UNIVERSIDAD TÉCNICA DE MANABÍ
Fundada en 1952

RIEMAT
Revista de Investigaciones en Energía, Medio Ambiente y Tecnología:

Impacto de la normalización de datos en la precisión de modelos de aprendizaje supervisado

Impact of data normalization on the accuracy of supervised learning models

Autores

Mariuxi Guillen-Intriago [1] *
https://orcid.org/0009-0005-6923-0538

Roberth Alcívar-Cevallos [1]
https://orcid.org/0000-0001-6282-8493

[1] Facultad de Posgrado, Universidad Técnica de Manabí. Portoviejo, Ecuador.

* Autor para correspondencia.
mariuxi.guillen@utm.edu.ec

Citacion sugerida: Guillen-Intriago, M. & Alcívar-Cevallos, R. (2025). Impacto de la normalización de datos en la precisión de modelos de aprendizaje supervisado. *Revista de Investigaciones en Energía, Medio Ambiente y Tecnología. RIEMAT, 10*(2), 59-79. https://doi.org/10.33936/riemat.v.10i2.7853

Recibido: 28/08/2025
Aceptado: 29/09/2025
Publicado: 12/10/2025

Resumen

La normalización de características, es un paso clave en clasificación supervisada, especialmente cuando los datos presentan escalas heterogéneas. Este estudio tiene objetivo evaluar el impacto de dos estrategias de normalización (MinMax y Z-Score) en el rendimiento de tres modelos: Regresión Logística, SVC/SVM y Árbol de Decisión, aplicados a cuatro datasets: Adult Income, Heart Disease, Student Performance Math y Student Performance Portuguese, obtenidos del repositorio Machine Learning Repository. Como metodología los modelos se entrenaron utilizando validación cruzada estratificada (k=5) y se compararon en términos de accuracy, precisión, recall, F1-score y ROC-AUC. Los resultados mostraron que la normalización con Z-Score tuvo un efecto significativo en el dataset de Adult Income, mejorando el rendimiento de la Regresión Logística (F1-score: 0.426 a 0.666; ROC-AUC: 0.641 a 0.904). En contraste, el dataset de Heart Disease mostró un buen rendimiento, incluso sin normalización, el SVC/SVM con Z-Score mejoró sus métricas con la normalización (F1-score: 0.741 a 0.881; ROC-AUC: 0.785 a 0.922). Sin embargo, estas diferencias no alcanzaron significancia estadística según el test de Wilcoxon ($p \approx 0.0625$), aunque si constituyen evidencia moderada. En los datasets de Student Performance los efectos de la normalización fueron mínimos y estadísticamente no significativos, lo cual puede explicarse porque las variables ya se encontraban en escalas comparables. Finalmente se confirman que la normalización no afecta por igual a todos los algoritmos: su impacto es más evidente en contextos socioeconómicos y clínicos, donde las variables suelen manejar escalas muy distintas. Esta evidencia aporta elementos prácticos para orientar el preprocesamiento de datos en áreas como salud, educación e industria.

Palabras clave: aprendizaje supervisado, normalización de datos, validación cruzada, clasificación binaria, desequilibrio de clases

Abstract

Feature normalization is a key step in supervised classification, especially when data are presented on heterogeneous scales. This study aims to evaluate the impact of two normalization strategies (MinMax and Z-Score) on the performance of three models: Logistic Regression, SVC/SVM, and Decision Tree, applied to four datasets: Adult Income, Heart Disease, Student Performance Math, and Student Performance Portuguese, obtained from the Machine Learning Repository. As a methodology, the models were trained using stratified cross-validation (k=5) and compared in terms of accuracy, precision, recall, F1-score, and ROC-AUC. The results showed that normalization with Z-Score had a significant effect on the Adult Income dataset, improving the performance of Logistic Regression (F1-score: 0.426 to 0.666; ROC-AUC: 0.641 to 0.904). In contrast, the Heart Disease dataset performed well even without normalization, but SVC/SVM with Z-Score improved its metrics with normalization (F1-score: 0.741 to 0.881; ROC-AUC: 0.785 to 0.922). However, these differences did not reach statistical significance according to the Wilcoxon test ($p \approx 0.0625$), although they do constitute moderate evidence. In the Student Performance datasets, the effects of normalization were minimal and statistically insignificant, which can be explained by the fact that the variables were already on comparable scales. Finally, it is confirmed that normalization does not affect all algorithms equally: its impact is more evident in socioeconomic and clinical contexts, where variables tend to use very different scales. This evidence provides practical elements to guide data preprocessing in areas such as health, education, and industry.

Keywords: supervised learning, data normalization, cross-validation, binary classification, class imbalance

CC BY NC SA

1. Introducción

En la era actual, los modelos de aprendizaje supervisado se han consolidado como pilares para la toma de decisiones y la predicción en diversos sectores. La eficacia de estos modelos no solo depende de la sofisticación algorítmica, sino fundamentalmente de la calidad y preparación de los datos de entrada (Yan, 2025). En este proceso, la normalización de características numéricas resulta critica, pues ajusta las escalas de variables para evitar que aquellas con valores más grandes dominen el aprendizaje del modelo. fomentando la estabilidad, velocidad de convergencia y precisión predictiva de los algoritmos sensitivos a escala, siendo esta última de mayor relevancia (Mahmud et al., 2024).

Diversos estudios han señalado la importancia la normalización dentro del preprocesamiento de datos. Por ejemplo, autores como de Amorim et al., (2023), demostraron que el uso de una técnica de escalado inadecuada puede afectar más negativamente el rendimiento del modelo que la ausencia total de la normalización. Este hallazgo se respalda con el trabajo de Ahsan et al., (2021), quienes evaluaron el impacto de varias técnicas de escalado, como Min-Max Scaling y Z-score Standardization, sobre los algoritmos de clasificación (Regresión Logística, Máquina de Vectores de Soporte (SVC/SVM) y Árboles de Decisión) en múltiples datasets, pero sin enfocar su análisis en condiciones controladas de desequilibrio de clases. Mas recientemente, Elik (2024) confirmó la eficiencia de estas técnicas para la clasificación entre variedades de arroz, aunque limitado a un caso de estudio específico.

Aunque muchos trabajos dan evidencia del beneficio de la normalización, aún persisten deficiencias que este estudio busca abordar. En primer lugar, la mayoría de estos trabajos se enfocan en escenarios muy generales o en comparaciones a gran escala; sin embargo, no se detienen a examinar de manera detallada el comportamiento específico de Min-Max Scaling y Z-score en un conjunto reducido de modelos. En segundo lugar, pocos estudios han evaluado de manera comparativa el rendimiento del modelo con datos de distribución normal frente a un conjunto de control sin normalización. Además, la mayoría no mantiene el desbalance natural de clases, pues recurre a técnicas de rebalanceo artificial como SMOTE o undersampling. Esta limitación impide comprender el efecto puro de la normalización. lo que resalta la necesidad de investigaciones metodológicamente sólidas, como la propuesta en este estudio, para llenar estas brechas específicas en la literatura.

Para abordar estas limitaciones, este trabajo presenta un análisis empírico sobre el efecto de las técnicas de normalización Min-Max Scaling y Z-Score Standardization en el rendimiento de tres algoritmos representativos: Regresión Logística, Máquinas de Vectores de Soporte (SVC/SVM) y Árboles de Decisión. El estudio se centra en cuatro conjuntos de datos públicos ampliamente utilizados: Adult Income, Heart Disease, Student Performance (Matemáticas) y Student Performance (portugués). Esta selección permite obtener un perfil detallado del comportamiento de las técnicas de normalización bajo condiciones de datos diversas, con distintas características de distribución y niveles de desequilibrio de clases. De esta manera, se busca aislar y evaluar el efecto real de la normalización, evitando la influencia de otros factores metodológicos.

Para este propósito fue planeada una estrategia experimental confiable y reproducible con el fin de permitir la comparabilidad del efecto de la normalización con variabilidad representativa de escenarios reales. Esta estrategia es coherente con criterios del marco CRISP-ML(Q), para garantizar trazabilidad y rigor del proceso analítico.

El objetivo principal de este trabajo es evaluar y comparar el efecto de las técnicas de normalización Min-Max Scaling y Z-score Standardization sobre modelos supervisados de aprendizaje, a partir de una experimentación replicable con datos reales. Los resultados buscan aportar evidencia empírica sobre cómo estas transformaciones inciden en el desempeño de los algoritmos y, de manera complementaria ofrecer

recomendaciones prácticas para expertos que enfrentan decisiones críticas de preprocesamiento en entornos reales de clasificación.

Para alcanzar el objetivo de esta investigación se plantearon las siguientes preguntas:

¿Cómo influye la aplicación de las técnicas de normalización Min-Max Scaling y Z-score Standardization en el rendimiento de los modelos de Regresión Logística, SVM y Árboles de Decisión, considerando las características intrínsecas de diversos conjuntos de datos?

¿Cuál de los modelos de aprendizaje supervisado analizados resulta más o menos afectado por el proceso de elección del tipo de normalización frente a la presencia de desbalance de clases natural?

