UNIVERSIDAD PERUANA UNIÓN

ESCUELA DE POSGRADO

Unidad de Posgrado de la Facultad de Educación

Patrones en la Educación Intercultural Bilingüe en instituciones educativas peruanas utilizando machine learning

Trabajo de investigación para obtener el Grado Académico de Maestra en Educación con mención en Investigación y Docencia Universitaria

Autor:

Jenny Raquel Tarrillo Vásquez
Lina Naida Mamani Mamani

Asesor:

Dr. Juan Jesús Soria Quijaite

Lima,Ñaña, 25 de junio 2026

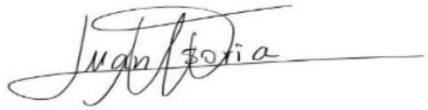
DECLARACIÓN JURADA DE ORIGINALIDAD DE TRABAJO DE INVESTIGACIÓN

Yo Juan Jesús Soria Quijaite, docente de la Unidad de Posgrado de la Facultad de Educación, Escuela de Posgrado de la Universidad Peruana Unión.

DECLARO:

Que la presente investigación titulada: **"PATRONES EN LA EDUCACION INTERCULTURAL BILINGÜE EN INSTITUCIONES EDUCATIVAS PERUANAS UTILIZANDO MACHINE LEARNING"** de los autores Jenny Raquel Tarrillo Vásquez y Lina Naida Mamani Mamani tiene un índice de similitud de 11 % verificable en el informe del programa Turnitin, y fue realizada en la Universidad Peruana Unión bajo mi dirección.

En tal sentido asumo la responsabilidad que corresponde ante cualquier falsedad u omisión de los documentos como de la información aportada, firmo la presente declaración en la ciudad de Ñaña, a los 25 días del mes de junio del año 2026



Juan Jesús Soria Quijaite

Asesor

ACTA DE SUSTENTACIÓN DE TRABAJO DE INVESTIGACIÓN

En Lima, Ñaña, Villa Unión, a los 25 días del mes de junio del año 2026, siendo las 17:00 horas se reunieron en la sala virtual https://teams.microsoft.com/meet/244100390791104?p=KtXvVGAdVn4BFufTNr bajo la dirección del señor presidente del Jurado: Mg. Josué Arturo Morán Condezo y los demás miembros siguientes:

Secretaria : Mtra. Karoline Elizabeth Chávez Vallejos
Asesor : Dr. Juan Jesús Soria Quijaite
Vocal : Dr. Esteban Tocto Cano
Vocal : Mtro. Carlos Daniel Abanto Ramírez

Con el propósito de llevar a cabo el acto público de la sustentación del trabajo de investigación de posgrado titulado: "Identificación de patrones en la Educación Intercultural Bilingüe en instituciones educativas peruanas mediante técnicas de aprendizaje automático", de las estudiantes Jenny Raquel Tarrillo Vasquez y Lina Naida Mamani Mamani, conducente a la obtención del Grado Académico de Maestro en Educación con mención en Investigación y Docencia Universitaria.

El presidente del Jurado dio por iniciado el acto académico, invitando a las candidatas a hacer uso del tiempo señalado para su exposición (20'). Concluida la misma, el presidente del Jurado invitó a los demás miembros a realizar las preguntas, cuestionamientos y aclaraciones pertinentes que fueron absueltas por las candidatas, el acto fue seguido de un receso de quince minutos para las deliberaciones y el dictamen de Jurado. Vencido el tiempo de las deliberaciones, el Jurado procedió a dejar constancia escrita del resultado en la presente acta, con dictamen siguiente:

APROBADAS por UNANIMIDAD calificación: APROBADO CON ESCALA VIGESIMAL DE 18 ESCALA CUALITATIVA CON NOMINACIÓN DE MUY BUENO, CON MÉRITO SOBRESALIENTE

El presidente del Jurado hizo alusión a las Maestrandas y solicitó a la secretaria la lectura correspondiente para poner en su conocimiento el resultado, terminado el mismo y sin objeción alguna, el presidente del jurado dio por concluido el acto, en fe de lo cual firman al pie.

Presidente Secretaria

Candidato

Vocal Vocal

Identificación de patrones en la Educación Intercultural Bilingüe en instituciones educativas peruanas mediante técnicas de aprendizaje automático

Jenny Tarrillo Vásquez[1]*, Lina Mamani[1] y Juan J. Soria[1]

[1] Escuela de Posgrado, Universidad Peruana Unión, Lima, Perú
* Correspondencia: jraqueltv@gmail.com

Resumen

La Educación Bilingüe Intercultural (EBI) en Perú busca brindar instrucción cultural y lingüísticamente relevante para estudiantes indígenas; sin embargo, su implementación sigue enfrentando desafíos estructurales, lingüísticos y de recursos.
Este estudio identificó patrones latentes en la implementación de la EBI mediante técnicas de aprendizaje automático y comparó el rendimiento predictivo de diferentes algoritmos de clasificación. Métodos: Se analizó un conjunto de datos que comprende 84,558 registros institucionales y 17 variables. Tras el preprocesamiento de datos, que incluyó imputación de valores faltantes, codificación categórica y reducción de dimensionalidad, se desarrollaron modelos de clasificación multiclase para predecir cinco escenarios de implementación de la EBI. Se evaluaron ocho algoritmos de machine learning bajo diferentes configuraciones de entrenamiento y prueba (80/20, 50/50, 25/75 y 10/90). El rendimiento del modelo se evaluó mediante precisión, exhaustividad, puntuación F1 ponderada y validación cruzada de cinco pliegues. Resultados: Los métodos basados en conjuntos superaron consistentemente a los clasificadores lineales y basados en distancia. El Gradient Boosting obtuvo el mejor rendimiento (precisión ≈ 0,77; puntuación F1 ponderada = 0,74), seguido de Random Forest y KNearest Neighbors. La validación cruzada confirmó la robustez del modelo, con precisiones medias que oscilaron entre 0,766 y 0,774 y bajas desviaciones estándar (< 0,007).
Los hallazgos demuestran que los modelos de machine learning pueden capturar eficazmente relaciones complejas y no lineales en sistemas educativos caracterizados por la diversidad lingüística e institucional. El marco propuesto ofrece un enfoque escalable y reproducible para el desarrollo de políticas basadas en evidencia, apoyando estrategias basadas en datos para mejorar la equidad y la calidad en la educación bilingüe intercultural.

