Sage Sphere in Artificial Intelligence

ISSN: 3103-1307

Vol. 4 Núm. 1 (2026)

SSAI-Vol.4. N1. 018

Factores familiares y culturales asociados al consumo de alcohol en adolescentes de educación secundaria mediante modelos de aprendizaje automático

Family and cultural factors associated with alcohol consumption in secondary school adolescents through machine learning models

Autores:
Wilder Mel Mamani Chaca
Universidad Pedagógica, Sucre, Bolivia
mamanimels3@gmail.com
https://orcid.org/0009-0003-4897-9932

Autor de correspondencia: *Israel Calani Huerta,* chi230191@gmail.com

Enviado: 26-04-2026 **Revisado:** 14-05-2026
Aceptado: 10-11-2026 **Publicado:** 18-11-2026

Cómo citar este artículo:
Mamani Chaca, W. M. (2026). Family and cultural factors associated with alcohol consumption in secondary school adolescents through machine learning models. *Sage Sphere in Artificial Intelligence*, *4*(1), 1-18. https://doi.org/10.63688/3wfr8616

© 2026; Los autores. Este es un artículo en acceso abierto, distribuido bajo los términos de una licencia Creative Commons (https://creativecommons.org/licenses/by/4.0) que permite el uso, distribución y reproducción en cualquier medio, siempre que la obra original sea correctamente citada.

ISSN: 3103-1307

Vol. 4 Núm. 1 (2026)

RESUMEN

El consumo de alcohol entre adolescentes se relaciona con condiciones de convivencia familiar, experiencias comunitarias y normas culturales que requieren un análisis contextualizado. El objetivo fue estructurar un enfoque de clasificación mediante aprendizaje automático para estudiar factores familiares y culturales asociados al consumo de alcohol en estudiantes de educación secundaria, tomando como referencia el contexto de la Unidad Educativa Máximo Novillo I, en Santa Cruz, Bolivia. Se integró la caracterización cualitativa descrita para las gestiones 2024–2025 con un ejercicio computacional reproducible sobre 100 registros sintéticos. Se operacionalizaron diez variables y una respuesta binaria de consumo reciente, y se compararon regresión logística, bosque aleatorio y potenciación de gradiente mediante validación cruzada estratificada de cinco particiones. Los resultados cualitativos permiten organizar factores de comunicación y supervisión familiar, modelos adultos, exposición en espacios productivos, normas comunitarias e influencia de pares. La comparación computacional muestra cómo interpretar discriminación, equilibrio de clasificación y calibración en un entorno controlado, sin atribuir sus cifras a participantes reales. Se concluye que el aprendizaje automático puede complementar la lectura sociocultural, siempre que la validación empírica, la protección de datos y la interpretación de las prácticas culturales acompañen la aplicación de los modelos.

Palabras clave: consumo de alcohol; adolescentes; aprendizaje automático; factores familiares; factores culturales.

ABSTRACT

Adolescent alcohol consumption is associated with family relationships, community experiences, and cultural norms that require context-sensitive analysis. This study aimed to structure a machine learning classification approach to examine family and cultural factors associated with alcohol consumption among secondary-school students, using the context of Máximo Novillo I Educational Unit in Santa Cruz, Bolivia, as a reference. The qualitative account of school activities during 2024–2025 was integrated with a reproducible computational exercise based on 100 synthetic observations. Ten explanatory variables and a binary recent-drinking outcome were operationalized. Logistic regression, random forest, and gradient boosting were compared using stratified five-fold cross-validation. The qualitative synthesis identified family communication and supervision, adult role models, exposure during productive activities, community norms, and peer influences as relevant analytical dimensions. The computational comparison illustrates the interpretation of discrimination, balanced classification, and calibration in a controlled setting; its numerical estimates are not outcomes observed among actual participants. Machine learning may complement a culturally grounded interpretation when empirical validation, data protection, and careful consideration of the meanings attached to community practices accompany model development and use.

