

Evaluation of Machine Learning Model Performance for Sentiment Analysis in Spanish Tweets under Different Class Imbalance Scenarios

ARTICLE HISTORY

Received 26 March 2026

Accepted 11 June 2026

Published 7 July 2026

Roly Steeven Cedeño Menéndez
Universidad Técnica de Manabí
Instituto de Lenguas Modernas
Portoviejo, Ecuador
roly.cedeno@utm.edu.ec
ORCID: 0009-0004-1571-9410

José Alberto León Alarcón
Universidad Técnica de Manabí
Instituto de Lenguas Modernas
Portoviejo, Ecuador
jose.leon@utm.edu.ec
ORCID: 0009-0004-6190-0990


Jandry Hernando Franco Cantos
Universidad Técnica de Manabí
Facultad de Ciencias Informáticas
Portoviejo, Ecuador
jandry.franco@utm.edu.ec
ORCID: 0009-0009-7848-9292





This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

Evaluación del Rendimiento de Modelos de Machine Learning para el Análisis de Sentimientos en Tweets en Español bajo Diferentes Escenarios de Desbalance de Clases

Evaluation of Machine Learning Model Performance for Sentiment Analysis in Spanish Tweets under Different Class Imbalance Scenarios

Roly Steeven Cedeño Menéndez 
 Universidad Técnica de Manabí
 Instituto de Lenguas Modernas
 Portoviejo, Ecuador
 roly.cedeno@utm.edu.ec

José Alberto León Alarcón 
 Universidad Técnica de Manabí
 Instituto de Lenguas Modernas
 Portoviejo, Ecuador
 jose.leon@utm.edu.ec

Jandry Hernando Franco Cantos 
 Universidad Técnica de Manabí
 Facultad de Ciencias Informáticas
 Portoviejo, Ecuador
 jandry.franco@utm.edu.ec

Resumen— El análisis de sentimientos ha adquirido gran relevancia en la clasificación de polaridades en textos no estructurados. Sin embargo, uno de sus principales desafíos lo constituye el desequilibrio de clases, el cual afecta de manera significativa el rendimiento de los modelos de aprendizaje automático. Por ello, el presente estudio compara el desempeño de seis algoritmos de clasificación (*Naive Bayes*, *SVM*, *Logistic Regression*, *Decision Tree*, *Random Forest* y *XGBoost*) en tweets en español, considerando tres escenarios: distribución equilibrada, moderadamente desequilibrada y totalmente desequilibrada. La evaluación se realizó mediante las métricas de exactitud, precisión, *recall* y *f1-score*. Los resultados demuestran que los modelos lineales logran un mejor rendimiento en escenarios balanceados, aunque su desempeño disminuye conforme aumenta el desequilibrio. Por otra parte, *Naive Bayes* mantiene un comportamiento más estable entre escenarios, y además, una alternativa competitiva es representada por *XGBoost*. Asimismo, se observa que el impacto del desbalance no se refleja adecuadamente en la exactitud, siendo el *f1-score* una métrica más representativa. En conjunto, los resultados resaltan la importancia de considerar el desequilibrio en la selección de modelos.

Palabras Clave— análisis de sentimientos en español, aprendizaje automático, desbalance de clases, clasificación de texto, *f1-score*

Abstract— *Sentiment analysis has gained significant relevance in the classification of polarities in unstructured texts; however, one of its main challenges is class imbalance, which significantly affects the performance of machine learning models. Therefore, this study compares the performance of six classification algorithms (Naive Bayes, SVM, Logistic Regression, Decision Tree, Random Forest, and XGBoost) on Spanish-language tweets, considering three scenarios: balanced distribution, moderately imbalanced, and highly imbalanced. The evaluation was conducted using the metrics of accuracy, precision, recall, and f1-score. The results show that linear models achieve better performance in balanced scenarios, although their effectiveness decreases as the imbalance increases. Meanwhile, Naive Bayes maintains a more stable behavior across*

scenarios, and XGBoost represents a competitive alternative. Additionally, it is observed that the impact of imbalance is not adequately reflected by accuracy, making the f1-score a more representative metric. Altogether, these findings highlight the importance of considering class imbalance in model selection.

Keywords— *Spanish sentiment analysis, machine learning, class imbalance, text classification, f1-score*

I. INTRODUCTION

En la última década, las redes sociales han emergido como uno de los ecosistemas digitales más prolíficos en cuanto a producción de contenido e intercambio de opiniones en la red, donde plataformas de microblogging como *X* posibilitan que una vasta comunidad de usuarios comparta sus pensamientos, valoraciones y estados emocionales sobre múltiples temáticas de manera inmediata [1]. Ante el volumen masivo de publicaciones generadas de forma continua, el procesamiento automatizado de dicha información ha adquirido una relevancia creciente dentro de las disciplinas del aprendizaje automático y la minería de texto. En este contexto, el análisis de sentimientos se ha posicionado como una metodología ampliamente adoptada para detectar y categorizar las opiniones contenidas en textos, facilitando así su clasificación en polaridad: positiva, negativa o neutral [2]. La utilización del análisis de sentimientos en las redes sociales ha alcanzado diversos temas, tales como la evaluación de la percepción ciudadana, el monitoreo de la reputación corporativa, el análisis del discurso político y el rastreo de fenómenos sociales emergentes [3]. En ese sentido, un rendimiento adecuado en la clasificación de información no estructurada como lo es el texto ha sido demostrado por los modelos de aprendizaje automático, entre los cuales destacan *Naive Bayes*, *Support Vector Machine*, *Logistic Regression*, *Decision Tree*, *Random Forest* y *XGBoost*, los cuales han sido utilizados de manera exitosa en problemas de clasificación de texto, gracias a su

capacidad para extraer y generalizar patrones a partir de conjuntos de datos previamente etiquetados [4].

No obstante, uno de los desafíos más comunes que enfrentan los modelos de clasificación es la presencia del desbalance en la distribución de clases dentro de los conjuntos de entrenamiento [5]. Esto sucede cuando en el conjunto de datos existe una predominancia de una clase en comparación a otra. En el análisis de sentimientos, este dilema es frecuente lo cual puede provocar que el modelo clasificador presente ciertos sesgos al realizar las predicciones, y de esta manera deteriore su capacidad para poder clasificar la clase minoritaria. Como resultado, métricas comúnmente usadas para medir el nivel de eficacia del modelo como la exactitud (*accuracy*) pueden reflejar de manera imprecisa el rendimiento real del modelo, lo que hace necesario optar por otras métricas tales como precisión, *recall* y *f1-score* para una evaluación más rigurosa [6].

Frente a este problema, resulta imprescindible examinar de qué manera los distintos niveles de distribución afectan el desempeño de los modelos. Es por esto que, el presente trabajo tiene como objetivo realizar una evaluación comparativa del rendimiento de seis modelos de aprendizaje automático en la clasificación de sentimientos sobre *tweets* en español, en distintos niveles de distribución de clases. Para ello, se establecen 3 escenarios: una distribución equilibrada, moderadamente desequilibrada y totalmente desequilibrada. El rendimiento de los modelos es medido mediante las métricas de *accuracy* y *f1-score*, las cuales permiten observar el rendimiento real de cada modelo en los distintos escenarios. A partir de este análisis, se pretende identificar cuales modelos tiene una mejor tolerancia frente al desbalance de las clases y determinar cómo este desequilibrio influye en los resultados de la clasificación de sentimientos.

II. MARCO TEÓRICO

A. Análisis de sentimientos

El análisis de sentimientos, denominado en la literatura anglosajona *sentiment analysis* u *opinion mining*, constituye una técnica propia del campo de la minería de texto y el procesamiento del lenguaje natural (NLP), cuyo propósito fundamental radica en identificar, extraer y categorizar las opiniones, emociones o actitudes manifiestas en un corpus textual [7]. De manera convencional, dicha categorización se estructura en torno a polaridades como positiva, negativa o neutral, si bien en determinados contextos puede ampliarse hacia una caracterización emocional de mayor granularidad [8].

