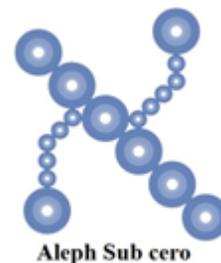


Reflexiones sobre la gestión de datos y el análisis ético en la era del Big Data



Sebastián André Pastén Fuenzalida¹

¹Facultad de Ingeniería, Arquitectura y Diseño, Universidad San Sebastián, Puerto Montt, Chile

E-mail: sebastian.pasten@uss.cl

(Recibido el 27/01/2025, aceptado el 17/02/2025)

Resumen

Este artículo aborda cómo la calidad de los datos en la era del Big Data es esencial para tomar decisiones correctas y más justas, ya que una gestión deficiente puede originar sesgos y afectar sectores sensibles como la educación y la justicia. Además, se presentan el concepto de Big Data para comprender la magnitud y complejidad de la información actual, compara de forma crítica el valor de los enfoques tradicionales frente a las técnicas emergentes, además de reflexionar sobre la integración de perspectivas cualitativas y cuantitativas para enriquecer el análisis de los datos. También se invita adoptar prácticas éticas y responsables en la gestión de datos para asegurar que el conocimiento generado contribuya a decisiones más justas y eficaces.

Palabras clave: Big Data, Gestión Ética de Datos, Sesgos en la Toma de Decisiones

Abstract

This article addresses how the quality of data in the Big Data era is essential for making accurate and fair decisions, as poor management can lead to biases and affect sensitive sectors such as education and justice. Furthermore, the concept of Big Data is presented to understand the magnitude and complexity of current information, critically comparing the value of traditional approaches versus emerging techniques, as well as reflecting on the integration of qualitative and quantitative perspectives to enrich data analysis. It also encourages the adoption of ethical and responsible practices in data management to ensure that the knowledge generated contributes to fairer and more effective decision-making.

Keywords: Big Data, Ethical Data Management, Decision-Making Bias

I. INTRODUCCIÓN

En la actualidad, marcada por el auge y uso del Big Data, las organizaciones y la sociedad en general enfrentan el desafío de gestionar, procesar y analizar grandes volúmenes de información [1]. Este enfoque basado en el uso de los datos parte del principio de que los datos pueden convertirse en una fuente invaluable de conocimiento, capaz de transformar la forma en que se toman decisiones en los diversos ámbitos de la actividad humana.

Sin embargo, como señala Manzano [2], el verdadero desafío no radica únicamente en el manejo de enormes cantidades de datos, sino también en garantizar la calidad de estos para asegurar que la información generada sea precisa y útil, evitando problemas como la redundancia, la inconsistencia y el llamado -ruido informativo-, es decir, datos irrelevantes o confusos que dificultan el análisis.

Un ejemplo a estos desafíos se observa, en la recolección de datos no estructurados en redes sociales (comentarios, publicaciones o imágenes compartidas) o en aquellos que provienen de dispositivos conectados a internet (sensores inteligentes, termostatos o relojes “inteligentes”), los cuales generan grandes volúmenes de

datos en tiempo real que, con frecuencia, resultan redundantes y producen -ruido informativo-, dificultando la extracción de información relevante. Por esta razón, la abundancia de datos por sí sola no garantiza su valor, pues la calidad, confiabilidad y pertinencia resultan esenciales para que dicha información se traduzca en conocimiento útil [3].

Es por esto, que la calidad de los datos se posiciona como un componente crítico para la construcción de modelos predictivos y descriptivos robustos. Una gestión inadecuada puede inducir interpretaciones sesgadas, poner en riesgo la validez de los resultados y, en última instancia, afectar decisiones estratégicas [4].

En este sentido, los marcos epistemológicos también influyen en la interpretación de los datos, y las decisiones humanas al seleccionar y procesar dichos datos afectan su objetividad [4]. Por lo tanto, se plantean las siguientes interrogantes clave: ¿qué implica garantizar la calidad de los datos? ¿Cómo evitar los sesgos y promover un análisis ético y riguroso?

