sus datos en Google Cloud Platgorm; utilizada para el aprendizaje automático, la exploración, análisis y la visualización de los datos. Permite ejecutar proyectos en la nube y escribir código para utilizar otros servicios de almacenamiento, mediante un amplio conjunto de bibliotecas creadas por Google y terceros (Google Cloud, 2023).
	Microsoft Azure Machine Learning	Es un servicio en la nube que permite acelerar y administrar el ciclo de vida de los proyectos de aprendizaje automático, el cual permite utilizar flujos de trabajo diarios como entrenamiento e implementación de modelos de minería de datos (Microsoft Azure, 2024).
	SAS: SAS Enterprise Miner	Es una solución de minería de datos que permite incorporar patrones inteligentes a los procesos de <i>marketing</i> , tanto operativos como estratégicos, provee acceso transparente a cualquier fuente de datos, incluyendo archivos planos, archivos jerárquicos, y los más importantes manejadores de bases de datos relacionales (Rodríguez y Díaz, 2009).

Fuente: Elaboración propia.

Por la gran variedad de herramientas, *software* y servicios de nube existentes actualmente para ser utilizadas en el análisis de grandes volúmenes de datos o *big data*, resulta indispensable efectuar comparativas y análisis sobre cuál herramienta o *software* se adapta a las necesidades que el usuario tenga para realizar un análisis de datos completo a una base de datos.

Dicho análisis comparativo de las herramienta o *software* debe basarse en los parámetros de calidad, asociados a las metodologías de calidad de *software*. Callejas-Cuervo *et al.* (2017) apunta a que el término de calidad de *software* se refiere al grado de desempeño de las principales características con las que debe cumplir un sistema computacional durante su ciclo de vida, las cuales de cierta manera garantizan al usuario o cliente que el sistema, herramienta o *software* es confiable, lo cual aumenta su satisfacción frente a la funcionalidad y eficiencia de este.

De acuerdo a lo anterior, una de las metodologías de calidad de *software* es la metodología bajo el acrónimo de FURPS, la cual fue desarrollada por Hewlett-Packard en el año 1987, y establece cinco factores de calidad que deben cumplir las herramientas o *softwares*, los cuales son: Funcionalidad (*Functionality*), Usabilidad (*Usability*), Confiabilidad (*Reliability*), Desempeño (*Performance*) y Soportabilidad (*Supportability*) o Capacidad de Soporte; además, cada factor proporciona criterios que son necesarios evaluar en las herramientas, como las características del programa, el tiempo de respuestas, eficiencia, entre otros.

Los atributos de calidad FURPS representan el objetivo de todo diseño de *software*; la cual se utilizará para analizar y ponderar las herramientas para el análisis de datos, para realizar la selección de ellas y saber cuál o cuáles son las que más destaquen durante el análisis. Cada atributo del modelo se define de la siguiente manera (Pressman, 2010):

- *Funcionalidad*. Se evalúa de acuerdo con el conjunto de características y capacidades del programa, la generalidad de las funciones que se entregan y la seguridad general del sistema.
- *Usabilidad*. Se evalúa tomando en cuenta factores humanos, la estética general, consistencia y documentación.
- *Confiabilidad*. Se evalúa con la medición de la frecuencia y gravedad de las fallas, la exactitud de los resultados, el tiempo medio para que ocurra una falla, la capacidad de recuperación ante ésta y lo predecible del programa.

- *Desempeño*. Se mide con base en la velocidad de procesamiento, tiempo de respuesta, consumo de recursos, rendimiento y eficiencia.
- *Soportabilidad*. Combina la capacidad del programa para ser ampliable (extensibilidad), adaptable y servicial, además que pueda probarse, ser compatible y configurable (capacidad de organizar y controlar los elementos de la configuración del *software*) y que cuente con la facilidad para instalarse en el sistema y se detecten los problemas.