1. Introducción

La Educación Intercultural Bilingüe (EIB) en el Perú tiene como objetivo garantizar una enseñanza cultural y lingüísticamente pertinente para los estudiantes indígenas, respetando su identidad y promoviendo una educación de calidad. En este sentido, durante las últimas décadas, el país ha logrado avances significativos a nivel normativo e institucional (Oliveros y Candia, 2022; Serrano y Quispe, 2023). Según reportes

recientes, existen más de 26,422 instituciones educativas en las que coexisten más de 40 lenguas indígenas junto con el castellano (Ministerio de Educación, 2024). Esta cifra evidencia que el Estado peruano ha realizado esfuerzos sustanciales para ampliar la educación bilingüe, transformándola de una iniciativa experimental en una política pública prioritaria durante la última década (Fornara, 2015).

No obstante, pese a estos avances institucionales, persisten múltiples desafíos relacionados con la implementación efectiva de esta modalidad educativa en las instituciones del país. Algunos de estos desafíos están asociados con la planificación, el acceso y la deserción estudiantil. Dada la diversidad étnica y lingüística del país, aún existe una necesidad urgente de adaptar y planificar las prácticas pedagógicas considerando los contextos culturales específicos de los estudiantes (Bedriñana y Gutiérrez, 2023).

En este contexto, se evidencia una falta de conocimiento integral por parte de los docentes respecto a las lenguas indígenas habladas por sus estudiantes, lo cual genera dificultades en la implementación de los planes curriculares, particularmente en zonas rurales (Liñán et al., 2023). Esta limitación se vincula con una débil articulación entre las políticas nacionales y las necesidades específicas de las comunidades locales. **Un** estudio realizado en instituciones interculturales del altiplano peruano encontró que la mayoría de los docentes considera que los recursos educativos en lenguas indígenas son insuficientes, lo que dificulta una enseñanza culturalmente pertinente y el fortalecimiento de la identidad lingüística de los estudiantes (Lara et al., 2024).

Asimismo, se observa que la EIB en el Perú enfrenta desafíos derivados de una implementación inadecuada de políticas pluralistas y democráticas, lo que perpetúa desigualdades entre las lenguas indígenas y debilita la identidad cultural de los estudiantes (Torres-Acurio y Turpo-Chaparro, 2024). Estas deficiencias parecen afectar negativamente el rendimiento académico dentro del sistema de educación intercultural bilingüe. Otro estudio señala que los docentes enfrentan dificultades relacionadas con la baja autoestima de los estudiantes, la desorganización y problemas en lectura, escritura y comprensión (Santos et al., 2025).

A nivel internacional, los avances en el análisis de grandes volúmenes de datos educativos han permitido identificar patrones y tendencias en la gestión pedagógica y el aprendizaje mediante técnicas de aprendizaje automático y minería de datos educativos (Romero y Ventura, 2020; Shoukath y Chakkaravarthy, 2025). Sin embargo, su aplicación en contextos de EIB sigue siendo limitada, particularmente en países latinoamericanos (Mager et al., 2023).

En el Perú, la literatura científica reciente se ha centrado principalmente en aspectos cualitativos y descriptivos de la EIB (Bardales et al., 2025; Loaiza et al., 2025; Vargas, 2021), lo que evidencia una brecha en estudios que integren análisis computacionales y enfoques predictivos.

Por lo tanto, el objetivo del presente estudio es identificar y analizar patrones en la educación intercultural bilingüe en instituciones peruanas mediante la aplicación de técnicas de aprendizaje automático. De este modo, la investigación busca superar las limitaciones tradicionales en la evaluación y diagnóstico de la EIB, aprovechando tecnologías emergentes para identificar patrones latentes en los datos educativos y proporcionar evidencia empírica que respalde la toma de decisiones.

2. Materiales y Métodos

Los conjuntos de datos educativos presentan desafíos particulares, entre ellos la presencia de datos faltantes, variables categóricas de alta cardinalidad y dependencias contextuales complejas. La predicción de escenarios institucionales, como el "Escenario 2019", puede contribuir a la planificación de políticas públicas y a una asignación más eficiente de recursos. Los métodos de aprendizaje automático, particularmente los algoritmos de tipo *ensemble*, han demostrado capacidad para manejar problemas de clasificación multiclase complejos y generalizar adecuadamente en conjuntos de datos heterogéneos. El presente estudio aplica un pipeline integral de aprendizaje automático al conjunto de datos **Registro_1.xlsx**, con el objetivo de predecir escenarios educativos, enfatizando las etapas de preprocesamiento, ingeniería de características, entrenamiento de modelos y validación cruzada.

2.1. Descripción del conjunto de datos

El conjunto de datos estuvo conformado inicialmente por 84,558 filas y 17 columnas, que contenían información sobre identificadores institucionales, características administrativas, descriptores geográficos y modalidades educativas. La variable objetivo, denominada "Escenario 2019", presentaba cinco categorías únicas, lo que configuró un problema de clasificación multiclase. El archivo ocupaba aproximadamente 75.49 MB.

2.2. Preprocesamiento de datos

Los datos faltantes y las variables de alta cardinalidad fueron tratados de manera sistemática con el fin de mejorar la robustez de los modelos. Las columnas con más del 50% de valores faltantes fueron eliminadas, reduciendo el conjunto de datos a 27,979 filas y 15 columnas. Los valores categóricos faltantes restantes fueron imputados con la categoría "Unknown", mientras que los valores numéricos faltantes fueron reemplazados por la mediana de la variable correspondiente. La variable objetivo fue convertida a formato numérico y se eliminaron los registros con valores inválidos. Las variables categóricas de alta cardinalidad, como "Nombre de la DRE o UGEL que supervisa la I.E." y "Centro poblado", fueron excluidas para evitar problemas de dimensionalidad y sobreajuste.

2.3. Ingeniería de características

Se generaron nuevas variables con el objetivo de mejorar la capacidad predictiva del modelo: **DRE_Dept** (concatenación del nombre de la DRE y el nombre del departamento), **Level_EIB** (combinación del nivel/modalidad con la forma de atención EIB), y **Populated_Center_Length** (longitud en caracteres del nombre del centro poblado). Las variables categóricas fueron codificadas mediante *label encoding*, agrupando las categorías poco frecuentes en la clase "Other". Las variables numéricas, como "Código modular" y "Anexo", fueron convertidas a tipo entero, asignando el valor cero a los datos faltantes. La matriz final de características estuvo compuesta por seis variables: cinco categóricas y una numérica.

