Aplicación del clasificador Naive Bayes para la predicción del riesgo de diabetes tipo 2 en pacientes ambulatorios del centro de salud metropolitano - Puno

Application of the Naive Bayes classifier for predicting type 2 diabetes risk in outpatients at the metropolitan health center - Puno

Jose Fernando Gomez Tacuri[1] https://orcid.org/0009-0009-7027-4545
Leonardo Sebastian Grimaldos Avila[1] https://orcid.org/0009-0009-1397-705X
Angel Javier Quispe Carita[1*] https://orcid.org/0000-0002-7357-4043
Leonid Alemán Gonzales[1] https://orcid.org/0000-0002-4072-6370
Renzo Apaza Cutipa[1] https://orcid.org/0000-0002-7089-2812

Recibido 28 de agosto de 2025, aceptado 29 de diciembre de 2025
Received: August 28, 2025 *Accepted: December 29, 2025*

RESUMEN

En este estudio se evaluó el desempeño del clasificador *Naive Bayes* para predecir el riesgo de diabetes tipo 2 en pacientes ambulatorios atendidos en el Centro de Salud Metropolitano de Puno. El análisis se realizó sobre un conjunto de datos de 121 pacientes recolectados en 2023, que incluyó información clínica, antropométrica y nutricional. Para fortalecer la confiabilidad del modelo, se incorporaron técnicas de manejo del desbalance mediante SMOTE y métodos de calibración probabilística aplicados a las predicciones generadas por *Naive Bayes*. En el conjunto de prueba, el modelo alcanzó una precisión global del 88%, con sensibilidad del 100% y especificidad del 85%, registrando 17 verdaderos negativos, 5 verdaderos positivos, 0 falsos negativos y 3 falsos positivos, lo que evidencia su utilidad como herramienta de tamizaje para la identificación temprana de pacientes en riesgo de diabetes tipo 2. Adicionalmente, se contextualizaron sus resultados frente a enfoques comúnmente utilizados en problemas similares, permitiendo situar el desempeño de *Naive Bayes* dentro del panorama actual del aprendizaje automático aplicado a la salud pública. Se concluye que *Naive Bayes* representa una alternativa viable y operativamente accesible para apoyar procesos de diagnóstico temprano y priorización de intervenciones preventivas en centros de atención primaria.

Palabras clave: Diabetes tipo 2, estratificación de riesgo, *Naive Bayes*, calibración de probabilidades, salud pública.

ABSTRACT

The performance of the Naive Bayes classifier was evaluated for predicting type 2 diabetes risk in outpatients treated at the Metropolitan Health Center of Puno. The analysis utilized a dataset of 121 patients collected in 2023, comprising clinical, anthropometric, and nutritional information. SMOTE was employed to mitigate class imbalance, and the Naive Bayes predictive probabilities were calibrated to improve reliability. In the test set, the model achieved an overall accuracy of 88%, with 100% sensitivity and 85% specificity, recording 17 true negatives, 5 true positives, 0 false negatives, and 3 false positives,

[1] Universidad Nacional del Altiplano. Escuela Profesional de Ingeniería Estadística e Informática. Puno, Perú.
Email: 73640829@est.unap.edu.pe; lgrimaldos@est.unap.edu.pe; aquispe@unap.edu.pe; laleman@unap.edu.pe; renzo@unap.edu.pe
* Autor de correspondencia: aquispe@unap.edu.pe

DOI: 10.64966/ingeniare.v33.41.

demonstrating its utility as a screening tool for early identification of patients at risk for type 2 diabetes. Additionally, the results were compared with commonly used approaches in similar studies, positioning the performance of Naive Bayes within the current landscape of machine learning applications in public health. Naive Bayes represents a viable and operationally accessible alternative to support early diagnostic processes and the prioritization of preventive interventions in primary care centers.

Keywords: Type 2 diabetes, risk stratification, Naive Bayes, probability calibration, public health.

INTRODUCCIÓN

La diabetes tipo 2 (DT2) es uno de los principales retos de salud pública por su elevada prevalencia, carga de morbilidad y costos asociados, especialmente en contextos con recursos limitados, donde la detección temprana y la estratificación de riesgo son cruciales para orientar intervenciones preventivas y optimizar el uso de recursos [1], [2]. En el Perú, y en particular en la región de Puno, donde los servicios de salud suelen enfrentar limitaciones operativas y alta demanda de atención ambulatoria, la identificación temprana de pacientes en riesgo es fundamental para orientar intervenciones preventivas más oportunas y eficientes [3].

En los últimos años, los enfoques de aprendizaje automático (ML) han mostrado un desempeño prometedor en la predicción del riesgo de DT2 a partir de variables clínicas, antropométricas y nutricionales, con resultados favorables reportados en diversos *datasets* [4] – [7]. Aunque varios estudios han explorado modelos más complejos como Random Forest o XGBoost, la literatura también destaca que algoritmos probabilísticos simples como *Naive Bayes* (NB) pueden ofrecer un desempeño competitivo con menor costo computacional, alta estabilidad en muestras reducidas y una interpretación más directa de las probabilidades clínicas [8]-[11].

Diversos algoritmos de aprendizaje automático han sido aplicados para la predicción del riesgo de diabetes tipo 2, reflejando un creciente interés por integrar métodos que permitan capturar patrones complejos entre variables clínicas. No obstante, dentro de este panorama metodológico, los enfoques probabilísticos como *Naive Bayes* (NB) continúan destacando por su simplicidad operativa, su eficiencia computacional y su rendimiento estable en muestras pequeñas, características particularmente relevantes para entornos de salud pública donde la disponibilidad de recursos tecnológicos y humanos puede ser limitada [12]-[16]. Estas propiedades consolidan a NB como una alternativa viable para fortalecer los procesos de tamizaje y estratificación de riesgo en atención primaria.