Esta disciplina ha evolucionado considerablemente en los últimos años, debido al crecimiento exponencial del volumen de datos generados en entornos digitales [9]. Las redes sociales se han posicionado como una fuente de gran valor para examinar la percepción colectiva de los usuarios en torno a una amplia variedad de temas, que van desde productos y servicios hasta eventos de relevancia o figuras de notoriedad social [10]. Mediante la aplicación de esta técnica, es posible convertir grandes masas de datos no estructurados en conocimiento accionable, susceptible de orientar procesos de toma de decisiones en distintos ámbitos [11].

B. Aprendizaje automático para la clasificación de texto

El aprendizaje automático ha adquirido un protagonismo indiscutible en el desarrollo de sistemas orientados al procesamiento y análisis de grandes volúmenes de información textual [12]. En el ámbito de la clasificación de texto, estos enfoques permiten asignar de manera automática una categoría a un documento a partir de su contenido semántico, lo que resulta de gran utilidad en aplicaciones como el análisis de sentimientos, la identificación de correo no deseado y la organización temática de contenidos informativos [13].

Dentro de este dominio, un lugar predominante es ocupado por los modelos de aprendizaje automático, gracias a su capacidad para inferir patrones relevantes a partir de datos previamente etiquetados [14]. Bajo este paradigma, el clasificador es entrenado sobre un conjunto de instancias, como publicaciones en redes sociales, en las que cada elemento se encuentra previamente asociado a una clase determinada, a través de lo cual el modelo construye una función de mapeo que le permite generalizar y predecir la categoría de nuevas observaciones no contempladas durante el entrenamiento [15].

En el campo de la clasificación textual supervisada, existe un conjunto consolidado de algoritmos de amplia aplicación [16], entre los cuales está *Naive Bayes* que fundamenta su funcionamiento en el cálculo de probabilidades condicionales bajo el supuesto de independencia entre características. Por su parte, *Support Vector Machine (SVM)* opera buscando el hiperplano de separación óptimo que maximice el margen entre las clases en el espacio de características. En tanto que, la *Logistic Regression* estima la probabilidad de asignación a una clase a través de una función logística. Los métodos basados en estructuras arbóreas, como el *Decision Tree* y el *Random Forest*, tienen la capacidad de capturar relaciones no lineales entre variables, mientras que algoritmos de mayor sofisticación como *XGBoost* emplean estrategias de ensamblado secuencial para optimizar la capacidad predictiva del modelo final [17].

La elección del modelo adecuado para una tarea en específico depende de múltiples factores, tales como el problema a resolver, la calidad y cantidad de los datos disponibles y su distribución. En los problemas de clasificación de texto, el rendimiento de los modelos puede verse comprometido por el desequilibrio en la representación de las clases dentro del conjunto de datos, por lo que se hace imprescindible realizar una evaluación de su comportamiento bajo distintos escenarios de distribución [18].

C. Desbalance de clases

El desbalance de clases se manifiesta cuando la distribución de las categorías en un conjunto de datos presenta una asimetría pronunciada, es decir, cuando una o más clases concentran una proporción de instancias notablemente superior a la de las restantes [19]. Este fenómeno aparece con frecuencia en múltiples dominios de aplicación, incluyendo el análisis de sentimientos, donde determinadas categorías, como las opiniones de polaridad negativa o neutral, tienden a estar sobrerrepresentadas respecto a otras [20].

La presencia de este desequilibrio puede comprometer de forma considerable la capacidad predictiva de los modelos de aprendizaje automático, debido a que gran parte de estos algoritmos supervisados presenta una notable tendencia a

favorecer la clase dominante durante el entrenamiento, dado que su función objetiva busca minimizar el error de clasificación global [21], lo que puede derivar en modelos que alcanzan valores elevados de exactitud (*accuracy*) simplemente mediante la predicción sistemática de la clase mayoritaria, sin haber aprendido representaciones significativas de las clases minoritarias, dando como resultado un desempeño marcadamente deficiente en la detección de estas últimas. En consecuencia, genera una circunstancia especialmente problemática en aplicaciones donde precisamente las clases subrepresentadas poseen una mayor relevancia práctica [22]. En el marco de este trabajo, con el propósito de cuantificar el impacto del desbalance de clases sobre el rendimiento de los modelos analizados, se definen tres escenarios experimentales diferenciados, a saber, un conjunto de datos con distribución equilibrada, un segundo conjunto con desbalance de intensidad moderada y un tercer conjunto caracterizado por un desbalance de alta severidad. De este modo, permite llevar a cabo un análisis comparativo sistemático de la capacidad de adaptación de cada algoritmo ante distintas distribuciones de datos, con el fin de identificar cuáles exhiben una mayor robustez y estabilidad frente a este tipo de condiciones adversas.

D. Métricas de evaluación

La cuantificación del desempeño de los modelos de aprendizaje automático constituye un componente esencial en cualquier tarea de clasificación, dado que posibilita valorar de manera objetiva su capacidad para generalizar y predecir correctamente las categorías de datos no observados durante el entrenamiento [23]. En el contexto específico del análisis de sentimientos orientado a discriminar textos entre categorías como positivo y negativo, resulta imprescindible recurrir a indicadores que capturen con fidelidad el comportamiento real del clasificador, particularmente en situaciones donde la distribución de las clases presenta desequilibrios significativos.

Entre los indicadores de rendimiento más difundidos se encuentra la exactitud (*accuracy*), que cuantifica la fracción de predicciones acertadas respecto al total de instancias evaluadas, calculándose como el cociente entre el número de clasificaciones correctas y el volumen total de observaciones [24]. No obstante, pese a su interpretabilidad inmediata y su amplia adopción, esta métrica puede inducir a conclusiones erróneas en conjuntos de datos con distribuciones asimétricas, puesto que un clasificador podría alcanzar valores aparentemente satisfactorios al limitarse a predecir de forma sistemática la categoría predominante, sin haber aprendido a distinguir adecuadamente entre las clases [25].

Por su parte, la precisión (*precision*) expresa la proporción de instancias clasificadas como positivas que corresponden efectivamente a dicha categoría, de modo que su utilidad radica en la evaluación de la calidad de las predicciones afirmativas emitidas por el modelo, siendo especialmente relevante, en particular, en escenarios donde los falsos positivos conllevan consecuencias adversas considerables [26].

El *recall* o sensibilidad, en cambio, mide la aptitud del modelo para recuperar correctamente la totalidad de las instancias pertenecientes a la clase positiva. Es decir que, refleja en qué medida el clasificador es capaz de detectar los casos verdaderamente positivos presentes en el conjunto de evaluación, siendo un indicador crítico cuando la omisión de casos positivos reales resulta particularmente costosa [27].

El *f1-score*, por otro lado, integra en un único valor tanto la precisión como el *recall*, combinándolos mediante su media armónica. Por lo tanto, esta propiedad lo convierte en una métrica especialmente adecuada para escenarios con desbalance de clases, ya que establece un balance entre la capacidad del modelo para minimizar tanto los falsos positivos como los falsos negativos, ofreciendo así una valoración más equilibrada e integral del desempeño clasificatorio [28].

En el presente estudio, se adoptan la exactitud y el *f1-score* como métricas primarias de evaluación, dado que en conjunto, permiten examinar tanto el rendimiento global del clasificador como su capacidad discriminativa frente a distribuciones de clases desiguales. Es por esto que la utilización conjunta de ambos indicadores facilita una comparación sistemática y rigurosa del comportamiento de los distintos algoritmos de aprendizaje automático en los diferentes escenarios experimentales contemplados.

III. METODOLOGÍA

A. Recolección de datos

El corpus empleado en el presente estudio proviene de publicaciones extraídas de la red social X. Dicho conjunto de datos fue compilado en el marco de una investigación precedente [29], cuyo objetivo central consistió en examinar las valoraciones y percepciones expresadas por los usuarios de la plataforma en relación con la gestión del mandatario ecuatoriano.