Keywords: alcohol consumption; adolescents; machine learning; family factors; cultural factors.

Doi: https://doi.org/10.63688/3wfr8616

ISSN: 3103-1307
Vol. 4 Núm. 1 (2026)

1. INTRODUCCIÓN

El consumo de alcohol durante la adolescencia constituye un problema de salud pública que atraviesa las relaciones familiares, la vida escolar y las normas de convivencia. Sus manifestaciones no dependen de una única variable: la disponibilidad de bebidas, las expectativas de aceptación, los modelos adultos y las experiencias emocionales operan conjuntamente. La Organización Mundial de la Salud (2024) ha insistido en la magnitud sanitaria de los daños asociados al consumo de alcohol y en la necesidad de adoptar enfoques preventivos que reconozcan las características de cada grupo poblacional. En este escenario, explicar el consumo adolescente exige examinar tanto la conducta individual como las oportunidades, restricciones y significados que configuran el entorno.

En el contexto latinoamericano, las desigualdades territoriales, los cambios en la estructura productiva y la movilidad de los hogares introducen variaciones en las experiencias escolares. Para Bolivia, Luwedde y Peltzer (2023) examinaron datos de una encuesta nacional de salud escolar de 2018 y mostraron que el consumo de alcohol de los adolescentes se vincula con factores conductuales y sociales. Este resultado ofrece una referencia poblacional, aunque no sustituye el conocimiento situado de cada unidad educativa. Una escuela de una comunidad agrícola y plurilingüe puede presentar itinerarios de socialización distintos de los observados en grandes ciudades, por lo que resulta relevante integrar los datos estructurados con información cualitativa del territorio.

El funcionamiento familiar constituye un eje central en la literatura. Ryan et al. (2010) sintetizaron investigaciones longitudinales y encontraron que diferentes conductas parentales pueden asociarse con menor consumo de alcohol en adolescentes. En una revisión posterior, Yap et al. (2017) identificaron factores parentales modificables vinculados con el consumo problemático, lo que respalda el estudio de la supervisión, la comunicación y las normas del hogar. Por su parte, Pizarro et al. (2017) documentaron la relevancia de las condiciones familiares en un vecindario peruano de alto riesgo. Estas relaciones deben interpretarse como asociaciones: no es correcto atribuir de forma automática el consumo de un adolescente a una configuración familiar específica.

Las amistades y las normas compartidas también forman parte del proceso. Ivaniushina y Titkova (2021) encontraron en un metaanálisis de modelos de redes sociales que los

Doi: https://doi.org/10.63688/3wfr8616

ISSN: 3103-1307
Vol. 4 Núm. 1 (2026)

mecanismos de influencia y selección de pares intervienen en el consumo adolescente. Eisenberg et al. (2014) distinguieron la importancia de las normas percibidas para explicar conductas relacionadas con sustancias; Nesi et al. (2017) mostraron que la exposición a publicaciones sobre alcohol en redes sociales se relaciona con trayectorias posteriores de consumo y con creencias sobre lo que hacen los compañeros. Estas evidencias justifican recoger información sobre pares y contextos digitales sin asumir que el acceso a internet, por sí solo, sea una causa suficiente.

La dimensión cultural requiere una aproximación diferenciada. MacArthur et al. (2020) mostraron, desde una perspectiva cualitativa, que los significados sociales del consumo se construyen en escenarios de interacción y pertenencia. En comunidades bolivianas pueden coexistir prácticas de reciprocidad, actividades productivas y ceremonias que emplean alcohol en momentos específicos, junto con procesos de normalización del consumo fuera de esos espacios. La hoja de coca, por su parte, posee significados tradicionales propios y no debe equipararse sin fundamento con bebidas alcohólicas ni con derivados ilícitos. El análisis escolar debe distinguir prácticas culturales, observación de modelos adultos y consumo efectivo en menores de edad.

