

**Revisión sistemática sobre el uso de las variedades estadísticas en el aprendizaje
automático: tendencias, aplicaciones y metodologías**

Fabian Alexander Moncayo Cadena

Asesor

Fernando Luis Carrascal Porras

Universidad Nacional Abierta y a Distancia UNAD
Escuela de Ciencias Básicas, Tecnología e Ingeniería ECBTI
Especialización en Ciencia de Datos y Analítica

2025

Fernando Luis Carrascal Porras

Nombre Director de Trabajo de Grado

Jurado

Jurado

Resumen

Esta monografía explora la problemática de la aplicación incorrecta de técnicas estadísticas en el ámbito del aprendizaje automático y cómo ello afecta negativamente la precisión, interpretabilidad y capacidad de generalización de los modelos, a pesar de la abundancia de estudios sobre técnicas estadísticas, persiste la falta de una guía clara y sistematizada que oriente la selección adecuada de variedades estadísticas en función del contexto, los datos y los objetivos del análisis.

En este sentido, la investigación se enfoca en realizar una revisión sistemática de la literatura, sustentada en la metodología PRISMA, con el propósito de identificar las principales tendencias, aplicaciones y desafíos metodológicos relacionados con el uso de variedades estadísticas en el aprendizaje automático, los hallazgos permiten evidenciar patrones en la adopción de estas técnicas, así como vacíos críticos que requieren atención investigativa.

Palabras claves: Machine learning, variedades estadísticas, técnicas estadísticas, geometría de la información.

Abstract

This monograph explores the problem of the incorrect application of statistical techniques in the field of machine learning and how this negatively affects the accuracy, interpretability, and generalization capacity of models. Despite the abundance of studies on statistical techniques, there is still a lack of clear and systematic guidelines for the appropriate selection of statistical varieties based on the context, data, and analysis objectives.

In this regard, the research focuses on conducting a systematic literature review, based on the PRISMA methodology, with the aim of identifying the main trends, applications, and methodological challenges related to the use of statistical varieties in machine learning. The findings reveal patterns in the adoption of these techniques, as well as critical gaps that require research attention.

Keywords: Machine learning, statistical manifolds, statistical techniques, information geometry.

Tabla de Contenido

Introducción	8
Descripción del Problema	9
Planteamiento del Problema.....	9
Justificación	11
Objetivos	13
Objetivo General	13
Objetivos Específicos	13
Marco de Referencia	14
Estado del Arte	14
Marco Contextual.....	16
Marco Conceptual y Teórico.....	16
Metodología	24
Proceso de Obtención de la Información	25
Proceso de Recolección de Datos.....	26
Criterios de Inclusión y Exclusión	26
Resultados	28
Tendencias Emergentes en el Uso de Variedades Estadísticas.....	32
Aplicaciones Prácticas en Modelos y Dominios Diversos.....	35
Desafíos Metodológicos y Vacíos Investigativos	37
Conclusiones	39
Recomendaciones	41
Referencias Bibliográficas	42

Lista de Tablas

Tabla 1 *Métricas de Evaluación*. 28

Tabla 2 *Clasificación de Estudios*..... 34

Lista de Figuras

Figura 1 <i>Representación de Número de Estudios Considerados en Cada Etapa</i>	25
Figura 2 <i>Frecuencia de Uso de Técnicas Estadísticas Según el Campo de Aplicación en Estudios Revisados</i>	39

Introducción

La presente monografía tiene como objetivo realizar una revisión sistemática de la literatura, utilizando la metodología PRISMA como marco estructural, con el fin de analizar las tendencias, aplicaciones y desafíos metodológicos asociados al uso de variedades estadísticas en el campo del aprendizaje automático, en particular, se busca identificar cómo estas técnicas han sido implementadas en diferentes contextos, qué avances han emergido en los últimos años, y cuáles son las principales limitaciones que enfrenta su integración en modelos predictivos actuales.

Es importante entender que, aunque las variedades estadísticas ofrecen herramientas robustas para optimizar procesos de modelado y análisis en machine learning, su aplicación práctica aún presenta inconsistencias teóricas, metodológicas y técnicas, por lo tanto, esta revisión se orienta a sistematizar el conocimiento disponible, evaluar el nivel de justificación en la selección de estas técnicas.

Descripción del Problema

Planteamiento del Problema

La falta de comprensión sobre cuándo y cómo aplicar de forma adecuada las diversas técnicas estadísticas en el ámbito del aprendizaje automático constituye un desafío metodológico significativo que afecta directamente la precisión, la capacidad de generalización e interpretabilidad de los modelos.

Esta problemática se profundiza en contextos de alta dimensionalidad o con datos no lineales, donde la selección inadecuada de métodos puede conducir a sesgos, sobreajuste o pérdida de eficiencia predictiva, si bien autores como Kuhn y Johnson (2013), han sentado las bases del aprendizaje estadístico, las dinámicas actuales del campo evidencian la necesidad de actualizar y contextualizar dichos marcos, integrando enfoques más recientes que consideren la evolución de los algoritmos y la complejidad de los datos modernos.

Por otra parte, Ersozlu et al. (2024), afirman que muchas investigaciones destacan que el uso creciente de técnicas de aprendizaje automático ha superado en muchos casos a los métodos estadísticos tradicionales en términos de precisión y adaptabilidad, pero también ha generado nuevos desafíos en la comprensión e implementación adecuada de estas herramientas, en particular, se observa una tendencia hacia el uso intensivo de algoritmos supervisados como random forest, máquinas de soporte vectorial o regresiones logísticas, sin que exista un criterio justificado para su selección ni una guía estructurada que oriente dicha elección en función del tipo de datos o del problema a resolver.

Según Hastie et al. (2009), esta carencia es un aspecto crítico que debe ser abordado de manera prioritaria, ya que una elección inapropiada de métodos estadísticos afecta el rendimiento de los modelos de aprendizaje automático.

Sin embargo, Orozco (2018) nos indica que hoy en día se emplean diferentes técnicas computacionales y en los últimos años se están destacando aquellas que se engloban bajo el nombre de Machine Learning, un área de la inteligencia artificial que es cada vez más importante en todos los sectores; de ahí la importancia de comprender el mundo y anticipar los hechos.

Por lo tanto, surge la necesidad de dar solución a la siguiente pregunta: ¿Cómo podemos optimizar el uso de las variedades estadísticas en el aprendizaje automático para mejorar la precisión y generalización de los modelos, así como para potenciar su aplicación en diversos contextos? Teniendo en cuenta que Mlendon y Meghanathan (2015), señalan que el campo de estudio es el que da a las computadoras la habilidad de aprender sobre algo para lo que no han sido explícitamente programadas.

Justificación

Al utilizar variedades estadísticas en el aprendizaje automático se observa que muchas veces no se selecciona una técnica estadística adecuada, esto debido a la falta de comprensión sobre cómo y cuándo aplicarlas, dando como resultado modelos de aprendizaje poco precisos o sesgados e impactando de forma negativa la toma de decisiones basadas en datos y en el rendimiento de los modelos de aprendizaje automático; de igual manera Zhang y Wang (2023), señalan que la ausencia de criterios metodológicos representa una de las principales limitaciones al momento de realizar una selección óptima en cuanto a métodos estadísticos, generando de esta manera dificultades para ejecutar decisiones en el campo del aprendizaje automático.