La calibración de probabilidades constituye un componente fundamental en la implementación de modelos predictivos en el ámbito clínico, ya que permite que las probabilidades estimadas reflejen de manera más fiel la frecuencia real del evento de interés [17]. En el caso de modelos probabilísticos como *Naive Bayes*, una calibración adecuada mejora la interpretabilidad y utilidad práctica de las predicciones, favoreciendo decisiones informadas en el seguimiento y priorización de pacientes [18]. Entre los métodos empleados para este fin, Platt scaling y la calibración isotónica han mostrado mejorar significativamente la confiabilidad de las predicciones en diversos escenarios, aportando solidez adicional a los modelos utilizados en procesos de tamizaje y estratificación de riesgo [19]-[20].

La aplicación de modelos predictivos basados en información clínica rutinaria ha mostrado resultados alentadores en entornos ambulatorios, especialmente para la identificación temprana de pacientes con riesgo de desarrollar diabetes tipo 2. Estudios realizados en atención primaria evidencian que enfoques probabilísticos como *Naive Bayes* pueden identificar individuos de riesgo utilizando variables de fácil acceso, tales como medidas antropométricas, indicadores nutricionales, análisis básicos y antecedentes clínicos [21]-[22]. Asimismo, la integración de este tipo de modelos en sistemas de historia clínica electrónica ha permitido la implementación de programas de cribado automatizados capaces de procesar grandes volúmenes de información de forma eficiente [23]. Esta capacidad resulta especialmente relevante en centros de salud metropolitanos como los de la región Puno, donde factores socioeconómicos y características poblacionales particulares influyen

en la aparición de DT2, haciendo necesario contar con herramientas de estratificación adaptadas al contexto local [24]-[25].

Este trabajo aporta: i) la implementación de un pipeline reproducible para la predicción del riesgo de diabetes tipo 2 en un contexto real de atención ambulatoria, incorporando técnicas de manejo explícito del desbalance, como SMOTE aplicado exclusivamente al conjunto de entrenamiento; ii) la integración de métodos de calibración probabilística para mejorar la confiabilidad de las predicciones generadas por *Naive Bayes*; y iii) un análisis conjunto de medidas de discriminación y calibración que permite evaluar la utilidad clínica del modelo para la estratificación de riesgo en salud pública. Estos elementos permiten determinar en qué medida un enfoque probabilístico como *Naive Bayes* puede apoyar la identificación temprana de pacientes en riesgo de DT2 y cómo la calibración contribuye a fortalecer la toma de decisiones clínicas y preventivas.

El propósito de este trabajo de investigación fue aplicar y evaluar el desempeño del clasificador *Naive Bayes* para predecir el riesgo de diabetes tipo 2 en pacientes ambulatorios del Centro de Salud Metropolitano de la región Puno, utilizando información clínica, antropométrica y nutricional recolectada durante el año 2023, e incorporando técnicas de manejo del desbalance y calibración probabilística para fortalecer la precisión y confiabilidad del modelo. Asimismo, y con el fin de contextualizar el alcance de sus predicciones dentro del panorama actual del aprendizaje automático, se incluyen referencias comparativas con enfoques ampliamente utilizados en problemas similares, como Random Forest y XGBoost, lo que permite valorar la pertinencia operativa de *Naive Bayes* en entornos de salud pública con recursos limitados.

METODOLOGÍA

La presente investigación es aplicada, cuantitativa, predictiva, no experimental y de corte transversal, en la que se busca aplicar y comparar los algoritmos *Naive Bayes, Random Forest y XGBoost* como recursos prácticos para la predicción del riesgo de diabetes tipo 2 en pacientes ambulatorios, en el ámbito de la salud pública, incorporando además calibración de probabilidades para mejorar la utilidad clínica de las salidas [4]-[11].

El presente estudio analiza una cohorte de 121 pacientes ambulatorios atendidos en 2023 en el Centro de Salud Metropolitano de Puno, con información clínica, antropométrica y nutricional incluyendo: género, edad, peso, talla, índice de masa corporal (IMC), circunferencia abdominal, frecuencia de consumo de alimentos, grado de desarrollo físico y diagnóstico de patologías [5]-[8], para abordar un problema de clasificación binaria de presencia/ausencia de DT2 (definición clínica conforme a la ADA) [26]. Dado el desbalance de clases característico de este tipo de problemas, se adoptó un pipeline que asegura la separación estricta entre entrenamiento y prueba y el manejo del desbalance mediante SMOTE aplicado exclusivamente al conjunto de entrenamiento [7], [8], [10]; adicionalmente, se incorpora calibración post-hoc de probabilidades (regresión isotónica y Platt) para mejorar su interpretabilidad clínica [8]-[11].

Metodológicamente, se comparan tres familias de algoritmos representativas y ampliamente utilizadas: *Naive Bayes* (NB), como modelo probabilístico simple, interpretable y eficiente con muestras pequeñas; *Random Forest* (RF), como ensamble basado en árboles; y *XGBoost* (XGB), como método de *gradient boosting* de alto desempeño [4], [5], [7]. La comparación contempla métricas de discriminación (AUC, exactitud, sensibilidad, especificidad, PPV) y de calibración (*Brier Score, Log-Loss*, ECE), junto con curvas ROC y curvas de confiabilidad [11].