El presente artículo explora la relevancia de la calidad de los datos en entornos del Big Data, examina no solo los sesgos interpretativos que pueden emerger, sino que

también evalúa las fortalezas y limitaciones de diferentes técnicas de análisis. Del mismo modo, se abordan los dilemas epistemológicos vinculados con las posturas positivistas y constructivistas, haciendo hincapié en la manera en que la combinación de metodologías cuantitativas y cualitativas enriquece la comprensión y contribuye a un examen más profundo e imparcial de los fenómenos estudiados [5].

II. DESARROLLO

A. La importancia de la calidad de los datos en entornos de Big Data.

En la actualidad, la calidad de los datos se ha convertido en un componente esencial dentro del Big Data [2]. Según un informe reciente de Merlin Data Quality [6], basado en estudios de Gartner, se estimó que en el año 2022 el 60% de las organizaciones consideraban la calidad de los datos como una prioridad crítica. Además, una investigación posterior indicó que el 47% de las empresas identificaban que la calidad de los datos era su principal desafío, seguido de la gobernanza y la seguridad de estos. Estos resultados destacan la importancia de que no solo la cantidad de datos es relevante, sino también la calidad de estos mismos para una correcta toma de decisiones.

Al respecto, Manzano [2], argumenta que la recopilación masiva de datos provenientes de diversas fuentes, como sensores, redes sociales y diversas transacciones de información, requiere controles rigurosos para evitar errores o sesgos en las interpretaciones. Existen diversos casos que evidencian la importancia del rigor en el trabajo con datos, como, por ejemplo, en el sistema de reclutamiento automatizado basado en Big Data implementado por la empresa Amazon. Dicho sistema, fue discontinuado tras su lanzamiento al descubrirse que otorgaba preferencia a candidatos masculinos por encima de otros criterios relevantes no relacionados con el género para el cargo, debido a un sesgo en los datos de entrenamiento, compuestos en su mayoría por perfiles de hombres [7].

En este sentido, Becerra y Castorina [4] analizan los dilemas epistemológicos vinculados al uso masivo de datos, advirtiendo los riesgos de interpretaciones sesgadas cuando no se ejerce un control adecuado sobre la calidad de estos. Estos ejemplos ilustran tanto el potencial como los desafíos éticos y metodológicos de la gestión de grandes volúmenes de datos, al requerir estrategias sólidas para salvaguardar la integridad de los datos y evitar la perpetuación de prejuicios históricos o la alteración de patrones relevantes.

Por su parte, Guerrero Ovejas [8] describe cómo los sistemas automatizados de evaluación de riesgo crediticio pueden reforzar sesgos discriminatorios al utilizar variables vinculadas a características socioeconómicas, generando patrones de desigualdad. La autora subraya la importancia de emprender auditorías continuas y mecanismos de revisión para resguardar la equidad en las decisiones y mitigar el riesgo de reproducir inequidades sociales. De

manera paralela, Gómez Carretero y Piattini Velthuis [3] señalan que la carencia de una gestión eficiente de la calidad de los datos no solo afecta la eficiencia operativa, sino que también puede incidir de forma negativa en la justicia y la objetividad de los sistemas de decisión.

En resumen, se observa que la gestión efectiva de la calidad de los datos se ha convertido en un foco estratégico, viéndose respaldada con la creación de normativas internacionales como el estándar ISO/IEC 25012 [9], que define un modelo general de calidad para datos estructurados dentro de un sistema informático. Por otro lado, estudios recientes, como los de Manzano & Avalos [10], destacan la aplicación de estos estándares en sectores con grandes volúmenes de datos, como la salud y la banca, para asegurar que las decisiones basadas en Big Data sean justas, equitativas y alineadas con los principios éticos y operativos de las organizaciones.

B. Definición y características de Big Data

Para comprender mejor el concepto de Big Data, conviene partir de la definición propuesta por De Mauro, Greco y Grimaldi [11], Big Data se describe como “activos de información caracterizados por un alto volumen, velocidad y variedad, que requieren tecnologías específicas y métodos analíticos para generar valor” (p. 8), permitiendo a las organizaciones transformar datos en conocimiento útil para la toma de decisiones estratégicas y operativas.

