



# Tendencias del Big Data con la ciencia estadísticas: Una revisión documental

Big Data trends in statistical science: A documentary review  
*Tendências do Big Data com a ciência estatística: uma revisão documental*

ARTÍCULO ORIGINAL

**Geovanny Desiderio Chancay-Quimis**   
geovanny.chancayq@ug.edu.ec

**Hilda Mercedes Blum-Alcívar**   
hilda.bluma@ug.edu.ec

**Omar Gabriel Mejía-Flores**   
omar.mejiaf@ug.edu.ec

**Eva Leonor Medrano-Freire**   
eva.medranof@ug.edu.ec

Universidad de Guayaquil. Guayaquil, Ecuador



Escanea en tu dispositivo móvil  
o revisa este artículo en:

<https://doi.org/10.33996/revistaingenieria.v10i26.141>

Artículo recibido 17 de noviembre 2025 / Aceptado 15 de diciembre 2025 / Publicado 05 de enero 2026

## RESUMEN

La convergencia entre el Big Data y la ciencia estadística ha redefinido los paradigmas de análisis de información en la era digital. El objetivo de este estudio consiste en analizar las tendencias del Big Data con la ciencia estadística mediante una revisión documental exhaustiva. La metodología empleada se basa en un enfoque cualitativo de tipo revisión documental, con diseño de revisión narrativa, consultando bases de datos como SciELO, Redalyc y Google Académico en el periodo 2018-2025. Se identificaron inicialmente 65 estudios, de los cuales se incluyeron 15 tras aplicar criterios de relevancia, vigencia y rigor metodológico. Los resultados indican que la integración de métodos estadísticos avanzados, como la minería de datos y la estadística bayesiana, permite procesar volúmenes masivos de datos estructurados y no estructurados para la toma de decisiones estratégicas. Se concluye que el Big Data actúa como un catalizador de la ciencia estadística, requiriendo una evolución constante de los modelos predictivos para enfrentar los desafíos de la datificación global.

**Palabras clave:** Analítica de datos; Big Data; Ciencia estadística; Integración de métodos; Tendencias tecnológicas

## ABSTRACT

The convergence between Big Data and statistical science has redefined the paradigms of information analysis in the digital era. The objective of this study is to analyze Big Data trends in statistical science through an exhaustive documentary review. The methodology employed is based on a qualitative approach of a documentary review type, with a narrative review design, consulting databases such as SciELO, Redalyc and Google Scholar in the period 2018-2025. Initially, 65 studies were identified, of which 15 were included after applying criteria of relevance, validity and methodological rigor. The results indicate that the integration of advanced statistical methods, such as data mining and Bayesian statistics, allows the processing of massive volumes of structured and unstructured data for strategic decision making. It is concluded that Big Data acts as a catalyst for statistical science, requiring constant evolution of predictive models to face the challenges of global data.

**Key words:** Data analytics; Big Data; Statistical science; Integration of methods; Technological trends

## RESUMO

A convergência entre o Big Data e a ciência estatística redefiniu os paradigmas de análise da informação na era digital. O objetivo deste estudo consiste em analisar as tendências do Big Data com a ciência estatística por meio de uma revisão documental exaustiva. A metodologia empregada se baseia em um enfoque qualitativo de tipo revisão documental, com design de revisão narrativa, consultando bases de dados como SciELO, Redalyc e Google Académico no período 2018-2025. Se identificaram inicialmente 65 estudos, dos quais se incluíram 15 para aplicar critérios de relevância, viglância e rigor metodológico. Os resultados indicam que a integração de métodos estatísticos avançados, como a mineração de dados e a estatística bayesiana, permite processar volumes enormes de dados estruturados e não estruturados para a tomada de decisões estratégicas. Conclui-se que o Big Data atua como um catalisador da ciência estatística, exigindo uma evolução constante dos modelos preditivos para enfrentar os desafios da datificação global.

**Palavras-chave:** Analítica de dados; Grandes Dados; Ciência estatística; Integração de métodos; Tendências tecnológicas

## INTRODUCCIÓN

La sociedad contemporánea atraviesa una transformación sin precedentes impulsada por la generación masiva de información, fenómeno ampliamente conocido como Big Data. Este concepto no solo se refiere al volumen creciente de datos, sino también a la capacidad de extraer valor significativo mediante técnicas avanzadas de análisis. Gutiérrez (2018) destaca, que el Big Data se ha popularizado debido a la producción de cantidades ingentes de información provenientes de sensores, dispositivos inteligentes y la actividad digital de los usuarios. Este escenario ha reconfigurado los fundamentos del análisis de datos, exigiendo marcos teóricos y metodológicos capaces de responder a la complejidad informacional de la era digital.

La problemática central de esta investigación radica en la brecha existente entre la generación exponencial de datos y la capacidad de las herramientas estadísticas tradicionales para procesarlos con precisión. Históricamente, la estadística evolucionó desde los censos rudimentarios en civilizaciones antiguas hasta la consolidación de la inferencia moderna con Pearson y Galton (Salazar y Del Castillo, 2018). Sin embargo, los supuestos clásicos de independencia, aleatoriedad y normalidad se ven desafiados por la magnitud, velocidad, heterogeneidad de los datos actuales, lo que demanda nuevas competencias analíticas y epistemológicas.