Se formuló un problema de clasificación binaria supervisada, donde la variable dependiente refleja la existencia (D = 1) o inexistencia (D = 0) de diabetes tipo 2, de acuerdo con criterios clínicos de la American Diabetes Association (ADA) [26]. Las variables independientes empleadas para la predicción fueron categorizadas y se refieren a: sexo (masculino o femenino), edad (menos de 60 años o 60 años y más), IMC (normal, sobrepeso, obesidad), perímetro abdominal (normal, elevado, muy elevado) y frecuencia de consumo alimentario (adecuada, necesita mejora, deficiente) [5]-[8].

El propósito del modelo bayesiano fue calcular la probabilidad de que un paciente se clasifique como diabético, basándose en sus valores observados en las variables predictivas, utilizando principios establecidos en investigaciones anteriores sobre

clasificación médica [17], [18] (D = 1: si el paciente tiene diabetes tipo 2; D = 0: si el paciente no tiene diabetes).

Las variables independientes utilizadas para la predicción fueron discretizadas y corresponden: *X*1: Sexo (Femenino o Masculino), *X*2: Edad (mayor o menor de 60 años), *X*3: IMC (normal, sobrepeso, obesidad), *X*1: Perímetro abdominal (normal o elevado), *X*5: Frecuencia de consumo alimentario.

El objetivo del modelo es estimar la probabilidad de que un paciente *i* pertenezca a la clase diabético (D = 1), dados sus valores observados *X* = (*X*1, *X*2, *X*3, *X*4, *X*5).

La distribución de la presencia de diabetes tipo 2 en el conjunto de datos es SI para 22 casos y No para 99 casos, siendo un total de 121 casos y siendo el cálculo de las probabilidades la siguientes:

Probabilidades previas:

$P\ (Enfermedad = SI) = 0{,}182$
$P\ (Enfermedad = NO) = 0{,}818$

Probabilidades condicionales:

Para enfermos (SI):

$P\ (X1 \mid SI) = 0{,}4550$
$P\ (X2 \mid SI) = 0{,}1847$
$P\ (X3 \mid SI) = 0{,}4550$
$P\ (X4 \mid SI) = 0{,}0045$
$P\ (X5 \mid SI) = 0{,}5450$

Para sanos (NO):

$P\ (X1 \mid NO) = 0{,}4446$
$P\ (X2 \mid NO) = 0{,}0111$
$P\ (X3 \mid NO) = 0{,}3639$
$P\ (X4 \mid NO) = 0{,}5756$
$P\ (X5 \mid NO) = 0{,}4748$

El clasificador *Naive Bayes* asume independencia condicional entre las variables dado la clase. Por tanto, la verosimilitud P(Datos|Clase) se calcula como el producto de las probabilidades condicionales de cada característica observada, conforme la ecuación (1):

$$P\left(X \middle| Clase\right) = \prod_{i=1}^{n} P\left(X \middle| Clase\right) \qquad (1)$$

Se realizó partición estratificada en dos subconjuntos: entrenamiento (n = 96; ≈79%) y prueba (n = 25; ≈21%), preservando la proporción de clases; se fijó semilla aleatoria = 42 para reproducibilidad. Como control interno se empleó validación cruzada estratificada (k = 5) sobre el conjunto de entrenamiento [5], [27].

Dado el desbalance observado (≈18% con DT2), se aplicó SMOTE exclusivamente en el conjunto de entrenamiento (parámetro por defecto $K = 5$), evitando fuga de información hacia el conjunto de prueba [8], [9], [10].

Sea X una instancia minoritaria y X_{NN} uno de sus k-vecinos más cercanos de la misma clase. SMOTE genera una nueva instancia sintética mediante interpolación lineal entre ambos puntos, definida en la ecuación (2).

$$x_{new} = x + \lambda\left(x_{NN} - x\right), \lambda \sim U\left(0,1\right)\ldots \qquad (2)$$

Modelos comparados y configuración

- **Naive Bayes (NB):** clasificador probabilístico con suavizado de Laplace (a = 1) para categorías de baja frecuencia.
- **Random Forest (RF):** ensamble basado en árboles (configuración parsimoniosa; número de árboles moderado).
- **XGBoost (XGB):** *gradient boosting* de alto desempeño con ajuste parco para evitar sobreajuste dada la muestra reducida. La comparación de estas familias es habitual en predicción de DT2 y escenarios multi-modelo [4]-[7].

Tras entrenar cada clasificador, se aplicó calibración post-hoc comparando regresión isotónica y *Platt scaling*. La elección del calibrador se basó en *Log-Loss, Brier Score y Expected Calibration Error* (ECE) estimados por validación interna; la evaluación final se realizó en el conjunto de prueba. La literatura respalda que la calibración mejora la confiabilidad de las probabilidades, en particular sobre NB, sin sacrificar la discriminación [9]-[11], y recomienda reportar curvas de confiabilidad [11].

Sobre el conjunto de prueba se reportaron: matriz de confusión (TP, TN, FP, FN), exactitud (*Accuracy*), sensibilidad (*Recall*), especificidad, precisión positiva (PPV) y AUC-ROC; para la calidad probabilística, *Log-Loss, Brier Score y ECE*, además de curvas ROC y curvas de calibración [14].

El pipeline se implementó en *Python* (por ejemplo, *scikit-learn, imbalanced-learn, xgboost*), documentando semillas, partición estratificada, la regla *SMOTE* solo en entrenamiento y el procedimiento de calibración para asegurar reproducibilidad [7], [8], [10]-[14].