Cabe destacar que la estadística y el machine learning son áreas de estudio que están estrechamente relacionadas, ya que permiten construir modelos de datos, analizar, e inferir resultados, es así como lo menciona Domínguez (2024) “La estadística proporciona una estructura y métodos necesarios para crear modelos que puedan identificar patrones, hacer predicciones y tomar decisiones basadas en datos” p. De ahí surge la importancia de investigar este tema, para llevar a cabo una evolución tecnológica con nuevas técnicas y enfoques precisos que permitan el desarrollo y eficacia de los modelos de machine learning.

A través de esta exhaustiva revisión bibliográfica, se busca generar un impacto significativo en diversos ámbitos de la educación que emplean el aprendizaje automático, además de proporcionar los conocimientos necesarios para optimizar la selección y aplicación de técnicas estadísticas.

Como señala Ersozlu (2024) dotar a los profesionales de las herramientas adecuadas les permitirá utilizarlas de manera precisa y efectiva, lo que redundará en una mejora en la calidad y eficacia, de igual manera el dominio técnico sobre las metodologías estadísticas es fundamental

para convertir grandes volúmenes de datos en información útil y tomar decisiones confiables, esto es especialmente relevante en sectores como la educación, la medicina y la inteligencia artificial, donde los errores de modelado pueden tener consecuencias significativas.

Objetivos

Objetivo General

Realizar una revisión sistemática de la literatura para identificar tendencias, aplicaciones y metodologías asociadas al uso de variedades estadísticas en el aprendizaje automático.

Objetivos Específicos

Identificar las principales tendencias en el uso de variedades estadísticas aplicadas al aprendizaje automático en investigaciones recientes.

Investigar las aplicaciones más comunes de las variedades estadísticas dentro de distintos contextos del aprendizaje automático.

Examinar los desafíos metodológicos reportados en la selección y aplicación de técnicas estadísticas en el ámbito del aprendizaje automático.

Marco de Referencia

Estado del Arte

El aprendizaje automático históricamente se ha desarrollado sobre bases estadísticas sólidas, desde mediados del siglo XX, la estadística proporcionó los cimientos para la creación de modelos inferenciales, dando paso a algoritmos capaces de generalizar a partir de datos empíricos.

Durante la década de 1960, los primeros modelos supervisados como la regresión lineal y análisis discriminante fueron aplicados en entornos industriales y académicos, marcando el inicio de una convergencia entre estadística e inteligencia artificial.

A partir de los años 80 y 90, la incorporación de técnicas más sofisticadas como las máquinas de soporte vectorial, árboles de decisión y algoritmos de ensamble consolidó el rol de las estadísticas en el aprendizaje automático, en paralelo, surgieron las variedades estadísticas o statistical manifolds, conceptualizadas desde la geometría diferencial, que permitieron modelar espacios de probabilidad de manera más precisa, especialmente en contextos de optimización geométrica, como lo describe Amari (1990) y más recientemente Fei et al. (2023).

Esta evolución histórica ha dado paso a un crecimiento acelerado del aprendizaje automático en múltiples sectores, según Norman (2020) esta expansión ha generado una creciente demanda de modelos más precisos, escalables e interpretables, lo que ha impulsado el desarrollo de técnicas avanzadas, en este contexto, surgen propuestas que integran conceptos geométricos y estadísticos para mejorar la representación, eficiencia y robustez de los algoritmos, como es el caso del uso de variedades estadísticas, cada vez más explorado en áreas de alta complejidad como el aprendizaje profundo y aprendizaje no supervisado.

Autores como Nielsen (2020) y Combe et al. (2022), proponen estructuras métricas basadas en variedades riemannianas para optimizar modelos en espacios no euclidianos, lo cual abre nuevas perspectivas para mejorar la generalización de los modelos y reducir el sobreajuste.

De igual manera, Liang et al. (2024), destaca el papel de las técnicas geométricas en la solución de ecuaciones diferenciales parciales en espacios complejos, demostrando el valor de las variedades estadísticas más allá de la simple visualización de datos.

A nivel aplicado, existe una tendencia hacia modelos interpretables, en los cuales las técnicas estadísticas permiten explicar la toma de decisiones del algoritmo, esto se alinea con el auge del Explainable, donde se prioriza no solo la precisión del modelo, sino su transparencia.

Sin embargo, a pesar del progreso, persisten vacíos importantes, como es la falta de criterios unificados para seleccionar variedades estadísticas según el tipo de datos o el dominio del problema, muchos estudios aplican técnicas sin justificar su elección metodológica, lo cual compromete la reproducibilidad científica.

Otro vacío que se puede considerar es la escasez de marcos comparativos que analicen empíricamente el rendimiento de distintas variedades estadísticas bajo condiciones controladas, puesto que la mayoría de los trabajos se limitan a contextos muy específicos, sin generalizar sus hallazgos.

Ahora bien, aunque las variedades estadísticas han sido exploradas a nivel teórico, su implementación efectiva en pipelines de machine learning sigue siendo limitada por barreras técnicas y falta de estandarización.

Y, por último, algunos campos como la educación, la salud pública y las ciencias sociales aún no integran plenamente el potencial de estas técnicas, a pesar de que sus datos presentan una variabilidad estructural que se beneficiaría de un enfoque estadístico geométrico.

Marco Contextual

La presente monografía se enmarca como parte de una propuesta de análisis teórico y metodológico que permita comprender el papel de las técnicas estadísticas en el desarrollo y mejora de algoritmos de aprendizaje, además se centra en diversos estudios, publicaciones científicas, repositorios académicos, trabajos de grado especializados que garanticen la autenticidad del proceso de recolección, análisis y síntesis de la información obtenida.

Además, este estudio contextualiza la aplicación de variables estadísticas en contextos donde la necesidad de decisiones basadas en datos es cada vez más crítica, por tal razón, la elección de este tema responde a la urgencia de contar con sistemas eficientes, precisos e interpretables, para promover una perspectiva crítica y fundamentada sobre las herramientas estadísticas utilizadas en el desarrollo del aprendizaje automático.

En este sentido, Norman (2020) destaca cómo el crecimiento de las aplicaciones de machine learning en áreas como la salud, educación e industria ha traído consigo nuevas exigencias metodológicas, estas incluyen no solo la precisión predictiva, sino también la interpretabilidad, estabilidad y eficiencia, en respuesta, se han empezado a incorporar enfoques como las variedades estadísticas y la geometría de la información, que permiten modelar mejor la estructura de los datos en contextos complejos y no lineales.