Esta tecnología ha sido utilizada en otros estudios similares, para analizar las características y calidad de *software* de diferentes herramientas y programas; un caso de estudio realizado fue por Núñez *et al.* (2018) en el cual efectuaron una búsqueda de las principales tecnologías utilizadas para el desarrollo *web*, donde posteriormente ponderaron las herramientas con base en la experiencia de utilización y cumplimiento de cada uno de los factores que contempla la metodología; obteniendo como resultado que las herramientas como *Angular JS*, *React JS* y *Bootstrap* son las que obtuvieron mayor ponderación y son las destacadas para ser utilizadas para el desarrollo *web*.

Por lo anterior descrito, el objetivo principal del presente trabajo es la realización de un análisis comparativo entre herramientas para la realización del análisis de datos, empleando la metodología de calidad de *software FURPS*, con la cual se analizan las características y calidad de las herramientas previamente descritas, las cuales serán ponderadas de acuerdo con los factores y criterios establecidos por la metodología *FURPS* y así poder visualizar cuál(es) es(son) la(s) herramienta(s) más adaptable(s) para ser utilizada(s) en análisis de datos.

DESARROLLO.

Durante la investigación fue necesario realizar la búsqueda y documentación pertinente de las diferentes herramientas y *softwares* que existen para el análisis de los datos, con el motivo para realizar un análisis comparativo que ayude a identificar y seleccionar cuáles son las herramientas con una mejor calidad de *software*; esto utilizando la metodología *FURPS*, que se compone por cinco factores primordiales y diversos

criterios. En la tabla 2 se presentan los factores y criterios de calidad que conforman la metodología FURPS (ver tabla 2).

Tabla 2. Factores y criterios de la metodología FURPS.

Factores	Criterios
Funcionalidad	Características y capacidades del programa. Generalidad de las funciones. Seguridad del sistema.
Usabilidad	Factores Humanos. Factores estéticos. Consistencia de la interfaz. Documentación.
Confiabilidad	Frecuencia y severidad de fallos. Exactitud de las salidas. Tiempo medio de fallos. Capacidad de recuperación ante fallos. Capacidad de predicción.
Rendimiento	Velocidad de procesamiento. Tiempo de respuesta. Consumo de recursos. Rendimiento efectivo total. Eficacia.
Capacidad de Soporte.	Extensibilidad. Adaptabilidad. Capacidad de prueba. Capacidad de configuración. Compatibilidad. Requisitos de instalación.

Fuente: Constanzo et al. (2014).

De acuerdo con lo descrito anteriormente, la metodología FURPS ha sido empleada en investigaciones donde se busca identificar la calidad de las herramientas o *software* para ser utilizadas, y un ejemplo de ello es la investigación realizada por Núñez *et al.* (2018), la cual es un trabajo similar al que se está realizando en esta investigación, donde se tiene el objetivo de seleccionar las herramientas adecuadas utilizadas. A continuación, se describen paso a paso la utilización de la metodología.

De los indicadores o factores que complementan la metodología, estos contemplan criterios que van desde las características, capacidades del programa, estética, hasta la compatibilidad y adaptabilidad en otros *softwares*. Tomando en cuenta los criterios, se realizó una sumatoria de ellos donde se obtuvo un total de 23 criterios que deben cumplir y analizarse de las herramientas, las cuales pueden tener una ponderación desde cero hasta 23. En la tabla 3 se muestran los factores, criterios y ponderación que pueden existir en cada una de las herramientas (ver tabla 3).

Tabla 3. Factores, criterios y ponderación de la metodología FURPS para la selección de las herramientas para análisis de datos.

Factores	Criterios	Ponderación
Funcionalidad	Características y capacidades del programa. Generalidad de las funciones. Seguridad del sistema.	3
Usabilidad	Factores Humanos. Factores estéticos. Consistencia de la interfaz. Documentación.	4
Confiabilidad	Frecuencia y severidad de fallos. Exactitud de las salidas. Tiempo medio de fallos. Capacidad de recuperación ante fallos. Capacidad de predicción.	5
Rendimiento	Velocidad de procesamiento. Tiempo de respuesta. Consumo de recursos. Rendimiento efectivo total. Eficacia.	5
Capacidad de Soporte.	Extensibilidad. Adaptabilidad. Capacidad de prueba. Capacidad de configuración. Compatibilidad. Requisitos de instalación.	6
Ponderación Total		23

Fuente: Elaboración propia con base en Constanzo et al. (2014).