La obtención de los datos se limitó a publicaciones redactadas en el idioma español, con el propósito de caracterizar el sentimiento manifestado por la comunidad hispanohablante, con un intervalo temporal de recopilación que comprende desde noviembre de 2023 hasta abril de 2024, periodo que permitió conformar un corpus representativo de la opinión pública durante los primeros meses del gobierno presidencial. En su versión original, el *dataset* comprende un total de 3.177 *tweets*, los cuales fueron sometidos a un proceso de etiquetado manual en torno a tres categorías de sentimiento: positivo, negativo y neutral. Este procedimiento de clasificación se diseñó con el fin de poder garantizar la integridad y la consistencia de las etiquetas asignadas, de esta manera contribuyendo a elevar la fiabilidad de los modelos de aprendizaje automático.

Para los fines del presente trabajo, se seleccionó un subconjunto del corpus original conformado exclusivamente por las categorías positiva y negativa. La exclusión de la clase neutral responde a tres razones fundamentales. En primer lugar, los *tweets* de polaridad neutral suelen presentar una menor carga semántica diferenciadora, lo que dificulta su separación respecto a las otras clases y puede introducir ambigüedad en el proceso de clasificación. En segundo lugar, la formulación binaria del problema permite aislar con mayor precisión el efecto del desbalance de clases sobre el rendimiento de los modelos, que constituye el objeto central de este estudio. Al incorporar una tercera clase, dicho efecto se vería condicionado adicionalmente por la complejidad multiclase, dificultando la interpretación de los resultados. En tercer lugar, el análisis de la polaridad positiva-negativa representa el enfoque más frecuente en la literatura sobre análisis de sentimientos en redes sociales, lo que facilita la comparabilidad de los hallazgos con estudios previos. El subconjunto resultante comprende un total de 644 *tweets*, de los cuales 117 corresponden a la clase positiva y 527 a la clase negativa. Esta distribución evidencia una marcada asimetría

entre las categorías, condición que resulta apropiada para analizar de manera controlada el efecto del desbalance de clases sobre el desempeño de los modelos de clasificación. Mediante la utilización de este corpus es posible evaluar el comportamiento de distintos algoritmos de aprendizaje automático en un entorno de aplicación real, caracterizado por datos provenientes de redes sociales y por las particularidades propias del registro lingüístico informal predominante en este tipo de plataformas.

B. Escenarios de balance de datos

Con el propósito de examinar la influencia del desequilibrio entre clases sobre el desempeño de los modelos de aprendizaje automático, se establecieron tres configuraciones experimentales que contemplan distintos niveles de distribución de los datos:

- **Distribución equilibrada:** En esta configuración se empleó un número idéntico de instancias para ambas categorías, conformando un subconjunto de 117 tweets positivos y 117 tweets negativos.
- **Distribución moderadamente desequilibrada:** Se consideraron 117 tweets positivos frente a 200 tweets negativos, generando así un nivel intermedio de asimetría entre clases.
- **Distribución severamente desequilibrada:** Se utilizó la totalidad del corpus disponible, integrado por 117 tweets positivos y 527 tweets negativos, reproduciendo fielmente la distribución original del conjunto de datos.

Para la construcción de los escenarios equilibrado y moderadamente desequilibrado, se llevó a cabo una selección aleatoria de instancias negativas a partir del conjunto original de 527 *tweets*, con la finalidad de asegurar la representatividad de las muestras resultantes y, al mismo tiempo, minimizar la introducción de sesgos sistemáticos en la conformación de los subconjuntos experimentales. Una vez definidos los tres escenarios, los datos correspondientes a cada configuración fueron particionados en dos subconjuntos mutuamente excluyentes, uno destinado al entrenamiento del modelo y otro reservado para su evaluación. Dicha partición se realizó siguiendo a una proporción del 80% para la fase de entrenamiento y del 20% para la fase de prueba, esquema que permite valorar el rendimiento de los clasificadores sobre datos no expuestos durante el proceso de aprendizaje.

La delimitación de estas tres configuraciones experimentales posibilita un análisis comparativo riguroso de la manera en que los distintos grados de desequilibrio entre clases condicionan el comportamiento de los algoritmos de aprendizaje automático, lo que facilita la identificación de aquellos que exhiben una mayor capacidad de adaptación y estabilidad ante este tipo de distribuciones asimétricas.

C. Modelos de aprendizaje automático evaluados

Los modelos considerados fueron *Naive Bayes*, *Support Vector Machine (SVM)*, *Logistic Regression*, *Decision Tree*, *Random Forest* y *XGBoost*, cuya selección responde tanto a su consolidada presencia en la literatura sobre análisis de sentimientos como a la diversidad de enfoques que representan para modelar distintos tipos de relaciones subyacentes en los datos [30]. La implementación de estos modelos se llevó a cabo mediante herramientas de programación especializadas

en análisis de datos, utilizando las representaciones vectoriales obtenidas tras el procesamiento previo del texto, y con el fin de garantizar la comparabilidad de los resultados, todos los algoritmos fueron entrenados y evaluados bajo condiciones experimentales homogéneas, asegurando así que las diferencias observadas en el rendimiento sean atribuibles a las características propias de cada modelo y no a variaciones en el entorno de evaluación.

En cuanto a la configuración de los clasificadores, se optó por mantener los valores de hiperparámetros predeterminados proporcionados por las bibliotecas utilizadas, con el propósito de examinar el rendimiento base de cada algoritmo sin la intervención de procesos de ajuste fino, lo que permite, en consecuencia, obtener una caracterización del comportamiento intrínseco de los modelos frente a los distintos niveles de desequilibrio entre clases, sin que los resultados se vean condicionados por optimizaciones específicas.

Sobre los conjuntos de datos correspondientes a las tres configuraciones experimentales definidas, cada uno de los clasificadores fue entrenado y, posteriormente, evaluado de forma sistemática mediante las métricas de rendimiento establecidas, lo que permitió, en consecuencia, realizar una comparación coherente y estructurada del desempeño de los algoritmos, así como examinar su capacidad de respuesta ante variaciones en la distribución de las categorías en el conjunto de datos.

D. Métricas de evaluación

Para cuantificar el rendimiento de los modelos de aprendizaje automático, se emplearon diversas métricas derivadas de la matriz de confusión, instrumento que posibilita un análisis detallado del comportamiento de los clasificadores en función de las predicciones emitidas [31]. En el marco de este estudio, se contemplaron cuatro resultados posibles de clasificación:

- **Verdaderos Positivos (True Positives o TP):** Ejemplos que han sido correctamente clasificados como positivos.
- **Falsos Positivos (False Positives o FP):** Ejemplos clasificados incorrectamente como positivos, y que en realidad son negativos.
- **Verdaderos Negativos (True Negatives o TN):** Ejemplos que han sido correctamente clasificados como negativos.
- **Falsos Negativos (False Negatives o FN):** Ejemplos clasificados incorrectamente como negativos, y que en realidad son positivos.

Basándonos en estos casos y mediante la matriz de confusión en la Tabla I, podemos calcular diversas métricas para evaluar el rendimiento del modelo, entre las métricas más utilizadas son:

TABLA I. MATRÍZ DE CONFUSIÓN

Matriz de confusión		Estimado por el modelo	
		Negativo (N)	Positivo (P)
Real	Negativo	TN	FP
	Positivo	FN	TP

- **Exactitud (accuracy):** Es el porcentaje de ejemplos correctamente clasificados sobre el total.

$$\text{Exactitud} = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

- **Precisión (precision):** Es el porcentaje de ejemplos correctamente clasificados como positivos sobre el total de ejemplos clasificados como positivos.

$$\text{Precisión} = \frac{TP}{TP+FP} \quad (2)$$

- **Exhaustividad (recall) o Sensibilidad (Sensitivity):** Es el porcentaje de ejemplos correctamente clasificados como positivos sobre el total de ejemplos que son realmente positivos.

$$\text{Exhaustividad} = \frac{TP}{TP+FN} \quad (3)$$

- **Valor-F (f1-score):** Es la media armónica de la precisión y la exhaustividad. Proporciona una medida equilibrada que toma en cuenta tanto la precisión como la exhaustividad, especialmente útil cuando se necesita un balance entre ambas métricas.

$$F1 - \text{Score} = 2 * \frac{\text{Precisión} * \text{Exhaustividad}}{\text{Precisión} + \text{Exhaustividad}} \quad (4)$$

Estas métricas fueron aplicadas de forma uniforme sobre la totalidad de los modelos evaluados y en cada una de las configuraciones experimentales definidas, lo que permitió establecer comparaciones objetivas y consistentes entre algoritmos, así como examinar con rigor el efecto del desequilibrio entre clases sobre los resultados de clasificación obtenidos.