En esta misma línea, Pizarro et al. [12] señalan que Big Data se refiere a la gestión de grandes volúmenes de información digital y, además, complementan la definición con las “5 V”: volumen, velocidad, variedad, veracidad y valor, cada una de las cuales representa un aspecto clave en la administración de datos:

- **Volumen:** Se refiere a la enorme cantidad de datos generados diariamente, lo que requiere capacidades avanzadas de almacenamiento y procesamiento [11]. Un ejemplo de esto son las plataformas de redes sociales que generan terabytes de datos en cuestión de minutos.
- **Velocidad:** Describe la rapidez con la que los datos se generan y necesitan ser procesados. En aplicaciones como el monitoreo en tiempo real de sistemas financieros o de salud, un retraso en el análisis puede tener consecuencias graves.
- **Variedad:** Indica la diversidad de formatos y fuentes de datos, que pueden ser estructurados (como bases de datos), semiestructurados (archivos JSON) o no estructurados (imágenes, videos, etc.). Esto requiere múltiples herramientas que sean capaces de integrar y analizar estos formatos simultáneamente.
- **Veracidad:** Se refiere a la calidad y fiabilidad de los datos. Los datos imprecisos, incorrectos o incompletos pueden conducir a decisiones erróneas. Un ejemplo de esto son los datos de sensores defectuosos que generan alertas falsas.

- **Valor:** Destaca la importancia de extraer información valiosa y práctica, de modo que los datos se conviertan en un activo estratégico para la toma de decisiones.

Estas cinco dimensiones ponen en relieve la complejidad de trabajar con datos de múltiples procedencias y formatos. A medida que la tecnología avanza, especialmente en áreas como el Internet de las Cosas (IoT) y el aprendizaje automático, se ha acelerado el ritmo de generación de información, lo que exige herramientas sólidas para garantizar la calidad de los datos y evitar que los errores se propaguen [2].

C. Sesgos de interpretación en el análisis de datos.

En el apartado anterior, se enfatizó cómo los sesgos de interpretación en el análisis de datos pueden afectar ámbitos tan diversos como la justicia, la educación o el marketing, subrayando su carácter transversal. Para hacer frente a estos desafíos, se requiere un enfoque más amplio y comparativo, que permita no solo identificar, sino también comprender las estrategias dirigidas a mitigar los sesgos, como la triangulación metodológica y la evaluación participativa. Dichos enfoques, reconocidos en estudios recientes, buscan no solo disminuir la distorsión que puedan generar los datos y los procesos de análisis, sino también sumar múltiples visiones que aporten una lectura más completa y fiel a la realidad [3, 13].

Asimismo, es importante recalcar la importancia de la calidad de la información y de la validación cruzada para evitar interpretaciones sesgadas y asegurar una toma de decisiones fundamentada. En este sentido, autores como Torres Ruiz [13] y Mercader Uguina [14] señalan la utilidad de instrumentos como las auditorías de datos históricos y la triangulación metodológica como herramientas que posibilitan la identificación y gestión de las subjetividades que surgen a lo largo del proceso analítico.

Reflexionar sobre cómo estas subjetividades influyen en diversos ámbitos lleva a adoptar una visión crítica que integre las dimensiones sociales y culturales de cada estudio, evitando así examinar la información de manera aislada y promoviendo una perspectiva más inclusiva [4, 14].

Es importante señalar que, los sesgos pueden manifestarse en distintas etapas del proceso de análisis, desde la recolección y selección de datos hasta la interpretación de los resultados. La calidad de las fuentes, las metodologías aplicadas y los criterios de exclusión pueden influir significativamente en los hallazgos. De Mauro et al. [11] destacan que estas complejidades exigen un abordaje riguroso para gestionar no solo el volumen de los datos, sino también la diversidad de las fuentes.

Complementando esto, Gaete Moreno [15] remarca la relevancia del uso de la triangulación metodológica para contrarrestar las subjetividades naturales del proceso analítico. Ambas perspectivas convergen en la urgencia de un análisis crítico que incorpore el trasfondo social y

cultural de cada investigación, a fin de evitar interpretaciones parciales y favorecer una comprensión más integral de la información.

Un ejemplo de los sesgos que pueden surgir en el ámbito educativo es cuando el proceso de recolección de información se enfoca únicamente en ciertos grupos poblacionales, dejando fuera a comunidades más pequeñas o de zonas rurales, tal como se señala en *“El transitar en la investigación cualitativa: un acercamiento a la triangulación”* [13]. Para abordar estas limitaciones, algunos programas de evaluación de desempeño estudiantil han optado por enfoques mixtos que combinan datos cuantitativos con entrevistas y observaciones, mitigando así interpretaciones sesgadas [5].