RESULTADOS

Análisis exploratorio

De la Figura 1, la multicolinealidad moderada-alta entre IMC y Perímetro es esperada y aceptable en *Naive Bayes*, ya que no requiere independencia estricta, solo condicional.

La edad de la cohorte fue 33,6 años en promedio, con mediana 27,0 años y rango 19-60 años. La relación *media* > *mediana* sugiere asimetría a la derecha (más concentración de pacientes jóvenes). Dado que el punto de corte operativo fue <60 vs. ≥60 años, la categoría ≥60 es minoritaria; por ello, la edad por sí sola podría no capturar bien el riesgo y debe complementarse con IMC y perímetro abdominal.

La prevalencia de DT2 fue 18,2% (22 de 121; NO = 99, SI = 22). Un intervalo de confianza del 95% para la prevalencia, calculado con el método de Wilson, es 12,3% - 26,0%, lo que confirma un desbalance de clases marcado. En consecuencia, la evaluación del modelo debe incluir métricas robustas al desbalance (curvas PR, PPV/VPN) y no limitarse a la exactitud global.

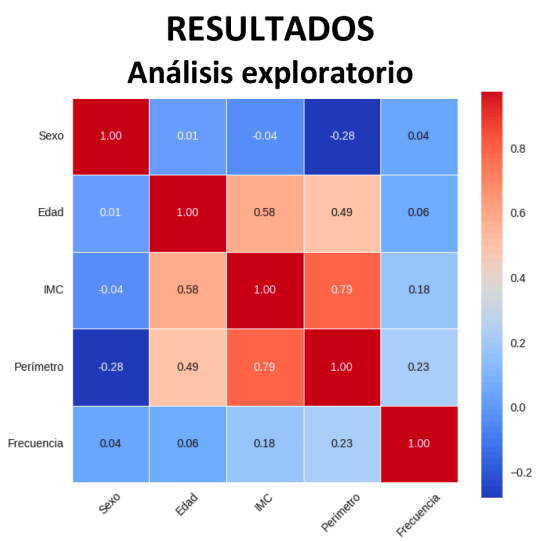

Figura 1. Matriz de correlación entre variables predictoras.

Se realizó una partición estratificada en entrenamiento (n = 96) y prueba (n = 25), preservando la proporción original de clases. En el conjunto de entrenamiento, la distribución inicial fue NO = 79 y SÍ = 17; sobre este subconjunto se aplicó SMOTE (k = 5) exclusivamente en entrenamiento hasta alcanzar una proporción 1:1 (NO = 79, SI = 79). El conjunto de prueba no fue sometido a ningún re-muestreo y conservó la prevalencia original (~18%), con el fin de obtener una evaluación realista del desempeño. Esta decisión metodológica evita fuga de información y permite reportar resultados comparables en ROC y, especialmente, en *Precision-Recall* para reflejar el desbalance subyacente.

El uso de SMOTE equilibró la clase minoritaria en el conjunto de entrenamiento, ayudando a mejorar el aprendizaje del modelo en contextos desbalanceados.

Evaluación del modelo

El modelo *Naive Bayes* aplicado a la base de datos de riesgo de personas con diabetes tipo 2 completa mostró los siguientes resultados en la matriz de confusión conforme la Figura 2.

Interpretación

- Verdaderos negativos (TN): 17
  El modelo predijo correctamente que 17 personas no tienen la enfermedad.
- Falsos positivos (FP): 3
  El modelo identifico 3 casos al clasificar como enfermas a personas sanas.

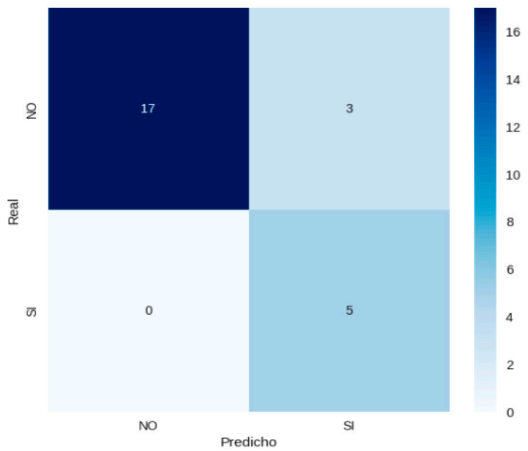

Figura 2. Matriz de confusión.

- Verdaderos positivos (TP): 5
  El modelo identificó correctamente 5 casos que sí tienen la enfermedad.
- Falsos negativos (FN): 0
  El modelo no detectó a ningún caso en el que indicara que realmente estaban enfermas.

El modelo demuestra una capacidad óptima para detectar correctamente los casos positivos (*Recall* = 100%), lo cual es crucial en contextos clínicos al evitar omitir pacientes en riesgo. No obstante, su precisión positiva es moderada (62,5%), indicando una proporción considerable de falsos positivos.

Métricas de evaluación

A partir de la matriz de confusión se calcularon las siguientes métricas de rendimiento, utilizando criterios estándar de evaluación establecidos en la literatura [28] conforme la Tabla 1 y las métricas de evaluación más relevantes mostradas en la Tabla 2.

- **Precisión (*Accuracy*):** 88,0%
- **Sensibilidad (*Recall*):** 100% (capacidad del modelo para detectar correctamente los casos con enfermedad - clase "SI")
- **Especificidad:** 85,0% (capacidad del modelo para identificar correctamente los casos sin enfermedad - clase "NO")

Tabla 1. Métricas de evaluación del modelo.