IV. ANÁLISIS Y DISCUSIÓN DE RESULTADOS

A. Rendimiento de los modelos en el escenario de distribución equilibrada

En el presente apartado se exponen los resultados obtenidos tras la evaluación de los distintos modelos de aprendizaje automático sobre el conjunto de datos con distribución equilibrada, conformado por una cantidad idéntica de instancias pertenecientes a las categorías positiva y negativa.

La Tabla II sintetiza los valores de exactitud (*accuracy*), precisión (*precision*), exhaustividad (*recall*) y *f1-score* registrados por cada modelo bajo esta configuración experimental.

TABLA II. RENDIMIENTO DE LOS MODELOS EN EL ESCENARIO DE DISTRIBUCIÓN EQUILIBRADA

Modelo	Exactitud	Precision	Recall	F1-Score
Naive Bayes	80,9%	78,3%	81,8%	80,0%
Support Vector Machine (SVM)	83,0%	80,0%	87,0%	83,3%
Logistic Regression	85,1%	83,3%	87,0%	85,1%
Decision Tree	61,7%	69,2%	39,1%	50,0%
Random Forest	74,5%	65,7%	100,0%	79,3%
XGBoost	74,5%	69,0%	87,0%	76,9%

Los resultados presentados en la Tabla II revelan un comportamiento diferenciado entre los modelos evaluados, donde la *Logistic Regression* emerge como el modelo de mayor rendimiento en este escenario al alcanzar un *f1-score* de

85,1% y una exactitud equivalente, lo que refleja una capacidad equilibrada para clasificar correctamente ambas polaridades. Le sigue en desempeño el *SVM* con un *f1-score* de 83,3%, caracterizado por un *recall* elevado del 87,0% que evidencia una notable sensibilidad para detectar instancias positivas. Por su parte, *Naive Bayes* obtiene resultados competitivos con un *f1-score* del 80,0%, consolidándose como una alternativa eficiente considerando su simplicidad computacional.

En el extremo opuesto, el *Decision Tree* registra el rendimiento más deficiente del conjunto evaluado, con un *f1-score* de tan solo 50,0%, como consecuencia directa de un *recall* reducido del 39,1%, lo que indica una capacidad limitada para identificar correctamente los casos positivos. En contraste, *Random Forest*, pese a alcanzar un *recall* perfecto del 100,0%, presenta una precisión notablemente inferior del 65,7%, lo que se traduce en una tendencia a clasificar como positivas instancias que no lo son y, por ende, refleja un sesgo hacia dicha categoría. En tanto que *XGBoost*, con un *f1-score* del 76,9%, se sitúa en una posición intermedia mostrando un comportamiento aceptable, aunque inferior al de los modelos lineales. En términos generales, los resultados obtenidos en este escenario sugieren que los clasificadores de naturaleza lineal, particularmente la *Logistic Regression* y el *SVM*, exhiben una mayor capacidad de adaptación ante distribuciones equilibradas de datos, logrando así un balance adecuado entre precisión y exhaustividad en la discriminación de ambas clases de sentimiento.

B. Rendimiento de los modelos en el escenario moderadamente desequilibrado

En esta configuración experimental, se analizó el comportamiento de los clasificadores ante un conjunto de datos que presenta un nivel intermedio de asimetría en la distribución de clases, conformado por 117 instancias positivas y 200 instancias negativas. Este escenario permite examinar la sensibilidad de los modelos ante una perturbación moderada respecto a las condiciones de equilibrio evaluadas previamente.

TABLA III. RENDIMIENTO DE LOS MODELOS EN EL ESCENARIO MODERADAMENTE DESEQUILIBRADO

Modelo	Exactitud	Precision	Recall	F1-Score
Naive Bayes	76,6%	64,5%	83,3%	72,7%
Support Vector Machine (SVM)	76,6%	62,5%	52,6%	57,1%
Logistic Regression	75,0%	100,0%	15,8%	27,3%
Decision Tree	71,9%	52,9%	47,4%	50,0%
Random Forest	76,6%	75,0%	31,6%	44,4%
XGBoost	71,9%	53,8%	36,8%	43,8%

Los resultados expuestos en la Tabla III ponen de manifiesto una degradación generalizada del rendimiento respecto al escenario equilibrado, evidenciando así la sensibilidad de varios algoritmos ante la introducción de desequilibrio en la distribución de los datos. El caso más notable es el de la *Logistic Regression*, que pese a alcanzar una precisión perfecta del 100,0%, registra un *recall* drásticamente reducido del 15,8%, lo que se traduce en un *f1-score* de apenas 27,3%, comportamiento que indica que el modelo adopta una estrategia de clasificación extremadamente conservadora, ya que emite predicciones positivas con alta certeza pero, al

mismo tiempo, omite la gran mayoría de los casos positivos reales, lo que finalmente lo convierte en un clasificador inadecuado bajo estas condiciones.

De manera similar, el *SVM* experimenta una caída pronunciada en su *f1-score*, descendiendo desde 83,3% en el escenario equilibrado hasta 57,1% en el presente escenario, lo que refleja una pérdida considerable de su capacidad discriminativa ante el incremento del desequilibrio entre clases. *Random Forest* y *XGBoost* también muestran deterioros significativos, con *f1-scores* de 44,4% y 43,8% respectivamente, como consecuencia de valores de *recall* reducidos que evidencian dificultades crecientes para recuperar correctamente las instancias de la clase minoritaria, en contraste, *Naive Bayes* se posiciona como el modelo de mayor estabilidad en este escenario al obtener el *f1-score* más elevado del conjunto con un 72,7%, sustentado en un *recall* del 83,3% que denota una capacidad relativamente sólida para identificar los casos positivos aun en presencia de desequilibrio moderado. Por su parte, el *Decision Tree* mantiene un *f1-score* del 50,0%, resultado idéntico al registrado en el escenario anterior, lo que sugiere una insensibilidad relativa a las variaciones en la distribución, aunque a costa de un rendimiento global limitado.

En términos generales, los resultados de este escenario revelan que la introducción de un desequilibrio moderado impacta de forma heterogénea sobre los distintos algoritmos evaluados, siendo los modelos lineales los más afectados, mientras que *Naive Bayes* demuestra una mayor capacidad de adaptación ante cambios en la distribución de los datos.

C. Rendimiento de los modelos en el escenario severamente desequilibrado

En el presente apartado, se exponen los resultados obtenidos al evaluar los clasificadores sobre el corpus original en su totalidad, cuya distribución natural refleja el mayor nivel de asimetría entre clases contemplado en este estudio, con 117 instancias positivas frente a 527 instancias negativas. Este escenario representa las condiciones reales del conjunto de datos y constituye el caso más exigente para los modelos evaluados, al reproducir fielmente el desequilibrio inherente a los datos provenientes de entornos de redes sociales.