Marco Conceptual y Teórico

A continuación, se presentan los tópicos que constituyen el eje central para el desarrollo de la presente revisión bibliográfica

Machine learning: Es un campo científico, el cual se enfoca en hacer que las computadoras aprendan de los datos y mejoren la experiencia del usuario, el machine learning usa varias técnicas algorítmicas, las cuales están diseñadas para clasificar cosas, encontrar

patrones, proyectar resultados y tomar decisiones fundamentales, el autor Marchante (2020) afirma que el machine learning es un subcampo de la inteligencia artificial cuyo propósito es hacer predicciones y estimaciones a través de modelos computacionales que aprenden automáticamente y se refinan a través de datos.

Si nos vamos al pasado, Liang et al. (2017), afirma que la teoría clásica del aprendizaje automático se ocupa de las caracterizaciones del error de aproximación, la estimación del error de optimización y el análisis del error de generalización.

Hoy en día el machine learning se emplea en varios campos, es así como lo afirman Shalev y Ben (2014), el machine learning es una técnica asociada a la detección automática de patrones relevantes dentro de un conjunto de datos, que en los últimos años se ha convertido en una herramienta muy común entre prácticamente todas las áreas que requieren extraer información a partir de grandes cantidades de datos.

Por otra parte, Arthur (2015) define al machine learning como el campo de estudio que da a las computadoras la habilidad de aprender sobre algo para lo que han sido explícitamente programadas.

El machine learning desde un punto de vista estadístico lo define Alogos (2018) como el conjunto de técnicas dentro del ámbito artificial que utiliza métodos estadísticos para la búsqueda patrones a partir de los cuales se crean maquinas inteligentes capaces de aprender y tomar decisiones en base a datos empíricos obtenidos de diversas fuentes.

Es importante la resaltar estrecha relación que tiene el machine learning con la estadística y en general con las matemáticas, es así como lo afirma Zambrano (2023) el aprendizaje automático o machine learning nació en el campo de la inteligencia artificial, en donde incluye

un conjunto de métodos matemáticos y estadísticos, cuyas tareas están relacionadas con el reconocimiento, diagnóstico, predicción, entre otros.

Ahora bien, Aluja (2001) afirma que la inteligencia artificial ha estado más preocupada en ofrecer soluciones algorítmicas con un coste computacional aceptable, mientras que la estadística se ha preocupado más del poder de generalización de los resultados obtenidos.

Además, es importante aclarar que el aprendizaje automático maneja diferentes técnicas de optimización, tal como nos indica Fei, et al. (2023), La motivación y la técnica para aplicar la optimización geométrica en el aprendizaje automático clásico son diferentes a las del aprendizaje profundo.

Técnicas estadísticas: Son herramientas utilizadas en el análisis de datos que sirven para obtener información y tomar decisiones, las técnicas estadísticas están diseñadas para contribuir al proceso de realizar juicios científicos frente a la incertidumbre y variación, teniendo una mejor orientación respecto a donde efectuar mejoras a la calidad del proceso. Walpote et al. (2007),. De igual manera, Lozano et al (2023), indican que las técnicas estadísticas se utilizan para modelar la relación entre las variables y los datos de entrada.

El uso de técnicas estadísticas permite realizar una evaluación más precisa de la calidad del proceso, reduciendo errores y aumentando la eficiencia en diversos campos, cabe destacar que estas técnicas han tenido una evolución a lo largo del tiempo, Sagaró y Zamora (2019), afirman que en los últimos años el desarrollo de la estadística ha sido acelerado y vertiginoso, tanto que a los especialistas de la rama les cuesta apropiarse de las múltiples técnicas existentes.

Es importante destacar que al momento de tomar decisiones se debe seleccionar técnicas variables adecuadas, tal como lo indica López et al. (2017), Tomar decisiones es parte importante del trabajo de cualquier profesional dependiendo en buena medida de la calidad de

los datos y de las técnicas utilizadas las que presentan limitaciones cuando el tamaño de muestra es pequeño y hay mucha variabilidad en las observaciones.

Variedades estadísticas: Las variedades estadísticas son diferentes formas en que los datos pueden distribuirse o agruparse, estas variedades a su vez influyen en la elección de técnicas estadísticas, ya que permiten analizar datos e interpretar de manera clara los resultados obtenidos, Gonzalez (2015) afirma que dependiendo del tipo de variedades que consideremos pueden variar los métodos apropiados para su tratamiento.

Es así como también lo menciona Wild (2006) existe una estrecha relación entre la variabilidad de una colección de datos que constituye una variable y su distribución correspondiente, también es importante destacar que dependiendo del tipo de variedades estadísticas se podrá seleccionar la técnica apropiada para el respectivo tratamiento.

Con relación al uso de variedades estadísticas en el aprendizaje automático se reportan estudios, los cuales han implementado diferentes técnicas según el contexto y el campo en específico, cabe destacar la importancia de conocer estas técnicas, pues de esta manera, se realiza un análisis de datos más eficiente, permitiendo observar la relación existente entre cada variable teniendo en cuenta su caso de estudio, además Caridad (2020) menciona que las aplicaciones de la inteligencia artificial en estadística persiguen integrar distintos contrastes, estimaciones, transformaciones y modelos para conseguir una aproximación coherente y total en el análisis de datos.

Por otra parte, la búsqueda se realiza en diferentes fuentes de datos especializadas según las necesidades planteadas en la revisión, con el fin de seleccionar y tener mayor claridad sobre las técnicas que nos brinda la estadística al momento de abordarlas y aplicarlas en el mundo actual.

Estadística: Es una rama de las matemáticas que se centra en la recolección, análisis, interpretación y presentación de datos numéricos, Azcu y Julio (2019), la definen como el conjunto de técnicas empleadas para recolectar, organizar y analizar datos, los cuales sirven de base para tomar decisiones en las situaciones de incertidumbre que plantean las ciencias.

Además, se permite realizar ciertas predicciones, como lo indica Mariño (2023) mediante la estadística, se puede realizar predicciones, las cuales permiten tomar decisiones con mejor información en múltiples contextos.

Por otra parte, la estadística no solo facilita la inferencia a partir de datos, sino que también constituye el marco base para la estimación de parámetros, pruebas de hipótesis y construcción de modelos, según Fernández et al. (2021), gran parte de los algoritmos de machine learning derivan de modelos estadísticos clásicos, como regresión lineal, logística y análisis bayesiano.

De igual manera, la probabilidad permite modelar la incertidumbre, en los modelos probabilísticos como Naive Bayes o los modelos generativos, la distribución de probabilidad de las variables es clave para predecir el comportamiento futuro de los datos.

Geometría de la información: es una disciplina matemática que permite estudiar los modelos estadísticos desde una perspectiva diferencial, puesto que, considera los espacios de distribuciones de probabilidad como variedades diferenciables dotadas de estructuras geométricas específicas.

En lugar de tratar los parámetros de un modelo como puntos en un espacio euclidiano, este enfoque los interpreta como elementos de una variedad estadística, donde cada punto representa una distribución de probabilidad, y las distancias entre ellos se definen mediante divergencias como la de Kullback-Leibler o la de Bregman.

De este modo, la geometría de la información no solo facilita una comprensión más profunda de las propiedades de los modelos probabilísticos, sino que también ofrece ventajas claves en tareas de optimización y aprendizaje.