Métrica	Valor
Exactitud (Accuracy)	0,88
Precisión (Precision)	0,63
Sensibilidad (Recall)	1,00
Especificidad	0,85
F1-Score	0,77
Balanced Accuracy	0,93
Cohen's Kappa	0,69
MCC	0,73
ROC-AUC	0,93

- **Valor predictivo positivo (VPP):** 62,5% (probabilidad de que un paciente predicho como "SI" realmente tenga la enfermedad)
- **Valor predictivo negativo (VPN):** 100% (probabilidad de que un paciente predicho como "NO" realmente no tenga la enfermedad)

La Figura 3 muestra el rendimiento general del clasificador *Naive Bayes*, evidenciando valores consistentes con su aplicación en entornos clínicos. El modelo alcanza una precisión global del 88%, una sensibilidad elevada y un AUC de 0,92, lo que indica buena capacidad discriminativa. Además, métricas como *Balanced Accuracy*, *Kappa* y MCC reflejan un desempeño estable y adecuado para procesos de tamizaje en atención primaria. Estos resultados coinciden con reportes previos sobre la eficacia de *Naive Bayes* en escenarios similares [29], [30].

El modelo reveló que no hay falsos negativos, que correspondían a pacientes que se consideraban saludables pero que en realidad padecían diabetes. Se detectaron 3 falsos positivos, señalando que el

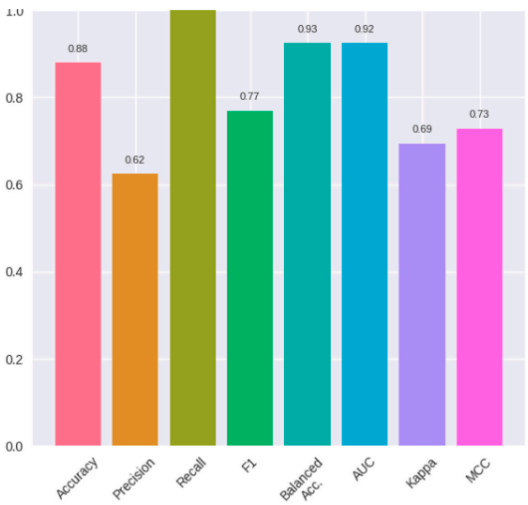

Figura 3. Grafica de métricas.

Tabla 2. Métricas de evaluación del modelo más relevantes.

Clase / Promedio	Precisión	Sensibilidad	F1-Score	Soporte
NO	1,00	0,85	0,92	20
SI	0,63	1,00	0,77	5
Accuracy	–	–	0,88	25
Promedio macro	0,81	0,93	0,84	25
Promedio ponderado	0,93	0,88	0,89	25

modelo es cauteloso en sus proyecciones positivas, una conducta parecida a la reportada por [31] y [32].

Según el modelo, los pacientes con mayor probabilidad de desarrollar diabetes tenían rasgos como edad avanzada, exceso de peso u obesidad, y un gran perímetro abdominal. En cambio, los pacientes con menor posibilidad de diabetes presentaban un Índice de masa corporal normal, edad inferior a 60 años y perímetro abdominal normal, características que se ajustan a los estudios epidemiológicos contemporáneos. [33], [34].

Análisis de errores probabilísticos

La Tabla 3, muestra que la calibración isotónica mejora consistentemente la calidad probabilística del clasificador *Naive Bayes*, obteniendo los valores más bajos de Log-Loss (0,2117) y *Brier Score* (0,0750), así como un ECE igual a 0.0000, lo que indica una alineación óptima entre probabilidades predichas y frecuencias observadas. Asimismo, la calibración isotónica produce una curva más cercana a la diagonal ideal, reflejando un ajuste probabilístico más confiable que las versiones original y basada en *Platt*.

En la Figura 4 se muestra la comparación de curvas ROC demuestra que *XGBoost* es el modelo superior con un AUC de 0,990, seguido por *Random Forest* (AUC = 0,960) y *Naive Bayes* (AUC = 0,925).

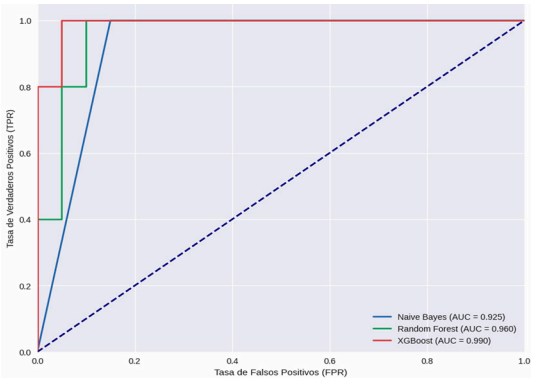

Figura 4. Comparación de modelos.

XGBoost: Curva casi rectangular, desempeño casi perfecto para distinguir pacientes diabéticos, *Random Forest*: Muy buen balance entre sensibilidad y especificidad, *Naive Bayes*: Buen desempeño discriminativo, con tendencia hacia alta sensibilidad. Todos los modelos superan significativamente el desempeño aleatorio (línea punteada), validando su utilidad clínica.