En este contexto, la estadística se posiciona como disciplina fundamental para organizar,

analizar e interpretar datos mediante herramientas probabilísticas. Rivadeneira et al. (2022) señalan que la estadística facilita la identificación de patrones y la generación de inferencias basadas en evidencia, mientras que Aceña et al. (2025) subrayan su capacidad predictiva y su aporte a la toma de decisiones en escenarios de incertidumbre. La estadística, por tanto, se convierte en el puente metodológico que permite transformar datos masivos en conocimiento accionable.

Desde una perspectiva metodológica, la estadística se divide en descriptiva e inferencial. La primera se orienta a la organización y síntesis de datos; la segunda, a la generalización de resultados mediante modelos probabilísticos Aceña et al. (2025). En el ámbito del Big Data, esta distinción se complejiza, pues la inferencia debe adaptarse a flujos continuos de información, estructuras no convencionales y bases de datos de alta dimensionalidad.

La integración entre Big Data y estadística ha impulsado el desarrollo de modelos predictivos, análisis experimentales y sistemas de apoyo a la toma de decisiones en sectores empresariales, gubernamentales y científicos (Salazar y Del Castillo, 2018). Con la consolidación de la industria 4.0, organizaciones y gobiernos emplean datos masivos para optimizar procesos, identificar patrones y generar valor estratégico (Manzano y Avalos, 2022). Este proceso ha generado un giro epistemológico en la estadística, que transita de un enfoque centrado en la inferencia hacia uno orientado también a la

predicción, automatización y analítica avanzada.

A nivel internacional, Luckin et al. (2016) destacan que la analítica de datos permite personalizar servicios y automatizar procesos, optimizando la experiencia del usuario en distintos sectores. En América Latina, Ávila et al. (2024) evidencian que la estadística apoyada en tecnologías emergentes optimiza la toma de decisiones en contextos académicos. Lasso et al. (2022) agregan que la evolución de la TIC, ha fortalecido el uso de la estadística en ingeniería, y Sigman y Bilinkis (2023) sostienen que el análisis de datos masivos revela patrones de comportamiento humano previamente invisibles. Estas perspectivas muestran que el Big Data no solo transforma los métodos sino también los objetos de estudio y las preguntas de investigación.

En el ámbito educativo y empresarial, Peláez (2025) señala que el Big Data permite evaluar el rendimiento estudiantil mediante análisis individualizados, mientras que Toscano et al. (2025) destacan su impacto en la toma de decisiones corporativas. En la administración pública Aranibar et al. (2023) evidencian su rol en el fortalecimiento del gobierno electrónico, y en el campo de las ciencias de la administración, la estadística ha sido fundamental en la evolución del gobierno electrónico, permitiendo al Estado obtener patrones de comportamiento de los ciudadanos para diseñar políticas más efectivas. Finalmente, Zúñiga et al. (2023) describen el Big Data como un fenómeno tecnológico derivado de la datificación del mundo. Este proceso de datificación constituye un cambio epistemológico

profundo, donde la realidad social se interpreta a través de trazas digitales.

La evolución de la estadística en el contexto del Big Data también implica una transformación en los métodos de recolección de datos. La proliferación del Internet de las Cosas (IoT), redes sociales y transacciones digitales ha generado un entorno de datos masivos, donde los métodos tradicionales de muestreo resultan insuficientes (Mayer-Schönberger y Cukier, 2013). En respuesta a ello, la estadística incorpora técnicas de computación distribuida y algoritmos de aprendizaje automático, para gestionar datos heterogéneos y de alta dimensionalidad (Provost y Fawcett, 2013; James et al., 2021). Esto marca un giro metodológico hacia enfoques híbridos que combinan inferencia, predicción y automatización.

No obstante, en la actualidad, esta rigurosidad se ve desafiada por la opacidad de los algoritmos y la complejidad de los datos no estructurados, lo que limita la interpretabilidad de los modelos analíticos y dificulta la toma de decisiones informadas. En consecuencia, la brecha digital ya no se limita al acceso de información, sino que implica una brecha de competencias cognitivas, relacionada con la capacidad de formular preguntas adecuadas a los datos e interpretar sus resultados dentro de un marco científico (Kitchin, 2014; Boyd y Crawford, 2012). Esta problemática se intensifica en el ámbito académico y profesional, donde la formación estadística aún se encuentra en proceso de adaptación frente a la era algorítmica, requiriendo nuevas competencias en análisis de

datos y pensamiento crítico (Davenport y Patil, 2012).

La veracidad de los datos constituye otro desafío, pues los conjuntos masivos suelen incluir ruido, sesgos y datos faltantes (Kelleher et al., 2020). La curaduría de datos, se vuelve entonces una fase crítica del proceso analítico (Batini y Scannapieco, 2016). La ausencia de estándares globales en metadatos y la necesidad de competencias en programación, ética algorítmica y bases de datos NoSQL evidencian la urgencia de actualizar la formación estadística.

Asimismo, el uso intensivo de datos plantea implicaciones éticas y socioeconómicas. La datificación de la vida cotidiana permite modelar comportamientos con alta precisión, lo que genera preocupaciones sobre privacidad y justicia algorítmica (Floridi et al., 2018). La tendencia hacia la inteligencia artificial explicable (XAI) busca garantizar modelos interpretables y auditables (Rudin, 2019). Este enfoque ético-metodológico es indispensable para legitimar el uso del Big Data en la toma de decisiones.