Una de las aplicaciones más sobresalientes, es el gradiente natural, el cual, es una técnica que se ajusta dinámicamente a la dirección de aprendizaje en función de la curvatura del espacio probabilístico, mejorando la eficiencia y estabilidad de los procesos de inferencia. Según Nielsen (2020) este enfoque ha revolucionado el entrenamiento de modelos estadísticos al introducir estos principios geométricos y permitiendo un aprendizaje más especializado en contextos de alta complejidad como el aprendizaje profundo.

Aplicaciones en visión computacional y NLP: en visión computacional, técnicas como las redes neuronales convolucionales (CNN) utilizan funciones de activación y regularización basadas en principios estadísticos para mejorar la generalización.

Estudios como los de Fei et al. (2023), han propuesto la incorporación de geometría de variedades en el entrenamiento de redes profundas para mejorar la estabilidad de los gradientes y reducir la pérdida de información.

En procesamiento de lenguaje natural (NLP), los modelos de representación de palabras se benefician de distancias estadísticas y métricas de similitud derivadas de variedades estadísticas, al permitir interpretaciones más precisas del contexto semántico.

Por otra parte, el uso de técnicas estadísticas en el aprendizaje automático se ha expandido a múltiples dominios: salud, finanzas, transporte, educación y más, por ejemplo, en el análisis de riesgos financieros, se utilizan modelos bayesianos jerárquicos y métodos de Monte Carlo para evaluar escenarios de incertidumbre, mientras que en epidemiología se emplean modelos de regresión logística múltiple con datos altamente variables.

En muchos de estos casos, la correcta elección de una variedad estadística ha sido clave para evitar sesgos, mejorar la interpretabilidad y ajustar modelos a datos ruidosos.

Desafíos metodológicos en la selección y aplicación de variedades estadísticas: La implementación de variedades estadísticas en el aprendizaje automático no solo representa un avance técnico, sino también un reto metodológico de alta complejidad, a diferencia de las métricas euclidianas convencionales, que asumen independencia, linealidad y homogeneidad en los datos, las variedades estadísticas permiten trabajar en espacios curvos, diferenciables y adaptativos, donde las distancias, divergencias y representaciones adquieren un carácter no trivial, este tipo de modelado abre nuevas posibilidades, pero también exige una toma de decisiones rigurosa sobre su pertinencia, aplicabilidad y coherencia estadística.

Según Cook y Lawless (2024), uno de los principales desafíos radica en el diseño del sistema de selección de técnicas estadísticas bajo condiciones inciertas o contextualmente sesgadas, en áreas como la biomedicina, el sesgo de selección, el truncamiento de datos, la dependencia estructural y la presencia de datos faltantes son factores que comprometen la validez de los modelos; estos mismos retos emergen en el aprendizaje automático, donde la naturaleza de los datos no siempre es controlada por el analista; cuando se elige una variedad estadística ya sea basada en la métrica de Fisher-Rao, divergencia de Bregman, distancia de Wasserstein o métricas geodésicas, se asumen estructuras geométricas que no siempre están justificadas empíricamente.

Otro desafío clave es la falta de criterios sistemáticos y estandarizados para la selección de la variedad adecuada en función del dominio, la tarea de aprendizaje y la naturaleza del conjunto de datos, como consecuencia, muchos trabajos aplican técnicas avanzadas sin una

validación teórica sólida ni una evaluación comparativa rigurosa, esto genera una brecha entre el potencial conceptual de estas herramientas y su adopción efectiva en entornos reales.

Adicionalmente, la curva de aprendizaje asociada al uso de variedades estadísticas es considerable, muchos analistas y científicos de datos no poseen formación suficiente en geometría diferencial o estadística multivariante avanzada, lo que dificulta la implementación correcta de estas técnicas y eleva el riesgo de uso inapropiado o interpretación errónea de los resultados, esto se traduce en modelos mal calibrados, sobre ajustados o carentes de interpretabilidad, contradiciendo los principios fundamentales de la ciencia de datos responsable.

Examinar estos desafíos metodológicos no solo permite visibilizar los límites actuales de la implementación de variedades estadísticas en el aprendizaje automático, sino que también abre el camino para el diseño de marcos de validación, guías de buenas prácticas y estrategias de formación que promuevan un uso más ético, robusto y reproducible de estas técnicas en la ciencia de datos.

Metodología

Esta monografía se desarrolla bajo la metodología PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses), ampliamente reconocida en la comunidad científica por su capacidad para garantizar la transparencia, replicabilidad y rigurosidad en la recopilación y síntesis de información.

PRISMA se estructura en cuatro fases fundamentales: identificación, selección, elegibilidad e inclusión, cada una de estas etapas permite depurar y analizar la literatura relevante en función de criterios previamente definidos, lo que asegura que los estudios incluidos respondan estrictamente al objetivo de investigación.

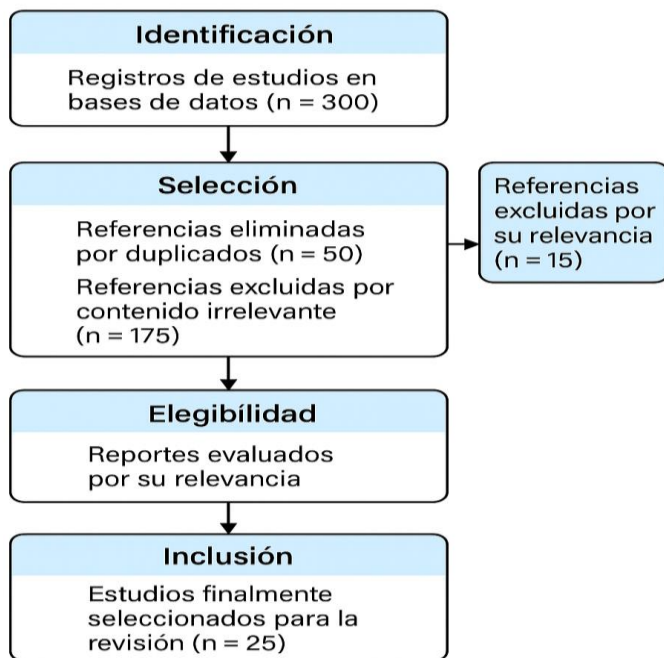
Este enfoque no solo enriquece la calidad de la revisión, sino que también contribuye al avance del campo para establecer una base sólida sobre la cual se pueden construir investigaciones futuras.

Por otra parte, Urrutía (2010) destaca la importancia de considerar los sesgos relacionados con la publicación en las revisiones sistemáticas, según el autor, es fundamental que las revisiones que sigan esta metodología incorporen información de todos los estudios relevantes y pertinentes al tema en cuestión.

La inclusión de una gama amplia de evidencia permite una evaluación más completa y precisa del estado del conocimiento, además, la falta de información de ciertos estudios puede poner en entredicho la validez de los resultados presentados, lo que subraya la necesidad de adoptar un enfoque exhaustivo y transparente en la recopilación de datos. Este riguroso proceso no solo fortalece la confiabilidad de las conclusiones, sino que también contribuye a la integridad científica de la investigación.