A estas perspectivas se agrega el aprendizaje automático, entendido como un conjunto de métodos capaces de reconocer regularidades en matrices de variables y producir clasificaciones probabilísticas. Breiman (2001) desarrolló el fundamento de los bosques aleatorios, que combinan múltiples árboles; Chen y Guestrin (2016) presentaron técnicas eficientes de potenciación de gradiente para relaciones complejas. La regresión logística continúa siendo una referencia interpretable frente a estos métodos. Tales procedimientos pueden identificar interacciones y patrones no lineales, pero no convierten las asociaciones en relaciones causales ni reemplazan la interpretación sustantiva de los factores culturales.

Las aplicaciones al consumo adolescente ya cuentan con antecedentes empíricos. Squeglia et al. (2017) utilizaron técnicas predictivas con información multimodal para estudiar el inicio del consumo; Afzali et al. (2019) contrastaron varios algoritmos y su transportabilidad entre estudios y culturas. Más recientemente, Lee (2025) evaluó modelos de aprendizaje automático y aprendizaje profundo para el consumo de alcohol entre adolescentes coreanos, mientras que Kim et al. (2025) examinaron predicción de uso de sustancias en cohortes

Doi: https://doi.org/10.63688/3wfr8616

internacionales. En conjunto, la literatura sugiere que el rendimiento y la relevancia de cada predictor dependen del contexto y de la definición precisa del desenlace.

El uso de modelos en investigación educativa debe acompañarse de precauciones metodológicas. Pedregosa et al. (2011) proporcionaron herramientas abiertas para desarrollar y evaluar algoritmos; Varma y Simon (2006) explicaron el sesgo que se produce cuando la selección de modelos se confunde con la evaluación de su desempeño. Collins et al. (2024) propusieron la guía TRIPOD+AI para informar investigaciones predictivas de manera transparente, y Moons et al. (2025) actualizaron la evaluación de calidad y sesgo mediante PROBAST+AI. Estos aportes aconsejan especificar variables y procedimientos, evitar filtraciones de información entre entrenamiento y prueba y reconocer los límites de muestras pequeñas.

La Unidad Educativa Máximo Novillo I, situada en la comunidad Integración del Norte, departamento de Santa Cruz, constituye un escenario de interés por su diversidad sociolingüística y por la participación de familias en agricultura, ganadería, transporte y construcción. El manuscrito institucional describe situaciones de convivencia entre generaciones, migración laboral, actividades comunitarias y presencia de adolescentes en espacios productivos. Tales elementos permiten construir una matriz de factores potencialmente asociados al consumo de alcohol. En lugar de reducir este fenómeno a una etiqueta de riesgo, se propone estudiar la interacción entre condiciones de protección, experiencias escolares y circunstancias culturales que contribuyen a su comprensión.

El objetivo del trabajo fue organizar los factores familiares y culturales descritos en el contexto escolar y demostrar un procedimiento de comparación de modelos de aprendizaje automático para la clasificación del consumo reciente de alcohol. De forma específica, se buscó identificar dimensiones relevantes para su medición, operacionalizar una matriz de variables, comparar algoritmos interpretables y no lineales en un ejercicio reproducible y establecer criterios de uso responsable de modelos predictivos en población adolescente.

2. METODOLOGÍA

Enfoque y diseño de investigación

Se integraron dos componentes metodológicos diferenciados. El primero consistió en una sistematización descriptiva e interpretativa del estudio escolar presentado para las gestiones

ISSN: 3103-1307
Vol. 4 Núm. 1 (2026)

2024–2025, centrado en factores familiares y culturales. El segundo fue una demostración computacional reproducible mediante un conjunto de observaciones sintéticas que reproduce la estructura de una matriz de encuestas, sin incorporar respuestas reales de estudiantes. Esta distinción permite analizar el diseño de clasificación y sus métricas sin confundir una prueba algorítmica con evidencia epidemiológica procedente de la institución.