Cada uno de estos indicadores o factores y criterios se aplicaron a las 16 herramientas, *software* o servicios en la nube para análisis de datos (ver tabla 4).

Tabla 4. Clasificación de las herramientas analizadas con el modelo FURPS.

Clasificación	Herramientas
Herramientas framework	RapidMiner Orange Weka
Lenguajes de programación	Lenguaje R Python Matlab
Herramientas de Microsoft	Microsoft SQL Server Microsoft Excel Microsoft Analysis Services
Servicios en nube	Oracle Business Intelligence Server IBM: SPSS Modeler Intel: DAAL Amazon EMR Google Cloud Datalab Microsoft Azure Machine Learning SAS Enterprise Miner

Fuente: Elaboración propia.

Resultados.

El análisis de las herramientas para análisis de datos se llevó a cabo haciendo uso de una tabla, en donde se describen los factores e indicadores que la metodología FURPS contempla; además, de que en la parte inferior se colocaron los nombres de las herramientas que se analizaron y si éstas cumplían con alguno(s) de los criterios de dicha metodología, se marcaba la casilla con una X y si no era el caso se dejaba en blanco. En la tabla 5 se muestra la tabla realizada para el análisis de las herramientas de acuerdo con los factores y criterios de la metodología FURPS, en donde la ponderación más alta para obtener por la herramienta o *software* era de 23 puntos (ver tabla 5).

Tabla 5. Análisis de la selección de las herramientas para análisis de datos aplicando la metodología

FURPS.

Factores / Criterios	Herramientas utilizadas para Análisis de Datos															
	RapidMiner	Orange	WEKA	Lenguaje R	Python	Matlab	Microsoft SQL Server	Microsoft Excel	Microsoft Analysis Services	Oracle Business Intelligence Server	SPSS Modeler	DAAL	Amazon EMR	Google Cloud Datalab	Microsoft Azure Machine Learning	SAS Enterprise Miner
Funcionalidad																
Características y capacidades del programa	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Generalidad de las funciones	x	x	x	x			x	x	x	x	x		x	x	x	x
Seguridad del sistema	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Usabilidad																
Factores humanos	x	x	x	x			x	x	x	x	x		x	x	x	x
Factores estéticos			x	x					x						x	
Consistencia de la interfaz	x	x	x	x	x	x	x	x	x	x	x		x	x	x	x
Documentación	x		x	x	x	x	x	x	x		x	x	x	x	x	x
Confiabilidad																
Frecuencia y severidad de fallos			x	x			x			x					x	
Exactitud de las salidas	x	x	x	x	x	x	x	x	x	x	x		x	x	x	x
Tiempo medio de fallos	x		x				x	x	x						x	

Factores / Criterios	Herramientas utilizadas para Análisis de Datos															
	RapidMiner	Orange	WEKA	Lenguaje R	Python	Matlab	Microsoft SQL Server	Microsoft Excel	Microsoft Analysis Services	Oracle Business Intelligence Server	SPSS Modeler	DAAL	Amazon EMR	Google Cloud Datalab	Microsoft Azure Machine Learning	SAS Enterprise Miner
Capacidad de recuperación ante fallos				x						x						
Capacidad de predicción		x	x	x				x	x		x			x	x	x
Rendimiento																
Velocidad de procesamiento	x		x	x	x	x	x	x	x	x	x	x	x	x	x	x
Tiempo de respuesta		x	x	x	x	x	x	x	x	x	x		x	x	x	x
Consumo de respuesta		x	x				x	x	x	x			x	x	x	x
Rendimiento efectivo total	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Eficacia	x	x	x	x	x	x	x	x	x	x	x		x	x	x	x
Capacidad de Soporte																
Extensibilidad	x	x	x	x			x	x		x	x	x	x	x	x	x
Adaptabilidad	x	x	x	x	x	x	x	x	x	x	x	x	x		x	x
Capacidad de prueba	x	x	x	x	x	x	x	x	x	x	x		x	x	x	x
Capacidad de configuración		x	x	x		x	x		x		x	x	x		x	x
Compatibilidad	x	x		x	x		x	x	x				x	x	x	
Requisitos de instalación	x		x	x			x	x	x	x			x			x

Fuente: Elaboración propia.