El Big Data se compone de datos estructurados, no estructurados y semi-estructurados ((Universidad Católica San Pablo, 2026), cuya gestión requiere arquitecturas computacionales especializadas. Sus características fundamentales –volumen, velocidad y variedad- determinan los métodos estadísticos aplicables. A estas dimensiones, la literatura reciente añade veracidad y el valor ampliando la comprensión integral del fenómeno.

Finalmente, la justificación del presente estudio se fundamenta en la necesidad de sistematizar las tendencias actuales que vinculan al Big Data con la ciencia estadística para orientar su uso ético y productivo. La pregunta que orienta esta revisión es: ¿Cuáles son las tendencias y métodos estadísticos más efectivos para el análisis de Big Data en la actualidad? En consecuencia, el objetivo general consiste en analizar las tendencias del Big Data con la ciencia estadística mediante una revisión documental exhaustiva, identificando sus alcances, limitaciones y aportes metodológicos. Este propósito busca fortalecer el dialogo entre teoría estadística, innovación tecnológica y responsabilidad social en el uso de datos masivos.

## **METODOLOGÍA**

El presente estudio se desarrolló bajo un enfoque cualitativo de tipo investigación documental, con diseño de revisión narrativa sistematizada. Este diseño permitió analizar, comparar y sintetizar la producción científica reciente sobre Big Data y ciencia estadística, siguiendo un proceso ordenado y verificable que garantiza la transparencia y replicabilidad del estudio.

La investigación se estructuró siguiendo un proceso sistemático de búsqueda, cribado y selección de información. Para ello, se consultaron las bases de datos científicas SciELO, Redalyc y Google Académico, priorizando artículos originales y revisiones publicadas entre el periodo 2018 y 2025. La selección temporal respondió a la necesidad de incluir literatura

actualizada en un campo caracterizado por su rápida evolución.

El proceso de búsqueda empleó descriptores y palabras clave en español, utilizando operadores booleanos para refinar los resultados. La cadena de búsqueda principal fue: ("Big Data" OR "Datos Masivos") AND ("Ciencia Estadística" OR "Estadística Aplicada") AND ("Tendencias"). Las fases de búsqueda se dividieron en: 1) Identificación inicial de registros en las bases de datos mencionadas, 2) Cribado por título y resumen para verificar la pertinencia temática, 3) Evaluación de elegibilidad a texto completo considerando el rigor metodológico y 4) Inclusión definitiva de los estudios que sustentan la revisión. Los criterios de elegibilidad incluyeron: (a) pertinencia temática con el Big Data y la estadística, (b) publicación en revistas indexadas, (c) vigencia temporal y (d) disponibilidad de DOI o URL verificable.

La cantidad de estudios encontrados inicialmente fue de 65 documentos. Tras el cribado y la aplicación de criterios de exclusión (estudios duplicados o sin rigor científico), se seleccionaron 15 estudios reales que son los que están citados en el estudio y se encuentran en la lista de referencias con su URL o DOI. El procesamiento de la información se realizó mediante una matriz de análisis que caracterizó cada estudio por autor, fecha, título, metodología y principales hallazgos. El análisis de la información se basó en la técnica de análisis de contenido, permitiendo identificar categorías emergentes relacionadas con los métodos estadísticos avanzados y las aplicaciones

sectoriales del Big Data.

## RESULTADOS Y DISCUSIÓN

La revisión realizada permite caracterizar de manera sistemática los estudios seleccionados, los cuales sustentan el análisis de las tendencias del Big Data en la ciencia estadística. En este sentido, se presenta la tabla de caracterización de los 15 estudios incluidos, considerando autor, metodología y principales hallazgos. Este proceso de sistematización permitió identificar patrones transversales, vacíos teóricos y líneas emergentes que configuran el estado del arte en la relación entre Big Data y estadística.

**Tabla 1.** Descripción de los estudios incluidos para la revisión

<b>Autor y año</b>	<b>Título</b>	<b>Metodología</b>	<b>Principales hallazgos</b>
Gutiérrez (2018)	Big Data: Conceptos y aplicaciones	Revisión conceptual	El Big Data requiere herramientas específicas para el tratamiento de datos masivos de sensores y redes.
Ávila et al. (2024)	Métodos estadísticos avanzados y minería de datos	Revisión bibliográfica	La minería de datos y la estadística apoyada en tecnología optimizan la toma de decisiones académicas.
Lasso et al. (2022)	Evolución de la tecnología y la estadística en ingeniería	Revisión documental	La rapidez tecnológica permite el desarrollo de la estadística en diversas áreas como la mecatrónica.
Peláez (2025)	Innovación y estrategias didácticas con Big Data	Estudio de caso	El Big Data permite evaluar el rendimiento estudiantil y crear estrategias de aprendizaje individualizadas.
Toscano et al. (2025)	Uso combinado de Big Data e IA en el mundo empresarial	Revisión estratégica	La combinación de Big Data e IA fortalece la toma de decisiones estratégicas y la inversión empresarial.
Aranibar et al. (2023)	Big Data en la evolución del gobierno electrónico	Revisión documental	El uso de Big Data permite al Estado obtener patrones de comportamiento ciudadano para políticas efectivas.
Zúñiga et al. (2023)	Big Data como fenómeno tecnológico en el Marketing	Revisión de mercado	La datificación permite predecir comportamientos de consumo y crear perfiles específicos de clientes.
Medina (2024)	Big Data en la integración de variables turísticas	Revisión sectorial	Modelos predictivos en turismo permiten establecer destinos más visitados y patrones de desplazamiento.
Salazar y Del Castillo (2018)	Fundamentos estadística y evolución	de Revisión su histórica	La estadística evoluciona desde censos rudimentarios hasta la inferencia moderna para fenómenos complejos.