El contexto de referencia fue la Unidad Educativa Máximo Novillo I, de la comunidad Integración del Norte, Santa Cruz, Bolivia. El manuscrito institucional señala un universo cercano a 597 estudiantes de secundaria comunitaria productiva y hace referencia a aproximadamente 100 estudiantes, además de 25 docentes y autoridades, en las actividades de investigación. La caracterización original incluye observación etnográfica, encuestas, entrevistas e historias orales. Se mantuvo el ámbito temporal de las gestiones 2024–2025 y se tomaron las descripciones del trabajo escolar como material de contextualización cualitativa.

Unidades de análisis y definición del desenlace

La unidad de observación del ejercicio computacional fue un registro estudiantil anonimizado de estructura equivalente a una fila de encuesta. Se trabajó con 100 filas sintéticas y diez predictores codificados, una cantidad elegida para ilustrar la escala muestral referida en el manuscrito. La variable objetivo se definió como consumo de alcohol en los últimos 30 días (0 = no; 1 = sí), un desenlace binario que debe diferenciarse del consumo alguna vez en la vida, la edad de inicio y el consumo excesivo episódico. Se evitó introducir como predictores variables equivalentes al resultado, porque ello produciría una clasificación artificialmente favorable.

Los predictores incluidos fueron comunicación familiar, supervisión de adultos responsables, normalización del consumo por modelos adultos, exposición a celebraciones con alcohol, participación laboral temprana, presión de pares, facilidad percibida de acceso a alcohol, estrés percibido, vinculación escolar y edad. Las nueve primeras dimensiones se representaron mediante escalas ordinales de uno a cinco para la simulación; la edad se generó como número entero entre 12 y 17 años. La estructura permite estudiar variables protectoras y de exposición sin codificar las prácticas culturales por pertenencia étnica o lengua, pues esos atributos no deben emplearse como sustitutos de riesgos individuales.

Tabla 1.

Doi: https://doi.org/10.63688/3wfr8616

ISSN: 3103-1307
Vol. 4 Núm. 1 (2026)

Operacionalización de variables familiares, culturales y escolares

Dimensión	Variable del modelo	Codificación utilizada
Familia	Comunicación y supervisión	Dos escalas ordinales (1–5)
Modelos adultos	Consumo adulto normalizado	Escala ordinal (1–5)
Contexto cultural	Exposición a celebraciones con alcohol	Escala ordinal (1–5)
Economía familiar	Participación en actividades laborales	Escala ordinal (1–5)
Pares y disponibilidad	Presión de pares y acceso a alcohol	Dos escalas ordinales (1–5)
Bienestar	Estrés percibido y vinculación escolar	Dos escalas ordinales (1–5)
Sociodemografía	Edad	12–17 años
Desenlace	Consumo de alcohol en últimos 30 días	Binaria: 0 = no; 1 = sí

Nota. La tabla presenta la especificación computacional. Los valores del ejercicio no proceden de encuestas institucionales.

Construcción del ejercicio de datos sintéticos

La matriz demostrativa se construyó con un generador seudoaleatorio de semilla fija 20261119 y cien observaciones. Las variables ordinales se generaron en el intervalo 1–5 y la edad entre 12 y 17 años. La probabilidad de la respuesta binaria se configuró mediante una función logística dependiente de los diez predictores, con incrementos y reducciones programados para mostrar la posibilidad de recuperar patrones multivariables. Después se realizó una extracción binomial. El resultado es una simulación técnicamente reproducible, no una estimación de prevalencia ni una inferencia acerca de los estudiantes del establecimiento.

Entrenamiento de modelos y validación

Se compararon tres estrategias: regresión logística con regularización L2 como línea de base interpretable; bosque aleatorio con 250 árboles, profundidad máxima cinco y un mínimo de cinco observaciones terminales; y potenciación de gradiente mediante histogramas con 80 iteraciones, tasa de aprendizaje 0,045, cinco hojas máximas y regularización L2 igual a dos. Se fijó el estado aleatorio en los algoritmos pertinentes para conservar la reproducibilidad. La elección de modelos de complejidad moderada respondió al tamaño reducido

Doi: https://doi.org/10.63688/3wfr8616

de la matriz, con el propósito de ilustrar diferencias de inductores sin introducir arquitecturas profundas innecesarias.