Una vez realizado el análisis de los criterios que cumplen o no las herramienta y *software*, se procedió a efectuar la sumatoria de los criterios para obtener el total que cumple cada una de las herramientas y observar cuál o cuáles son las que obtuvieron mejor ponderación. Lo mencionado anteriormente, se presentada en la tabla 6, donde se observa que cada una de las herramientas obtuvieron una puntuación medio a alta (ver tabla 6).

Tabla 6. Ponderación de las herramientas de acuerdo con la metodología FURPS.

Factores	Herramientas utilizadas para Análisis de Datos															
	RapidMiner	Orange	WEKA	Lenguaje R	Python	Matlab	Microsoft SQL Server	Microsoft Excel	Microsoft Analysis Services	Oracle Business Intelligence Server	SPSS Modeler	DAAL	Amazon EMR	Google Cloud Datalab	Microsoft Azure Machine Learning	SAS Enterprise Miner
Funcionalidad	3	3	3	3	2	2	3	3	3	3	3	2	3	3	3	3
Usabilidad	3	2	4	4	2	2	3	3	4	2	3	1	3	3	4	3
Confiabilidad	2	2	4	4	1	1	3	3	3	3	2		1	2	4	2
Rendimiento	3	4	5	4	4	4	5	5	5	5	4	2	5	5	5	5
Capacidad de soporte	5	5	5	6	3	3	6	5	5	4	4	3	6	3	5	5
Total	16	16	21	21	12	12	20	19	20	17	16	8	18	16	21	18

Fuente: Elaboración propia.

De acuerdo con la tabla anterior, se puede observar, que las herramientas con mejor ponderación para la aplicación de análisis de datos son las herramientas: *Weka*, *Lenguaje R* y *Microsoft Azure Machine Learning*, las cuales cumplen con 21 de los 23 criterios que contempla la metodología FURPS.

A continuación, se describen con detalle las características de las herramientas con mejor ponderación obtenidas en el análisis anterior.

- **Lenguaje R.**

Es un entorno de *software* libre con licencia GNU GLP y lenguaje de programación interpretado; además, tiene compatibilidad con sistemas operativos como *UNIX*, *Windows*, *MacOS* y *Linux* (Unir Revista, 2019).

Las características principales de esta herramienta para el análisis de datos son Unir Revista (2019):

- Manejo y almacenamiento efectivo de datos.
- Operadores para realizar cálculos con matrices.
- Colección de modelos para el análisis de datos.
- Utilización de gráficos para la visualización de los datos.
- Dispone de herramientas de análisis estadísticos para profundizar en el conocimiento de los datos.
- Creación de *dashboards* para visualizar y analizar datos.
- Crear informes automáticos.

De igual forma, Lenguaje R cuenta con una herramienta o interfaz que permite acceder y visualizar de manera más sencilla los códigos, figuras, tablas o gráficos; esta herramienta es *RStudio*, la cual es un entorno de desarrollo integrado (IDE) para R.

- **Microsoft Azure Machine Learning.**

Es un servicio en la nube que permite acelerar y administrar el ciclo de vida de los proyectos de aprendizaje automático; esta herramienta permite a los individuos y equipos de trabajo, implementar MLOps dentro de su organización en un entorno de producción seguro y auditable (Microsoft Azure, 2024).

Las características principales de la herramienta son las siguientes (Microsoft Azure, 2024):

- Notebooks para escribir y ejecutar código en servidores *Jupyter Notebook* administrados que están integrados directamente.
- Visualización de métricas de ejecución.
- Diseñador para entrenar e implementar modelos de *Machine Learning* son escribir código, solo arrastrando y soltando el conjunto de datos y componentes para crear las visualizaciones.