TABLA IV. RENDIMIENTO DE LOS MODELOS EN EL ESCENARIO SEVERAMENTE DESEQUILIBRADO

Modelo	Exactitud	Precision	Recall	F1-Score
Naive Bayes	82,2%	55,3%	77,8%	64,6%
Support Vector Machine (SVM)	82,2%	66,7%	16,0%	25,8%
Logistic Regression	80,6%	0,0%	0,0%	0,0%
Decision Tree	80,6%	50,0%	36,0%	41,9%
Random Forest	81,4%	100,0%	4,0%	7,7%
XGBoost	85,3%	75,0%	36,0%	48,6%

Los resultados consignados en la Tabla IV revelan una degradación generalizada y pronunciada del rendimiento en la mayoría de los modelos evaluados, siendo este el escenario donde el impacto del desequilibrio entre clases se manifiesta con mayor intensidad. *Logistic Regression* se destaca como el caso más extremo, que registra valores nulos tanto en precisión como en *recall* y *f1-score*, lo que indica que el modelo ha

colapsado hacia una estrategia de predicción sistemática de la clase mayoritaria. En consecuencia, resulta completamente incapaz de identificar instancia alguna de la categoría positiva, comportamiento que, conocido en la literatura como *majority class bias*, ilustra de manera contundente el efecto devastador que el desequilibrio severo puede ejercer sobre clasificadores lineales sin mecanismos de compensación [32].

Random Forest presenta una situación análoga desde una perspectiva diferente, ya que si bien alcanza una precisión perfecta del 100,0%, su *recall* desciende a un valor residual del 4,0%, derivando así en un *f1-score* de apenas 7,7%, resultado que indica que el modelo emite predicciones positivas con absoluta certeza en los escasísimos casos en que se aventura a hacerlo, pero que, al mismo tiempo, prácticamente renuncia a detectar la clase minoritaria, lo que lo convierte en un clasificador de utilidad marginal en este contexto, mientras que, de forma similar, el *SVM* experimenta una caída severa en su *f1-score*, reduciéndose al 25,8% como consecuencia de un *recall* de tan solo 16,0%.

En el extremo opuesto, *Naive Bayes* se posiciona como el modelo que posee una mayor solidez en este escenario, por el cual al tener el *f1-score* más elevado del conjunto con un 64,6%, seguido de un *recall* del 77,8%. Estos resultados evidencian que su enfoque probabilístico le confiere una mayor tolerancia ante distribuciones asimétricas lo que le permite mantener un nivel de rendimiento relativamente superior al resto de los modelos evaluados, mientras que *XGBoost*. Por otro lado, muestra una estabilidad destacable con un *f1-score* del 48,6% y una precisión del 75,0%, consolidándose como el segundo modelo con mejor rendimiento en este escenario, lo que reafirma que su arquitectura de ensamblado secuencial le otorga una capacidad de adaptación superior frente a condiciones de desequilibrio severo.

En términos generales, los resultados de este escenario ponen en evidencia que el desequilibrio severo entre clases constituye en un factor crítico que compromete sustancialmente la capacidad discriminativa de la mayoría de los modelos evaluados, siendo *Naive Bayes* y *XGBoost* los únicos que logran mantener un nivel de desempeño funcional bajo estas condiciones adversas.

D. Análisis comparativo del rendimiento entre escenarios

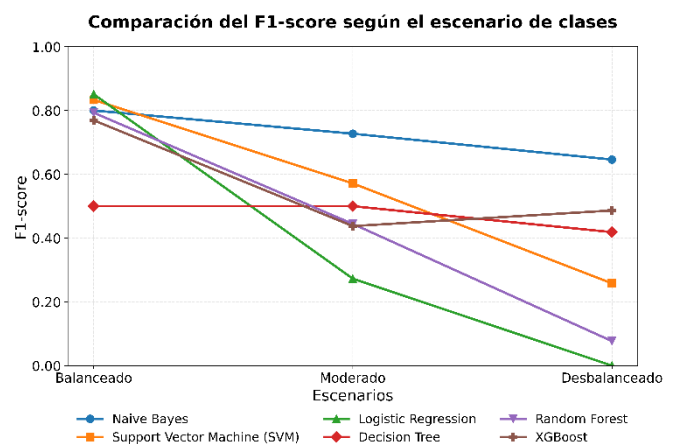


Fig. 1. Comparación del *f1-score* según el escenario de clases

La Figura 1 ilustra la evolución del *f1-score* de cada modelo evaluado a lo largo de los tres escenarios experimentales considerados, permitiendo visualizar de manera integral cómo el rendimiento de los clasificadores se ve condicionado por el nivel de desequilibrio presente en la distribución de los datos.

De manera general, se aprecia una tendencia descendente en el desempeño de la mayoría de los modelos a medida que el grado de asimetría entre clases se incrementa, siendo el escenario equilibrado el que concentra los valores de *f1-score* más elevados en prácticamente todos los casos, lo que a su vez refuerza la premisa de que el desequilibrio entre clases constituye un factor determinante en la capacidad predictiva de los algoritmos de aprendizaje automático aplicados a tareas de clasificación textual.

No obstante, la magnitud y la forma en que dicha degradación se produce varía considerablemente entre modelos, observándose que la *Logistic Regression* y el *SVM* exhiben las caídas más pronunciadas y abruptas, ya que la primera desciende desde un *f1-score* de 85,1% en el escenario equilibrado hasta un valor nulo en el escenario severamente desequilibrado, mientras que el *SVM* reduce su rendimiento desde 83,3% hasta 25,8%. Este comportamiento sugiere que ambos clasificadores lineales son altamente sensibles a las variaciones en la distribución de clases y que, en consecuencia, pierden progresivamente su capacidad discriminativa a medida que la clase minoritaria se vuelve menos representada en el conjunto de entrenamiento.

En contraste, la trayectoria de degradación más gradual y controlada del conjunto evaluado es presentada por *Naive Bayes*, con valores de *f1-score* de 80,0%, 72,7% y 64,6% para los escenarios equilibrado, moderadamente desequilibrado y severamente desequilibrado, respectivamente. Esto refleja, en consecuencia, una evolución sostenida que lo posiciona como el algoritmo de mayor estabilidad a lo largo de los tres escenarios y, por ende, evidencia una notable capacidad de adaptación ante condiciones de distribución adversas.

XGBoost, por su parte, presenta un comportamiento singular que lo distingue del resto de los modelos, ya que, si bien experimenta una reducción en su *f1-score* al transitar del escenario equilibrado al moderadamente desequilibrado, descendiendo de 76,9% a 43,8%, logra recuperar parcialmente su rendimiento en el escenario severamente desequilibrado al alcanzar un *f1-score* de 48,6%, patrón que sugiere que la arquitectura de ensamblado secuencial propia de *XGBoost* le confiere una mayor capacidad de resistencia ante desequilibrios extremos. Por ende, lo posiciona como una alternativa robusta en contextos donde los datos presentan una asimetría pronunciada, condición habitual en aplicaciones reales sobre datos provenientes de redes sociales. En síntesis, el análisis comparativo entre escenarios evidencia que la elección del algoritmo de clasificación no debe considerar únicamente su rendimiento bajo condiciones ideales de distribución, sino también su comportamiento ante escenarios de desequilibrio, siendo *Naive Bayes* y *XGBoost* los modelos que demuestran mayor solidez y consistencia a lo largo del espectro de condiciones evaluadas.