Autor y año	Título	Metodología	Principales hallazgos
Aceña et al. (2025)	Estadística y toma de decisiones en la era digital	Revisión técnica	El manejo de la incertidumbre mediante inferencia estadística es crucial en el análisis de datos masivos.
Quiroz et al. (2024)	Componentes y gestión del Big Data en las organizaciones	Revisión técnica	Identificación de componentes clave: lagos de datos, almacenes y canalizaciones para la estandarización.
Universidad Católica San Pablo (2026)	Tipos y elementos del Big Data	Revisión técnica	Clasificación de datos en estructurados, no estructurados y semi-estructurados para su procesamiento.
Luckin et al. (2016)	Intelligence Unleashed: An argument for AI in Education	Revisión conceptual	La analítica de datos permite una personalización de los servicios y evaluación automatizada.
Sigman y Bilinkis (2023)	Artificial: La nueva inteligencia y lo humano	Revisión crítica	El análisis de datos masivos permite rescatar patrones de comportamiento humano invisibles.
González-Calatayud et al. (2021)	El potencial de la IA para revolucionar el aprendizaje	Revisión documental	La IA y el Big Data transforman la interacción con el material de estudio y la gestión del conocimiento.

### Tendencias generales identificados en los estudios incluidos en la revisión

El análisis de los estudios evidencia que el Big Data se ha consolidado como un eje transversal en múltiples disciplinas, desde la ingeniería y la educación hasta la administración pública y el marketing. Una tendencia destacada es la convergencia entre estadística, inteligencia

artificial y minería de datos, lo que redefine los métodos tradicionales de análisis y amplía el alcance de la inferencia estadística hacia modelos predictivos y automatizados.

Los estudios de Ávila et al. (2024), Toscano et al. (2025) y González-Calatayud et al. (2021) coinciden en que la integración de técnicas estadísticas con herramientas tecnológicas

avanzadas optimiza la toma de decisiones en contextos académicos y empresariales. Este hallazgo confirma la transición de la estadística desde un enfoque descriptivo/inferencial hacia un paradigma híbrido que incorpora aprendizaje automático y analítica avanzada.

Asimismo, investigaciones como las de Aranibar et al. (2023) y Zúñiga et al. (2023) muestran que el Big Data no solo transforma los procesos analíticos, sino también los modelos de gestión pública y comercial, permitiendo identificar patrones de comportamiento ciudadano y de consumo con mayor precisión. Esto evidencia un cambio epistemológico en la forma de comprender los fenómenos sociales, ahora mediados por trazas digitales y modelos algorítmicos.

### Tipologías y elementos del Big Data

La revisión confirma que la ciencia estadística se enfrenta a tres grandes tipologías de datos: estructurados, no estructurados y semi-

estructurados (Universidad Católica San Pablo, 2026). Esta clasificación no es meramente técnica, sino que determina el tipo de modelos estadísticos aplicables y las arquitecturas computacionales necesarias para su procesamiento.

### Componentes para el manejo y procesamiento de datos

Para gestionar el Big Data de manera eficiente, la literatura revisada identifica cuatro componentes esenciales que permiten almacenar, transformar y construir grandes volúmenes de información. La tabla 2 sintetiza estos elementos, los cuales constituyen la infraestructura mínima para habilitar procesos analíticos avanzados.

**Tabla 2.** Componentes para el manejo y procesamiento de datos

Componente	Descripción
Lago de datos	Espacio para almacenar datos (estructurados o no) de forma centralizada, sin ser procesados, provenientes de diferentes fuentes.
Almacén de datos (Warehouse)	Almacena gran cantidad de datos ya extraídos y transformados, listos para el análisis estadístico.
Data Pipeline	Proceso de transformación (filtrado, enmascaramiento, agregación) para estandarizar los datos antes de su consumo.
Business intelligence	Técnicas y herramientas de software para consumir datos masivos de manera amigable y entendible para la toma de decisiones.

**Nota:** (Quiroz et al., 2024)

Estos componentes evidencian que el análisis estadístico en entornos de Big Data depende no solo de técnicas analíticas, sino también de arquitecturas tecnológicas robustas que garanticen la calidad, disponibilidad y trazabilidad de los datos.