- Creación de experimentos de aprendizaje automático automatizado con una interfaz fácil de usar.
- Etiquetado de datos para coordinar de forma eficaz los proyectos.
- Incluye herramienta para crear aplicaciones de IA generativa.

- **WEKA.**

Es una herramienta de código abierto que proporciona una colección de algoritmos de aprendizaje automático para tareas de minería de datos. Está desarrollado en Java y se utiliza ampliamente con fines educativos y de investigación en los campos de la estadística, análisis de datos y ciencia de datos (Learn Statistics Easily, s.f.).

Las principales características de WEKA son las siguientes (Learn Statistics Easily, s.f.):

- Amplia biblioteca de algoritmos de aprendizaje automático.
- Utilización fácilmente de algoritmos como árboles de decisión, máquinas de vectores de soporte y redes neuronales a sus conjuntos de datos.
- Incluye herramienta para el preprocesamiento de datos, lo que permite limpiar y preparar los datos antes del análisis.
- Admite la validación cruzada y otras técnicas para evaluar el rendimiento de los modelos de aprendizaje automático, lo que garantiza que los usuarios puedan tomar decisiones informadas en función de sus análisis.
- Interfaz gráfica de usuario (GUI) está diseñada para ser intuitiva y fácil de usar.

CONCLUSIONES.

Los avances tecnológicos tienen un impacto importante en la vida de las personas y en las organizaciones, ya que en la actualidad cada día durante las 24 horas del día se crean e intercambian datos, lo que conlleva un crecimiento rápido en las bases de datos de las organizaciones y que se cree la necesidad de analizar dichos datos para una mejor comprensión de los usuarios; es decir, identificar patrones o tendencias que ellos tengan.

Por ello, SAP España (2023) menciona, que el análisis de datos es fundamental para el funcionamiento y éxito de una empresa en la actualidad, porque la importancia del análisis de datos radica en su capacidad para proporcionar información valiosa y basada en evidencias, para ayudar a las organizaciones en la toma de decisiones informadas; por lo cual, la necesidad de analizar dichos datos ha llevado al desarrollo de diversas herramientas y *softwares* para el análisis de datos, las cuales pueden ser de software libre, lenguaje de programación que dentro de sus librerías se encuentran opciones para realizar análisis de datos o servicios en nube que cada vez toman más fuerza entre las organizaciones para ser utilizadas.

De acuerdo con lo anterior, en la presente investigación se realizó el análisis de diversas herramientas y *softwares* que se encuentran disponibles para ser utilizadas para la exploración y análisis de grandes volúmenes de datos; esto utilizando la metodología de calidad de *software* FURPS. La importancia de dicho estudio radica en contar con herramientas y *softwares* eficientes, y que faciliten el análisis de los datos, para transformarlos en conocimiento útil para la toma de decisiones en las organizaciones.

Los resultados obtenidos después de evaluar las 16 herramientas clasificadas en *framework*, lenguajes de programación, herramientas de Microsoft y servicio en nube, de acuerdo con la metodología de la calidad de software FURPS, se logró determinar que las herramientas de Lenguaje R, *Microsoft Azure Machine Learning* y WEKA, presentan un mejor rendimiento y desempeño al momento de realizar el análisis de los datos y destacan en el cumplimiento de los criterios establecidos por la metodología FURPS.

El lenguaje de programación R se destaca por ser un software de código abierto y la amplia biblioteca especializada para la estadísticas, análisis y visualización de los datos, además de la compatibilidad entre diferentes sistemas operativos existentes; así también, *Microsoft Azure Machine Learning* se distingue por ser un servicio en nube, por la integración que posee al poder trabajar con proyectos de aprendizaje automático y análisis de datos, además de facilitar la automatización y la gestión en entornos empresariales y equipos de trabajo; y por último, WEKA es una herramienta *framework*, que posee un entorno de trabajo amigable con el usuario, además de ser una herramienta de código abierto, por su versatilidad y la gran

colección de algoritmos de aprendizaje automático con las que cuenta para realizar tareas de análisis de datos.