2. Materiales y Métodos

Esta investigación utilizó una metodología estructurada según el marco CRISP-ML(Q) (Studer et al., 2021), con una variación del proceso CRISP-DM, aplicable a proyectos de aprendizaje automático que incorpora mecanismos de aseguramiento de la calidad en todas las fases. El objetivo principal fue evaluar el efecto de dos técnicas de normalización: Min-Max Scaling y Z-score Standardization sobre modelos supervisados de aprendizajes aplicados a cuatro conjuntos de datos de distintos dominios. La aplicación de estas técnicas permitió compara su impacto de manera sistemática y reproducible.

2.1. Comprensión del problema

La normalización de variables es una técnica clave en el preprocesamiento de datos. Esta transformación ajusta las escalas de las variables para que sean comparables entre sí, lo cual puede tener un impacto significativo en el rendimiento de ciertos algoritmos, especialmente aquellos sensibles a la magnitud de los valores de entrada (Aksu et al., 2019; Singh & Singh, 2020). En este estudio se cuantificó el efecto de dos métodos de normalización ampliamente utilizados: Min-Max Scaling y Z-score Standardization, sobre el desempeño de tres algoritmos clásicos de clasificación supervisada: Regresión Logística, Máquina de Vectores de Soporte (SVC/SVM) y árboles de decisión. La elección de estas técnicas se fundamenta en su uso frecuente en la literatura científica para estudios comparativos, lo que permite situar los resultados de este trabajo dentro del marco de investigaciones previas (de Amorim et al., 2023).

2.2. Comprensión de los datos

Se trabajó con cuatro conjuntos de datos públicos multivariados, extraídos del UCI Machine Learning Repository (Dua, 2019). Estos son: Adult Income Dataset de Kohavi & Becker, (1996), Student Performance Dataset de Cortez & Silva, (2008), Heart Disease Dataset de Detrano et al. (1989). Cada conjunto incluye variables numéricas y categóricas, con estructuras heterogéneas propias de su dominio. En el caso del Student Performance Dataset, se trabajó con dos subconjuntos independientes: uno corresponde al curso de matemáticas y otro al curso de portugués. Ambos comparten estructura y origen, pero fueron procesados por separado para preservar la homogeneidad del análisis. En la tabla 1 se muestran las variables clave que se usaron en cada conjunto de datos.

Tabla 1

Variables claves utilizadas por conjunto de datos.

Dataset	Variable	Tipo de Dato	Unidad / Valores	Descripción
Adult Income	Age	Numérica	Años	Edad del individuo
Adult Income	Education	Categórica	Nivel educativo (e.g., HS-grad, Bachelors)	Nivel educativo alcanzado
Adult Income	Occupation	Categórica	Tipo de ocupación	Actividad laboral principal

Cont... Tabla 1

Variables claves utilizadas por conjunto de datos.

Dataset	Variable	Tipo de Dato	Unidad / Valores	Descripción
Adult Income	capital.gain	Numérica	USD	Ganancia de capital
Adult Income	hours.per.week	Numérica	Horas	Horas trabajadas por semana
Adult Income	Income	Binaria	<=50K, >50K	Variable objetivo: nivel de ingreso
Heart Disease	Age	Numérica	Años	Edad del paciente
Heart Disease	ChestPainType	Categórica	ATA, NAP, ASY, TA	Tipo de dolor torácico
Heart Disease	Cholesterol	Numérica	mg/Dl	Nivel de colesterol en sangre
Heart Disease	MaxHR	Numérica	Latidos por minute	Frecuencia cardíaca máxima alcanzada
Heart Disease	HeartDisease	Binaria	0: No, 1: Sí	Presencia de enfermedad cardíaca (variable objetivo)
Student Performance	Sex	Categórica	M, F	Género del estudiante
Student Performance	Studytime	Ordinal	1 a 4	Tiempo de studio
Student Performance	Failures	Numérica	0–3	Número de asignaturas no aprobadas anteriormente
Student Performance	G3	Numérica	0–20	Nota final del estudiante. Base para derivar la variable objetivo.
Student Performance	Pass	Binaria	1:Aprobado, 0: Reprobado	Variable objetivo: 1 si G3 ≥ 10, 0 en caso contrario
Student Performance	Internet	Categórica	yes, no	Acceso a internet en casa

Nota: la estructura de variables del Student Performance Dataset aplica tanto al subconjunto de matemáticas como al de portugués.

2.3. Preparación de los datos

El proceso de preparación de los datos fue estandarizado para todos los datasets. Y se desarrollo de la siguiente manera:

Tratamiento de valores faltantes: En el conjunto de datos Adult Income, los registros con datos faltantes en columnas clave (por ejemplo, ‘tipo de trabajo’, ‘ocupación’) que representaban apenas el 7,37% del total,

fueron eliminadas para evitar imputaciones arbitrarias. En columnas menos relevantes, como, por ejemplo, 'país de origen' se imputó el valor más frecuente (moda). En los conjuntos de datos Student Performance Dataset y Heart Disease Dataset, no se detectaron datos faltantes, por lo tanto, no se aplicó ningún tratamiento.

Codificación de variables categóricas: Se aplicó codificación One-Hot a todas las variables categóricas, excluyendo la primera categoría para evitar multicolinealidad. Tras este proceso de preprocesamiento, se definió el tamaño muestral final de cada dataset, así como el número de variables resultantes. La Tabla 2 (Sección de Resultados) resume estos valores, junto con la distribución de clases.

Conversión de valores booleanos: Las variables de tipo booleano con valores (verdadero/falso o sí/no) fueron convertidas a 1/0 para garantizar su compatibilidad con los modelos de aprendizaje supervisado.

Definición de la variable objetivo: En el conjunto Student Performance Dataset, se creó una variable categórica llamada pass, la cual considera como aprobados a los estudiantes que obtuvieron una nota final $G3 \geq 10$, según el sistema escolar portugués. En el conjunto de datos Heart Disease Dataset, la variable Heart Disease ya venía establecida como categórica, indicando presencia (1) o ausencia (0) de la enfermedad.

2.4. Transformación de datos

La transformación de datos fue un paso clave para evaluar el efecto de distintas técnicas de normalización sobre los modelos de clasificación seleccionados. Solo se aplicó a las variables numéricas predictoras, manteniendo sin modificar la variable objetivo y las variables categóricas ya codificadas.

Se implementaron dos métodos clásicos:

Min-Max Scaling: Cada característica se reescala al rango [0,1] aplicando la fórmula:

$$X' = \frac{X - Xmin}{Xmax - Xmin} \tag{1}$$

Donde: X'es el valor original y (Xmin;Xmax) son los valores mínimo y máximo de la columna. Esta técnica es muy útil cuando los algoritmos dependen de la escala de los datos, como SVM y Regresión Logística.

Z-score Standardization: Transformó los datos a una a una distribución con medida (0-1) donde la media es 0 y la desviación estándar 1, utilizando la fórmula:

$$Z = \frac{X - u}{\sigma} \tag{2}$$

más adecuada cuando se desea neutralizar la influencia de valores extremos (outliers) y se asume una distribución aproximadamente normal de los datos.

Para evaluar cada técnica por separado, se diseñaron escenarios experimentales paralelos. Incluyendo un grupo de control sin normalizar. La normalización se realizó en cada fold de la validación cruzada estratificada, asegurando que los parámetros de escalado se calcularan únicamente sobre el conjunto de entrenamiento, evitando así la fuga de datos y manteniendo la proporción de clases en cada división. Además, se decidió conservar la distribución original de clases en todos los conjuntos de datos, incluso al identificar cierto desbalance en uno de los dataset, como en el Student Performance Dataset. Esto permitió evaluar el efecto real de la normalización sin introducir sesgos mediante técnicas de rebalanceo artificial (como SMOTE o undersampling). Para mitigar los posibles efectos adversos del desbalance, se incorporaron métricas de evaluación robustas, tales como el F1-score y el ROC-AUC.

2.5. Modelado

En esta fase se entrenaron tres algoritmos de clasificación ampliamente utilizados y respaldados en la literatura, con el objetivo de obtener resultados variados que nos permitan realizar un análisis integral del impacto de la normalización. Se incluyó:

Regresión Logística: Un modelo lineal clásico, interpretable y base en muchas aplicaciones. Para este

modelo, se aplicó regularización L2 (Ridge) por defecto, buscando mejorar la generalización y mitigar el sobreajuste.

Máquina de Vectores de Soporte (SVC/SVM): Un algoritmo potente y eficaz en espacios de alta dimensión, pero sensible a la escala de los datos.

Árbol de Decisión: Este es un modelo jerárquico no lineal, robusto a datos mixtos, pero propenso al sobreajuste.

La selección de estos modelos (lineal, de margen y basado en reglas) facilita la comparación del efecto de la normalización en diferentes tipos de algoritmos, siguiendo la metodología de estudios comparativos similares (de Amorim et al., 2023; Singh & Singh, 2020).

Para asegurar la robustez y comparabilidad de los resultados, todos los modelos fueron evaluados mediante validación cruzada estratificada k-fold (k=5), manteniendo la proporción de clases en cada partición.

2.6. Evaluación

Para evaluar el rendimiento de los modelos de clasificación, se emplearon métricas estándar para tareas binarias. En todos los conjuntos de datos se definió como clase positiva el valor 1, correspondiente a ingresos mayores a 50K, presencia de enfermedad cardíaca y aprobación escolar, según el caso.