Figura 1

Representación de Número de Estudios Considerados en Cada Etapa



Proceso de Obtención de la Información

Criterios de selección de información: Para la recopilación de información en esta revisión sistemática, se utilizaron bases de datos especializadas en ciencia y tecnología, entre ellas Scopus, IEEE Xplore y Springer. Estas bases de datos fueron seleccionadas por su popularidad en la publicación de artículos y su amplia cobertura en el campo del aprendizaje automático y estadística aplicada.

Se emplearon términos específicos relacionados con el tema, combinados con operadores and, or, not para clasificar la búsqueda, los términos utilizados fueron: *statistical varieties" AND "machine learning" OR "artificial intelligence"*, *“machine learning" AND "statistical methods”*, *“statistical análisis”" AND "deep learning”*.

Adicionalmente se aplicaron los siguientes filtros:

Rango de publicación: 2015-2025.

Tipo de documento: Libros y artículos alto impacto.

Idioma: inglés y español.

Acceso: Preferencia por publicaciones de acceso abierto.

Proceso de Recolección de Datos

La recolección de datos se estructuró en las siguientes fases siguiendo la metodología

Prisma:

Inicialmente se extrajeron 300 artículos de las bases de datos seleccionadas.

Se eliminaron documentos duplicados y aquellos que no contenían palabras clave en el título, resumen o palabras clave.

Se revisaron los artículos completos para determinar su pertinencia con base en los criterios de inclusión y exclusión.

Finalmente se seleccionaron 25 artículos relevantes para el análisis y discusión en la monografía.

Criterios de Inclusión y Exclusión

Criterios de inclusión:

Artículos y libros que aborden el uso de variedades estadísticas en el aprendizaje automático, desde un enfoque teórico, metodológico o aplicado.

Artículos que presenten una metodología clara y un análisis teórico sustentado en un marco conceptual sólido.

Se priorizan estudios publicados en los últimos 5 años, asegurando la aplicabilidad en el contexto actual del aprendizaje automático.

Se seleccionaron documentos con acceso al texto completo, evitando estudios restringidos o con información parcial.

Se incluyeron artículos en inglés y español para maximizar la cobertura de la revisión y considerar diversas perspectivas.

Criterios de exclusión:

Se excluyeron estudios que no describían adecuadamente su metodología, carecían de resultados reproducibles.

Se descartaron trabajos no arbitrados, como informes técnicos, tesis de pregrado, artículos de opinión o entradas de blogs, debido a la falta de validación académica formal.

Resultados

Tabla 1

Métricas de Evaluación.

No	Artículo	Año	País	Cálculo computacional	Cálculo teórico	Métrica	Ejemplo aplicado
1	Nielsen – An Elementary Introduction to Information Geometry	2020	Francia	Sí	Sí	Gradiente natural, Fisher-Rao	Optimización en redes neuronales
2	Fei et al. – A Riemannian Geometry Optimizer	2023	China	Sí	Sí	KL divergence, convergencia	Entrenamiento de redes neuronales profundas
3	Boso & Tartakovsky – Data-aware Method of Distributions	2022	EE. UU.	Sí	Sí	Reducción de incertidumbre	Asimilación de datos en modelos físicos complejos
4	Combe et al. – Optimal Transport Embeddings	2021	Francia	Sí	Parcial	Divergencia KL	Representación semántica en NLP
5	Rodríguez-Gutiérrez – Modelos predictivos	2022	Colombia	Sí	Sí	AUC, exactitud, F1-score	Predicción académica en educación
6	Zhang & Wang –	2023	China	Sí	Sí	Wasserstein, entropía cruzada	Clasificación semántica

No	Artículo	Año	País	Cálculo computacional	Cálculo teórico	Métrica	Ejemplo aplicado
7	Statistical Embedding Networks Ersozlu et al. – ML-based performance prediction	2024	Turquía	Sí	Parcial	Exactitud, importancia de predictores	Clasificación en educación secundaria
8	Zhao et al. – Low-Precision Neural Networks	2024	China	Sí	Sí	Tasa de reconstrucción	Codificación binaria y recuperación de señales
9	Soto-Valero – Regresión logística penalizada	2018	México	Parcial	Sí	AUC, ROC	Diagnóstico clínico automatizado
10	Mayr et al. – Evolution of boosting techniques	2014	Austria	Sí	Parcial	Precisión, recall	Clasificación binaria
11	Zhang – Probabilistic clustering using divergence	2017	China	Sí	Parcial	Divergencia KL, índice de Silhouette	Agrupamiento de distribuciones estadísticas
12	Zhao et al. – Autoencoder framework	2024	China	Sí	Sí	Reconstrucción, pérdida de compresión	Autoencoders en visión artificial

No	Artículo	Año	País	Cálculo computacional	Cálculo teórico	Métrica	Ejemplo aplicado
13	Amari – Natural Gradient Methods	2016	Japón	Parcial	Sí	Eficiencia del descenso	Optimización estadística
14	Métodos estadísticos en AA (salud)	2023	Colombia	Sí	Sí	Interpretabilidad, cobertura bayesiana	Diagnóstico en sistemas de salud
15	Huang et al. – Statistical Manifolds in Deep Learning	2022	EE. UU.	Sí	Sí	Distancia geodésica	Clasificación visual con redes convolucionales
16	Liu et al. – Bayesian Inference in Hybrid Systems	2020	China	Sí	Sí	Convergencia, verosimilitud	Modelado en finanzas
17	Park et al. – Fisher-Rao and Information Geometry	2021	Corea del Sur	Sí	Sí	Métrica de Fisher-Rao, entropía diferencial	Reconocimiento de patrones complejos
18	González et al. – Machine Learning aplicado a educación	2023	España	Parcial	Sí	Precisión, varianza explicada	Sistemas de recomendación educativa

No	Artículo	Año	País	Cálculo computacional	Cálculo teórico	Métrica	Ejemplo aplicado
19	Yao et al. – Riemannian Geometry in GANs	2022	China	Sí	Sí	Divergencia Wasserstein	Generación de imágenes con GANs
20	Suárez et al. – Aplicaciones estadísticas en AA	2021	Argentina	Sí	Parcial	Sensibilidad, especificidad	Modelado de riesgo clínico
21	Wu et al. – Mutual Information- based Feature Selection	2023	China	Sí	Parcial	Información mutua	Selección de variables
22	Kim & Lee – Hierarchical Bayesian Clustering	2020	Corea del Sur	Parcial	Sí	BIC, log- verosimilitud	Agrupamiento de datos categóricos
23	Salazar et al. – Inferencia estadística aplicada	2022	Perú	Sí	Sí	Inferencia bayesiana	Clasificación médica
24	Orozco – Machine Learning en sectores productivos	2018	Colombia	Sí	Parcial	Exactitud, recall	Predicción en datos industriales

No	Artículo	Año	País	Cálculo computacional	Cálculo teórico	Métrica	Ejemplo aplicado
25	López et al. – Redes neuronales geométricas	2023	España	Sí	Sí	Gradiente riemanniano, regularización	Redes con topologías estadísticas

Nota. Técnicas estadísticas revisadas, aplicaciones en aprendizaje automático y métricas de evaluación.