Por lo anterior descrito, se puede concluir, que las herramientas mencionadas anteriormente son herramientas potentes y permiten a los usuarios o profesionales en el área de análisis de datos, realizar todas las tareas necesarias, y de igual forma, visualizar los resultados que ellos pueden obtener.

Con la realización de la investigación, se pone de manifiesto que es importante el realizar pruebas y aplicar metodologías para evaluar o revisar cuáles herramientas son las adecuadas para efectuar diversas actividades que el usuario desee, así como poder identificar cuáles son las fortalezas y las debilidades que presenta cada herramienta y *software*.

Finalmente, para investigaciones futuras se recomienda ampliar el número de herramientas y *softwares* a analizar, ya sean nuevas o emergentes, que cada día van surgiendo en el mundo de la ciencia de datos, análisis de datos y el aprendizaje automático.

REFERENCIAS BIBLIOGRAFICAS.

1. Amazon Web Services [AWS]. (s.f.). Amazon EMR. <https://aws.amazon.com/es/emr/>
2. Ballesteros, A., Sánchez-Guzmán, D. y García, R. (2013). Minería de datos educativa: Una herramienta para la investigación de patrones de aprendizaje sobre un contexto educativo. Lat. Am. J. Phys. Educ., 7(4), 662-668. http://www.lajpe.org/dec13/22-LAJPE_814_bis_Alejandro_Ballesteros.pdf
3. Callejas-Cuervo, M., Alarcón-Aldana, A. y Álvarez-Carreño, A. (2017). Modelos de calidad del software, un estado del arte. Entramado, 13(1), 236-250. <http://dx.doi.org/10.18041/entramado.2017v13n1.25125>
4. Constanzo, M., Casas, S. y Marcos, C. (2014). Comparación de modelos de calidad, factores y métricas en el ámbito de la Ingeniería de Software. Informes Científicos Técnicos-UNPA, 6(1), 1-36. <https://doi.org/10.22305/ict-unpa.v6i1.89>

5. Contreras, J., Molina, E. y Arteaga, P. (2010). Introducción a la programación estadística con R para profesores. (1ª ed.). Grupo de Educación Estadística, Universidad de Granada.
6. García, F. (2013). Aplicación de técnicas de Minería de Datos a datos obtenidos por el Centro Andaluz de Medio Ambiente (CEAMA) [Trabajo Fin de Máster]. Universidad de Granada. Granada, España.
7. Google Cloud. (01 de junio de 2023). Resumen de los servicios de Google Cloud Platform. <https://cloud.google.com/terms/services/index-20230601>
8. IBM. (s.f.). IBM SPSS Modeler. <https://www.ibm.com/mx-es/products/spss-modeler>
9. Fedorov, G. (16 de junio de 2016). How to Install the Python* Version of Intel® Data Analytics Acceleration Library (Intel® DAAL). Intel. <https://www.intel.com/content/www/us/en/developer/articles/troubleshooting/how-to-install-the-python-version-of-intel-data-analytics-acceleration-library-intel-daal.html>
10. Learn Statistics Easily. (s.f.). ¿Qué es Weka? <https://es.statisticseasily.com/glossario/what-is-weka-comprehensive-overview/>
11. López, F., Macías, M. y Sandoval, E. (2019). Minería de datos: identificando causas de deserción en las instituciones públicas de educación superior de México. Revista de Tecnología e Innovación en Educación Superior, 1(2), 1-14. <https://doi.org/10.22201/dgtic.26832968e.2019.2.4>
12. MathWorks. (s.f.a). Matlab para análisis de datos. <https://la.mathworks.com/products/matlab/data-analysis.html>
13. MathWorks. (s.f.b). Statistics and Machine Learning Toolbox. <https://la.mathworks.com/products/statistics.html>
14. Microsoft Azure. (19 de septiembre de 2024). What is Azure Machine Learning? <https://docs.microsoft.com/en-us/azure/machine-learning/overview-what-is-azure-machine-learning>
15. Microsoft. (s.f.). Análisis de datos en Excel. <https://support.microsoft.com/es-es/office/an%C3%A1lisis-de-datos-en-excel-3223aab8-f543-4fda-85ed->