La primera métrica utilizada fue Accuracy (Precisión Global), que indicó la proporción de predicciones totales que fueron correctas. Se calculo con la ecuación:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{3}$$

Donde: TP (True Positivo) son las instancias positivas correctamente clasificadas, TN (True Negatives) son las instancias negativas correctamente clasificadas, FP (False Positives) son las instancias negativas clasificadas erróneamente como positivas, FN (False Negatives) son las instancias positivas clasificadas erróneamente como negativas. Nota: a partir de estas mismas definiciones se derivan otras métricas que también serán utilizadas y posteriormente detalladas, tales como el Recall (Ecuación 4), la Precision (Ecuación 5) y finalmente, ambas métricas pueden combinarse en el F1-score (Ecuación 6).

No todos los errores tienen el mismo impacto sobre el rendimiento del modelo, por lo que se incluyeron métricas más sensibles a distintos tipos de fallos. Una de ellas fue el Recall (también conocido como sensibilidad), que mide la proporción de casos positivos reales correctamente identificados por el modelo. Esta métrica es importante en situaciones donde no detectar un caso positivo puede tener consecuencias graves, como un paciente con enfermedad. Se calculó con la fórmula:

$$Recall = \frac{TP}{TP+FN} \tag{4}$$

Junto con ella, se empleó la precisión, que indica la proporción de predicciones positivas que fueron correctas. Esta métrica es relevante cuando los falsos positivos implican un costo alto y se calcula con la ecuación:

$$Precision = \frac{TP}{TP+FP} \tag{5}$$

Como síntesis de las dos anteriores, se utilizó el F1-score, una medida armónica que balancea la sensibilidad y la precisión. Esta métrica es particularmente útil cuando existe un desbalance entre clases, ya que penaliza al modelo por equivocarse en cualquiera de los dos sentidos, calculada con la ecuación.

$$\mathrm{F1-score} = 2.\frac{Precision.Recall}{Precision+Recall} \tag{6}$$

Finalmente, se consideró la métrica ROC-AUC (Área Bajo la Curva ROC), que evalúo la capacidad del modelo para distinguir entre clases sin depender de un umbral específico de clasificación. Esta se construye a partir de la curva ROC, que relaciona la tasa de verdaderos positivos con la de falsos positivos a lo largo de distintos umbrales. Un área mayor indica un mejor desempeño en términos de discriminación entre clases

Estas métricas fueron seleccionadas por ser las más relevantes en tareas de clasificación binaria, especialmente en contextos con posible desbalance de clases. Su uso está alineado con (Mohammed et al., 2022), quienes destacan la importancia de evaluar con métricas robustas en contextos de calidad de datos.

2.7. Entorno y herramientas

Los experimentos fueron desarrollados en el entorno de Google Colab empleando el lenguaje de programación Python (versión 3.10), aprovechando los recursos de cómputo compartido (CPU y GPU) asignados dinámicamente por la plataforma. Para el procesamiento y análisis de datos se utilizaron las bibliotecas pandas (McKinney, 2010) y numpy (Harris et al., 2020) mientras que scikit-learn (Pedregosa et al., 2011) fue clave en las tareas de preprocesamiento, normalización (con MinMaxScaler y StandardScaler), validación cruzada, implementación de algoritmos de clasificación (LogisticRegression, Máquina de Vectores de Soporte SVC/SVM y DecisionTreeClassifier) y evaluación de modelos mediante métricas como accuracy_score, precision_score, recall_score, f1_score y roc_auc_score.

Las visualizaciones y gráficos comparativos se generaron con matplotlib (Hunter, 2007) y seaborn (Waskom, 2011). Adicionalmente, se integraron algunas bibliotecas como warnings, random y os, estas fueron usadas para controlar la aparición de advertencias, configuración de semillas aleatorias y gestionar el sistema de archivos. Para facilitar la integración con Google Drive y la carga de archivos se empleó la funcionalidad de Google.colab, y opcionalmente se utilizó la biblioteca tabulate para presentar las tablas de manera más ordenada y legible.

El código fue estructurado de forma modular para garantizar la reproducibilidad de los resultados. Para ello, se fijaron semillas mediante random.seed(42) y numpy.random.seed(42), asegurando así la consistencia en los experimentos realizados.

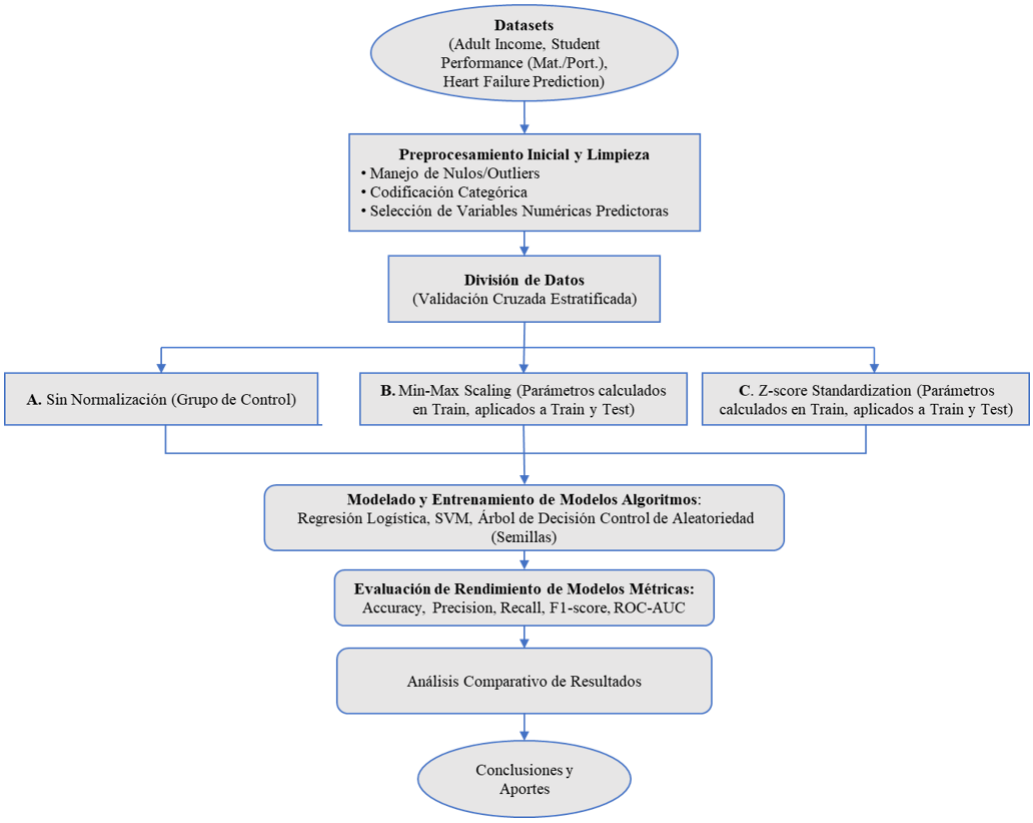
2.8. Aseguramiento de calidad (CRISP-ML (Q))

En cada etapa del proceso se aplicaron principios de aseguramiento tomados de la metodología CRISP-ML(Q). Desde el inicio del proyecto, se formularon preguntas de investigación coherentes con el enfoque supervisado y con los objetivos experimentales planteados. Durante la fase de exploración de los datos se identificaron posibles sesgos, estructuras no uniformes y presencia de clases desequilibradas. Se trabajó con cuatro conjuntos de datos, tratando por separado los subconjuntos de matemáticas y portugués del Student Performance Dataset, con el fin de preservar la homogeneidad de cada dominio y evitar combinaciones que pudiesen dificultar la interpretación de los resultados.

La preparación y transformación de los datos se realizó de forma estandarizada, documentada y trazable. Las técnicas de normalización se aplicaron exclusivamente sobre las variables numéricas predictoras, y se ejecutaron en cada partición (fold) de la validación cruzada estratificada, con el fin de evitar la fuga de información. En la fase de modelado se empleó validación cruzada estratificada con control de aleatoriedad, lo que permitió garantizar tanto la comparabilidad como la reproducibilidad de los experimentos. La evaluación se basó en métricas robustas como Accuracy, Precision, Recall y F1-score, apropiadas para problemas de clasificación binaria. Todos los experimentos fueron desarrollados en Google Colab, con un código modular y documentación clara sobre cada decisión tomada. La Figura 1 resume el trabajo metodológico empleado en esta investigación.

Figura 1

Diagrama metodológico del estudio.



Nota: Esquema de las fases de preprocesamiento, normalización, entrenamiento, validación cruzada y evaluación de modelos, aplicadas sobre los cuatro conjuntos de datos analizados.

3. Resultados y Discusión

Los hallazgos cuantitativos se obtuvieron a partir de la evaluación comparativa de los algoritmos de clasificación (Regresión Logística, Support Vector Classifier y Árbol de Decisión) bajo los esquemas de normalización (ninguna, MinMax y Z-Score). Los valores presentados corresponden a los promedios que se obtuvieron mediante la validación cruzada estratificada k-fold (k=5) mismos que se presentan por separados según el conjunto de datos. En la fase experimental se calcularon cinco métricas de rendimiento (Accuracy, Precision, Recall, F1-score y ROC-AUC), pero para el análisis detallado y la discusión se priorizaron el F1-score y el ROC-AUC, ya que estas permitieron capturar de forma más precisa el rendimiento de los modelos en tareas donde las escalas están desbalanceadas.