En el campo de la justicia, los sistemas de predicción de reincidencia han sido cuestionados debido a que refuerzan sesgos raciales y socioeconómicos al basarse en bases de datos históricas con patrones discriminatorios. Mercader Uguina [14] analiza cómo estas herramientas automatizadas pueden influir negativamente en decisiones judiciales, destacando la necesidad de un análisis más riguroso en su diseño y aplicación.

Por otro lado, el sector del marketing enfrenta desafíos similares en el uso de algoritmos para la segmentación de audiencias. Aunque estas herramientas han logrado mayor precisión en la identificación de públicos específicos, también han reforzado estereotipos de consumo. Los algoritmos, al basarse en datos históricos, pueden perpetuar categorías limitantes y llevarnos a decisiones sesgadas, generando preocupaciones sobre su impacto en los patrones de compra y comportamiento del consumidor, como también se advierte en el contexto judicial [14].

Torres Ruiz [13] también señala que la subjetividad del investigador puede influir en el análisis de los datos, lo que destaca la importancia de adoptar estrategias de triangulación y el uso de múltiples fuentes para disminuir este riesgo.

Asimismo, resulta crucial emplear enfoques comparativos que permitan identificar posibles inconsistencias y determinar hasta qué punto las decisiones metodológicas modelan los resultados obtenidos.

Tradicionalmente, el análisis de datos se ha respaldado en el enfoque positivista, que privilegia la objetividad y la medición cuantitativa. Sin embargo, desde el paradigma constructivista se cuestiona esta visión, al recalcar la importancia de los factores socioculturales y las subjetividades inherentes a la interpretación de datos. Por ejemplo, Martínez Musiño [16] expone cómo un mismo conjunto de datos puede llevar a interpretaciones diferentes según el marco teórico empleado. Esta tensión epistemológica evidencia la necesidad de adoptar un enfoque crítico y reflexivo que reconozca la pluralidad de interpretaciones posibles.

D. Tipos de sesgos en la interpretación de datos

Como se ha ido desarrollando, la interpretación de datos en entornos de Big Data afronta una de sus principales

<http://www.unet.edu.ve/matyfis/alphsub.html>

dificultades en la presencia de sesgos, que pueden impedir la obtención de resultados realmente objetivos y precisos. Estos sesgos no solo provienen de la calidad o selección de la información, sino también de las metodologías y decisiones influenciadas por valores éticos, políticos y sociales, así como de prejuicios conscientes e inconscientes de quienes investigan.

Becerra y Castorina [4] resaltan que los datos distan de ser neutrales, pues están marcados por prácticas y procesos que reflejan marcos epistemológicos subyacentes, de modo que las interpretaciones se ven influidas tanto por la información disponible como por las creencias y valores que guían la producción de conocimiento científico.

En la misma línea, Boyd y Crawford [17] advierten sobre la “mitología del Big Data”, en la cual se tiende a considerar que manejar grandes cantidades de datos garantiza una objetividad intrínseca. Sin embargo, estos autores señalan que cada registro surge de decisiones interpretativas, lo que introduce inevitablemente sesgos.

Por tanto, procesar enormes volúmenes de información no implica neutralidad en sí misma, y se hace necesario cuestionar críticamente las suposiciones, los valores y las limitaciones vinculadas al análisis de datos.

A continuación, se presentan los principales tipos de sesgos que afectan la interpretación de datos en contextos de Big Data:

- **Sesgo de selección:** Esto ocurre cuando los datos recopilados no reflejan con exactitud a la población de estudio, lo que deriva en conclusiones parciales o desajustadas [7]. Un ejemplo se da en el ámbito educativo, si solo se recaban datos de centros urbanos, excluyendo así a las comunidades rurales y sesgando los hallazgos hacia la realidad urbana.
- **Sesgo de confirmación:** Se produce cuando el investigador busca exclusivamente evidencias que respalden sus hipótesis previas [14]. En estudios de mercado, esto ocurre cuando se omite información contraria a las expectativas del investigador, centrando el análisis únicamente en resultados que respaldan dichas suposiciones.
- **Sesgo algorítmico:** Surge cuando los algoritmos replican prejuicios presentes en los datos de entrenamiento, perpetuando desigualdades o errores en los resultados [4]. Un ejemplo ilustrativo es el proceso automatizado de reclutamiento de Amazon, que excluía de forma discriminatoria ciertos perfiles al basarse en parámetros sesgados.