A continuación se presenta el diagnóstico para un caso específico utilizando el modelo propuesto, siendo las características de este caso, Datos: [Sexo = F, Edad = 60, IMC = Sobrepeso, Perímetro = Normal, Frecuencia = Necesita mejora].
Se tiene dos clases:

Para la clase enfermo (SI):

Calculando el productorio de probabilidades condicionales:

$$P(X1 = \text{True} \mid clase) = 0{,}454955$$
$$P(X2 = \text{True} \mid clase) = 0{,}1884685$$
$$P(X3 = \text{True} \mid clase) = 0{,}454955$$
$$P(X4 = \text{True} \mid clase) = 0{,}004505$$
$$P(X5 = \text{True} \mid clase) = 0{,}545045$$

$P(SI)$, $P(SI)$, $P(SI)$, $P(SI)$,
$P\,(PERIMETRO\_NORMAL = True \mid SI) \cdot$
$P\,(FREC\_NECESITA\_MEJORA = True \mid SI)$

La probabilidad de los mismos datos si la persona está enferma.

$$= 0{,}454955 \cdot 0{,}184685 \cdot 0{,}454955 \cdot 0{,}004505 \cdot 0.545045 = 9{,}385 \text{ X } 10^{-5}$$

$$Productorio\ final\ (verosimilitud) = 9{,}38527e^{-5}$$

Para clase sano (NO):

Calculando el productorio de probabilidades condicionales:

Tabla 3. Métricas de evaluación de errores probabilísticos.

Métrica	Original	Isotónica	Platt	Mejor resultado
Log-Loss	0,8537	0,2117	0,3718	Isotónica (↓ menor mejor)
Brier Score	0,1827	0,0750	0,1161	Isotónica (↓ menor mejor)
ECE (Expected Calibration Error)	0,3269	0,0000	0,1891	Isotónica (↓ ideal: 0.0)
Calibración visual	Lejana a la diagonal	Cercana a la diagonal	Intermedia	Isotónica

$$P(X1 = True \mid clase) = 0{,}444556$$
$$P(X2 = True \mid clase) = 0{,}011089$$
$$P(X3 = True \mid clase) = 0{,}363911$$
$$P(X4 = True \mid clase) = 0{,}575605$$
$$P(X5 = True \mid clase) = 0{,}474798$$

$$P\,(NO) = P\,(NO) \cdot P\,(NO),\ P\ (IMC\_SOBREPESO = True \mid NO) \cdot P\ (PERIMETRO\_NORMAL = True \mid NO) \mid \cdot P\,(NO)$$

La probabilidad de los datos si la persona está sana:

$$= 0{,}444556 \cdot 0{,}011089 \cdot 0{,}363911 \cdot 0{,}575605 \cdot 0{,}474798 = 4{,}903\ X\ 10^{-4}$$

Productorio final (verosimilitud): $4.902723e^{-4}$

$$P\ (SI\ datos) = 1{,}7064140287e - 05/4{,}1819581382e - 04 = 0{,}040804$$
$$P\ (NO\ datos) = 4.0113167353E - 04/4.1819581382e - 04 = 0{,}959196$$

Resultado: No tiene enfermedad

Probabilidad enfermo: 0,040804
Probabilidad sano: 0,959196

DISCUSIÓN

Los resultados obtenidos muestran que el clasificador *Naive Bayes* presenta un desempeño sólido para la predicción del riesgo de diabetes tipo 2 en el contexto ambulatorio del Centro de Salud Metropolitano de Puno. La sensibilidad del 100% evidencia su capacidad para identificar todos los casos positivos sin omisiones, una cualidad fundamental en procesos de tamizaje donde la prioridad es no pasar por alto pacientes en riesgo. No obstante, la precisión moderada sugiere la presencia de falsos positivos, lo cual indica un comportamiento conservador en la clasificación que debe interpretarse en el marco clínico correspondiente.

El valor del AUC (0,925) indica una adecuada capacidad discriminativa del modelo, reforzando su utilidad para apoyar decisiones iniciales de evaluación en atención primaria. Si bien en la literatura se reportan enfoques más sofisticados, como métodos *ensemble* basados en árboles, que suelen obtener métricas superiores, *Naive Bayes* continúa destacando por su simplicidad, interpretabilidad y bajo costo computacional, características particularmente valiosas en ámbitos de salud pública con recursos limitados. Esta perspectiva contextual permite situar sus resultados dentro del panorama actual del aprendizaje automático aplicado al diagnóstico temprano en entornos reales.

Las limitaciones del estudio, especialmente el tamaño reducido de la muestra (121 pacientes) y el desbalance natural entre clases, son desafíos comunes en conjuntos de datos clínicos regionales [35], [36]. La aplicación de técnicas como SMOTE y la calibración probabilística permitió mitigar parcialmente estos efectos, aunque investigaciones futuras podrían considerar estrategias adicionales como remuestreos más robustos y la ampliación de las cohortes para fortalecer la estabilidad y generalización del modelo [37].

CONCLUSIONES

El presente estudio demostró que *Naive Bayes* constituye una herramienta valiosa para la predicción del riesgo de diabetes tipo 2 en el contexto ambulatorio del Centro de Salud Metropolitano de Puno, alcanzando una precisión del 88% y un AUC de 0,925. La sensibilidad del 100% resalta su utilidad como método de tamizaje inicial, ya que garantiza que ningún caso positivo sea omitido durante la evaluación, un aspecto esencial en intervenciones de medicina preventiva.

Aunque se reconoce que la literatura reporta modelos más complejos con métricas potencialmente superiores, *Naive Bayes* mantiene ventajas operativas importantes, como su simplicidad, interpretabilidad y bajo costo computacional, factores clave en entornos de salud pública con recursos limitados. Su desempeño observado en este estudio respalda su pertinencia como modelo de primera línea para la identificación temprana de pacientes en riesgo.

Se recomienda ampliar futuras investigaciones mediante la incorporación de un mayor número de pacientes y variables clínicas relevantes, con el fin de robustecer la generalización del modelo. Asimismo, la integración progresiva de este tipo de herramientas en los sistemas electrónicos de salud del centro permitiría evaluar su funcionamiento en tiempo real y contribuir al fortalecimiento de estrategias de detección temprana y prevención de la diabetes tipo 2 en la región Puno.