La Tabla 2, presenta la estructura final de los conjuntos de datos tras la aplicación técnicas de preprocesamiento, incluyendo la normalización y codificación de variables. Se detalla el número de filas eliminadas por la presencia de valores nulos, la cantidad de nuevas columnas generadas mediante One-Hot Encoding y el balance de clases para la variable objetivo. Estos aspectos proporcionan un marco de referencia para la interpretación de los resultados posteriores.

Tabla 2

Tamaño muestral por dataset tras el preprocesamiento.

Dataset	Filas Originales	Columnas Originales	Filas Tras Preprocesamiento	Columnas Tras Preprocesamiento	Filas Eliminadas	Clase Mayoritaria (%)	Clase Minoritaria (%)	Columnas Numéricas Escalares	Columnas Categóricas OHE
Adult Income	32561	15	30162	97	2399	0 (75.11%)	1 (24.89%)	age, fnlwgt, education.num, capital.gain, capi…	90 columnas generadas
Heart Disease	918	12	918	16	0	1 (55.34%)	0 (44.66%)	Age, RestingBP, Cholesterol, FastingBS, MaxHR,...	9 columnas generadas
Student Performance (Math)	395	33	395	40	0	1 (67.09%)	0 (32.91%)	age, Medu, Fedu, traveltime, studytime, failur...	26 columnas generadas
Student Performance (Portuguese)	649	33	649	40	0	1 (84.59%)	0 (15.41%)	age, Medu, Fedu, traveltime, studytime, failur...	26 columnas generadas

Nota: Resumen de dimensiones y variables finales obtenidas tras el preprocesamiento en cada dataset. One-Hot Encoding.

Dataset Adult Income

En el conjunto de datos Adult Income, la evaluación de los modelos de clasificación reveló que la Regresión Logística y el SVC/SVM superaron de manera significativa al Árbol de Decisión. Esta superioridad se cuantificó utilizando el F1-score y el ROC-AUC. La Figura 2 presenta las curvas ROC correspondientes a los modelos más destacados, con y sin normalización. Los valores numéricos detallados de todas las métricas evaluadas para este conjunto de datos se muestran en la Tabla 3.

Tabla 3

Métricas de rendimiento F1-score y ROC-AUC por modelo y tipo de normalización para el Dataset Adult Income.

Modelo	Normalización	Accuracy	Precision	Recall	F1-score	ROC-AUC
Decision Tree	MinMax	0.811087	0.618646	0.628529	0.623520	0.750118
Decision Tree	None	0.811054	0.618596	0.628396	0.623427	0.750051
Decision Tree	Z-Score	0.810921	0.618349	0.628130	0.623170	0.749874
Logistic Regression	MinMax	0.846131	0.733692	0.599359	0.659649	0.902355
Logistic Regression	None	0.795770	0.706317	0.315414	0.426485	0.641398
Logistic Regression	Z-Score	0.847988	0.735503	0.607884	0.665520	0.904376
SVC/SVM	MinMax	0.833930	0.716924	0.550211	0.622464	0.889649
SVC/SVM	None	0.787680	0.979994	0.150106	0.260317	0.629974
SVC/SVM	Z-Score	0.843644	0.745935	0.564200	0.642309	0.892034

Nota: Valores promedio obtenidos mediante validación cruzada estratificada (k=5), considerando el

preprocesamiento aplicado (MinMax, Z-Score y ninguno).

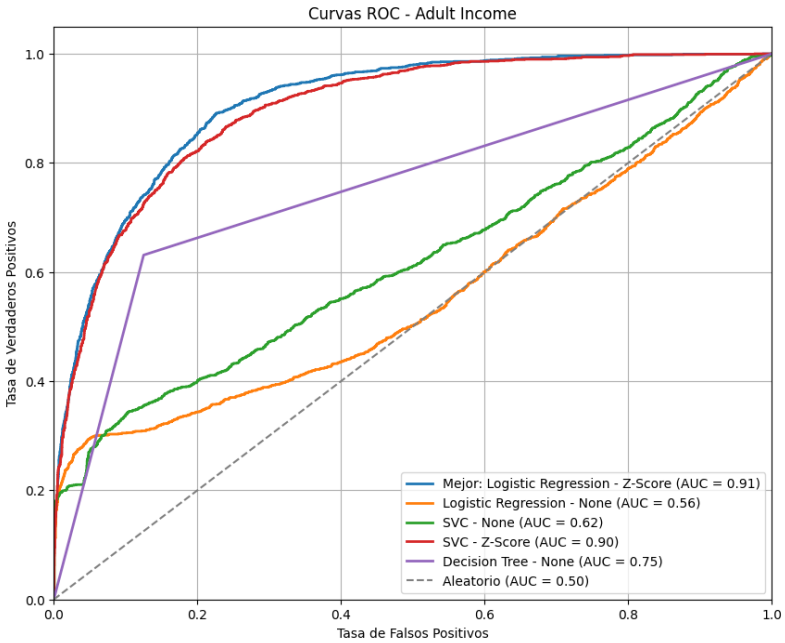
En el dataset Adult Income, el modelo de Regresión Logística sin normalización alcanzó un F1-score de 0.426 y un ROC-AUC de 0.641. Al aplicar la normalización por estandarización (Z-Score), el rendimiento mejoró de forma significativa, alcanzando un F1-score de 0.666 (+56.2%) y un ROC-AUC de 0.904 (+41.0%). Este incremento confirmó la alta dependencia de este modelo respecto a la escala de las variables. Comparado con el modelo SVC/SVM bajo las mismas condiciones, la Regresión Logística con Z-Score logró un F1-score ligeramente mayor (0.666 frente a 0.642).

El Support Vector Classifier (SVC/SVM) mostró un comportamiento similar. Sin normalización obtuvo un F1-score de 0.260 y un ROC-AUC de 0.630. Al aplicar la estandarización (Z-Score), el rendimiento mejoró significativamente, obteniendo un F1-score que subió a 0.642 (+146.9%) y un ROC-AUC de 0.892 (+41.6%). Esta diferencia valida la importancia de la normalización para el (SVC/SVM), ya que, al ser un modelo basado en distancias, se beneficia de que las características tengan escalas homogéneas, esto le permite definir mejor sus fronteras de decisión. Aunque su F1-score fue ligeramente inferior al de la Regresión Logística con Z-Score, superaron claramente al Árbol de Decisión en este conjunto de datos.

El Árbol de Decisión, sin normalización, obtuvo un F1-score de 0.623 y un ROC-AUC de 0.750. A diferencia de la Regresión Logística y SVC/SVM, este modelo mostró un rendimiento estable y notablemente menos sensible a las técnicas de normalización. Las variaciones en las métricas fueron mínimas: la puntuación F1 mostró cambios de +0.02% con MinMax y -0.04% con Z-Score, mientras que el ROC-AUC mostró cambios de +0.01% con MinMax y -0.02% con Z-Score, tal como se presenta en la figura 2.

Figura 2

Imagen de las curvas ROC para el dataset Adult Income



Nota: Se compara el desempeño de los modelos de clasificación con y sin normalización Z-Score. La Regresión Logística y SVC, con preprocesamiento Z-Score, alcanzaron los valores AUC más altos (0.91).

Este comportamiento confirma su estabilidad frente a la escala de las variables, dado que su funcionamiento

Estos resultados mostrado figura 2, confirma la estabilidad del modelo frente a la escala de las variables, dado que su funcionamiento se basa en particiones jerárquicas y no en cálculos de distancia o gradientes. Sin embargo, su rendimiento absoluto fue inferior a los modelos de Regresión Logística y SVC/SVM en este dataset,

3.1. Dataset Heart Disease

En el conjunto de datos Heart Disease, los modelos de Regresión Logística y SVC/SVM mostraron un rendimiento superior respecto al Árbol de Decisión. En la figura 3 se representan las curvas ROC generadas por cada modelo, y los resultados numéricos de todas las métricas evaluadas se presentan en la tabla 4.

Tabla 4

Rendimiento promedio de los modelos en el dataset Heart Disease.

Modelo	Normalización	Accuracy	Precision	Recall	F1-score	ROC-AUC
Decision Tree	MinMax	0.771205	0.792895	0.793322	0.792681	0.768612
Decision Tree	None	0.769019	0.792188	0.789361	0.790468	0.766632
Decision Tree	Z-Score	0.767932	0.791812	0.787401	0.789287	0.765651
Logistic Regression	MinMax	0.863833	0.866660	0.893826	0.879192	0.925464
Logistic Regression	None	0.868187	0.868999	0.899709	0.883234	0.924430
Logistic Regression	Z-Score	0.867100	0.871441	0.893807	0.881805	0.924839
SVC/SVM	MinMax	0.862746	0.861763	0.897729	0.878906	0.917748
SVC/SVM	None	0.717855	0.755416	0.728441	0.740713	0.785553
SVC/SVM	Z-Score	0.864926	0.861179	0.903650	0.880992	0.921914

Nota: Valores promedio obtenidos mediante validación cruzada estratificada (k=5), considerando el preprocesamiento aplicado (MinMax, Z-Score y ninguno).