Finalmente, Lemus-Delgado y Pérez Navarro [5] subrayan que el análisis de Big Data está condicionado por la forma en que se seleccionan los datos y se diseñan los modelos de interpretación, basándose en una determinada visión epistemológica. Esto puede derivar en explicaciones que reflejen más un punto de vista previo que la complejidad genuina de los fenómenos estudiados.

E. Perspectivas cualitativas y cuantitativas: un enfoque complementario

Las investigaciones recientes resaltan la importancia de combinar enfoques cualitativos y cuantitativos para lograr un análisis más completo de los fenómenos estudiados. Torres Ruiz [13] analiza cómo las metodologías mixtas y su aplicación en educación han demostrado que el uso de datos cuantitativos para identificar patrones se complementa con entrevistas y observaciones para comprender de manera más profunda el contexto que rodea dichos patrones.

Por otro parte, investigaciones de Pizarro et al. [12] muestran que la aplicación de enfoques mixtos en estudios sobre la transformación digital, destacando la manera en que las narrativas cualitativas añaden profundidad interpretativa a los análisis estadísticos. Este tipo de estrategia evita explicaciones reduccionistas y favorece una visión más completa, ofreciendo un panorama más completo de la realidad.

En el ámbito de los estudios globales y el análisis de datos, las metodologías cuantitativas facilitan la identificación de patrones y correlaciones en grandes volúmenes de información. Según Gómez Carretero y Piattini [3], estos autores subrayan que el uso del modelo MAMD (Modelo Alarcos de Mejora de Datos) es una herramienta esencial para garantizar datos de alta calidad en entornos digitales, destacando que una adecuada gobernanza y gestión de datos optimiza la toma de decisiones estratégicas en diversas industrias.

Sin embargo, Lemus-Delgado y Pérez Navarro [5] comentan que los resultados numéricos requieren también de enfoques cualitativos que ayuden a interpretarlos de manera crítica, introduciendo así narrativas y perspectivas que revelen los entramados humanos, sociales y culturales subyacentes.

Un ejemplo claro se encuentra en los estudios migratorios, donde los datos cuantitativos pueden revelar volúmenes y direcciones de flujo, mientras que los enfoques cualitativos aportan una visión más profunda al explorar las razones subyacentes detrás de las decisiones de movilidad [4, 16]. La combinación de enfoques cualitativos y cuantitativos permite un análisis más completo, considerando tanto las tendencias generales como los contextos específicos que influyen en el fenómeno migratorio.

En esta misma línea, las investigaciones en ciencia de datos e informática han mostrado cómo conjuntos de datos similares pueden ser interpretados de manera distinta dependiendo de los marcos teóricos aplicados, resaltando la importancia del contexto en la interpretación [16]. Esta conclusión muestra que en el estudio de fenómenos complejos no basta con conocer los resultados cuantitativos; también se debe comprender el entorno cultural, político y social que les da forma. Sin embargo, la calidad de las fuentes, las metodologías aplicadas y los criterios de exclusión pueden influir significativamente en los hallazgos.

Por su parte, Pizarro et al. [12] exponen cómo la transformación digital impacta el análisis de datos, permitiendo lecturas diversas según los objetivos de cada investigación. Por lo tanto, integrar ambas perspectivas de análisis se vuelve esencial para lograr una comprensión más profunda y evitar explicaciones reduccionistas. Adicionalmente, Becerra y Castorina [4] hacen hincapié en la importancia de un enfoque dialógico entre cifras y contextos, al describir cómo los valores epistemológicos influyen en la interpretación de los datos y resaltando la importancia de integrar metodologías tanto cuantitativas como cualitativas. Estas consideraciones muestran cómo cada enfoque aporta una pieza clave en la construcción de un entendimiento más completo de los fenómenos analizados.