E. Análisis comparativo entre exactitud y *f1-score*

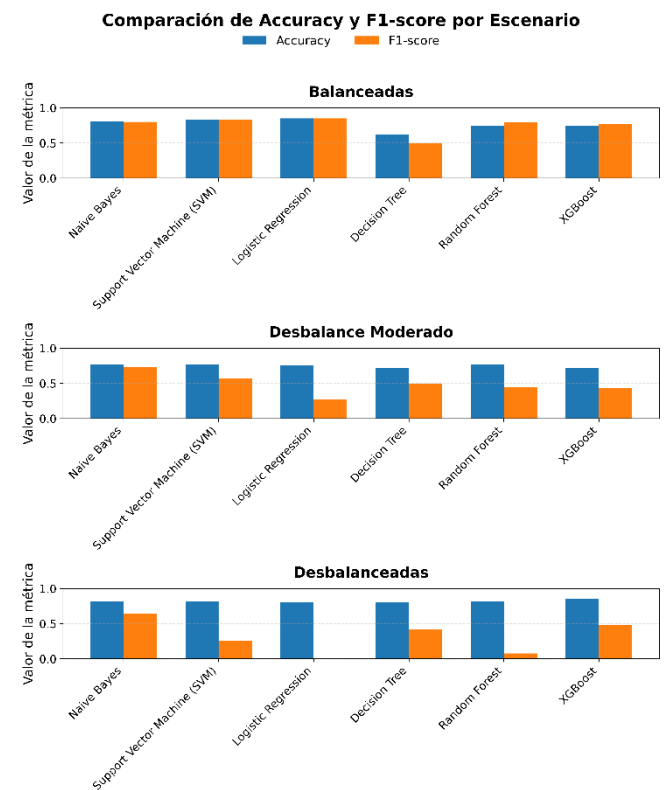


Fig. 2. Comparación de *accuracy* y *f1-score* por escenario

La Figura 2 presenta la comparación entre los valores de exactitud y *f1-score* obtenidos por cada modelo en los tres escenarios experimentales, organizados en grupos de barras que permiten visualizar simultáneamente ambas métricas para cada configuración de distribución de datos, lo que resulta especialmente relevante para examinar la divergencia entre ambos indicadores a medida que el nivel de desequilibrio entre clases se incrementa. En el escenario equilibrado, se observa una correspondencia relativamente estrecha entre la exactitud y el *f1-score* en la mayoría de los modelos, reflejando que bajo condiciones de distribución balanceada ambas métricas ofrecen una caracterización consistente del rendimiento clasificatorio, siendo la *Logistic Regression* un ejemplo paradigmático de este comportamiento al registrar valores prácticamente idénticos en ambas métricas, con 85,1% tanto en exactitud como en *f1-score*.

Sin embargo, a medida que el desequilibrio entre clases se acentúa, emerge una divergencia progresiva y sistemática entre ambos indicadores que se vuelve especialmente pronunciada en el escenario severamente desequilibrado, siendo los casos más ilustrativos de este fenómeno la *Logistic Regression* y *Random Forest*, ya que el primero mantiene una exactitud del 80,6% en dicho escenario mientras su *f1-score* colapsa a 0,0%, en tanto que el segundo alcanza una exactitud del 81,4% con un *f1-score* residual de tan solo 7,7%. Esto pone de manifiesto que, en presencia de desequilibrio severo, la exactitud puede reflejar valores aparentemente satisfactorios como consecuencia de la predicción sistemática de la clase mayoritaria, sin que ello implique una capacidad real de discriminación entre categorías.

Este comportamiento constituye una evidencia empírica contundente de las limitaciones de la exactitud como indicador único de rendimiento en contextos con distribuciones

asimétricas. En efecto, un modelo que clasifica la totalidad de las instancias como pertenecientes a la clase negativa, la predominante en el *dataset*, puede alcanzar una exactitud global cercana al 81,8%, correspondiente a la proporción natural de dicha clase, sin haber aprendido patrón discriminativo alguno, por lo que el *f1-score*. Por ello, el *recall* mediante su media armónica, resulta sensible a este tipo de comportamiento degenerativo y lo penaliza de forma proporcional, ofreciendo así una valoración más fidedigna de la capacidad real del clasificador.

La importancia de adoptar un enfoque multimétrico en la evaluación de modelos de clasificación es resaltada por los resultados de este análisis, especialmente en tareas donde existe un desequilibrio entre clases. La utilización exclusiva de la exactitud como criterio único de evaluación en estos escenarios puede conducir, en consecuencia, a conclusiones erróneas sobre la calidad real de los modelos, siendo la métrica *f1-score*, en particular, un indicador indispensable para lograr una valoración con mayor rigurosidad a la hora de medir el desempeño clasificatorio de los modelos.

V. CONCLUSIONES

Los hallazgos derivados del presente estudio demuestran de manera inequívoca que el grado de desequilibrio entre clases representa un factor de influencia crítica sobre el rendimiento de los modelos de aprendizaje automático aplicados al análisis de sentimientos en tweets en español. De modo que, a lo largo de las tres configuraciones experimentales evaluadas, se constata que el comportamiento de los algoritmos no únicamente fluctúa en términos de desempeño absoluto, sino también en su capacidad intrínseca de adaptación ante distribuciones de datos progresivamente más asimétricas, poniendo así de manifiesto la complejidad inherente a este tipo de problemas en contextos reales.

En el escenario de distribución equilibrada, los clasificadores de naturaleza lineal, en particular la *Logistic Regression* y el *SVM*, evidenciaron una superioridad manifiesta frente al resto de los algoritmos evaluados, logrando un balance adecuado entre precisión y *recall* que se tradujo en valores elevados de *f1-score*. Esto sugiere que la arquitectura de estos modelos resulta especialmente apropiada cuando las categorías se encuentran representadas de forma proporcional en el conjunto de datos, hallazgo que, además, guarda consistencia con lo reportado en la literatura especializada, donde los clasificadores lineales tienden a destacar en tareas de clasificación textual bajo condiciones ideales de distribución [33].

No obstante, una degradación considerable en el rendimiento de estos modelos es desencadenada por la introducción de un desequilibrio moderado, siendo el caso más ilustrativo de esta vulnerabilidad la *Logistic Regression*, pues, a pesar de preservar valores elevados de precisión, es su *recall* el que experimenta una reducción drástica. En consecuencia, esto revela una tendencia progresiva a favorecer la clase dominante en detrimento de la capacidad para recuperar instancias de la categoría minoritaria, situación que se agudiza de forma extrema en el escenario severamente desequilibrado, donde en un colapso clasificatorio total incurre el modelo, y evidencia una incapacidad absoluta para identificar ejemplos positivos. A su vez, de manera análoga, contracciones pronunciadas en su desempeño son registradas por el *SVM* y

Random Forest, lo que corrobora, por ende, su elevada sensibilidad ante las perturbaciones introducidas por el desequilibrio entre clases.

En contraposición, como el algoritmo de mayor estabilidad a lo largo del espectro de escenarios evaluados emerge *Naive Bayes*, muestra que la degradación de su *f1-score* sigue una trayectoria progresiva y notablemente más contenida en comparación con los restantes modelos, lo que sugiere, en consecuencia, que su fundamento probabilístico le proporciona una resiliencia inherente para sostener una capacidad clasificatoria consistente incluso ante distribuciones adversas. Además, pone de relieve su valor práctico en entornos reales donde el desequilibrio entre categorías constituye una condición estructural de los datos más que una anomalía excepcional.

A lo largo del análisis, se identifica un aspecto de relevancia metodológica en la relación dinámica entre las métricas de evaluación empleadas. Se observa que, en condiciones de distribución equilibrada, la exactitud y el *f1-score* convergen hacia valores similares y, por tanto, ofrecen representaciones concordantes del rendimiento de los modelos. Sin embargo, a medida que el desequilibrio entre clases se intensifica, emerge, en consecuencia, una divergencia sistemática y creciente entre ambos indicadores. De forma particularmente llamativa, varios modelos preservan valores de exactitud aparentemente satisfactorios al mismo tiempo que su *f1-score* se desploma hacia valores mínimos. Este fenómeno expone la naturaleza engañosa de la exactitud como criterio único de evaluación en contextos desequilibrados, puesto que este indicador puede inflarse artificialmente por la predicción sistemática de la clase mayoritaria sin que ello implique capacidad discriminativa real alguna.

En este sentido, la métrica *f1-score* se posiciona como el indicador de referencia para la evaluación del desempeño en presencia de desequilibrio entre clases, al integrar simultáneamente la precisión y el *recall*. Además, penaliza de forma proporcional los errores asociados a la clase minoritaria, lo cual refuerza la necesidad de adoptar estrategias de evaluación multimétrica en las tareas de clasificación aplicadas a problemas reales, donde la uniformidad en la distribución de los datos suele ser más la excepción que la norma.