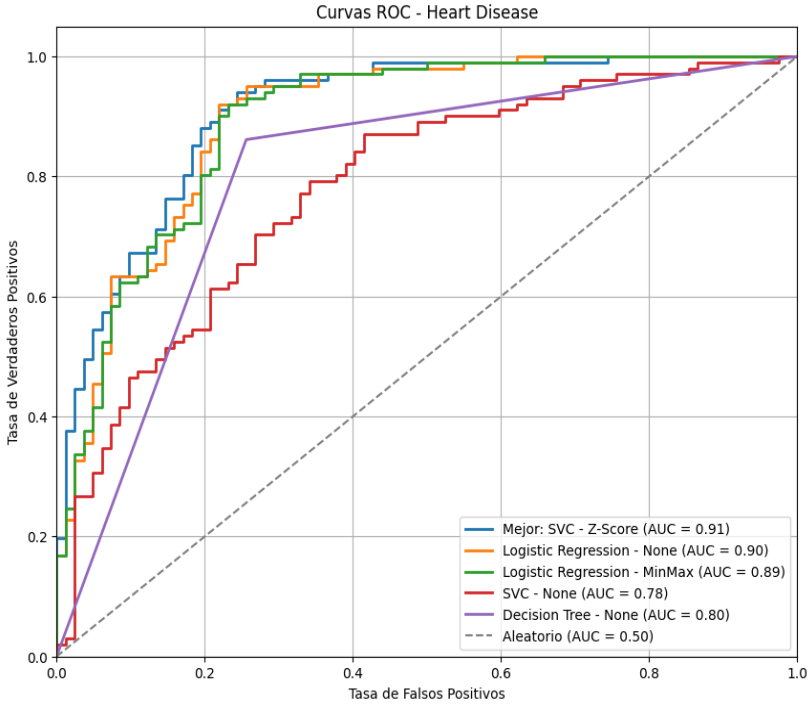
Para el conjunto de datos Heart Disease, el modelo de Regresión Logística sin normalización alcanzó un F1-score de 0.883 y un ROC-AUC de 0.924. Aplicando la normalización este modelo tuvo un impacto mínimo en su rendimiento: el F1-score se mantuvo estable, disminuyendo ligeramente (por ejemplo, 0.879 con MinMax, -0.5%; 0.882 con Z-Score, -0.1%), mientras que el ROC-AUC mostró un incremento poco significativo (0.925 con MinMax, +0.1%; 0.925 con Z-Score, +0.04%). Este comportamiento indica que las características de este conjunto de datos poseen una escala intrínsecamente homogénea, lo que permitió identificar que este modelo se adapta bien incluso sin normalizar las variables. Su rendimiento fue similar al del modelo (SVC/SVM) en este contexto.

Por su parte, el Support Vector Classifier (SVC/SVM) sin normalización registró un F1-score de 0.741 y un ROC-AUC de 0.786. A diferencia de la Regresión Logística, el SVC/SVM mostró mejoras más claras tras la normalización. Al aplicar la estandarización Z-Score, el F1-score mejoró a 0.881 (un incremento del 18.9%) y su ROC-AUC a 0.922 (un incremento del 17.4%). Este resultado evidencia lo mucho que este modelo depende de una escala homogénea, al tratarse de un modelo basado en distancias, incluso en conjuntos de datos donde otros modelos lineales son menos sensibles al escalado. El SVC/SVM obtiene beneficios significativos. En consecuencia, su rendimiento final resultó equiparable al de la Regresión Logística.

El Árbol de Decisión sin normalización obtuvo un F1-score de 0.790 y un ROC-AUC de 0.767. Al igual que en el conjunto de datos Adult Income, la normalización tuvo un efecto mínimo en su desempeño: con MinMax el F1-score aumentó apenas un 0.3% y el ROC-AUC un 0.3%, mientras que con Z-Score ambos indicadores mostraron descensos marginales de 0.1%, tal como se presenta en la figura 3.

Figura 3

Curvas ROC para los modelos de clasificación evaluados en el dataset Heart Disease.



Nota: Se compara el desempeño de los modelos de clasificación con y sin normalización. El SVC con preprocesamiento Z-Score demostró el mayor impacto, alcanzando el valor AUC más alto (0.91).

En la figura 3 se presenta la capacidad para manejar datos no escalados contrastados con la sensibilidad de la Regresión Logística y el SVC/SVM. A pesar de esto, su rendimiento general se mantuvo por debajo del de estos modelos, lo que sugiere una menor capacidad para modelar relaciones complejas en los datos,

3.2. Dataset Student Performance (Math)

los resultados del conjunto de datos Student Performance (Math) reflejaron que la normalización no afectó a todos los modelos por igual. En la figura 4 se pueden observar las curvas ROC generadas por cada modelo. La Tabla 5, recoge los valores promedio obtenidos por cada combinación de modelo y técnica de normalización.

En el conjunto de datos Student Performance (Math), con la Regresión Logística sin normalización se obtuvo un F1-score de 0.776 y un ROC-AUC de 0.680. La normalización, tanto con MinMax como con Z-Score, mostró cambios mínimos: el F1-score subió ligeramente con MinMax (0.791, +1.9%) pero bajó con Z-Score (0.768, -1.0%). En el ROC-AUC, estos cambios fueron poco relevantes (0.682 con MinMax, +0.4%; 0.672 con Z-Score, -1.1%). Probablemente se relaciona con que las variables ya están en rangos similares, por lo que aplicar escalado no ofrece mejoras significativas. Aun así, su rendimiento fue similar al del SVC/SVM.

El Support Vector Classifier (SVC/SVM) sin normalización logró un F1-score de 0.799 y un ROC-AUC de 0.713. Tras la aplicación de Z-Score, el F1-score aumentó a 0.811 (+1.5%), mejorando la detección de la clase minoritaria. Sin embargo, el ROC-AUC disminuyó a 0.662 (-7.2%). Este contraste evidencia que la normalización puede favorecer algunas métricas, pero perjudicar otras, reflejando una relación compleja entre el escalado y la capacidad del modelo para diferenciar clases.

Tabla 5

Rendimiento promedio de los modelos en el dataset Student Performance (Math).

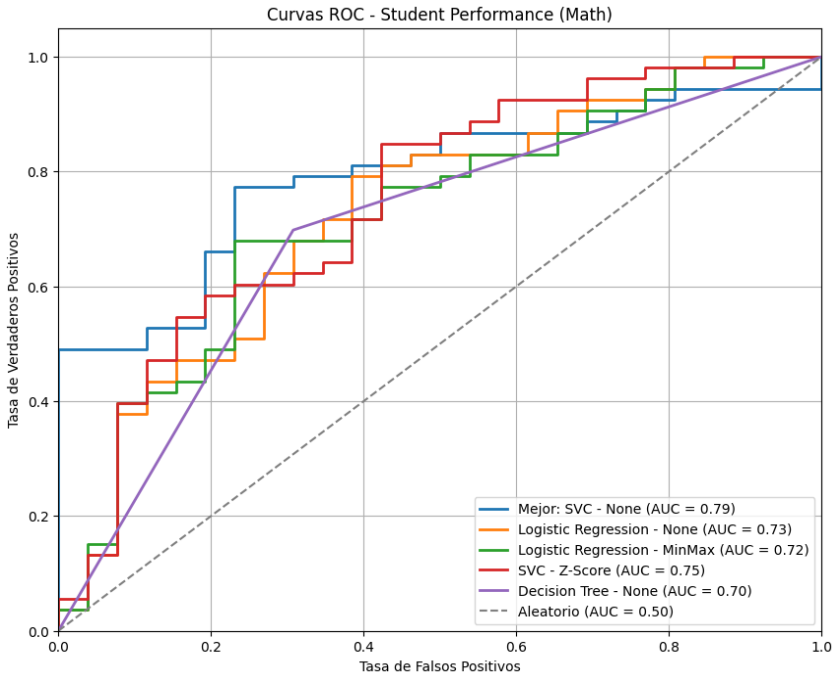
Modelo	Normalización	Accuracy	Precision	Recall	F1-score	ROC-AUC
Decision Tree	MinMax	0.640506	0.740311	0.720755	0.728939	0.598839
Decision Tree	None	0.640506	0.740311	0.720755	0.728939	0.598839
Decision Tree	Z-Score	0.643038	0.743287	0.720755	0.730298	0.602685
Logistic Regression	MinMax	0.688608	0.719251	0.879245	0.791215	0.682438
Logistic Regression	None	0.678481	0.728937	0.830189	0.776051	0.679971
Logistic Regression	Z-Score	0.668354	0.723660	0.818868	0.768138	0.672279
SVC/SVM	MinMax	0.678481	0.685461	0.962264	0.800547	0.669158
SVC/SVM	None	0.665823	0.669198	0.992453	0.799375	0.713062
SVC/SVM	Z-Score	0.703797	0.707666	0.950943	0.811379	0.661684

Nota: Valores promedio obtenidos mediante validación cruzada estratificada (k=5), considerando el preprocesamiento aplicado (MinMax, Z-Score y ninguno).