### **Métodos estadísticos avanzados utilizados**

Los estudios revisados muestran que la estadística ha evolucionado para adaptarse a la complejidad del Big Data, integrándose con tecnologías emergentes y ampliando su alcance en disciplinas como ingeniería, ciencias sociales, economía y educación. Lasso et al. (2022) destacan que la acelerada evolución tecnológica ha impulsado el desarrollo de métodos estadísticos aplicados al análisis de datos masivos, especialmente en áreas como electrónica y mecatrónica. Esta convergencia entre estadística y tecnología constituye una tendencia clave de transformación de los procesos analíticos contemporáneos. La revisión permitió identificar los métodos estadísticos avanzados más utilizados en el análisis de Big Data, los cuales se describen a continuación.

### **Regresión avanzada y análisis multivariante**

La regresión avanzada y el análisis multivariante constituyen herramientas fundamentales en el tratamiento de datos masivos, ya que permiten modelar relaciones complejas entre múltiples variables en contextos de alta dimensionalidad. En el marco del Big Data, estos métodos han evolucionado para integrarse con tecnologías de procesamiento masivo de datos, facilitando la identificación de patrones y tendencias relevantes para la toma de

decisiones.

En este sentido, Ávila et al. (2024) sostienen que el uso de métodos estadísticos avanzados permite optimizar el análisis de grandes volúmenes de información en entornos académicos y científicos. Asimismo, Aceña et al. (2025) destacan que la inferencia estadística es clave para el manejo de la incertidumbre en la era digital, lo cual refuerza la aplicabilidad de modelos multivariantes en contextos complejos.

### **Análisis de conglomerados (Clustering)**

El análisis de conglomerados o clustering es una técnica ampliamente utilizada en el contexto del Big Data para segmentar grandes volúmenes de datos en grupos homogéneos, permitiendo identificar patrones ocultos sin necesidad de variables dependientes predefinidas. Esta técnica resulta especialmente útil en áreas como el marketing, la educación y la gestión organizacional.

Sobre esto, Zúñiga et al. (2023) señalan que el Big Data permite predecir comportamientos de consumo mediante la identificación de perfiles específicos de clientes, lo cual se fundamenta en técnicas de segmentación como el clustering. De igual manera, Quiroz et al. (2024) destacan que la adecuada gestión de los datos mediante infraestructuras como los lagos de datos y pipelines facilita la aplicación de técnicas analíticas avanzadas, incluyendo métodos de agrupamiento para la toma de decisiones estratégicas.

### **Minería de datos (Data Mining).**

La minería de datos, apoyada en el Big Data, permite a los investigadores de nivel superior

identificar patrones emergentes mediante algoritmos capaces de procesar grandes volúmenes de información. Ávila et al. (2024) señalan que el uso de esta técnica posibilita analizar grandes cantidades de información, ofreciendo a los investigadores la posibilidad de evaluar patrones que antes eran inaccesibles. La minería de datos constituye un puente entre la estadística tradicional y los algoritmos de aprendizaje automático, permitiendo análisis más profundos y predictivos.

### **Estadística Bayesiana**

La aplicación de la estadística bayesiana en el Big Data ha ganado terreno debido a su capacidad para incorporar conocimiento previo (prior) y actualizar las probabilidades a medida que se reciben nuevos datos. A diferencia de la estadística frecuentista, el enfoque bayesiano permite modelar la incertidumbre de manera más flexible, lo que es crucial en entornos de datos masivos donde la variabilidad es la norma.

Los modelos de redes bayesianas, por ejemplo, se utilizan para identificar relaciones de causalidad en sistemas complejos, como en el análisis de riesgos financieros o en la medicina personalizada (Aceña et al., 2025). La capacidad de estos modelos para manejar datos faltantes y su interpretabilidad los convierte en una herramienta preferida frente a otros algoritmos de "caja negra", lo que permite mejorar la toma de decisiones en contextos de incertidumbre.

### **Simulaciones y modelado estocástico.**

Las simulaciones y el modelado estocástico permiten analizar escenarios complejos mediante la generación de múltiples trayectorias posibles.

Medina (2024) evidencia su utilidad en el turismo para anticipar comportamientos de desplazamiento, mientras que Lasso et al. (2022) su aplicación en ingeniería para la toma de decisiones bajo incertidumbre. Las simulaciones de Monte Carlo se han vuelto indispensables para evaluar riesgos, planificar escenarios urbanos y analizar fenómenos dinámicos en tiempo real.

### **Modelos de ecuaciones estructurales.**

Los modelos de ecuaciones estructurales (SEM) permiten analizar relaciones causales complejas entre variables observadas y latentes. González-Calatayud et al. (2021) destacan que estas técnicas son fundamentales en el análisis de procesos educativos y organizacionales, ya que permiten comprender estructuras subyacentes en grandes volúmenes de datos. En el contexto del Big Data, los SEM se fortalecen gracias a su capacidad computacional para estimar modelos complejos con miles de observaciones.

### **Aprendizaje automático (Machine Learning)**

Por otro lado, el aprendizaje automático (Machine Learning) ha revolucionado la estadística aplicada al proporcionar algoritmos capaces de aprender de los datos sin ser programados explícitamente. Técnicas como los bosques aleatorios (Random Forests) y las máquinas de vectores de soporte (SVM) se han integrado en el arsenal estadístico para tareas de clasificación y regresión en grandes conjuntos de datos.

Por su parte, Toscano et al. (2025) destacan que la validación cruzada y el ajuste de hiperparámetros se han convertido en procesos estándar para garantizar la robustez de los

modelos.