2.4. Entrenamiento y evaluación de modelos

Se evaluaron ocho modelos de aprendizaje automático: Árbol de Decisión, Bosque Aleatorio, Regresión Logística, Máquina de Soporte Vectorial (SVM) con kernel lineal, SVM con kernel RBF, K-Nearest Neighbors (k = 5), Perceptrón Multicapa (MLP) con capas ocultas de 100 y 50 neuronas, Potenciación de Gradiente (50 estimadores, profundidad máxima de 5 y tasa de aprendizaje de 0.1). Los modelos fueron entrenados utilizando cuatro particiones entrenamiento-prueba: 80/20, 50/50, 25/75 y 10/90. Adicionalmente, se realizó validación cruzada de cinco pliegues (*five-fold cross-validation*) para los modelos Árbol de Decisión, Bosque de Decisión, MLP, Potenciación de Gradiente, XGBoost y LightGBM, calculando la exactitud media (*mean accuracy*) y el puntaje F1 ponderado junto con sus respectivas desviaciones estándar.

2.4.1. Árbol de Decisión

Un Árbol de Decisión particiona recursivamente el espacio de características en regiones $R * m$ seleccionando variables y umbrales que maximizan una métrica de reducción de impureza, como el Índice de Gini o la Entropía (Hussain et al., 2024). Para clasificación, la clase predicha en la región $R * m$ es:

$$\hat{y} = \arg\max * c \ \in 1, \dots, Kp * m, c, \quad p_{m,c} = \frac{1}{N_m} \sum_{i \in R*m} \mathbf{1}(y_i = c), \qquad (1)$$

donde N_m es el número de muestras en el nodo m y $p * m, c$ es la proporción de la clase c en dicho nodo.

2.4.2. Bosque Aleatorio

El Bosque Aleatorio construye Bárboles de decisión utilizando muestras *bootstrap* y subconjuntos aleatorios de características (Iftikhar, Khan, et al., 2024). La predicción final se obtiene mediante votación mayoritaria:

$$\hat{y} = \text{mode}\hat{y}^{(b)}b = 1^B. \qquad (2)$$

2.4.3. Regresión Logística

Para clasificación multiclase con Kclases, la Regresión Logística utiliza la función *softmax*:

$$P(y = c \mid x) = \frac{\exp(\beta_c^T x)}{\sum_{k=1}^{K} \exp(\beta_k^T x)}, \qquad (3)$$

donde el vector de parámetros $\beta_c$se estima mediante máxima verosimilitud:

$$L(\beta) = \prod_{i=1}^{N} P(y_i \mid x_i). \qquad (4)$$

2.4.4. Máquina de Soporte Vectorial (SVM)

Kernel lineal:

Encuentra el hiperplano $w^T x + b = 0$ que maximiza el margen resolviendo:

$$\min_{w,b} \frac{1}{2} \| w \|^2 + C \sum_{i=1}^{N} \xi_i, \qquad \text{s.a. } y_i(w^T x_i + b) \geq 1 - \xi_i, \xi_i \geq 0 \qquad (5)$$

Kernel RBF:

Mapea la entrada usando $\phi(x)$ con kernel $K(x_i, x_j) = \exp(-\gamma |x_i - x_j|^2)$ y resuelve el problema dual:

$$\max_{\alpha} \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{N} \alpha_i \alpha_j y_i y_j K(x_i, x_j), \qquad 0 \leq \alpha_i \leq C. \qquad (6)$$

2.4.5. K-Nearest Neighbors (KNN)

La predicción para una nueva muestra x se basa en la votación mayoritaria entre sus k vecinos más cercanos $N * k(x)$ (Iftikhar, Zywiołek, et al., 2024):

$$\hat{y} = \arg\max * c \in 1, \dots, K \sum_{i \in N_k(x)} \mathbf{1}(y_i = c), \qquad k = 5. \qquad (7)$$

2.4.6. Perceptrón Multicapa (MLP)

Un MLP con L capas transforma la entrada x en la salida $\hat{y}$ mediante activaciones no lineales:

$$h^{(l)} = f\left(W^{(l)} h^{(l-1)} + b^{(l)}\right), \qquad l = 1, \dots, L \qquad (8)$$

$$\hat{y} = \text{softmax}\left(W^{(L)} h^{(L-1)} + b^{(L)}\right), \qquad (9)$$

donde $f(\cdot)$ es una función de activación (ReLU). La red utilizada incluyó 100 y 50 neuronas en las capas ocultas y empleó *early stopping* (Qureshi et al., 2025).

2.4.7. Potenciación de Gradiente

La Potenciación de Gradiente ajusta secuencialmente modelos débiles $h_m(x)$ para minimizar una función de pérdida diferenciable L:

$$F_m(x) = F_{m-1}(x) + \eta h_m(x) \qquad (10)$$

donde η es la tasa de aprendizaje (0.1) y h_m es típicamente un árbol poco profundo (profundidad máxima 5) (Iftikhar et al., 2025). Cada árbol se entrena sobre el gradiente negativo de la pérdida:

$$r_{im} = -\left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}\right] * F = F_{m-1} \qquad (11)$$

La predicción final es $\hat{y} = \arg\max\ F_M(x)$ después de $M = 50$ iteraciones.

2.5. Métricas de Evaluación

Para evaluar de manera integral el desempeño de los modelos propuestos, se emplearon múltiples métricas: Exactitud (*Accuracy*), Precisión (*Precision*), Exhaustividad (*Recall*) y Puntaje F1 (Gonzales et al., 2025). Estas métricas se derivan de la matriz de confusión y proporcionan perspectivas complementarias sobre la efectividad del modelo, especialmente en presencia de desbalance de clases. Sea TP, TN, FP y FN el número de verdaderos positivos, verdaderos negativos, falsos positivos y falsos negativos, respectivamente.

2.5.1. Exactitud (Accuracy)

La exactitud representa la proporción total de instancias correctamente clasificadas respecto al número total de muestras. Proporciona una medida general del desempeño del modelo; sin embargo, puede resultar engañosa en conjuntos de datos desbalanceados.

$$\text{Exactitud} = \frac{TP + TN}{TP + TN + FP + FN} \qquad (12)$$

Un valor más alto de exactitud indica un mejor desempeño global en la clasificación.

2.5.2. Precisión (Precision)

La precisión mide la confiabilidad de las predicciones positivas al cuantificar la proporción de instancias positivas correctamente predichas respecto al total de instancias clasificadas como positivas.

$$\text{Precisión} = \frac{TP}{TP + FP} \qquad (13)$$

Una alta precisión indica una baja tasa de falsos positivos, lo cual es particularmente importante en aplicaciones donde las falsas alarmas son costosas.