F. Técnicas de análisis de datos en el contexto de la Big Data.

Las técnicas de análisis de datos han evolucionado considerablemente en el último tiempo, gracias al desarrollo de una mayor capacidad de procesamiento de información, nuevas herramientas y enfoques como la minería de datos, el aprendizaje automático y los sistemas de visualización avanzada. Estas herramientas han permitido transformar datos en conocimiento útil de manera más eficiente, como lo resalta Moreno Salinas [18], quien enfatiza el papel de estas tecnologías en la optimización de la toma de decisiones mediante la generación de información procesada y estructurada. Estas herramientas permiten identificar patrones, correlaciones y anomalías en grandes volúmenes de información de manera eficiente, aplicándose en campos tan diversos como el financiero o la salud pública. En este sentido, Pizarro Gurrola et al. [12] y Manzano [2] evidencian cómo los sistemas de Big Data pueden detectar tendencias de comportamiento en mercados globales en tiempo real, ofreciendo una visión más profunda de dinámicas complejas.

No obstante, existen limitaciones importantes. Becerra y Castorina [4], en su análisis epistemológico del Big Data, señalan la posible pérdida de interpretabilidad en los modelos muy complejos y la necesidad de contar con una infraestructura sólida. Asimismo, Lemus-Delgado y Pérez Navarro [5] alertan sobre la propensión a sesgos cuando los conjuntos de datos de entrenamiento no son representativos o presentan errores. Además, la integración de datos provenientes de fuentes heterogéneas exige habilidades avanzadas para mantener la coherencia y la relevancia contextual de los análisis, lo cual resulta esencial en ámbitos sociales y de política pública.

G. Ventajas y desventajas claves que presentan las técnicas de análisis de datos en diversos entornos de Big Data.

La evolución de los métodos de análisis ha generado un debate constante sobre el uso de técnicas tradicionales frente a enfoques más recientes en diversos campos. En

lugar de verlas como polos opuestos, conviene examinar sus fortalezas y limitaciones considerando el objetivo del estudio y el contexto. Tal como señalan Gaete Moreno [15] y Martínez Musiño [16], las técnicas estadísticas clásicas, como la regresión lineal, han demostrado gran utilidad en la economía, mientras que los métodos basados en aprendizaje automático han revolucionado la forma de predecir resultados en áreas como la salud pública. La clave radica en elegir la técnica adecuada según el propósito de la investigación y la naturaleza de los datos, sin perder de vista los sesgos y limitaciones que pueden presentarse en cada enfoque.

A continuación, se desarrollan los diferentes grupos de técnicas con sus respectivas ventajas y desventajas para una mayor comprensión:

Técnicas tradicionales: Dentro de este grupo de técnicas se encuentran métodos estadísticos como la regresión lineal y el análisis de varianza. No obstante, aunque son muy usados, presentan limitaciones cuando los datos son no estructurados (por ejemplo, textos o imágenes). Además, en campos como la educación y la salud pública, recientes investigaciones sugieren que, para abordar datos complejos, se requieren estrategias más flexibles. Un caso ilustrativo es el de Gaete Moreno [15], quien usó un análisis de varianza para examinar diferencias de ingresos en distintos sectores de la economía, demostrando que estos métodos pueden evidenciar desigualdades relevantes y aportar información valiosa. En consecuencia, dado que pueden modelar relaciones lineales y evaluar diferencias significativas entre variables, son muy comunes en disciplinas como la economía o la biomedicina.

Ventajas:

- Ofrecen resultados fáciles de interpretar y están respaldadas por un marco teórico bien establecido, lo cual facilita su validación.
- Sus fundamentos matemáticos están ampliamente documentados, tal como menciona Villegas Zamora [19] al destacar su utilidad en la enseñanza de la estadística para abordar problemas gerenciales complejos y reducir la incertidumbre

Desventajas:

- Presentan dificultades para manejar datos no estructurados, lo que limita su aplicación en ámbitos que requieren analizar imágenes, texto u otros formatos no cuantitativos.

Técnicas emergentes: Dentro de este grupo de técnicas encontramos algoritmos de aprendizaje automático, redes neuronales y métodos de agrupamiento (clustering). Aunque su capacidad para detectar patrones complejos y generar predicciones en tiempo real ofrece múltiples beneficios, también plantea retos éticos y riesgos en la transparencia [4]. Un ejemplo del uso de estas técnicas es el trabajo de Pizarro et al. [12], donde se emplearon redes neuronales para predecir tendencias de consumo en el sector minorista, logrando identificar patrones detallados y

mejorar la personalización de productos y servicios. Estas técnicas permiten modelar relaciones no lineales, analizar grandes volúmenes de datos y procesar información proveniente de diversas fuentes.