Se aplicó validación cruzada estratificada de cinco particiones, con mezcla aleatoria y semilla 31. Para cada partición, el tratamiento de valores y el escalamiento se situaron dentro del proceso de entrenamiento y no antes de dividir los datos; la regresión logística incorporó estandarización, mientras que los modelos de árboles conservaron la escala original. Las predicciones fuera de la partición de entrenamiento se reunieron para calcular área bajo la curva ROC (AUC), exactitud equilibrada, puntuación F1 y Brier. El umbral de clasificación se fijó en 0,50 para todas las comparaciones y no se optimizó sobre las mismas predicciones utilizadas para informar desempeño.

La AUC mide discriminación entre observaciones clasificadas como positivas y negativas; la exactitud equilibrada promedia sensibilidad y especificidad; F1 combina precisión y sensibilidad; y Brier resume el error cuadrático de las probabilidades, donde valores menores representan mejor ajuste probabilístico. Estos criterios permiten discutir varios aspectos del comportamiento de los modelos y evitar una lectura aislada de la exactitud global. En aplicaciones empíricas, la selección de hiperparámetros y la evaluación requieren separar estrictamente etapas de desarrollo y validación, en línea con las consideraciones de Varma y Simon (2006).

Análisis cualitativo complementario

La caracterización del entorno escolar se organizó en cuatro dimensiones: relaciones familiares y normas de cuidado; incorporación de adolescentes a actividades productivas; significados atribuidos a celebraciones y rituales comunitarios; y experiencias de pares, escolaridad y medios digitales. La interpretación distinguió entre exposición a conductas adultas y consumo adolescente efectivamente medido. Además, se evitó confundir el uso ritual de elementos tradicionales con patrones de consumo que requieren una definición conductual independiente.

Consideraciones éticas y de protección de datos

El procedimiento computacional no utilizó nombres, entrevistas individualizadas ni registros personales de menores de edad: todas sus filas se generaron de manera artificial. Para una aplicación con información escolar real, el análisis requiere autorización institucional, consentimiento de representantes y asentimiento de los adolescentes cuando

ISSN: 3103-1307
Vol. 4 Núm. 1 (2026)

corresponda, minimización de variables, control de acceso, anonimización y restricciones explícitas de uso. Una puntuación de riesgo no constituye diagnóstico y no debe utilizarse para sancionar, estigmatizar o clasificar públicamente a estudiantes. El tratamiento culturalmente pertinente exige incorporar a la comunidad educativa en la interpretación de los hallazgos.

3. RESULTADOS

Caracterización del contexto de investigación

La descripción institucional sitúa el estudio en un entorno rural y plurilingüe con prácticas de agricultura, ganadería, transporte, construcción civil y transformación de lácteos. Las familias proceden de distintas regiones del país y conviven en escenarios comunitarios con uso de castellano y lenguas originarias. En este territorio, las actividades de reciprocidad y celebración, junto con el trabajo familiar, generan situaciones en las que los adolescentes observan comportamientos adultos relacionados con bebidas alcohólicas. Tales observaciones proporcionan categorías para la elaboración de variables, pero no equivalen a tasas de consumo ni a asociaciones estadísticas demostradas.

Las descripciones del manuscrito atribuyen relevancia a la comunicación entre cuidadores y estudiantes, la supervisión de las actividades juveniles, la exposición a hábitos adultos y las normas del grupo de pares. También señalan situaciones de migración laboral y redistribución de responsabilidades domésticas. Estas experiencias se interpretan como condiciones de convivencia que merecen exploración y no como un inventario de causas inevitables. La lectura de las prácticas rituales considera su función comunitaria y diferencia su significado de la conducta de ingerir alcohol en menores.