La sinergia entre machine learning y estadística inferencial permite no solo predecir resultados, sino también cuantificar la confianza en dichas predicciones, lo cual es esencial en sectores como salud, seguridad y finanzas.

### **El Big Data como herramienta de transformación social**

El término Big Data, popularizado en los últimos años, hace referencia a la producción de cantidades ingentes de datos generados por múltiples redes de sensores, dispositivos y la actividad de los usuarios en Internet (Gutiérrez, 2018). Estos datos aportan información sobre multitud de procesos, desde el crecimiento de las ciudades y la contaminación del aire hasta la movilidad de la población y el consumo. Tienen en común su carácter masivo y su naturaleza distinta a los datos convencionales, requiriendo herramientas específicas para su tratamiento.

Finalmente, la estadística es el eje que posibilita convertir estos datos en conocimiento útil, superando las limitaciones de las herramientas tradicionales. Por otro lado, las simulaciones y el modelado estocástico, según Medina (2024) y Lasso et al. (2022), permiten anticipar escenarios complejos, mejorar la planificación y fortalecer la toma de decisiones estratégicas en sectores como turismo, ingeniería y gestión de riesgos.

### **DISCUSIÓN**

El análisis crítico de las tendencias del Big Data en relación con la ciencia estadística revela

una transformación profunda en la manera en que se produce, valida y aplica el conocimiento. La revisión muestra un desplazamiento desde la estadística tradicional, centrada en la descripción y la inferencia, hacia un enfoque orientado a la predicción prescriptiva, la automatización y la toma de decisiones basada en datos masivos. Este cambio implica una reconfiguración epistemológica que redefine los límites y alcances de la disciplina.

Una de las tensiones más relevantes identificadas en la literatura se relaciona con los supuestos clásicos del muestreo frente a la lógica del Big Data. La estadística tradicional se fundamenta en muestras representativas, mientras que los sistemas contemporáneos suelen trabajar con poblaciones completas o con volúmenes tan grandes que los conceptos de error muestral y significancia estadística requieren reinterpretación. En este sentido, lo planteado por Gutiérrez (2018) sobre la capacidad del Big Data para procesar datos en tiempo real coincide con lo expuesto por Ávila et al. (2024), quienes sostienen que la minería de datos puede identificar patrones emergentes que optimizan la toma de decisiones.

La significancia basada en el valor  $p$  pierde relevancia en contextos donde los tamaños muestrales son extremadamente grandes, ya que incluso efectos mínimos pueden resultar estadísticamente significativos, por ello, la tendencia actual que orienta hacia la evaluación del tamaño del efecto, la relevancia práctica y la estabilidad de los modelos, tal como sugieren los enfoques contemporáneos discutidos Aceña et al. (2025).

El avance tecnológico ha impulsado una convergencia acelerada entre estadística, ingeniería e informática. La integración de técnicas estadísticas con arquitecturas computacionales avanzadas ha dado lugar a la Ciencia de Datos como campo interdisciplinario. Esta convergencia es coherente con lo señalado por Lasso et al. (2022), quienes explican que la evolución tecnológica ha fortalecido el desarrollo estadístico en áreas como la electrónica y la mecatrónica. En este nuevo escenario, la estadística aporta el rigor metodológico, mientras que la informática provee la capacidad de cómputo necesaria para procesar datos masivos.

En el ámbito educativo, los estudios realizados muestran que el Big Data permite una personalización del aprendizaje sin precedentes. La analítica educativa facilita la identificación de patrones de rendimiento y la creación de estrategias individualizadas, como lo demuestra Peláez (2025). Esta visión se alinea con la propuesta de Luckin et al. (2016) quienes consideran que la analítica de datos es un motor de innovación pedagógica. Sin embargo, las advertencias de Sigman y Bilinkis (2023) sobre la deshumanización del análisis de datos revelan riesgos asociados a sesgos algorítmicos y decisiones automatizadas que pueden reproducir desigualdades.

Desde una perspectiva práctica, el sector empresarial también refleja esta transformación. La combinación de Big Data e Inteligencia Artificial genera ventajas competitivas significativas, permitiendo optimizar inversiones,

anticipar tendencias y dirigir estrategias con mayor precisión, como señalan Toscano et al. (2025). Este fenómeno se extiende al sector público Aranibar et al. (2023) documentan cómo el gobierno electrónico utiliza datos masivos para diseñar políticas basadas en evidencia. No obstante, esta eficiencia administrativa se enfrenta a la tensión entre el uso intensivo de datos y la protección de la privacidad ciudadana.

Desde una perspectiva metodológica, la transición hacia métodos avanzados como la estadística bayesiana, el aprendizaje automático y las simulaciones estocásticas representa un salto cualitativo. La precisión predictiva descrita por Zúñiga et al. (2023) en ámbito del marketing confirma la potencia de estas técnicas. Sin embargo, la revisión de Quiroz et al. (2024) sobre los componentes de gestión, como los lagos de datos, almacenes y pipelines, evidencia que la validez del análisis depende de la calidad de los datos y de la infraestructura que los soporta. El principio “Garbage In, Garbage Out” adquiere una relevancia crítica en entornos donde la automatización amplía errores y sesgos.