2.5.3. Exhaustividad (Recall)

La exhaustividad, también conocida como sensibilidad o tasa de verdaderos positivos, mide la capacidad del modelo para identificar correctamente todas las instancias positivas relevantes.

$$\text{Recall} = \frac{TP}{TP + FN} \qquad (14)$$

Un valor alto de exhaustividad indica que el modelo minimiza eficazmente los falsos negativos.

2.5.4. Puntaje F1

El Puntaje F1 es la media armónica entre la Precisión y la Exhaustividad, proporcionando una métrica equilibrada que considera tanto los falsos positivos como los falsos negativos.

$$\text{Puntaje F1} = 2 \times \frac{\text{Precisión} \times \text{Recall}}{\text{Precisión} + \text{Recall}} \qquad (15)$$

El Puntaje F1 es particularmente adecuado para evaluar modelos de clasificación en conjuntos de datos desbalanceados, ya que equilibra el compromiso entre precisión y exhaustividad.

2.5.5. Métricas de Validación Cruzada

Además de la evaluación mediante partición *hold-out*, se empleó validación cruzada de cinco pliegues para evaluar la robustez y la capacidad de generalización del modelo. La media y la desviación estándar de la Exactitud y el Puntaje F1 a través de los pliegues se calcularon de la siguiente manera:

$$\bar{M} = \frac{1}{K}\sum_{i=1}^{K} M_i \qquad (16)$$

$$\sigma = \sqrt{\frac{1}{K}\sum_{i=1}^{K} (M_i - \bar{M})^2} \qquad (17)$$

donde M_i representa la métrica de evaluación obtenida en el pliegue i y K es el número de pliegues. Estas estadísticas proporcionan información sobre la estabilidad y consistencia del desempeño del modelo a través de diferentes particiones de los datos. Por lo tanto, la combinación de estas métricas garantiza una evaluación integral y confiable de los modelos propuestos.

3. Resultados

La Tabla 1 resume el desempeño de los modelos para la partición 80/20. La Potenciación de Gradiente alcanzó la mayor exactitud (0.7706) y el mayor Puntaje F1 (0.7415). KNN, MLP y Árbol de Decisión también mostraron un desempeño competitivo.

Tabla 1. Desempeño de los modelos en la partición entrenamiento-prueba 80/20.

Modelo	Exactitud	Puntaje F1
Árbol de Decisión	0.7680	0.7413
Bosque Aleatorio	0.7645	0.7274
Regresión Logística	0.7266	0.6787
SVM Lineal	0.7207	0.6260
SVM RBF	0.7450	0.7029
KNN	0.7695	0.7328
MLP	0.7686	0.7359
Potenciación de Gradiente	0.7706	0.7415

El desempeño de los modelos se mantuvo consistente en todas las particiones (50/50, 25/75 y 10/90), con valores de exactitud entre 0.72 y 0.78, y Puntaje F1 entre 0.62 y 0.75. Notablemente, incluso utilizando solo el 10% de los datos para entrenamiento, el MLP y los modelos *ensemble* mantuvieron un alto rendimiento predictivo, demostrando una sólida capacidad de generalización.

La validación cruzada de cinco pliegues confirmó la robustez de los modelos *ensemble*. La Potenciación de Gradiente obtuvo la mayor exactitud media (0.7738), mientras que LightGBM presentó el mayor Puntaje F1 medio (0.7513). Las desviaciones estándar para ambas métricas fueron consistentemente bajas (< 0.007), lo que indica una generalización estable a través de los distintos pliegues (Tabla 2).

Modelo	Media Exactitud	Std (Exactitud)	Media Puntaje F1	Std (Puntaje F1)
Árbol de Decisión	0.7664	0.0063	0.7446	0.0052
Bosque Aleatorio	0.7723	0.0049	0.743	0.0052

MLP	0.7699	0.0023	0.7375	0.0024
Potenciación de Gradiente	0.7738	0.0055	0.7457	0.0062
XGBoost	0.773	0.0036	0.7436	0.0037
LightGBM	0.7711	0.0047	0.7513	0.0049

La comparación del desempeño de los modelos de aprendizaje automático evaluados en la partición entrenamiento-prueba 80/20 se presenta en las Figuras 1, 2 y 3. En términos generales, los resultados indican que los métodos basados en *ensemble* superan a los modelos individuales tanto en exactitud como en Puntaje F1.

La Potenciación de Gradiente alcanza la mayor exactitud de clasificación (0.7706), seguido de cerca por K-Nearest Neighbors (0.7695), el Perceptrón Multicapa (0.7686) y el Árbol de Decisión (0.7680). Esta variación marginal en la exactitud entre los modelos con mejor desempeño sugiere que múltiples clasificadores pueden aprender de manera efectiva la distribución subyacente de los datos. En contraste, la Regresión Logística y las Máquinas de Soporte Vectorial con kernel lineal presentan menores niveles de exactitud, lo que evidencia su limitada capacidad para capturar relaciones no lineales complejas en los datos.