REFERENCIAS

[1] P. Saeedi *et al.*, "Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas, 9th edition," *Diabetes Research and Clinical Practice*, vol. 157, Art. no. 107843, 2019, doi: 10.1016/j.diabres.2019.107843.

[2] World Health Organization, "Global report on diabetes," World Health Organization, Geneva, Switzerland, 2016. [Online]. Available: https://www.who.int/publications/i/item/9789241565257

[3] R.L. Japura Espinal, "Factores de riesgo asociados a desarrollar diabetes tipo 2 en usuarios ambulatorios del Centro de Salud Metropolitano - Puno, 2023", Tesis de Licenciatura, Universidad Nacional del Altiplano, Puno, Perú, 2024.

[4] Y. Huang, J. Li, F. Macheret, R.A. Gabriel, and L. Ohno-Machado, "A tutorial on calibration measurements and calibration models for clinical prediction models," *Journal of the American Medical Informatics Association*, vol. 27, no. 4, pp. 621-633, 2020, doi: 10.1093/jamia/ocz228.

[5] W. Chen, B. Sahiner, F. Samuelson, A. Pezeshk, and N. Petrick, "Calibration of medical diagnostic classifier scores to the probability of disease," *Statistical Methods in Medical Research*, vol. 27, no. 5, pp. 1394-1409, 2018, doi: 10.1177/0962280216661371.

[6] F.M. Ojeda *et al.*, "Calibrating machine learning approaches for probability estimation: A comprehensive comparison," *Statistics in Medicine*, vol. 42, no. 29, pp. 5451-5478, 2023, doi: 10.1002/sim.9921.

[7] A. Niculescu-Mizil and R. Caruana, "Predicting good probabilities with supervised learning," in *Proceedings of the 22nd International Conference on Machine Learning (ICML '05)*, Bonn, Germany, 2005, pp. 625-632, doi: 10.1145/1102351.1102430.

[8] H. Zhou, S. Rahman, M. Angelova, C.R. Bruce, and C. Karmakar, "A robust and generalized framework in diabetes classification across heterogeneous environments," *Computers in Biology and Medicine*, vol. 186, Art. no. 109720, 2025, doi: 10.1016/j.compbiomed.2025.109720.

[9] H. Shao, X. Liu, D. Zong, and Q. Song, "Optimization of diabetes prediction methods based on combinatorial balancing algorithm," *Nutrition & Diabetes*, vol. 14, Art. no. 63, 2024, doi: 10.1038/s41387-024-00324-z.

[10] M. Alghamdi, M. Al-Mallah, S. Keteyian, C. Brawner, J. Ehrman, and S. Sakr, "Predicting diabetes mellitus using SMOTE and ensemble machine learning approach: The Henry Ford exercise testing (FIT) project," *PLoS One*, vol. 12, no. 7, Art. no. e0179805, 2017, doi: 10.1371/journal.pone.0179805.

[11] M. Matboli *et al.*, "Machine learning-based stratification of prediabetes and type 2 diabetes progression," *Diabetology & Metabolic Syndrome*, vol. 17, Art. no. 227, 2025, doi: 10.1186/s13098-025-01786-6.

[12] B. Alsadi, S. Musleh, H. Al-Absi, M. Refaee, R. Qureshi, N. El Hajj, and T. Alam, "An ensemble-based machine learning model for predicting type 2 diabetes and its effect on bone health," *BMC Medical Informatics and Decision Making*, vol. 24, Art. no. 144, 2024, doi: 10.1186/s12911-024-02540-0.

[13] S. Uddin *et al.*, "Intelligent type 2 diabetes risk prediction from administrative claim data," *Informatics for Health and Social Care,* vol. 47, no. 3, pp. 243-257, 2022, doi: 10.1080/17538157.2021.1988957.

[14] M.A. Talukder *et al.*, "Toward reliable diabetes prediction: Innovations in data engineering and machine learning applications," *Digital Health*, vol. 10, Art. no. 20552076241271867, 2024, doi: 10.1177/20552076241271867.

[15] M. Lugner, A. Rawshani, E. Helleryd, and B. Eliasson "Identifying top ten predictors of type 2 diabetes through machine learning analysis of UK Biobank data," *Scientific Reports*, vol. 14, Art. no. 2102, 2024, doi: 10.1038/s41598-024-52023-5.

[16] S. Wang *et al.*, "Comparative study on risk prediction model of type 2 diabetes based on machine learning theory: A cross-sectional study," *BMJ Open*, vol. 13, no. 8, Art. no. e069018, 2023, doi: 10.1136/bmjopen-2022-069018.

[17] A. Leiherer *et al.*, "Machine learning approach to metabolomic data predicts type 2 diabetes mellitus incidence," *International Journal of Molecular Sciences*, vol. 25, no. 10, Art. no. 5331, 2024, doi: 10.3390/ijms25105331.

[18] M.A. Mizani *et al.*, "Identifying subtypes of type 2 diabetes mellitus with machine learning: Development, internal validation, prognostic validation and medication burden in linked electronic health records in 420 448 individuals," *BMJ Open Diabetes Research & Care*, vol. 12, no. 3, Art. no. e004191, 2024, doi: 10.1136/bmjdrc-2024-004191.

[19] B. Coles, K. Khunti, S. Booth, F. Zaccardi, M.J. Davies, and L.J. Gray, "Prediction of type 2 diabetes risk in people with non-diabetic hyperglycaemia: Model derivation and validation using UK primary care data," *BMJ Open*, vol. 10, no. 10, Art. no. e037937, 2020, doi: 10.1136/bmjopen-2020-037937.