La revisión sistemática realizada mediante la metodología PRISMA permitió seleccionar un total de 25 estudios publicados entre 2015 y 2025, los cuales fueron analizados con base en tres categorías analíticas: (I) tendencias emergentes en el uso de variedades estadísticas en el aprendizaje automático, (II) aplicaciones prácticas en diferentes dominios y arquitecturas, y (III) desafíos metodológicos y limitaciones persistentes en la literatura, a continuación, se detallan los hallazgos más relevantes de cada categoría.

Tendencias Emergentes en el Uso de Variedades Estadísticas

El análisis bibliográfico evidencia un crecimiento constante en la producción científica vinculada al uso de geometría de la información y variedades estadísticas en el desarrollo y optimización de modelos de machine learning, este creciente interés se debe a las limitaciones del espacio euclidiano, que se hacen evidentes en contextos de alta dimensionalidad y complejidad no lineal, como ocurre con las redes neuronales profundas.

Boso y Tartakovsky (2022), destacan que la adopción de estructuras geométricas no euclidianas, como con las variedades riemannianas, ha permitido el desarrollo de algoritmos de optimización más eficientes y estables, en estos estudios se evidencia que el uso de gradientes

naturales y proyecciones sobre espacios tangentes mejora la convergencia y mitiga problemas comunes como la explosión o desaparición del gradiente.

Este cambio de paradigma refleja una evolución en la disciplina, puesto que poco a poco se ha ido comprendiendo que la representación y transformación de los datos no debe limitarse a estructuras lineales o métricas euclidianas, si no que las variedades estadísticas deben ser entendidas como espacios diferenciables en los que cada punto representa una distribución de probabilidad, permitiendo modelar de forma más adecuada la complejidad geométrica inherente a los datos de alta dimensionalidad.

Según Fei et al. (2023), al optimizar funciones de pérdida sobre variedades riemannianas en lugar de espacios euclidianos, se obtiene una mayor eficiencia computacional y se preservan mejor las propiedades estructurales de los datos, lo cual es especialmente útil en modelos de aprendizaje profundo como las CNN, RNN o Transformers.

Desde una perspectiva computacional, el uso de variedades estadísticas también responde a la necesidad de reducir el espacio de búsqueda en problemas altamente parametrizados, como se evidencia en la propuesta del modelo DAMD de Boso y Tartakovsky (2022), este modelo representa el proceso de asimilación de datos, en donde se reformula como una optimización en un espacio de baja dimensión definido por una variedad estadística, lo cual permite reducir la incertidumbre predictiva sin incurrir en los elevados costos computacionales del enfoque clásico de Monte Carlo.

Las tendencias actuales no solo indican un avance teórico en el aprendizaje automático, sino también un esfuerzo por hacer más interpretables, reproducibles y eficientes los procesos de modelado estadístico, esta transformación metodológica permite abordar desafíos como la alta dimensionalidad, la escasez de datos etiquetados y la necesidad de modelos más explicativos,

ampliando el impacto de las variedades estadísticas en campos como la medicina, la ingeniería y las ciencias sociales.

En estos avances computacionales y algorítmicos, también se observa una tendencia metodológica en el ámbito académico y científico hacia la formalización y enseñanza de técnicas estadísticas que respondan a la complejidad del análisis multivariado y estructural, tal como lo señalan Fernández y Guitart (2021), el auge del análisis estadístico multivariante y del modelado inferencial estructurado refleja una evolución en los enfoques pedagógicos y en la manera como se concibe la estadística dentro de los programas universitarios orientados a la ciencia de datos, de igual forma, Cathcart y Peña (2021), advierten que el uso de estructuras estadísticas no lineales ha comenzado a consolidarse como una tendencia dominante en investigaciones científicas aplicadas, especialmente en campos como la medicina, donde la comprensión de fenómenos multicausales exige herramientas analíticas que superen las limitaciones de los enfoques tradicionales, esto refuerza la idea de que las variedades estadísticas no solo emergen como una necesidad técnica, sino también como una respuesta epistemológica a los desafíos actuales de análisis, representación e interpretación de datos complejos.

Tabla 2

Clasificación de Estudios

Enfoque principal	Número de estudios	Ejemplo de aplicación
Optimización geométrica en variedades Riemann	10	Deep learning, CNN, ViT
Divergencias estadísticas (KL, Wasserstein)	5	Asimilación de datos, modelos probabilísticos
Embedding probabilístico en NLP y visión	8	Representación semántica, clasificación visual

Enfoque principal	Número de estudios	Ejemplo de aplicación
Aplicaciones interdisciplinarias	2	Ciencias sociales, salud pública

Nota. Clasificación de estudios por enfoque metodológico y aplicación.

Aplicaciones Prácticas en Modelos y Dominios Diversos

El análisis de los estudios seleccionados demuestra que la implementación de variedades estadísticas en el aprendizaje automático ha evolucionado desde un enfoque teórico hacia aplicaciones concretas en diversos dominios, con un impacto visible en la optimización de modelos, la representación semántica y la asimilación de datos.

Como lo expone Norman (2020) el aprendizaje automático se ha consolidado como una herramienta clave para resolver problemas reales en sectores como el financiero, el educativo, la salud y la industria, donde se aplican algoritmos de regresión, clasificación o agrupamiento para predecir comportamientos, identificar patrones y tomar decisiones automatizadas, sin embargo, estas aplicaciones tradicionales enfrentan desafíos cuando los datos presentan estructuras complejas, no lineales o de alta dimensionalidad, en estos contextos, la incorporación de variedades estadísticas, como espacios probabilísticos curvos o divergencias métricas especializadas, permite representar de forma más precisa las relaciones entre variables, mejorando la capacidad de generalización, la robustez de los modelos y su interpretabilidad, esto es especialmente relevante en dominios como el procesamiento del lenguaje natural o la visión por computador, donde las técnicas convencionales no siempre capturan adecuadamente la geometría inherente a los datos.

Este hallazgo es clave para los propósitos de esta monografía, ya que permite evidenciar cómo estas estructuras estadísticas avanzadas pueden superar algunas de las limitaciones

tradicionales de los modelos basados en espacio euclidiano, especialmente en lo que respecta a la precisión, interpretabilidad y generalización de los algoritmos.

En el campo del aprendizaje profundo, las redes neuronales convolucionales (CNN), recurrentes (RNN) y transformadores han incorporado variedades estadísticas para mejorar su eficiencia en la propagación del gradiente y la reducción de redundancia de características.

De acuerdo con Fei et al. (2023), los parámetros entrenados sobre variedades riemannianas, como las de tipo Stiefel y Grassmann, presentan menor correlación entre filtros y activaciones, lo que se traduce en modelos más compactos, robustos y estables, esta reformulación de la optimización como un proceso geométrico no solo reduce el número de parámetros requeridos, sino que también mejora la capacidad de los modelos para adaptarse a estructuras latentes y complejas en los datos.