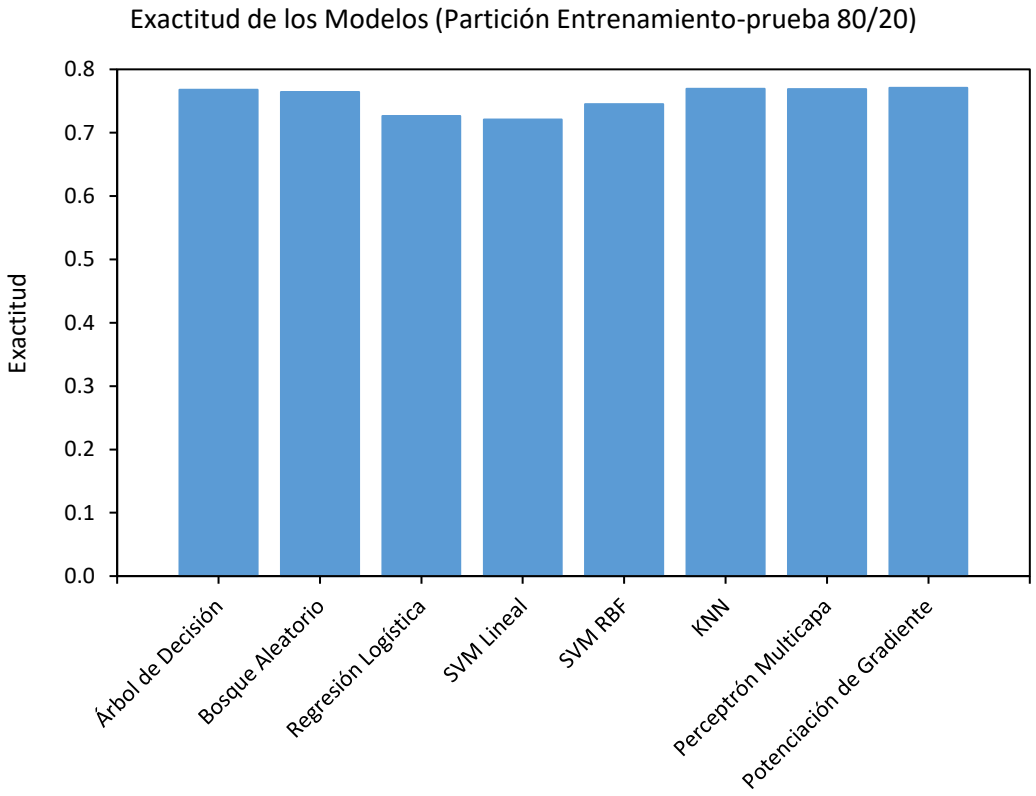


Figura 1. Comparación de la exactitud de los modelos en la partición entrenamiento-prueba 80/20.

La comparación del puntaje F1 refuerza aún más la robustez de los métodos *ensemble* y de los modelos basados en árboles. La Potenciación de Gradiente (0.7415) y el Árbol de Decisión (0.7413) registran los valores más altos de puntaje F1, lo que indica un equilibrio adecuado entre precisión y exhaustividad. El Perceptrón Multicapa también muestra un desempeño sólido con un puntaje F1 de 0.7359, confirmando la efectividad de las redes neuronales para esta tarea de clasificación. Notablemente, la SVM con kernel lineal reporta el puntaje F1 más bajo (0.6260), lo que sugiere dificultades para manejar el desbalance de clases o fronteras de decisión no lineales. Estos hallazgos implican que los modelos capaces de capturar interacciones de orden superior logran un desempeño predictivo superior.

La Figura 4 presenta los resultados del análisis de validación cruzada de cinco pliegues, proporcionando información sobre la capacidad de generalización y la estabilidad de los modelos. La potenciación de Gradiente alcanza la mayor exactitud media (0.7738), seguido de cerca por XGBoost (0.7730) y el Bosque Aleatorio (0.7723), lo que demuestra la ventaja consistente de las técnicas *ensemble* basadas en boosting.

Las desviaciones estándar relativamente bajas reportadas en la Tabla 2 indican un desempeño estable a través de los diferentes pliegues, reforzando la confiabilidad de estos modelos.

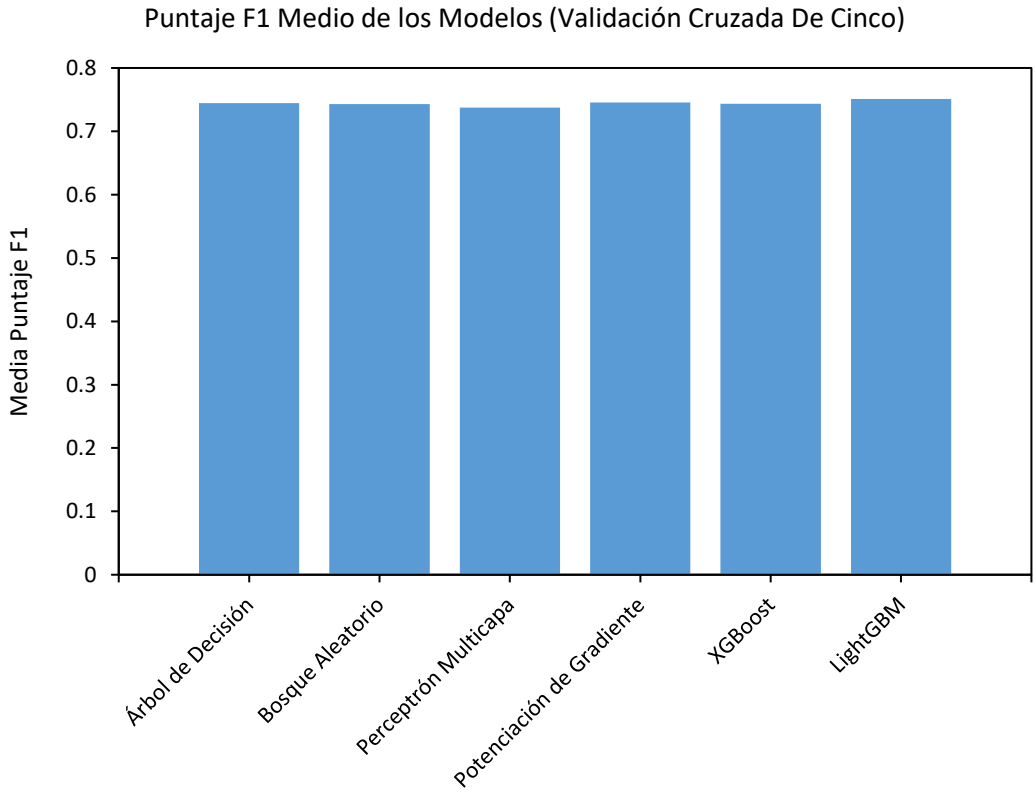


Figura 2. Puntaje F1 medio de los modelos seleccionados mediante validación cruzada de cinco pliegues.