[20] A.M. Tuppad and S.D. Patil, "An efficient type 2 diabetes classification framework incorporating feature interactions," *Expert Systems with Applications*, vol. 239, Art. no. 122138, 2024, doi: 10.1016/j.eswa.2023.122138.

[21] M.T. Moghaddam *et al.*, "Predicting diabetes in adults: Identifying important features in unbalanced data over a 5-year cohort study using machine learning algorithm," *BMC Medical Research Methodology*, vol. 24, Art. no. 220, 2024, doi: 10.1186/s12874-024-02341-z.

[22] S. Ghiasi Hafezi *et al.*, "Prediction of the 10-year incidence of type 2 diabetes mellitus based on advanced anthropometric indices using machine learning methods in the Iranian population," *Diabetes Research and Clinical Practice*, vol. 214, Art. no. 111755, 2024, doi: 10.1016/j.diabres.2024.111755.

[23] A. Site, J. Nurmi, and E.S. Lohan, "Machine-learning-based diabetes prediction using multisensor data," *IEEE Sensors Journal*, vol. 23, no. 22, pp. 28370-28377, 2023, doi: 10.1109/JSEN.2023.3319360.

[24] X. Luo *et al.*, "Establishment and health management application of a prediction model for high-risk complication combination of type 2 diabetes mellitus based on data mining," *PLoS One*, vol. 18, no. 8, Art. no. e0289749, 2023, doi: 10.1371/journal.pone.0289749.

[25] H.M. Deberneh *et al.*, "1233-P: Prediction of type 2 diabetes occurrence using machine learning model," *Diabetes*, vol. 69, Supplement 1, Art. no. 1233-P, 2020, doi: 10.2337/db20-1233-P.

[26] Y.-T. Huang, A. Steptoe, L. Wei, and P. Zaninotto, "Polypharmacy difference between older people with and without diabetes: Evidence from the English longitudinal study of ageing," *Diabetes Research and Clinical Practice*, vol. 176, Art. no. 108842, 2021, doi: 10.1016/j.diabres.2021.108842.

[27] K.S. Praveenkumar and R. Gunasundari, "Optimizing type ii diabetes prediction through hybrid Big Data analytics and H-SMOTE tree methodology," *International Journal of Computational and Experimental Science and Engineering*, vol. 11, no. 1, 2025, doi: 10.22399/ijcesen.727.

[28] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Proceedings of the 14th International Joint Conference on Artificial Intelligence,* Montreal, QC, Canada, 1995, pp. 1137-1143.

[29] S. Perveen, M. Shahbaz, A. Guergachi, and K. Keshavjee, "Performance analysis of data mining classification techniques to predict diabetes," *Procedia Computer Science*, vol. 82, pp. 115-121, 2016, doi: 10.1016/j.procs.2016.04.016.

[30] O. Iparraguirre-Villanueva, K. Espinola-Linares, R. O. Flores Castañeda, and M. Cabanillas-Carbonell, "Application of machine learning models for early detection and accurate classification of type 2 diabetes," *Diagnostics*, vol. 13, no. 14, Art. no. 2383, 2023, doi: 10.3390/diagnostics13142383.

[31] S. Bhatia, P. Prakash, and G.N. Pillai, "SVM based decision support system for heart disease classification with integer-coded genetic algorithm to select critical features," in *Proceedings of the World Congress on Engineering and Computer Science*, San Francisco, USA, 2008, pp. 34-38.

[32] M.A. Naji, S. El Filali, K. Aarika, E.H. Benlahmar, R.A. Abdelouhahid, and O. Debauche, "Machine learning algorithms for breast cancer prediction and diagnosis," *Procedia Computer Science*, vol. 191, pp. 487-492, 2021, doi: 10.1016/j.procs.2021.07.062.

[33] A. Dinh, S. Miertschin, A. Young, and S. D. Mohanty, "A data-driven approach to predicting diabetes and cardiovascular disease with machine learning," *BMC Medical Informatics and Decision Making*,

vol. 19, Art. no. 211, 2019, doi: 10.1186/s12911-019-0918-5.

[34] K.M. Tsiouris, V.C. Pezoulas, M. Zervakis, S. Konitsiotis, D.D. Koutsouris, and D.I. Fotiadis, "A long short-term memory deep learning network for the prediction of epileptic seizures using EEG signals," *Computers in Biology and Medicine*, vol. 99, pp. 24-37, 2018, doi: 10.1016/j.compbiomed.2018.05.019.

[35] A. Dutta, M.K. Hasan, M. Ahmad, M.A. Awal, M.A. Islam, M. Masud, and H. Meshref, "Early prediction of diabetes using an ensemble of machine learning models," *International Journal of Environmental Research and Public Health*, vol. 19, no. 19, Art. no. 12378, 2022, doi: 10.3390/ijerph191912378.

[36] W. Zhang, M. Xu, Y. Feng, Z. Mao, and Z. Yan "The effect of procrastination on physical exercise among college students-the chain effect of exercise commitment and action control," *International Journal of Mental Health Promotion*, vol. 26, no. 8, pp. 611-622, 2024, doi: 10.32604/ijmhp.2024.052730.

[37] I. Abousaber, H.F. Abdallah, and H. El-Ghaish, "Robust predictive framework for diabetes classification using optimized machine learning on imbalanced datasets," *Frontiers in Artificial Intelligence*, vol. 7, 2025, doi: 10.3389/frai.2024.1499530.