La integridad de los datos no solo implica veracidad técnica, sino también representatividad ética. Los modelos estadísticos entrenados con datos sesgados pueden perpetuar desigualdades históricas, en especial en ámbitos como la educación, salud, seguridad y empleo. Por ello, la estadística contemporánea debe incorporar principios de justicia algorítmica y responsabilidad social en todas las etapas del análisis, tal como sugieren las reflexiones éticas de Floridi et al. (2018) y la propuesta de modelos explicables de Rudin (2019).

Las implicaciones teóricas de este estudio sugieren una reconfiguración de la enseñanza de la estadística. Ya no es suficiente con dominar las pruebas de hipótesis básicas; el estadístico moderno debe ser capaz de navegar en entornos de datos no estructurados y semi-estructurados. La evolución histórica descrita por Salazar y Del Castillo (2018) nos recuerda que la estadística siempre ha sido una ciencia adaptativa. En la era del Big Data, esta adaptabilidad se traduce en la capacidad de integrar simulaciones y modelado estocástico para predecir fenómenos en sistemas complejos, como el turismo analizado por Medina (2024).

Los hallazgos del estudio señalan que el Big Data necesita de la estadística para poder procesar todo el conjunto de datos tan complejos que con herramientas tradicionales no puede, convirtiéndose en una sinergia indispensable para el progreso científico y social. Esta sinergia debe estar guiada por una visión humanista que priorice el bienestar social sobre la mera eficiencia algorítmica, asegurando que la ciencia estadística siga siendo una herramienta de emancipación y no de control.

Finalmente, las limitaciones de esta revisión documental incluyen la exclusión de literatura no indexada que podría contener innovaciones técnicas recientes, así como la dificultad de generalizar hallazgos en un campo que cambia mensualmente. Las implicaciones prácticas para los investigadores sugieren la necesidad de una formación interdisciplinaria que combine la estadística con habilidades de programación y ética de datos.

El presente estudio ha permitido analizar de manera exhaustiva las tendencias del Big Data en relación con la ciencia estadística, confirmando que esta convergencia es el pilar fundamental de la toma de decisiones en la sociedad contemporánea. Se concluye que el Big Data ha dejado de ser una simple acumulación de información para convertirse en un activo estratégico que, bajo el rigor de la ciencia estadística, permite transformar datos masivos en conocimiento accionable. Los hallazgos resaltan que la integración de métodos avanzados como la minería de datos, la estadística bayesiana y el aprendizaje automático ha permitido superar las limitaciones de los modelos tradicionales, ofreciendo una capacidad predictiva y prescriptiva sin precedentes en sectores tan diversos como la educación, el marketing, el gobierno y la ingeniería.

La relevancia de esta investigación radica en la sistematización de los componentes y tipologías de datos que rigen el ecosistema del Big Data. Se ha demostrado que la gestión exitosa de la información depende de una infraestructura sólida (lagos y almacenes de datos) y de una comprensión profunda de la naturaleza de los datos (estructurados y no estructurados). No obstante, se enfatiza que la tecnología por sí sola es insuficiente; la ciencia estadística aporta la estructura lógica y el manejo de la incertidumbre necesarios para validar los resultados y evitar interpretaciones erróneas derivadas de la complejidad algorítmica.

Como líneas de investigación futura, se recomienda explorar el impacto de la ética algorítmica y la transparencia en los modelos estadísticos aplicados al Big Data social. Asimismo, es imperativo investigar cómo la automatización del análisis de datos afectará la formación de los futuros estadísticos y científicos de datos.

En definitiva, la sinergia entre el Big Data y la ciencia estadística representa una evolución irreversible que demanda la actualización constante de los marcos teóricos, metodológicos y éticos. Asegurar que el análisis de datos masivos se mantenga al servicio del bienestar humano y el desarrollo científico constituye el desafío centra de esta nueva era datificada, donde la estadística continúa siendo la herramienta esencial para dotar de sentido, rigor y responsabilidad al vasto universo de información disponible. con rigor y ética.