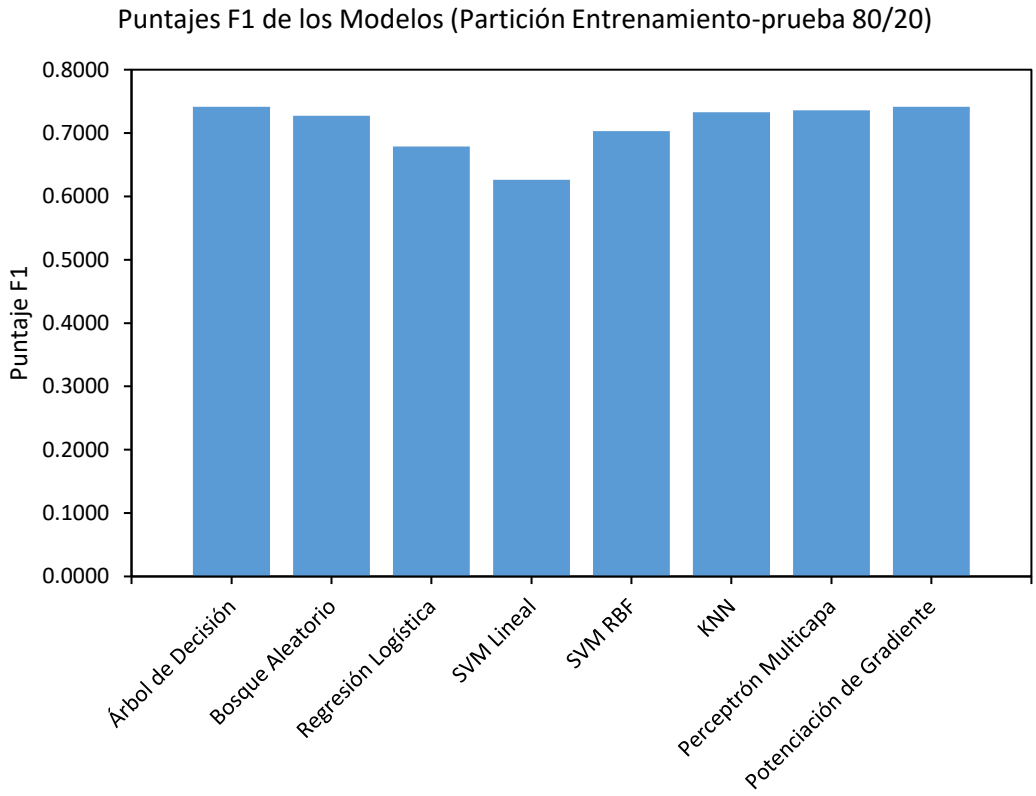


Figura 3. Comparación de los Puntajes F1 de los modelos en la partición entrenamiento-prueba 80/20.

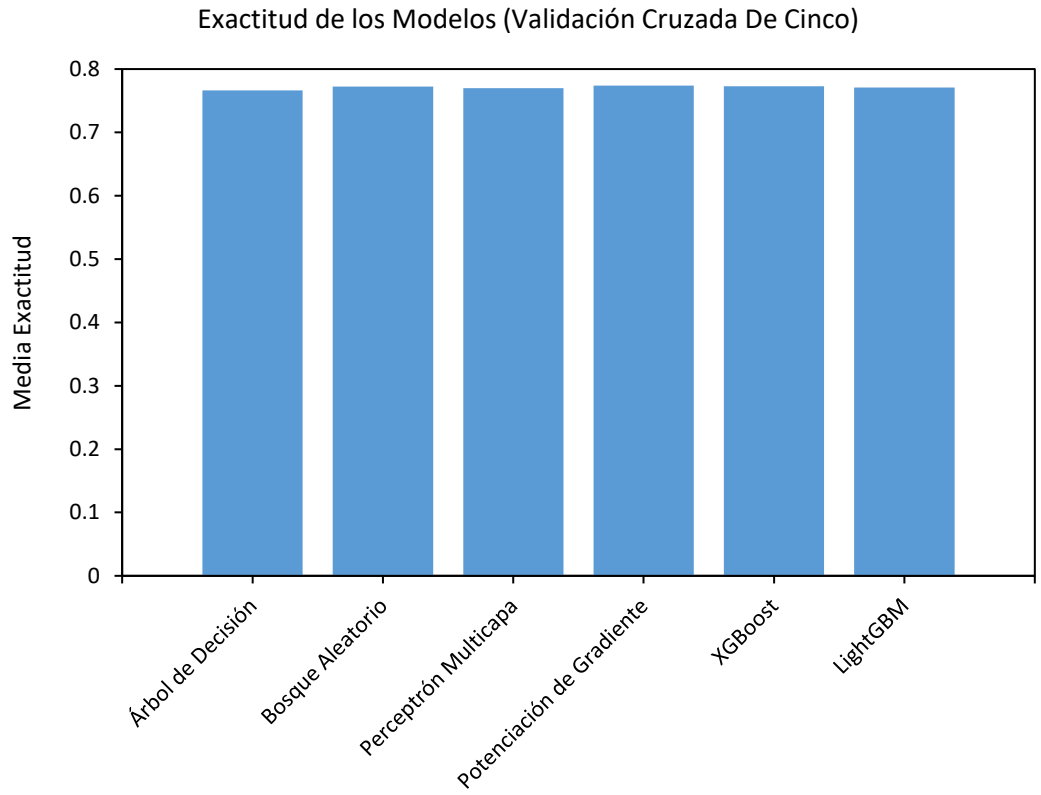


Figura 4. Exactitud media de los modelos seleccionados mediante validación cruzada de cinco pliegues.

En términos del puntaje F1 medio, LightGBM supera a todos los demás modelos con un valor de 0.7513, lo que resalta su capacidad superior para equilibrar precisión y exhaustividad a través de los distintos pliegues. La Potenciación de la Gradiente (0.7457) y el Árbol de Decisión (0.7446) también muestran un desempeño competitivo, lo que confirma la efectividad del aprendizaje basado en árboles en este contexto. La consistencia entre los resultados de la validación cruzada y los obtenidos en el conjunto de prueba (*hold-out*) sugiere que los modelos seleccionados generalizan adecuadamente y no presentan sobreajuste. En conjunto, los métodos *ensemble*, particularmente la Potenciación de Gradiente y LightGBM, emergen como los candidatos más adecuados para su implementación, debido a su sólido desempeño predictivo y su robustez.

4. Discusión

Los resultados demuestran que un *pipeline* de aprendizaje automático adecuadamente diseñado puede clasificar de manera efectiva escenarios educativos, incluso en presencia de desafíos metodológicos significativos como datos faltantes y variables categóricas de alta cardinalidad. En este contexto, la ingeniería de características

desempeñó un papel fundamental, particularmente mediante la construcción de variables administrativas y contextuales combinadas. Estas características diseñadas mejoraron la capacidad de generalización de los modelos y permitieron una representación más precisa de la complejidad estructural subyacente al fenómeno analizado.

Desde una perspectiva comparativa, los modelos basados en *ensemble* —especialmente la Potenciación de Gradiente (Exactitud = 0.7706; Puntaje F1 = 0.7415)— junto con Bosque Aleatorio, XGBoost y LightGBM, superaron consistentemente a los clasificadores lineales en todas las particiones entrenamiento-prueba y esquemas de validación cruzada. Este patrón indica que los escenarios educativos analizados ("Escenario 2019") no están determinados por relaciones estrictamente lineales, sino por interacciones complejas entre variables institucionales, administrativas y de modalidad educativa. En consecuencia, el proceso generador de datos parece presentar estructuras no lineales que son capturadas de manera más efectiva por modelos capaces de aprender interacciones de orden superior y efectos conjuntos entre predictores. El hecho de que la Potenciación de Gradiente haya alcanzado la mayor exactitud media en validación cruzada, mientras que LightGBM obtuvo el puntaje F1 más alto, respalda adicionalmente la idoneidad de los métodos de aprendizaje *ensemble* como candidatos robustos para su implementación en entornos reales.