[76bb0295ffc4#:~:text=An%C3%A1lisis%20de%20datos%20en%20Excel%20le%20permite%20comprender%20los%20datos,alto%20nivel%2C%20tendencias%20y%20patrones.](#)

16. Microsoft. (13 de marzo de 2014). Microsoft SQL Server. [https://docs.microsoft.com/es-es/previous-versions/bb545450\(v=msdn.10\)](https://docs.microsoft.com/es-es/previous-versions/bb545450(v=msdn.10))
17. Microsoft (22 de diciembre de 2023). What is Analysis Services? <https://learn.microsoft.com/en-us/analysis-services/analysis-services-overview?view=asallproducts-allversions>
18. Núñez, A., Ferreira, H., Ayala, K., Ruiz, J. y Cuin, J. (18 de mayo de 2018). Construcción de modelo de referencia para el desarrollo de aplicaciones WEB: propuesta metodológica [Ponencia]. Congreso Internacional de Investigación de Academia Journals Morelia 2018, Morelia, Michoacán.
19. Oracle. (s.f.). Oracle Business Intelligence. <https://www.oracle.com/mx/middleware/technologies/bi.html>
20. Pressman, R. (2010). Ingeniería de Software, Un enfoque práctico. (7ma ed.). McGraw-Hill.
21. Parada, P. (22 de noviembre de 2024). ¿Qué es el Data Mining o minado de datos? Ventajas, Tipos y Técnicas de minado. IEBS. <https://www.iebschool.com/blog/data-mining-mineria-datos-big-data/>
22. Rodríguez, Y. y Díaz, A. (2009). Herramientas de Minería de Datos. Revista Cubana de Ciencias Informáticas, 3(3-4), 73-80. <https://rcci.uci.cu/?journal=rcci&page=article&op=view&path%5B%5D=78>
23. SAP España. (02 de agosto de 2023). La importancia del análisis de datos en una empresa. SAP Spain Centro de Noticias. <https://news.sap.com/spain/2023/08/la-importancia-del-analisis-de-datos-en-una-empresa/>
24. Unir Revista. (29 de noviembre de 2019). Lenguaje R, ¿qué es y por qué es tan usado en Big Data? uniR: La Universidad en Internet. <https://www.unir.net/ingenieria/revista/lenguaje-r-big-data/>
25. Vargas, L., Farfán, J., Rodríguez, M., Aramayo, F., Flores, H., Bustamante, S. y López, V. (2016). Comparación de las principales herramientas de Data Mining y Análisis de Sábanas Telefónicas. En L.

Malbernat, J. Finochietto y G. Bacigalup (comps.). II Jornadas Argentinas de Tecnología, Innovación y Creatividad – II JATIC 2016. Mar del Plata, Buenos Aires.

DATOS DE LOS AUTORES.

1. **Zully Kristel Guzmán Caraveo.** Maestra en Administración de Tecnologías de la Información. Estudiante de posgrado (División Académica de Ciencias y Tecnologías de la Información) en la Universidad Juárez Autónoma de Tabasco. México. Correo electrónico: caraveokristel@gmail.com
2. **Martha Patricia Silva Payró.** Doctorado en Administración Educativa. Universidad Juárez Autónoma de Tabasco. Profesora-Investigadora de la División Académica de Ciencias y Tecnologías de la Información. México. Correo electrónico: patypayro@gmail.com
3. **Guillermo De los Santos Torres.** Doctorado en Sistemas Computacionales. Universidad Juárez Autónoma de Tabasco. Profesor-Investigador de la División Académica de Ciencias y Tecnologías de la Información. México. Correo electrónico: guillermo.delossantos@ujat.mx

RECIBIDO: 10 de febrero del 2025.

APROBADO: 22 de marzo del 2025.