El análisis comparativo transversal entre escenarios permite establecer que no existe un algoritmo universalmente óptimo para todas las condiciones de distribución, sino que la superioridad relativa de cada modelo se encuentra fuertemente condicionada por las características propias de la distribución de los datos. En este sentido, como contribución principal del presente estudio se destaca la evaluación empírica y estructurada del comportamiento de múltiples modelos de aprendizaje automático ante distintos niveles de desequilibrio de clases en un contexto aplicado de análisis de sentimientos sobre datos provenientes de redes sociales. Todo esto evidencia que *Naive Bayes* y *XGBoost* representan opciones preferentes en escenarios con distribuciones asimétricas, mientras que un rendimiento superior bajo condiciones de equilibrio entre categorías es ofrecido por los modelos lineales. Estos resultados constituyen un aporte de valor práctico para investigadores y profesionales que enfrentan problemas de clasificación en dominios donde el desequilibrio de clases es una realidad estructural ineludible.

A pesar de los aportes descritos, el presente estudio no está exento de limitaciones, puesto que el corpus utilizado se circunscribe a un dominio temático específico y a un período temporal determinado, lo que podría restringir la generalización de los hallazgos a otros contextos discursivos o a datos provenientes de distintas épocas o plataformas. Asimismo, el análisis se realizó sin la aplicación de técnicas de balanceo de clases como *SMOTE* o *undersampling*, ni de estrategias de preprocesamiento avanzado, aspectos que potencialmente habrían permitido mitigar el impacto del desequilibrio sobre el rendimiento de los modelos, en este sentido, investigaciones futuras podrían explorar el efecto combinado de dichas técnicas de balanceo junto con modelos de lenguaje preentrenados como *BERT* o *RoBERTa* en español, así como ampliar el análisis a corpus multidominio y multiclase, con el objetivo de obtener una comprensión más robusta y generalizable del comportamiento de los clasificadores ante distribuciones asimétricas en tareas de análisis de sentimientos.

REFERENCES

- [1] S. Giménez, “Redes Sociales, estado actual y tendencias 2023 OBSbusiness.school,” 2023.
- [2] A. Albladi, M. Islam, and C. Seals, “Sentiment Analysis of Twitter Data Using NLP Models: A Comprehensive Review,” *IEEE Access*, vol. 13, pp. 30444–30468, 2025, doi: 10.1109/ACCESS.2025.3541494.
- [3] J. J. Moreno and R. M. Nicolás, “Evaluación de la percepción ciudadana en la red social X mediante técnicas de minería y analítica de datos para el fortalecimiento institucional de la Secretaría Distrital de Hacienda,” Oct. 2025, Accessed: Mar. 16, 2026. [Online]. Available: <http://repository.unad.edu.co/handle/10596/78213>
- [4] I. J. Girón, “Análisis comparativo de modelos de aprendizaje supervisado para el reconocimiento de emociones en texto,” 2025, Accessed: Mar. 16, 2026. [Online]. Available: <https://uvadoc.uva.es/handle/10324/79497>
- [5] M. E. Jonatan, “Clasificación de datos desbalanceados,” May 2022, Accessed: Mar. 16, 2026. [Online]. Available: <http://sedici.unlp.edu.ar/handle/10915/147410>
- [6] J. J. Campo Yepes, “Evaluación de métricas en modelos predictivos de clasificación en machine learning,” Accessed: Mar. 16, 2026. [Online]. Available: <https://repository.universidadean.edu.co/entities/publication/81ddd265-219d-480f-8422-1e5486ad0c75>
- [7] E. M. Porras, A. R. Fernández, L. P. S. Bastos, and J. A. D. González, “Uso de Procesamiento de Lenguaje Natural para procesar respuestas abiertas de una encuesta de Opinión Pública,” *Revista Latinoamericana de Metodología de la Investigación Social: ReLMIS*, no. 29, pp. 51–67, 2025.
- [8] J. E. Lozano González, “Revisión sistemática sobre el análisis de sentimientos en interacciones por chat en videojuegos,” Feb. 2025, Accessed: Mar. 16, 2026. [Online]. Available: <http://repository.unad.edu.co/handle/10596/67163>
- [9] N. M. Sanchez Posada, “Optimización del marketing digital: revisión sistemática de técnicas de Big Data y monitoreo de redes sociales,” Aug. 2025, Accessed: Mar. 16, 2026. [Online]. Available: <http://repository.unad.edu.co/handle/10596/73389>
- [10] L. Rivadeneira, “Análisis del comportamiento de decisión usando -ProQuest,” *Revista Ibérica de Sistemas e Tecnologias de Informação*, 2023.
- [11] E. G. Mita Arancibia, “Revisión sistemática sobre análisis de datos en tiempo real: Herramientas para tomar decisiones estratégicas,” *Panel - Revista de Administración*, Jul. 2024.
- [12] A. D. Jiménez Alfaro and J. V. Díaz Ospina, “Revisión sistemática de literatura: Técnicas de aprendizaje automático (Machine Learning),” *Cuaderno activa*, vol. 13, no. 1, pp. 113–121, 2021.
- [13] A. Cevallos-Culqui, C. Pons, and G. Rodriguez, “Semi-supervised learning models for document classification: A systematic review and meta-analysis,” *Inteligencia Artificial*, vol. 26, no. 72, pp. 81–111, Jun. 2023, doi: 10.4114/intartif.vol26iss72pp81-111.
- [14] M. A. Hernández Castañeda and M. F. Forero Dorado, “Uso de algoritmos Machine Learning en la clasificación de objetos astronómicos: una revisión sistemática,” *UNAD*, Dec. 2024.
- [15] J. L. Romero Ibarra, “Análisis integral de algoritmos de clasificación en aprendizaje automático: perspectivas, comparaciones y aplicaciones,” *Serie Científica de la Universidad de las Ciencias Informáticas*, vol. 18, no. 1, pp. 283–304, 2025.
- [16] A. F. Ruiz Delgado, “Revisión sistemática de modelos de machine learning y deep learning aplicados a la detección temprana de depresión en redes sociales,” *UNAD*, Dec. 2025.
- [17] R. Tobar-Díaz, Y. Gao, J. F. Mas, and V. H. Cambrón-Sandoval, “Classification of land use and land cover through machine learning algorithms: a literature review,” *Revista de Teledetección*, vol. 2023, no. 62, pp. 1–19, Jul. 2023, doi: 10.4995/raet.2023.19014.
- [18] V. H. Bustamante Morán, C. E. Quiroz Calle, J. R. Oquendo Silva, and J. R. Oquendo Silva, “Inteligencia artificial en el diagnóstico diferencial de patologías de tejidos blandos orales,” *RECIAMUC*, vol. 9, no. 4, pp. 472–492, Dec. 2025, doi: 10.26820/reciamuc/9.(4).diciembre.2025.472-492.
- [19] L. J. Montesdeoca Espinoza, S. J. Zambrano Rojas, V. J. Pinargote-Bravo, and L. Cedeño-Valarezo, “Redes generativas para balanceo de datos en imágenes agrícolas: una revisión sistemática de la literatura,” *Revista Científica de Informática ENCRIPAR*, vol. 8, no. 16, pp. 153–168, Oct. 2025, doi: 10.56124/encriptar.v8i16.008.
- [20] H. P. Segovia Granda, “Revisión sistemática y análisis de metodologías que utilizan técnicas de minería de datos y aprendizaje automático para detección del trolling en las redes sociales,” 2022.
- [21] C. Huamañi Ninahuanca, C. J. Quintana-Castro, and N. E. Tovar Soto, “Modelo predictivo sobre pérdida de beca por motivos académicos en beneficiarios de Beca 18,” 2025.
- [22] G. Poquechoque Foronda and C. W. Pacheco Lora, “Evaluación de la experiencia de usuario ante interfaces web de software de gestión a través del análisis de emociones,” *Revista Ciencia y Tecnología Digital*, vol. 1, no. 1, pp. 1–20, Oct. 2025.
- [23] F. E. Garza, Y. M. Ramírez, A. R. Noriega, and I. N. Á. Sánchez, “Una revisión sistemática sobre la precisión de modelos de aprendizaje automático aplicados a la tasación de bienes raíces,” *RITI*, vol. 12, no. 28, pp. 4–16, 2024, doi: 10.36825/RITI.12.28.002.
- [24] J. E. Pino Cotillo, “Comparación de modelos de machine learning para la predicción temprana de diabetes mellitus tipo 2,” *UNAD*, Dec. 2025.
- [25] O. Ali, W. Abdelbaki, A. Shrestha, E. Elbasi, M. A. A. Alryalat, and Y. K. Dwivedi, “A systematic literature review of artificial intelligence in the healthcare sector,” *Journal of Innovation & Knowledge*, vol. 8, no. 1, p. 100333, Jan. 2023, doi: 10.1016/j.jik.2023.100333.
- [26] S. A. Montero Franco, “Revisión Teórica De Las Redes Neuronales Profundas Para La Detección De Malware,” *UNAD*, Aug. 2025.
- [27] J. I. Valdés Espinoza, “Clasificación automatizada de actividad cerebral normal en pacientes neurocríticos para mejorar capacidad diagnóstica,” 2022, doi: 10.58011/X5MP-TG04.
- [28] E. Cruz, M. González, and J. C. Rangel, “Técnicas de machine learning aplicadas a la evaluación del rendimiento y a la predicción de la deserción de estudiantes universitarios, una revisión,” *Prisma Tecnológico*, vol. 13, no. 1, pp. 77–87, Feb. 2022, doi: 10.33412/pri.v13.1.3039.
- [29] R. S. Cedeño Menéndez, J. A. León Alarcón, and J. H. Franco Cantos, “Análisis de Sentimientos en la Red Social ‘X’, Percepción Pública sobre el Presidente del Ecuador, Daniel Noboa (noviembre 2023 - abril 2024),” *Latin-American Journal of Computing*, vol. 12, no. 2, pp. 40–48, Jul. 2025, doi: 10.33333/lajc.vol12n2.03.
- [30] V. A. García Meza and S. H. Lázaro Barrera, “Revisión teórica de modelos de Machine Learning para la predicción del comportamiento de pago en clientes gestionados desde contact-center, sector cobranzas,” *UNAD*, Dec. 2025.
- [31] J. Urrego Piedrahita and J. J. Acosta Jiménez, “Machine Learning aplicado a la predicción de pacientes en EPS: una revisión de literatura,” *Cuaderno activa*, vol. 16, no. 1, May 2024, doi: 10.53995/20278101.1574.
- [32] H. Ali, M. Najib, M. Salleh, R. Saedudin, K. Hussain, and M. Faheem Mushtaq, “Imbalance class problems in data mining: a review,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 14, no. 3, pp. 1560–1571, 2019, doi: 10.11591/ijeecs.v14.i3.pp1560-1571.
- [33] L. U. Hidalgo Vargas, J. A. León Borges, J. C. Ramírez Pacheco, H. Toral Cruz, T. G. Makita Balcorta, and I. Osuna Galán, “Análisis de