Desde una perspectiva aplicada, estos hallazgos tienen implicancias directas para la gestión de la Educación Intercultural Bilingüe (EIB) en el Perú. Las estrategias de planificación institucional y asignación de recursos no deberían basarse únicamente en la contribución aislada de predictores individuales, sino considerar configuraciones multivariadas específicas —como combinaciones de departamento, nivel o modalidad educativa y tipo de centro poblado— que influyen conjuntamente en la clasificación de los escenarios educativos. Asimismo, la estabilidad del desempeño predictivo a través de múltiples esquemas de partición de datos (80/20, 50/50, 25/75 y 10/90), incluyendo escenarios con reducciones sustanciales del conjunto de entrenamiento, junto con desviaciones estándar consistentemente bajas en la validación cruzada (< 0.007), proporciona evidencia sólida de la robustez del conjunto de características diseñadas y de la capacidad de generalización interna de los modelos seleccionados.

En conjunto, el marco metodológico propuesto constituye un enfoque escalable y transferible para el modelamiento predictivo en analítica educativa, y puede extenderse a aplicaciones relacionadas, tales como la identificación de riesgos institucionales o la evaluación prospectiva del impacto de políticas públicas.

5. Limitaciones y Líneas Futuras de Investigación

A pesar del rigor metodológico y del tamaño relativamente grande de la muestra (84,558 registros), deben reconocerse varias limitaciones. En primer lugar, los resultados

dependen de la calidad y completitud de los registros administrativos utilizados, lo que puede introducir sesgos asociados con inconsistencias en el reporte o con información contextual faltante. Aunque se implementaron estrategias sistemáticas de preprocesamiento e imputación, no puede descartarse completamente el impacto potencial de las limitaciones en la calidad de los datos.

En segundo lugar, las variables categóricas de alta cardinalidad fueron excluidas para prevenir problemas de dimensionalidad y sobreajuste. Si bien esta decisión estuvo metodológicamente justificada, pudo haber reducido la riqueza contextual del espacio de características y limitado la profundidad explicativa de los modelos. Investigaciones futuras podrían explorar estrategias alternativas de codificación, como *target encoding* o representaciones basadas en *embeddings*, con el fin de conservar información relevante de variables de alta cardinalidad sin comprometer la estabilidad del modelo.

Asimismo, el análisis se restringió a un único corte temporal (2019), lo que limita la posibilidad de realizar inferencias longitudinales sobre la evolución de los escenarios educativos. En consecuencia, la validez externa de los hallazgos en distintos años o contextos regionales requiere validación independiente. Estudios futuros deberían incorporar datos longitudinales para examinar dinámicas temporales y evaluar la estabilidad de los patrones predictivos a lo largo del tiempo.

Finalmente, aunque los modelos demostraron una sólida capacidad de generalización interna, la validación externa mediante conjuntos de datos independientes fortalecería aún más la robustez del marco propuesto. La ampliación del conjunto de variables para incluir indicadores socioeconómicos o a nivel comunitario, así como la integración de enfoques cualitativos, podría proporcionar una comprensión más integral de los mecanismos estructurales que subyacen a la clasificación de escenarios educativos y mejorar la relevancia de la analítica predictiva para el diseño de políticas públicas en el contexto de la Educación Intercultural Bilingüe.

6. Conclusiones

Los resultados demuestran la superioridad consistente de los modelos basados en *ensemble*, particularmente la Potenciación de Gradiente, que alcanzó los valores más altos de exactitud y puntaje F1 tanto en la partición 80/20 como en la validación cruzada de cinco pliegues. Este desempeño confirma la efectividad de los enfoques *ensemble* para modelar relaciones complejas, heterogéneas y no lineales presentes en datos educativos —características especialmente relevantes en el contexto de la Educación Intercultural Bilingüe en el Perú, donde convergen múltiples variables lingüísticas, culturales, territoriales e institucionales.

Los modelos no lineales, como la Red Neuronal Multicapa, K-Nearest Neighbors y la Máquina de Soporte Vectorial con kernel RBF, también mostraron un desempeño competitivo, evidenciando su capacidad para identificar patrones asociados a distintos

escenarios de educación intercultural, incluso en contextos con información limitada. En contraste, los modelos lineales, como la Regresión Logística y la SVM lineal, obtuvieron resultados inferiores, lo que sugiere limitaciones para capturar la complejidad estructural inherente a los sistemas educativos interculturales.

La estabilidad observada en los resultados de validación cruzada, reflejada en valores bajos de desviación estándar, respalda la robustez y capacidad de generalización de los modelos evaluados, fortaleciendo su potencial uso como herramientas de apoyo a la toma de decisiones educativas basadas en evidencia. En este sentido, la aplicación de modelos predictivos en la Educación Intercultural Bilingüe puede contribuir directamente al Objetivo de Desarrollo Sostenible 4, al respaldar el diseño de políticas orientadas a mejorar la calidad, equidad y pertinencia de la educación, y al Objetivo de Desarrollo Sostenible 10, al facilitar la identificación temprana de brechas educativas asociadas a contextos de vulnerabilidad, diversidad lingüística y desigualdad territorial.

Asimismo, el uso del puntaje F1 ponderado como métrica principal resulta particularmente apropiado en este marco, ya que permite una evaluación más equitativa del desempeño predictivo en presencia de posibles desbalances entre escenarios educativos, alineándose con los principios de inclusión y reducción de desigualdades promovidos por la Agenda 2030.

Más allá de los hallazgos empíricos, este estudio aporta una contribución metodológica al demostrar el valor de integrar preprocesamiento sistemático, ingeniería de características dirigida y aprendizaje *ensemble* dentro de un marco analítico unificado y reproducible. Los resultados resaltan que el desempeño predictivo en sistemas educativos complejos depende no solo de la selección del algoritmo, sino también del diseño cuidadoso del espacio de características y de la estrategia de validación. En consecuencia, el *pipeline* propuesto constituye un modelo escalable y transferible para la analítica educativa en otros contextos multilingües y socioeconómicamente diversos. Desde una perspectiva más amplia de gobernanza basada en datos, estos hallazgos refuerzan el papel del aprendizaje automático responsable y contextualizado como instrumento estratégico para avanzar hacia políticas educativas orientadas a la equidad.

Financiamiento: Esta investigación no recibió financiamiento externo.