Ventajas:

- Pueden procesar grandes volúmenes de datos, incluso cuando estos proceden de fuentes diversas.
- Analizan relaciones no lineales y capturan matices que métodos más tradicionales podrían pasar por alto [12].

Desventajas:

- Su complejidad puede obstaculizar la interpretación de resultados y favorecer la aparición de sesgos [17].

III. CONCLUSIONES

Como se puede apreciar en este artículo, la calidad de los datos en entornos de Big Data es un pilar fundamental en nuestros días para asegurar la precisión de los análisis y la correcta toma de decisiones. Por ejemplo, en 2018, un sistema de análisis predictivo en el sector salud en EE. UU. falló al identificar correctamente pacientes prioritarios debido a datos incompletos, lo que generó retrasos críticos en el acceso a tratamientos [2]. Este caso refuerza la necesidad de un control de calidad riguroso para evitar consecuencias negativas o no esperadas. Otro caso documentado muestra cómo la mala calidad de los datos en sistemas de salud pública puede generar diagnósticos imprecisos, como lo evidenció el estudio de Moreno Salinas [18], donde se analizan los riesgos de información incompleta en bases de datos de monitoreo médico. Sin embargo, más allá de aplicar técnicas avanzadas, se requiere una visión crítica y reflexiva que considere los posibles sesgos y las limitaciones de los modelos empleados. Esto implica entender que los algoritmos y métodos no son infalibles y que pueden amplificar desigualdades o llevarnos a tomar decisiones erradas si no se contextualizan de manera adecuada.

Es por esto, que los tomadores de decisiones deben implementar pasos concretos para minimizar los riesgos de sesgo y mejorar la transparencia. Dentro de las principales recomendaciones incluyen:

- Auditorías periódicas de los modelos: Evaluar continuamente la precisión, equidad y relevancia de los algoritmos, identificando posibles sesgos y actualizando los parámetros según nuevas realidades.
- Transparencia en los procesos: Documentar y comunicar de manera clara los criterios utilizados en el análisis de datos y los resultados obtenidos, promoviendo un acceso abierto a las metodologías.
- Interdisciplinariedad: Formar equipos de trabajo que incluyan expertos técnicos, científicos sociales y especialistas en ética para enriquecer las perspectivas y garantizar decisiones más completas.
- Capacitación continua: Promover programas de formación para que los líderes comprendan tanto las

- Su implementación requiere infraestructura sólida y personal con alto nivel de especialización [20].

Por otra parte, herramientas como Python y R han impulsado nuevas formas de análisis y visualización de datos, facilitando la construcción de modelos predictivos y descriptivos [18]. Sin embargo, sacar el máximo provecho de estas tecnologías exige una base sólida de conocimientos técnicos y una infraestructura adecuada para manejar grandes cantidades de información. De ahí la relevancia de un enfoque interdisciplinario, donde colaboren especialistas en ciencia de datos, expertos en la temática de estudio y profesionales que velen por una gestión ética de la información.

capacidades como las limitaciones de las herramientas de Big Data.

Por ejemplo, en el sector salud, un modelo predictivo debe ser monitoreado de forma continua para evitar inequidades en los tratamientos recomendados. De igual forma, en la política pública, los análisis deben ser revisados para evitar conclusiones que perpetúen brechas sociales, como se pudo apreciar en los ejemplos citados en este artículo.

En resumen, la integración responsable de enfoques de análisis de datos debe ir más allá de la aplicación de técnicas avanzadas para estar guiada por principios de justicia, precisión, transparencia y responsabilidad social. Esto implica no solo ver el potencial de la información, sino también reconocer y entender las limitaciones propias de los modelos utilizados, como los sesgos y la pérdida de interpretabilidad en sistemas complejos. La combinación de métodos cualitativos y cuantitativos, junto con la incorporación de perspectivas interdisciplinarias, nos permitirá enriquecer los análisis y garantizar que nuestras decisiones sean más completas y equitativas.