Para el Árbol de Decisión, sin normalización obtuvo un F1-score de 0.729 y un ROC-AUC de 0.599. Como en los otros conjuntos de datos, la normalización tuvo un efecto casi nulo (F1: 0% con MinMax, +0.2% con Z-Score; ROC-AUC: 0% con MinMax, +0.6% con Z-Score). Este modelo mantiene un buen desempeño con datos sin escalar, aunque su rendimiento global fue inferior al de los otros modelos, indicando una menor capacidad para capturar relaciones complejas en este dominio, tal como se presenta en la figura 4.

Figura 4

Curvas ROC para–Student Performance (Math).



Nota: Se compara el desempeño de los modelos de clasificación con y sin normalización. El SVC sin normalización demostró el mejor desempeño, alcanzando el valor AUC más alto (0.79).

3.3. Dataset Student Performance (Portuguese)

Para el dataset Student Performance (Portuguese) se observaron tendencias de rendimiento similares a las del conjunto de Matemáticas, destacando que los modelos de Regresión Logística y SVC/SVM superaron en desempeño al Árbol de Decisión. La Figura 5 presenta las curvas ROC de los modelos evaluados. Los valores numéricos detallados se encuentran en la Tabla 6.

Tabla 6

Rendimiento promedio de los modelos en el dataset Student Performance (Portuguese).

Modelo	Normalización	Accuracy	Precision	Recall	F1-score	ROC-AUC
Decision Tree	MinMax	0.802767	0.893832	0.870709	0.881877	0.650354
Decision Tree	None	0.802767	0.893832	0.870709	0.881877	0.650354
Decision Tree	Z-Score	0.804305	0.894029	0.872527	0.882880	0.651264
Logistic Regression	MinMax	0.839726	0.869312	0.954479	0.909750	0.790000
Logistic Regression	None	0.845903	0.877722	0.950859	0.912584	0.788921
Logistic Regression	Z-Score	0.839726	0.877943	0.941718	0.908619	0.778600
SVC/SVM	MinMax	0.841300	0.846278	0.992727	0.913659	0.813957
SVC/SVM	None	0.845915	0.845915	1.000.000	0.916527	0.782966
SVC/SVM	Z-Score	0.855134	0.863048	0.985438	0.920103	0.836133

Nota: Valores promedio obtenidos mediante validación cruzada estratificada (k=5), considerando el preprocesamiento aplicado (MinMax, Z-Score y ninguno).

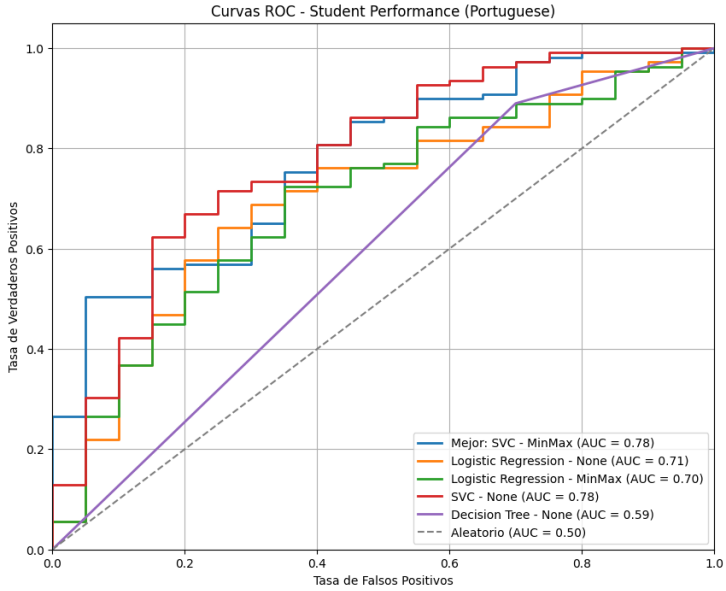
En el conjunto Student Performance (Portuguese), la Regresión Logística sin normalización obtuvo un F1-score de 0.913 y un ROC-AUC de 0.789. Al aplicar MinMax, el F1-score bajó a 0.910 (-0.3%) y el ROC-AUC subió apenas a 0.790 (+0.1%). Con Z-Score, el F1-score fue de 0.909 (-0.4%) y el ROC-AUC bajó a 0.779 (-1.3%). Interpretó que este modelo mantiene un rendimiento estable aun cuando se aplican técnicas de normalización, algo que también se observó en el conjunto de Matemáticas.

El Support Vector Classifier (SVC/SVM) sin normalización logró un F1-score de 0.917 y un ROC-AUC de 0.783. Al usar Z-Score, el F1-score subió ligeramente a 0.920 (+0.4%) y el ROC-AUC mejoró más notablemente hasta 0.836 (+6.8%). Esto sugiere que la normalización ayudó al SVC/SVM a mejorar su capacidad para separar clases, incluso cuando los datos ya tienen escalas similares.

En cuanto al Árbol de Decisión, obtuvo un F1-score de 0.882 y un ROC-AUC de 0.650 sin aplicar normalización. Al igual que en los otros conjuntos, escalar los datos prácticamente no modificó su rendimiento (0% con MinMax, +0.1% con Z-Score). Aunque este modelo no necesita escalar los datos para funcionar correctamente, en este conjunto volvió a mostrar un desempeño inferior al de los otros algoritmos, lo que refuerza la idea de que tiene limitaciones para modelar relaciones más complejas.

Figura 5

Curvas ROC para Student Performance (Portuguese)



Nota: Se compara el desempeño de los modelos de clasificación con y sin normalización. El SVC con el preprocesamiento MinMax demostró el mejor desempeño, alcanzando el valor AUC más alto (0.78).

3.4. Análisis estadístico de las diferencias entre técnicas de normalización

Para validar si las mejoras observadas en las métricas de rendimiento fueron estadísticamente significativas, se aplicó la prueba no paramétrica de Wilcoxon para muestras pareadas, considerando los valores de F1-score obtenidos en los cinco folds de validación cruzada. Aunque se analizaron otras métricas (Accuracy, Precision, Recall, ROC-AUC), solo el F1-score evidenció diferencias relevantes entre las condiciones de normalización. Los resultados mostraron tendencias a significancia ($p \leq 0.10$) en tres comparaciones específicas, detalladas en la Tabla 7.

Tabla 7

Comparaciones con evidencia moderada ($p \leq 0.10$)

Dataset	Modelo	Δ F1-score	p-value	IC 95% Inferior	IC 95% Superior
Adult Income	Logistic Regression	+0.239	0.0625	0.131	0.347
Adult Income	SVC/SVM	+0.382	0.0625	0.365	0.399
Heart Disease	SVC/SVM	+0.140	0.0625	0.111	0.170

Nota: ΔF1-score corresponde a la diferencia entre Z-score y las demás condiciones de preprocesamiento (MinMax y None), evaluada con la prueba de Wilcoxon. Se reportan solo comparaciones con evidencia moderada ($p \leq 0.10$).

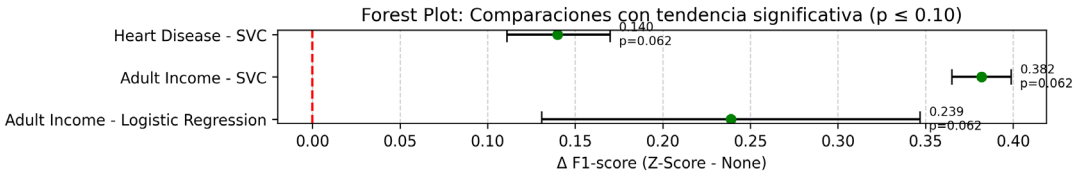
En particular, el modelo SVC/SVM presentó mejoras sustanciales con la técnica Z-Score en los conjuntos de datos Adult Income y Heart Disease, mientras que la Regresión Logística también evidenció una mejora relevante en Adult Income. Estas diferencias fueron consistentes, con intervalos de confianza que no cruzaron el cero, lo cual se presenta en la figura 6.

De los cuatro conjuntos de datos evaluados, solo Adult Income y Heart Disease mostraron diferencias

estadísticamente relevantes. En los datasets de desempeño estudiantil, las variaciones en las métricas no fueron significativas, por lo que se excluyeron de la tabla resumen.

Figura 6

Forest Plot: Comparaciones con tendencia significativa (p ≤ 0.10)



Nota: Diferencias en F1-score (Z-score - None) con IC 95%; se muestran solo comparaciones con p ≤ 0.10.

Discusión

Los hallazgos presentados en los resultados se interpretan considerando las propiedades algorítmicas, las características de los datos y la literatura científica relevante.

Rendimiento General de los Modelos

En términos generales, se observó que los modelos de SVC/SVM y Regresión Logística tendieron a superar al Árbol de Decisión en la mayoría de escenarios. En conjuntos de datos complejos como Adult Income y Heart Disease, estos dos modelos alcanzaron las mejores puntuaciones en las métricas de ROC-AUC y F1-score, con resultados más estables entre las distintas particiones de validación cruzada (ver Figuras 2-5). Esto coincide con estudios previos que documentan su sensibilidad a la escala de las características (de Amorim et al., 2023; Shantal et al., 2023). Este comportamiento no es aislado; en otros trabajos se han reportado resultados similares. Por ejemplo, en tareas de predicción médica como la detección de tumores cerebrales o la predicción de diabetes, el SVC/SVM ha superado a la Regresión Logística en métricas como precisión y F1-score (Rao et al., 2024; Salian et al., 2024).