Declaración del Comité de Ética (Institutional Review Board): No aplica.

Consentimiento Informado: No aplica.

Conflictos de Interés: Los autores declaran no tener conflictos de interés.

Referencias

Bardales, S. E. O., Acurio, J. T., Apaza, E. V. B., & Cachicatari, G. M. (2025). Gestión de la participación comunitaria y su impacto en la educación intercultural bilingüe en el altiplano peruano. *Revista Espacios*, 46, 38–51. Disponible en: https://ve.scielo.org/scielo.php?script=sci_arttext&pid=S0798-10152025000200038 (consultado el). https://doi.org/10.48082/ESPACIOS-A25V46N02P04.

Bedriñana, K. A., & Gutiérrez, G. H. (2023, 6). La protección de la cultura y el derecho a la educación intercultural bilingüe en Pichari (VRAEM) de Perú. *Revista Electrónica Iberoamericana*, 17, 120–154. https://doi.org/10.20318/reib.2023.7804.

Ministerio de Educación. (2024, 4). Educación intercultural bilingüe (EIB) - proceso de caracterización de instituciones educativas interculturales - orientación - Ministerio de Educación - Plataforma del Estado Peruano. Disponible en: https://www.gob.pe/41377-educacion-intercultural-bilingue-eib-proceso-de-caracterizacion-de-instituciones-educativas-interculturales (consultado el).

Fornara, M. L. (2015, 10). Educación Intercultural Bilingüe: nuevos avances y desafíos | UNICEF. UNICEF. Disponible en: https://www.unicef.org/peru/historias/educacion-intercultural-bilingue-nuevos-avances-y-desafios (consultado el).

Gonzales, W., Cordero, Z., Abanto-Ramírez, C. D., Torres Armas, E. A., Iftikhar, H., & Lopez-Gonzales, J. L. (2025). A hybrid AI approach for predicting academic performance in RBE students. *Frontiers in Artificial Intelligence*, 8, 1651100.

Hussain, I., Qureshi, M., Ismail, M., Iftikhar, H., Zywiołek, J., & López-Gonzales, J. L. (2024). Optimal features selection in the high dimensional data based on robust technique: Application to different health database. *Heliyon*, 10(17).

Iftikhar, H., Hashem, A. F., Qureshi, M., & Rodrigues, P. C. (2025). Clinical application of machine learning models for early-stage chronic kidney disease detection. *Diagnostics*, 15(20), 2610.

Iftikhar, H., Khan, N., Raza, M. A., Abbas, G., Khan, M., Aoudia, M., Touti, E., & Emara, A. (2024). Electricity theft detection in smart grid using machine learning. *Frontiers in Energy Research*, 12, 1383090.

Iftikhar, H., Zywiołek, J., López-Gonzales, J. L., & Albalawi, O. (2024). Electricity consumption forecasting using a novel homogeneous and heterogeneous ensemble learning. *Frontiers in Energy Research*, 12, 1442502.

Lara, T. E. B., Quispe, G. D. M., Apaza, E. V. B., Bardales, S. E. O., & Acurio, J. T. (2024, 11). Material educativo en lenguas originarias para la Educación Intercultural Bilingüe

en el altiplano puneño, Perú. *Espacios*, 45, 97–105. https://doi.org/10.48082/espacios-a24v45n06p08.

Liñán, H. F., Sano, J. G., Paucar, C. J. M., & Tafur, L. V. (2023, 3). Intercultural Bilingual Education (EIB) in Peru. *Research in Subject-matter Teaching and Learning (RISTAL)*, 6, 84–95. https://doi.org/10.2478/ristal-2023-0006.

Loaiza, G. A., Brasileiro, T. S. A., & Cueva, M. A. L. (2025, 1). Políticas de educación intercultural bilingüe en el Perú. *Amazônica*, 18, 1028–1047. Disponible en: //periodicos.ufam.edu.br/index.php/amazonica/article/view/17547 (consultado el).

Mager, M., Bhatnagar, R., Neubig, G., Vu, N. T., & Kann, K. (2023). Neural Machine Translation for the Indigenous Languages of the Americas: An Introduction. En *Proceedings of the Workshop on Natural Language Processing for Indigenous Languages of the Americas (AmericasNLP)* (pp. 109–133). Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.americasnlp-1.13.

Oliveros, N. V., & Candia, E. S. (2022, 12). Dificultades y avances de la EIB en la zona andina del Cusco en cuanto al reconocimiento de los derechos lingüísticos de sus hablantes. *Boletín de la Academia Peruana de la Lengua*. https://doi.org/10.46744/bapl.202202.013.

Qureshi, M., Ishaq, K., Daniyal, M., Iftikhar, H., Rehman, M. Z., & Salar, S. A. (2025). Forecasting cardiovascular disease mortality using artificial neural networks in Sindh, Pakistan. *BMC Public Health*, 25(1), 34.

Romero, C., & Ventura, S. (2020, 5). Educational data mining and learning analytics: An updated survey. *WIREs Data Mining and Knowledge Discovery*, 10. https://doi.org/10.1002/widm.1355.

Santos, R. C., Marín, Y. R., Bardales, E. S., Caro, O. C., & Rituay, A. M. C. (2025, 6). The Analysis of Intercultural Bilingual Education in Indigenous Communities: A Case Study in Regular Basic Education. *Journal of Intercultural Communication*, 218–238. https://doi.org/10.36923/jicc.v25i2.1138.

Serrano, U. M., & Quispe, L. E. T. (2023, 3). La participación de los pueblos indígenas en las políticas públicas de las escuelas rurales en el Perú y los desafíos de la Educación Intercultural Bilingüe en el contexto actual. *Revista Iberoamericana de Educación*, 91, 135–144. https://doi.org/10.35362/rie9115411.

Shoukath, T. K., & Chakkaravarthy, M. (2025, 3). Predictive analytics in education: machine learning approaches and performance metrics for student success – a systematic literature review. *Data and Metadata*, 4, 730. https://doi.org/10.56294/dm2025730.

Torres-Acurio, J., & Turpo-Chaparro, J. E. (2024). Gestión educativa intercultural bilingüe en instituciones rurales peruanas en tiempos de Covid-19. *Praxis Educativa*, 19, 1–19. https://doi.org/10.5212/PraxEduc.v.19.22824.024.

Vargas, D. R. (2021, 7). Origen y perspectivas de las políticas de la Educación Intercultural Bilingüe en el Perú: utopía hacia una EIB de calidad. *Revista Histórica de la Educación Latinoamericana*, 23, 205–225. https://doi.org/10.19053/01227238.10831.