REFERENCIAS

- [1] V. Mayer-Schönberger & K. Cukier, *Big Data: A revolution that will transform how we live, work, and think*. Mariner Books, 2014.
- [2] F. Manzano, “La calidad de los datos estadísticos en la era del Big Data”, 2021. <https://doi.org/10.13140/RG.2.2.27930.16323/1>
- [3] A. I. Gómez Carretero & M. Piattini, “Importancia de la calidad de los datos en la transformación digital”. *Revista Ruidera*, no. 13, 2018. <https://revista.uclm.es/index.php/ruiderae/article/view/183>
- [4] G. Becerra & J. Castorina, “Hacia un análisis de los marcos epistémicos del Big Data”, *Cinta de Moebio*, no. 76, 50–63, 2023. <https://doi.org/10.4067/s0717-554x2023000100050>
- [5] D. Lemus-Delgado & R. Pérez Navarro, “Ciencia de datos y estudios globales: aportaciones y desafíos metodológicos”, *Colombia Internacional*, 102, 45–60, 2020. <https://doi.org/10.7440/colombiaint102.2020.03>

- [6] Merlin Data Quality, 5 tendencias en Data Quality a 2027. Congreso Tecnología y Negocios América Digital México, 2023. <https://merlindataquality.com/>
- [7] A. Coddou Mc Manus & R. Padilla Parga, “Discriminación algorítmica en los procesos automatizados de reclutamiento y selección de personal”, *Revista Chilena de Derecho y Tecnología*, 13, 1–34, 2024. <https://doi.org/10.5354/0719-2584.2024.71312>
- [8] M. Guerrero Ovejas, Score automatizado en la concesión de crédito, Universidad de Barcelona, 2024. https://diposit.ub.edu/dspace/bitstream/2445/214987/1/WP%20Marta%20Guerrero%20Ovejas_2024_5-1.pdf
- [9] International Organization for Standardization, ISO/IEC 25012:2008 - Software engineering - Software product Quality Requirements and Evaluation (SQuaRE) - Data quality model, 2008. <https://www.iso.org/standard/35736.html>
- [10] F. A. Manzano, & D. Avalos, “Análisis de calidad de los datos en las estadísticas públicas y privadas, ante la implementación del Big Data”, *Ciencias Administrativas*, 22, 1–15, 2023. <https://doi.org/10.24215/23143738e119>
- [11] A. De Mauro, M. Greco & M. Grimaldi, ¿What is Big Data? A consensual definition and a review of key research topics. AIP Conference Proceedings. Preprint, 2014. <https://doi.org/10.1063/1.4907823>
- [12] R. Pizarro Gurrola et al., *Ciencia de los datos. Propuestas y casos de uso*, Instituto Tecnológico de Durango, 2020.
- [13] A. E. Torres Ruiz, “El transitar en la investigación cualitativa: un acercamiento a la triangulación”, *Revista Científica*, Vol 6, no. 20, 275–295, 2021. <https://doi.org/10.29394/Scientific.issn.2542-2987.2021.6.20.15.275-295>
- [14] J. R. Mercader Uguina, “Discriminación algorítmica y derecho granular: nuevos retos para la igualdad en la era del big data”, *LABOS Revista De Derecho Del Trabajo Y Protección Social*, Vol 2, No. 2, 4-10, 2021. <https://doi.org/10.20318/labos.2021.6213>
- [15] A. Gaete Moreno, “Análisis e interpretación de datos cualitativos”. *Tema De Investigación Central De La Academia*, 151–165, 2017. <https://publicacionesacague.cl/index.php/tica/article/view/171>
- [16] C. Martínez Musiño, “Ciencia de datos versus ciencia de la información: informetría y análisis textual”, *Scire*, Vol. 27, no. 2, 31–42, 2021. <https://doi.org/10.54886/scire.v27i2.4731>
- [17] D. Boyd & K. Crawford, “Critical questions for big data. *Information, Communication & Society*, Vol 15, no. 5, 662–679, 2012. <https://doi.org/10.1080/1369118X.2012.678878>
- [18] J. G. Moreno Salinas, “Científico de datos: codificando el valor oculto e intangible de los datos”, *Revista Digital Universitaria*, Vol. 18, no. 7, 1–15, 2017. <https://doi.org/10.22201/codeic.16076079e.2017.v18n7.a2>
- [19] D. A. Villegas Zamora, (2019). “La importancia de la estadística aplicada para la toma de decisiones en