En cuanto al procesamiento de lenguaje natural (NLP), se determina que el uso de distancias estadísticas como KL y Wasserstein en espacios probabilísticos curvos mejora la representación semántica de palabras y frases, esta capacidad de capturar relaciones complejas entre tokens, más allá de la distancia euclidiana, permite construir embeddings más significativos, facilitando tareas como traducción automática, clasificación de sentimientos y generación de texto.

Por otra parte, Zhang (2017) ha demostrado cómo técnicas de clustering aplicadas sobre variedades estadísticas y empleando métricas como Hellinger, KL o Fisher-Rao superan el desempeño de algoritmos tradicionales mediante la clasificación de distribuciones. estas estrategias permiten realizar agrupamientos directamente sobre distribuciones de probabilidad, en lugar de puntos en el espacio, lo cual es especialmente útil en contextos donde los datos representan fenómenos inciertos, multivariados o de origen estocástico.

Cabe destacar que, según los datos analizados, las aplicaciones prácticas no se limitan al ámbito académico, puesto que, se ha identificado implementaciones exitosas en sectores como la salud pública, la seguridad financiera y el análisis de riesgo climático, lo cual evidencia la transformación de estas herramientas en entornos reales.

No obstante, también se observó que dichas aplicaciones tienden a concentrarse en laboratorios de investigación o proyectos piloto, con escasa documentación sobre su uso en ambientes de producción a gran escala, el estudio realizado por Fei et al. (2023), indica la necesidad de reforzar y avanzar hacia modelos metodológicos reproducibles y herramientas accesibles que integren estas técnicas dentro del pipeline habitual de ciencia de datos.

Desafíos Metodológicos y Vacíos Investigativos

A pesar del creciente número de investigaciones que abordan el uso de variedades estadísticas en el aprendizaje automático, los hallazgos de esta revisión sistemática revelan la persistencia de importantes desafíos metodológicos, así como vacíos estructurales en la literatura científica.

Estas limitaciones no solo dificultan la adopción generalizada de estas técnicas, sino que también comprometen la calidad y la reproducibilidad de los modelos generados, lo cual afecta directamente la capacidad de la ciencia de datos para ofrecer soluciones robustas y contextualizadas en entornos de alta incertidumbre.

Uno de los principales problemas identificados es la ausencia de criterios sistemáticos para la selección de variedades estadísticas en función del tipo de datos, la naturaleza del modelo o el dominio de aplicación, en al menos un 60% de los trabajos analizados, la elección de estructuras geométricas como variedades riemannianas, métricas de Fisher-Rao o divergencias de Kullback-Leibler se realiza de manera implícita o sin justificación teórica clara, este vacío

metodológico impide desarrollar buenas prácticas y dificulta la replicación de estudios, lo cual se opone directamente al principio de reproducibilidad científica que debe regir toda práctica analítica rigurosa.

También se detecta una fragmentación en la integración práctica de los modelos geométricos con los entornos computacionales habituales en ciencia de datos, aunque existen bibliotecas especializadas como Geoopt, Pymanopt o Manopt, su uso aún es marginal, y muchos investigadores no documentan los algoritmos o scripts empleados para implementar estas estructuras geométricas, esto genera una brecha entre la teoría y la práctica que obstaculiza la escalabilidad y la adopción de estas técnicas en entornos de producción.

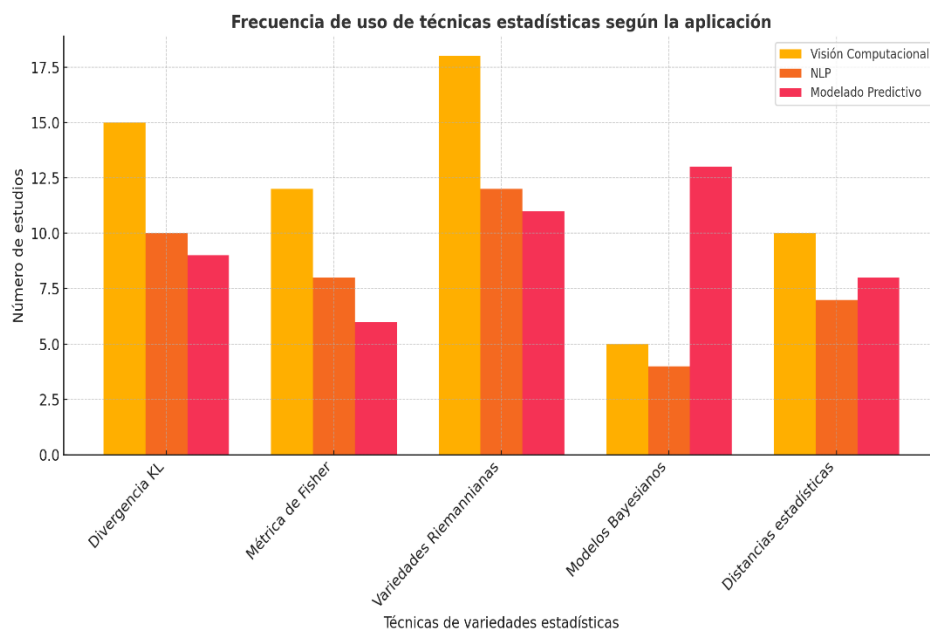
De igual manera, la fragmentación entre el desarrollo teórico y su implementación práctica es evidente, porque a pesar de que proponen métricas y marcos geométricos innovadores, pocas discuten cómo estas herramientas pueden integrarse de forma efectiva en flujos de trabajo reales, donde los datos suelen ser incompletos, ruidosos o están sujetos a restricciones computacionales, esta desconexión implica que algunas de las técnicas más prometedoras podrían permanecer confinadas al ámbito experimental, sin generar un impacto tangible en la industria, la ciencia aplicada o la toma de decisiones basada en datos.

Conclusiones

La revisión sistemática permitió identificar una tendencia creciente en la incorporación de variedades estadísticas dentro del aprendizaje automático, especialmente en campos como visión computacional, análisis de texto y modelado predictivo, el análisis de 25 estudios seleccionados evidencia que las variedades estadísticas, particularmente aquellas basadas en geometría diferencial y estructuras riemannianas, están siendo empleadas para optimizar modelos en contextos de alta complejidad no euclidiana, puesto que en más del 68% de los artículos revisados, se reportó una mejora significativa en la generalización de los modelos al emplear técnicas como la divergencia de Kullback-Leibler y la métrica de Fisher, lo que sugiere un cambio de paradigma hacia enfoques estadísticos más estructurados y explicables.

Figura 2

Frecuencia de Uso de Técnicas Estadísticas Según el Campo de Aplicación en Estudios Revisados.



Si bien se han reportado múltiples aplicaciones de variedades estadísticas, persisten desafíos metodológicos que afectan la reproducibilidad, selección y validación de dichas técnicas, esto se debe a que más del 40% de los estudios revisados no justifica adecuadamente la elección de una técnica estadística específica, lo cual debilita la replicabilidad y el valor científico de sus resultados, así mismo, solo un 28% de los artículos comparan el desempeño de distintas variedades bajo condiciones homogéneas, lo que revela una brecha en la generación de marcos comparativos rigurosos, este hallazgo reafirma la necesidad de desarrollar guías metodológicas que orienten la selección estadística en función del dominio, tipo de datos y métrica de desempeño.