El Árbol de Decisión presentó los resultados más bajos en los datasets analizados. Si bien este modelo es fácil de interpretar y puede capturar relaciones no lineales simples, su rendimiento disminuye cuando se enfrenta a datos con muchas variables o con interacciones más complejas, confirmando los hallazgos de de Amorim et al. (2023). Por ello, otras investigaciones han propuesto para este modelo variantes como los árboles aditivos (Goedhart et al., (2025), que intentan mantener la interpretabilidad mientras mejoran la precisión.

En cuanto a operatividad, la Regresión Logística se mantiene como una alternativa sencilla y eficiente; al aplicarse con regularización, ofrece buena capacidad de generalización. Por su parte, El SVC/SVM, aunque requiere mayor ajuste de hiperparámetros y recursos computacionales, ha demostrado mejorar su rendimiento y escalabilidad en grandes volúmenes de datos con técnicas como los modelos en cascada o los ensambles (Bailly et al., 2022; Dudzik et al., 2024).

En lo que respecta al preprocesamiento, la normalización de las características demostró ser un factor clave para ambos modelos. En particular, el SVC/SVM mostró una mayor sensibilidad a la escala de las variables, un hallazgo que coincide con estudios previos donde se destaca el impacto positivo de técnicas como Z-Score y MinMax en el rendimiento de este clasificador y otros algoritmos basados en distancia (Rataj et al., 2023; Shantal et al., 2023).

Impacto de la Normalización y Aplicaciones Prácticas.

La tabla 8 resume de forma comparativa los efectos de MinMax y Z-Score sobre las métricas de rendimiento,

junto con su utilidad práctica en distintos ámbitos de aplicación.

Los patrones observados en cada conjunto de datos mostraron que la eficacia de los modelos dependió tanto de la arquitectura como de las propiedades estadísticas del conjunto de datos. En Adult Income y Heart Disease, los modelos de regresión logística y SVC/SVM alcanzaron valores más altos en F1-score y ROC-AUC, incluso en escenarios con desequilibrio de clases. Estos resultados confirman una capacidad solida de discriminación en contextos socioeconómicos y clínicos, lo que coincide con lo que se ha publicado en la literatura sobre su sensibilidad a la escala de las variables y su eficacia en problemas con estructuras complejas (Bailly et al., 2022; Modhugu et al., 2024). En la práctica, para el sector industrial, comprender el impacto de la normalización es fundamental para el desarrollo de modelos predictivos que ayuden a la clasificación de perfiles de riesgo socioeconómico, control de calidad, y el análisis de riesgo financiero. La precisión y estabilidad de estos modelos pueden traducirse en una reducción de costos, una mejora de los procesos y una mitigación de los riesgos operativos y económicos.

Tabla 8

Resumen comparativo del impacto de la normalización en el rendimiento de los modelos por dataset.

Dataset	Modelo	Mejor Normalización	Mejora Observada	Utilidad Práctica (Aplicación Real)
Adult Income	Logistic Regression	Z-Score	F1-score: 0.666 (↑ 56%) ROC-AUC: 0.904 (↑ 41%)	Evaluación para beneficios sociales o análisis de ingresos, con alta sensibilidad y equilibrio.
	SVC/SVM	Z-Score	F1-score: 0.642 (↑ 147%) ROC-AUC: 0.892 (↑ 41%)	Clasificación de perfiles de riesgo en instituciones públicas o privadas.
	Decision Tree	(Sin Normalizar)	F1-score: 0.624 (≈ sin cambios) ROC-AUC: 0.750 (≈ sin cambios)	Aplicaciones rápidas donde se prioriza la interpretabilidad sobre la precisión.
Heart Disease	Logistic Regression	MinMax	F1-score: 0.879 (↑ ~5%) ROC-AUC: 0.925 (↑ 0.1%)	Diagnóstico clínico con variables diversas (edad, presión, colesterol).
	SVC/SVM	Z-Score	F1-score: 0.881 (↑ ~19%) ROC-AUC: 0.921 (↑ 17%)	Sistemas inteligentes para predecir eventos cardíacos.
	Decision Tree	(Sin Normalizar)	F1-score: 0.790 (≈ sin cambios) ROC-AUC: 0.767 (≈ sin cambios)	Escenarios con recursos limitados, donde lo simple y comprensible es lo más útil.
Student Perf. (Math)	SVC/SVM	Z-Score	F1-score: 0.811 (↑ 1.5%) ROC-AUC: 0.661 (↓ 7%)	Identificación temprana de estudiantes con bajo rendimiento en matemáticas.
	Logistic Regression	MinMax	F1-score: 0.791 (↑ 2%) ROC-AUC: 0.682 (↑ 0.3%)	Programas de refuerzo académico para estudiantes con riesgo moderado.
	Decision Tree	(Sin Normalizar)	F1-score: 0.728 (≈ sin cambios) ROC-AUC: 0.599 (≈ sin cambios)	Análisis simple de desempeño estudiantil para docentes o directivos.
Student Perf. (Portuguese)	SVC/SVM	Z-Score	F1-score: 0.921 (↑ 0.5%) ROC-AUC: 0.836 (↑ 6.7%)	Sistemas de alerta académica y predicción dAe abandono escolar.
	Logistic Regression	Z-Score	F1-score: 0.908 (↑ 1.6%) ROC-AUC: 0.779 (↓ 1%)	Selección de candidatos para becas o programas de apoyo escolar.
	Decision Tree	Z-Score	F1-score: 0.882 (↑ 0.2%) ROC-AUC: 0.651 (↑ 0.2%)	Herramientas de visualización y diagnóstico educativo.

Nota: La eficacia de los modelos mostró variaciones influenciadas tanto por la arquitectura utilizada como

por las propiedades estadísticas del conjunto de datos.

En el cojunto de Heart Disease, aunque los modelos obtuvieron buenos resultados aun sin normalización, el SVC/SVM mejoró significativamente con el escalado, obteniendo una puntuación Z-Score, con aumentos cercanos al 19 % en el F1-score y del 17 % en la curva ROC-AUC. Este hallazgo respalda la utilidad y la importancia de aplicar la normalización en sistemas que requieran precisión para diagnósticos tempranos, coincidiendo con Bailly et al. (2022).

En los conjuntos Student Performance (Math y Portuguese) la normalización tuvo un impacto más limitado. Debido a la naturaleza de sus variables, muchas categóricas o discretas con poca variabilidad, se observaron mejoras mínimas en métricas como ROC-AUC, lo cual sustenta lo ya reportado en la literatura (Brooks et al., 2023; Uddin & Lu, 2024). Sin embargo, el F1-score se mantuvo elevado en modelos como el SVC/SVM y la Regresión Logística, especialmente en el conjunto de datos portugués, donde esta métrica alcanzó un valor cercano a 0.91. Este resultado es crucial en el ámbito educativo, porque permite identificar de manera temprana a los estudiantes en riesgo de abandono, reducir el coste institucional de los falsos negativos y facilitar la implementación de programas preventivos (Adnan Aslam et al., 2025; Bujang et al., 2021).

La comparación entre modelos muestra un patrón claro: el árbol de decisión presentó menor sensibilidad a la normalización y mantuvo resultados estables, aunque generalmente inferiores. Su sencillez e interpretabilidad lo convierten en una herramienta útil en contextos en los que la aplicabilidad es una prioridad y la precisión puede quedar en un segundo plano frente a la facilidad de uso. Por otro lado, SVC/SVM y regresión logística se presentan como opciones más robustas cuando se pretende optimizar la capacidad predictiva, con un desempeño diferencial según el campo de aplicación: salud, educación o industria.

Limitaciones del Estudio y direcciones futuras

Este estudio se restringió a tres clasificadores (Regresión Logística, SVC/SVM y Árbol de Decisión), sin considerar modelos de ensamble ni redes neuronales, y no incluyó una optimización exhaustiva de hiperparámetros, lo que pudo influir en el rendimiento observado. Además, el análisis se limitó a datasets estructurados y problemas de clasificación binaria, por lo que los hallazgos no son directamente generalizables a otros contextos. No se aplicaron técnicas de balanceo de clases, aunque se emplearon métricas robustas como F1-score y ROC-AUC. Finalmente, no se descarta la presencia de sesgos en la selección de variables, ni se abordó el costo computacional diferencial de los algoritmos, que en el caso de SVC/SVM puede ser considerable en bases de datos extensas.

Futuras investigaciones podrían incorporar algoritmos más complejos (ensambles o redes neuronales), optimización rigurosa de hiperparámetros, técnicas de rebalanceo de clases, análisis de importancia de variables y la extensión a datasets no estructurados o problemas multiclase.

4. Conclusiones

En esta investigación se abordó la evaluación comparativa del rendimiento y la estabilidad de los modelos de la Regresión Logística, Máquina de Vectores de Soporte (SVC/SVM) y Árbol de Decisión, así como el impacto de las técnicas de normalización (MinMax, Z-Score) en cuatro datasets diversos: Adult Income, Heart Disease, y Student Performance (Math y Portuguese). Los resultados que se obtuvieron sirvieron para identificar a los modelos de Regresión Logística y SVC/SVM como los clasificadores más robustos y con mayor rendimiento en la mayoría de los escenarios. Estos modelos, que se caracterizan por su capacidad para modelar relaciones lineales y no lineales mediante kernels, demostraron una superioridad consistente en ROC-AUC y F1-score, junto con una baja variabilidad en su rendimiento a través de los folds de validación cruzada. Por el contrario, el Árbol de Decisión mostró un rendimiento inferior.