## REFERENCIAS

- Aceña, V., Martín, I., y Lancho, C. (2025). Inferencia estadística. Escuela Técnica Superior de Ingeniería Informática. Obtenido de <https://burjcdigital.urjc.es/bitstreams/fdfd1671-9292-4a93-9d36-4cb009f2e5b7/download>
- Aranibar, E., Salinas, F., & Seguil, N. (2023). Explorando tendencias del devenir público: cienciometría y revisión sistemática. *Revista de Ciencias Sociales y Humanas*, XXI(39), 109-135. doi:<https://doi.org/10.17163/uni>
- Ávila, L., Alvarado, J., Munayco, J., y Reinoso, D. (2024). Métodos avanzados de estadística para la investigación en nivel superior: nuevas perspectivas y aplicaciones. *Revista Social Fronteriza*, 4(5). doi:[https://doi.org/10.59814/resofro.2024.4\(5\)494](https://doi.org/10.59814/resofro.2024.4(5)494)
- Batini, C., y Scannapieco, M. (2016). *Data and information quality: Dimensions, principles and techniques*. Springer. <https://doi.org/10.1007/978-3-319-24106-7>
- Boyd, d., & Crawford, K. (2012). Critical questions for Big Data. *Information, Communication & Society*, 15(5), 662-679. <https://doi.org/10.1080/1369118X.2012.678878>
- Davenport, T. H., & Patil, D. (2012). Data scientist: The sexiest job of the 21st century. *Harvard Business Review*, 90(10), 70-76. <https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century>
- Floridi, L., Cows, J., Beltrametti, M., Chatila, R., & Chazerand, P. (2018). AI4People—An ethical framework for a good AI society. *Minds and Machines*, 28, 689-707. <https://doi.org/10.1007/s11023-018-9482-5>
- González-Calatayud, V., Prendes-Espinosa, M. P., & Roig-Vila, R. (2021). El potencial de la IA para revolucionar el aprendizaje universitario. *Revista Iberoamericana de Educación a Distancia*, 24(3), 201-218. <https://doi.org/10.5944/ried.24.3.30235>
- Gutierrez, J. (2018). Big Data y nuevas geografías: la huella digital de las actividades humanas. *Documents d'Anàlisi Geogràfica*, 64(2), 195-217. doi:<https://doi.org/10.5565/rev/dag.526>
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An introduction to statistical learning* (2nd ed.). Springer. <https://doi.org/10.1007/978-1-0716-1418-1>
- Kelleher, J. D., Mac Namee, B., & D'Arcy, A. (2020). *Fundamentals of machine learning for predictive data analytics* (2nd ed.). MIT Press. <https://mitpress.mit.edu/9780262044691/fundamentals-of-machine-learning-for-predictive-data-analytics/>
- Lasso, L., Franco, D., y Estrada, R. (2022). Aplicaciones de la datificación y Big Data en América Latina entre el 2015 y 2019. *Revista Logos Ciencia & Tecnología*, 14(2), 125-143. Obtenido de <https://revistalogos.policia.edu.co:8443/index.php/rllct>
- Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). *Intelligence Unleashed: An argument for AI in Education*. Pearson

- Education.  
<https://www.pearson.com/content/dam/one-dot-com/one-dot-com/global/Files/about-pearson/innovation/open-ideas/Intelligence-Unleashed.pdf>
- Manzano, F., y Avalos, D. (2022). La utilidad del Big Data en las estadísticas públicas y empresas privadas. *Cad Metrop. Sao Paulo*, 24(55), 989-1005. doi:<http://dx.doi.org/10.1590/2236-9996.2022-5506>
- Medina, J. (2024). Big data en la predicción de flujos turísticos. *Visión Académica*, 2(4), 10-19. doi:<https://doi.org/10.70577/rmgp0d53>
- Mayer-Schönberger, V., & Cukier, K. (2013). *Big Data: A revolution that will transform how we live, work, and think*. Houghton Mifflin Harcourt.  
<https://www.hmhco.com/shop/books/Big-Data/9780544002692>
- Peláez, J. (2025). Implementación de big data para mejorar el análisis de indicadores de eficiencia. *Revista INVECOM*, 5(1). Obtenido de [https://ve.scielo.org/scielo.php?pid=S2739-00632025000102074&script=sci\\_arttext](https://ve.scielo.org/scielo.php?pid=S2739-00632025000102074&script=sci_arttext)
- Provost, F., & Fawcett, T. (2013). *Data Science for Business*. O'Reilly Media.  
<https://www.oreilly.com/library/view/data-science-for/9781449374273/>
- Quiroz, J., Pérez, A., Román, G., Castro, M., y Aguilar, M. (2024). Explorando la ciencia de datos: desde la estadística hasta el Big Data. *Contactos, Revista de Educación en Ciencias e Ingeniería*(137), 115-121. doi:<https://contactos.izt.uam.mx/index.php/contactos/article/view/450>
- Rivadeneira, J., Herrera, M., Casanova, C., & Bueno, F. (2022). Estadística y crecimiento empresarial: análisis bibliométrico. *AlfaPublicaciones*, 4(2), 6-20. doi:<https://doi.org/10.33262/ap.v4i2.195>
- Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1, 206–215. <https://doi.org/10.1038/s42256-019-0048-x>
- Salazar, C., y Del Castillo, S. (2018). *Fundamentos básicos de estadística* (Primera Edición ed.). Obtenido de [https://gc.scalahed.com/recursos/files/r161r/w24899w/Trabajo%20Final/Fundamentos\\_Basicos\\_de\\_Estadistica.pdf](https://gc.scalahed.com/recursos/files/r161r/w24899w/Trabajo%20Final/Fundamentos_Basicos_de_Estadistica.pdf)
- Sigman, M. y Bilinkis, S. (2023). Artificial: La nueva inteligencia y el contorno de lo humano. *Debate*. Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*, 104, 333–339. <https://doi.org/10.1016/j.jbusres.2019.07.039>
- Universidad Católica San Pablo. (2026). Big Data: definición, tipos, características y beneficios. Obtenido de [postgrado.ucsp.edu.pe/articulos/que-es-big-data/](https://postgrado.ucsp.edu.pe/articulos/que-es-big-data/)
- Zuñiga, F., Mora, D., y Llerena, W. (2023). El Big Data y su implicación en el Marketing. *Revista de Comunicación de la SEECI*(56), 302-321. doi:<http://doi.org/10.15198/seeci.2023.56.e832>