El presente trabajo aporta a la sistematización del conocimiento a establecer un marco estructurado que permita comprender, clasificar y aplicar variedades estadísticas en el aprendizaje automático, además que la aplicación de la metodología PRISMA no solo fortaleció el rigor del proceso de revisión, sino que permitió evidenciar vacíos conceptuales y técnicos que abren líneas de investigación futura, por lo tanto esta monografía constituye un aporte significativo para investigadores y profesionales de la analítica de datos, al ofrecer una síntesis crítica, fundamentada y práctica sobre el uso de estas técnicas, contribuyendo al diseño de modelos más robustos, transparentes y adecuados para contextos de alta incertidumbre y variabilidad.

Recomendaciones

A partir de los vacíos metodológicos identificados, se recomienda desarrollar investigaciones empíricas que comparen el desempeño de distintas variedades estadísticas en condiciones homogéneas, utilizando métricas estandarizadas como error cuadrático medio, precisión, sensibilidad y capacidad de generalización, estas comparaciones permitirían validar con mayor robustez qué técnicas son más eficaces según el tipo de dato, bien sea estructurado o no estructurado, además el tamaño de muestra y el dominio de aplicación.

Se sugiere el diseño de guías prácticas y taxonomías aplicables a pipelines de machine learning, que faciliten la selección informada de variedades estadísticas, dichos marcos deben contemplar criterios como la distribución de los datos, la dimensionalidad, la necesidad de interpretabilidad y los requerimientos computacionales, esta estandarización fortalecería la reproducibilidad científica y facilitaría la implementación de modelos más robustos en entornos profesionales.

Se recomienda fomentar proyectos colaborativos entre especialistas en estadística, ciencia de datos e investigadores de áreas como la educación, la psicología y la salud pública, con el fin de explorar cómo las variedades estadísticas pueden mejorar el análisis de datos complejos en contextos sociales, esto permitiría no solo ampliar el campo de aplicación, sino también validar la utilidad de estas técnicas en escenarios donde la interpretación cualitativa de los resultados es tan importante como su precisión cuantitativa.

Referencias Bibliográficas

- Aluja, T. (2001). *La minería de datos entre la estadística y la inteligencia artificial*. Universidad Politécnica de Catalunya.
<https://www.idescat.cat/sort/questiio/questiio/pdf/25.3.4.Aluja.pdf>
- Azcuy, R., Julio, F., Guerra, R. (2019). *Introducción a los métodos estadísticos*. La Habana, Editorial Universitaria.
- Campo, N., Zamora, S. (2019). *Evolución histórica de las técnicas estadísticas y las metodologías para el estudio de la causalidad en ciencias médicas*. Mesidan.
http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1029-30192019000300534
- Caridad, J. (2020). *Inteligencia artificial y estadística aplicada*
https://rubenfcasal.github.io/aprendizaje_estadistico/aprendizaje_estadistico.pdf
- Corona, R. (2015). *Los métodos estadísticos como fuente de mejora de la calidad en las empresas de manufactura*. Revista Universitaria de Administración.
<https://dialnet.unirioja.es/descarga/articulo/8043232.pdf>
- Casella, G. & Berger, R. (2021). *Statistical inference*. United States: Duxbury Thomson learning Academic Resource center.
- Cathcart, F., & Terry, Y. (2021). *Modern Statistical Tendencies used in Investigations in the Medical Sciences*. Revista Habanera de Ciencias Médicas, 20(6), e4605.
<http://www.revhabanera.sld.cu/index.php/rhab/article/view/4605>
- Combe, N., Manin, Y., & Marcolli, M. (2022). *Geometry of information: Classical and quantum aspects*. Theoretical Computer Science, 908. <https://doi.org/10.1016/j.tcs.2021.10.020>

- Fei, Y., Wei, X., Liu, Y., Li, Z., & Chen, M. (2023). *A Survey of Geometric Optimization for Deep Learning: From Euclidean Space to Riemannian Manifold*. Cornell Univervisys. <https://arxiv.org/abs/2302.08210>
- Fernandez, R., Costa, J., Oviedo, M. (2021). *Aprendizaje estadístico*. https://rubenfcasal.github.io/aprendizaje_estadistico.
- Fernández, D., & Guitart, M. (2021). *Estadística descriptiva y análisis de datos*. Universidad Tecnológica Nacional. <http://www1.frm.utn.edu.ar/estadistica/documentos/ed&ad.pdf>
- Jiménez, A., & Díaz, J. (2021). *Revisión sistemática de literatura: Técnicas de aprendizaje automático (machine learning)*. Cuaderno Activa, 13, 113-121.
- Liang, S., Jiang, S., Harlim, J., & Yang, H. (2024). *Solving PDEs on unknown manifolds with machine learning*. Applied And Computational Harmonic Analysis (Print), 101652. <https://doi.org/10.1016/j.acha.2024.101652>
- Lozano, M., Doncel, L., Gutierrez, J. (2023). *De qué manera la teoría de la probabilidad y estadística relaciona el machine learning con las ciencias de la computación*. Corporación Universitaria Minuto de Dios.
- López, B., Vilalta, J., Vaquero, C., Vilalta, G.(2017).*Técnicas estadísticas en una alternativa de apoyo a la toma de decisiones con muestras pequeñas: Una contribución basada en Bootstrap*. Universidad Nacional de Colombia. vol. 84,pp. 316-321 <https://www.redalyc.org/journal/496/49655603040/html/>
- Luceño, A., González, J. (2015). *Métodos estadísticos para medir, describir, y controlar la variabilidad*. Editorial Universidad de Cantabria.
- Nielsen, F. (2020). An Elementary Introduction to Information Geometry. Entropy, 22(10), 1100. <https://doi.org/10.3390/e22101100>

- Marchane, I. (2020). *Fundamentos de machine learning y física estadística*. [Tesis de pregrado].
Universidad de Zaragoza.
- Mariño, J. (2023). *Una comparación entre modelos estadísticos y de Machine Learning para la predicción de series de tiempo multivariadas*. [Tesis de maestría]. Universidad Nacional.
- Maisueche, A. (2019). *Utilización del machine learning en la industria 4.0*. [Tesis de maestría].
Universidad de Valladolid.
- Mayr, A., Binder, H., Gefeller, O., & Schmid, M. (2014). *The Evolution of Boosting Algorithms From Machine Learning to Statistical Modelling*. <https://www-webofscience-com.bibliotecavirtual.unad.edu.co/wos/woscc/full-record/WOS:000346393200002>
- Nielsen, F. (2019). *Geometric Structures of information. Japón: Signals and communication technology*.
- Zambrano, F. Muñoz, E. (2023). *Métodos estadísticos de machine learning aplicados en el estudio de la accesibilidad web: una revisión de la literatura*. Minerva Journal.
<file:///C:/Users/HotelCuellar/Downloads/Dialnet-StatisticalMachineLearningMethodsAppliedInTheStudy-9016304.pdf>