La normalización de características en el proceso de transformación de los datos, demostró ser muy

importante, especialmente para el datasets Adult Income, donde la estandarización Z-Score mejoró la capacidad predictiva de los modelos. En los otros datasets, su impacto fue menos pronunciado, lo que sugiere que la necesidad de normalización depende de las propiedades de escala inherentes de las características. Se puede concluir que este estudio proporciona una base empírica para la selección de modelos y estrategias de normalización en problemas de clasificación con datos tabulares. Confirma la eficacia de SVC/SVM y la Regresión Logística para una amplia gama de aplicaciones, destacando la necesidad de un preprocesamiento de datos informado para maximizar el rendimiento predictivo. Estos hallazgos son cruciales para la práctica profesional, ya que orientan a los científicos de datos y analistas a tomar decisiones más informadas sobre la aplicación de la normalización. Sugiere que, en lugar de una aplicación universal, la elección de la estrategia de normalización debe ser un proceso deliberado y basado en las características específicas del dataset y el modelo a emplear, optimizando así los recursos computacionales y mejorando la confiabilidad de los sistemas predictivos en escenarios reales.

5. Referencias

Adnan Aslam, M., Murtaza, F., Ehatisham Ul Haq, M., Yasin, A., & Ali, N. (2025). SAPEx-D: A Comprehensive Dataset for Predictive Analytics in Personalized Education Using Machine Learning. Data 2025, Vol. 10, Page 27, 10(3), 27. https://doi.org/10.3390/DATA10030027

Ahsan, M. M., Mahmud, M. A. P., Saha, P. K., Gupta, K. D., & Siddique, Z. (2021). Effect of Data Scaling Methods on Machine Learning Algorithms and Model Performance. Technologies, 9(3). https://doi.org/10.3390/technologies9030052

AKSU, G., GÜZELLER, C. O., & ESER, M. T. (2019). The Effect of the Normalization Method Used in Different Sample Sizes on the Success of Artificial Neural Network Model. International Journal of Assessment Tools in Education, 6(2), 170–192. https://doi.org/10.21449/ijate.479404

Bailly, A., Blanc, C., Francis, É., Guillotin, T., Jamal, F., Wakim, B., & Roy, P. (2022). Effects of dataset size and interactions on the prediction performance of logistic regression and deep learning models. Computer Methods and Programs in Biomedicine, 213, 106504. https://doi.org/10.1016/J.CMPB.2021.106504

Brooks, C., Kovanović, V., & Nguyen, Q. (2023). Predictive modeling of student success. Handbook of Artificial Intelligence in Education, 350–369. https://doi.org/10.4337/9781800375413.00027

Bujang, S. D. A., Selamat, A., Ibrahim, R., Krejcar, O., Herrera-Viedma, E., Fujita, H., & Ghani, N. A. M. (2021). Multiclass Prediction Model for Student Grade Prediction Using Machine Learning. IEEE Access, 9, 95608–95621. https://doi.org/10.1109/ACCESS.2021.3093563

Cortez, P., & Silva, A. M. G. (2008). Using Data Mining to Predict Secondary School Student Performance. https://archive.ics.uci.edu/ml/datasets/student+performance

de Amorim, L. B. V., Cavalcanti, G. D. C., & Cruz, R. M. O. (2023). The choice of scaling technique matters for classification performance. Applied Soft Computing, 133. https://doi.org/10.1016/j.asoc.2022.109924

Detrano, R., Janosi, A., Steinbrunn, W., Pfisterer, M., & Bloedow, D. (1989). Heart Disease Dataset [Dataset]. UCI Machine Learning Repository. https://archive.ics.uci.edu/dataset/45/heart+disease

Dua, D. & G. C. (2019). Machine Learning Repository. UCI Machine Learning Repository. https://archive.ics.uci.edu/

Dudzik, W., Nalepa, J., & Kawulok, M. (2024). Ensembles of evolutionarily-constructed support vector machine cascades. Knowledge-Based Systems, 288. https://doi.org/10.1016/J.KNOSYS.2024.111490

Elik, A. C. ¸. (2024). Acadlore Transactions on AI and Machine Learning Evaluating the Impact of Data Normalization on Rice Classification Using Machine Learning Algorithms. Acadlore Trans. Mach. Learn, 3(3), 162–171. https://doi.org/10.56578/ataiml030

Goedhart, J. M., Klausch, T., Janssen, J., & van de Wiel, M. A. (2025). Adaptive Use of Co-Data Through

Empirical Bayes for Bayesian Additive Regression Trees. Statistics in Medicine, 44(5), e70004. https://doi.org/10.1002/SIM.70004;PAGE:STRING:ARTICLE/CHAPTER

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., & Smith, N. J. (2020). Array programming with NumPy. Nature, 585, 357–362. https://doi.org/10.1038/s41586-020-2649-2

Hunter, J. D. (2007). Matplotlib: A 2D Graphics Environment. Computing in Science & Engineering. https://doi.org/10.1109/MCSE.2007.55

Kohavi, R., & Becker, B. (1996). UCI Machine Learning Repository: Adult Data Set (Census Income). https://archive.ics.uci.edu/ml/datasets/adult

Mahmud Sujon, K., Binti Hassan, R., Tusnia Towshi, Z., Othman, M. A., Abdus Samad, M., & Choi, K. (2024). When to Use Standardization and Normalization: Empirical Evidence from Machine Learning Models and XAI. IEEE Access, 12, 135300–135314. https://doi.org/10.1109/ACCESS.2024.3462434

McKinney, W. (2010). Data Structures for Statistical Computing in Python. Proceedings of the 9th Python in Science Conf. https://doi.org/10.25080/Majora-92bf1922-00a

Modhugu, V. R., & Ponnusamy, S. (2024). Comparative Analysis of Machine Learning Algorithms for Liver Disease Prediction: SVM, Logistic Regression, and Decision Tree. Asian Journal of Research in Computer Science, 17(6), 188-201. https://doi.org/10.9734/ajrcos/2024/v17i6467

Mohammed, S., Budach, L., Feuerpfeil, M., Ihde, N., Nathansen, A., Noack, N., Patzlaff, H., Naumann, F., & Harmouch, H. (2022). The Effects of Data Quality on Machine Learning Performance. 1. https://doi.org/10.1145/nnnnnnn.nnnnnnn

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 2012, 2825–2830. http://jmlr.csail.mit.edu/papers/v12/pedregosa11a.html

Rao, P., … A. R. I. S. and C. (ICISC, & 2024, undefined. (2024). Machine Learning Approaches for Diabetes Prediction: Comparative Analysis and Pre-processing Insights. Ieeexplore.Ieee.OrgPVK Rao, AS Rao2024 8th International Conference on Inventive Systems and Control, 2024•ieeexplore.Ieee.Org. https://ieeexplore.ieee.org/abstract/document/10677564/

Rataj, M., Zhang, X., Wang, J.-Q., Shantal, M., Othman, Z., Abu Bakar, A., & My, A. A. B. (2023). A Novel Approach for Data Feature Weighting Using Correlation Coefficients and Min–Max Normalization. Symmetry 2023, Vol. 15, Page 2185, 15(12), 2185. https://doi.org/10.3390/SYM15122185

Salian, S., Cherishma, S., & & Powar, O. S. (2024). Enhanced Brain Tumor Detection using Support Vector Classifier and Logistic Regression with Principal Component Analysis. In 2024 Control Instrumentation System Conference (CISCON) (pp. 1-5). IEEE. https://ieeexplore.ieee.org/document/10442059

Shantal, M., Othman, Z., & Bakar, A. A. (2023). A Novel Approach for Data Feature Weighting Using Correlation Coefficients and Min-Max Normalization. Symmetry, 15(12). https://doi.org/10.3390/SYM15122185

Singh, D., & Singh, B. (2020). Investigating the impact of data normalization on classification performance. Applied Soft Computing, 97, 105524. https://doi.org/10.1016/J.ASOC.2019.105524

Studer, S., Bui, T. B., Drescher, C., Hanuschkin, A., Winkler, L., Peters, S., & Müller, K. R. (2021). Towards CRISP-ML(Q): A Machine Learning Process Model with Quality Assurance Methodology. Machine

Learning and Knowledge Extraction, 3(2), 392–413. https://doi.org/10.3390/make3020020

Uddin, S., & Lu, H. (2024). Dataset meta-level and statistical features affect machine learning performance. Scientific Reports, 14(1). https://doi.org/10.1038/S41598-024-51825-X

Waskom, M. L. (2011). seaborn: statistical data visualization. GitHub / Zenodo (Según Fuente Que Uses). https://doi.org/10.5281/zenodo.592845

Yan, Y. (2025). The optimization and impact of public sports service quality based on the supervised learning model and artificial intelligence. Scientific Reports, 15(1). https://doi.org/10.1038/s41598-025-94613-x