Revisión Sistemática de la Aplicación de Algoritmos de Aprendizaje Automático en Sistemas de Detección de Intrusión en Internet de las Cosas para Ciudades Inteligentes,” *Ciencia Latina Revista Científica Multidisciplinar*, vol. 8, no. 6, pp. 11500–11517, Feb. 2024, doi: 10.37811/cl_rcm.v8i6.15929.

DECLARACIÓN DE USO DE HERRAMIENTAS DE INTELIGENCIA ARTIFICIAL

Durante la preparación de este manuscrito, por los autores se declara que una herramienta basada en inteligencia artificial fue utilizada exclusivamente con fines de traducción y mejora de la redacción, siendo empleada para optimizar la calidad lingüística y la legibilidad del texto, mientras que, en contraste, el contenido científico, el análisis de datos, las interpretaciones y las conclusiones son responsabilidad íntegra de los autores.

AUTHORS

Roly Steeven Cedeño Menéndez



Ingeniero en Sistemas de Información por la Universidad Técnica de Manabí y Magíster en Sistemas de Información con mención en Data Science por la Pontificia Universidad Católica del Ecuador. Su formación académica y experiencia profesional se enfocan en el análisis de datos, el aprendizaje automático y la aplicación de técnicas avanzadas para la extracción de conocimiento a partir de grandes volúmenes de información. Actualmente se desempeña como técnico docente en la Universidad Técnica de Manabí y cuenta con un año de experiencia adicional como docente en modalidad online.

Ha participado en proyectos de investigación vinculados a la ciencia de datos, destacando su trabajo de tesis de posgrado titulado “Análisis de sentimientos utilizando la red social X (Twitter) para medir el nivel de aceptación del nuevo presidente del Ecuador, Daniel Noboa (noviembre 2023 - abril 2024)”. También cuenta con dos artículos académicos publicados. Sus áreas de interés incluyen la inteligencia artificial, la minería de datos y el desarrollo de soluciones basadas en ciencia de datos. Sus objetivos profesionales actuales se centran en mejorar continuamente como docente y consolidarse como investigador en el área, contribuyendo con nuevas publicaciones científicas.

José Alberto León Alarcón



José León Alarcón es un profesional especializado en Ciencia de Datos, posee un máster en Sistemas de Información con mención en Data Science por la Pontificia Universidad Católica del Ecuador (PUCE Quito). Su formación académica se complementa con una sólida experiencia en el ámbito de la inteligencia artificial, especialmente en el aprendizaje automático (machine learning) y el aprendizaje profundo (deep learning). A lo largo de su trayectoria profesional, se ha enfocado en el análisis de imágenes médicas, contribuyendo al desarrollo de modelos capaces de apoyar el diagnóstico clínico mediante técnicas avanzadas de procesamiento de imágenes. Además, ha trabajado en la extracción y análisis de información a partir de datos complejos, aplicando metodologías estadísticas y herramientas computacionales modernas. Sus áreas de interés incluyen la inteligencia artificial, el análisis predictivo y el desarrollo de soluciones innovadoras que permitan transformar grandes volúmenes de datos en conocimiento útil para la toma de decisiones. Se caracteriza por su compromiso con la investigación aplicada y el desarrollo tecnológico orientado a resolver problemas reales.

AUTHORS

Jandry Franco Cantos



Ingeniero en Sistemas de Información con una Maestría en Ingeniería en Sistemas de Información, mención en Data Science. Ha formado parte de diversos proyectos enfocados en el desarrollo de software e implementación de soluciones basadas en inteligencia artificial, aplicadas al análisis de datos, la optimización de procesos y la automatización de tareas.

Actualmente se desempeña como docente universitario en la Universidad Técnica de Manabí, Ecuador, donde combina la formación académica con la investigación aplicada. Sus principales áreas de interés incluyen la inteligencia artificial, el aprendizaje automático, la visualización de datos y la ciencia de datos orientada a la toma de decisiones.

Cuenta con experiencia en la integración de herramientas tecnológicas en entornos educativos y productivos, participando activamente en iniciativas interdisciplinarias que promueven la innovación tecnológica con impacto real. Su enfoque profesional se basa en el desarrollo de soluciones prácticas y eficientes, alineadas con los avances actuales en ciencia y tecnología.

Comprometido con la formación de nuevas generaciones de profesionales, busca contribuir al avance del conocimiento científico y al desarrollo de tecnologías sostenibles que respondan a las necesidades actuales de la sociedad.

R. S. Cedeño Menéndez, J. A. León Alarcón, and J. H. Franco Cantos, "Evaluación del rendimiento de modelos de Machine Learning para el análisis de sentimientos en tweets en español", Latin-American Journal of Computing (LAJC), vol. 13, no. 2, 2026.