Distribución del ejercicio computacional

En la matriz sintética de 100 registros, 32 observaciones recibieron la etiqueta de consumo reciente y 68 la etiqueta de no consumo. La proporción correspondiente se deriva exclusivamente de los parámetros del generador y no representa una prevalencia de la comunidad escolar. La presencia de las dos clases permitió calcular las métricas de discriminación y clasificación mediante predicciones obtenidas por validación cruzada.

Tabla 2.

Características del ejercicio algorítmico

Doi: https://doi.org/10.63688/3wfr8616

ISSN: 3103-1307
Vol. 4 Núm. 1 (2026)

Elemento	Especificación
Registros	100 observaciones sintéticas
Clasificación binaria	32 positivos; 68 negativos
Predictores	10 variables: 9 ordinales y edad
Validación	5 particiones estratificadas, predicciones fuera de muestra
Algoritmos	Regresión logística, bosque aleatorio y gradient boosting
Indicadores	AUC, exactitud equilibrada, F1 y puntuación Brier

Nota. Estas cifras describen exclusivamente una simulación reproducible y no deben trasladarse como resultados de encuestas.

Comparación del desempeño de los modelos

El procedimiento produjo estimaciones de desempeño a partir de predicciones generadas para cada fila cuando esta se encontraba fuera del subconjunto de entrenamiento. La comparación se presenta como demostración de funcionamiento, no como evidencia de capacidad predictiva en la población escolar. Las diferencias entre algoritmos permiten ilustrar que una mayor complejidad no asegura mejores resultados en conjuntos pequeños, y que la selección responsable debe considerar simultáneamente discriminación, errores por clase y calidad de probabilidades.

Tabla 3.

Métricas obtenidas en los 100 registros sintéticos

Modelo	AUC	Exactitud equilibrada	F1	Brier
Regresión logística	0.791	0.700	0.590	0.171
Bosque aleatorio	0.718	0.581	0.348	0.191
Gradient boosting	0.688	0.599	0.429	0.204

Nota. Valores derivados de la simulación, con validación cruzada estratificada de cinco particiones y umbral 0,50. Brier: menor es mejor. No son estimaciones empíricas de la Unidad Educativa Máximo Novillo I.

Doi: https://doi.org/10.63688/3wfr8616

ISSN: 3103-1307
Vol. 4 Núm. 1 (2026)

En esta especificación, la regresión logística ofrece un punto de comparación útil porque muestra la relación direccional entre predictores estandarizados y probabilidad de clasificación. El bosque aleatorio admite interacciones y relaciones no lineales, aunque su importancia de variables puede ser inestable con un tamaño muestral limitado. La potenciación de gradiente combina reglas sucesivas, pero también requiere controlar su complejidad. En los tres casos, los resultados deben interpretarse con las restricciones de un diseño transversal y con atención a la posibilidad de sobreajuste.

Lectura de predictores e implicaciones educativas

La matriz de variables organizada para el componente predictivo muestra una correspondencia sustantiva entre los factores identificados en la descripción cualitativa y las preguntas susceptibles de medición. Las escalas de comunicación y supervisión familiar permiten representar experiencias de acompañamiento; las variables de modelos adultos, celebraciones y actividad productiva distinguen espacios de exposición; y la presión de pares, disponibilidad y vínculo escolar aportan información sobre la socialización adolescente. La edad actúa como variable de ajuste, sin sustituir la medición de conductas.

Tabla 4.

Interpretación educativa de dimensiones incorporadas al análisis

Dimensión	Lectura para el contexto escolar	Uso preventivo
Comunicación y supervisión	Acompañamiento y establecimiento de normas en el hogar	Diálogo familia–escuela y educación preventiva
Modelos adultos y celebraciones	Contextos de observación de conductas y significados culturales	Prevención contextualizada sin estigmatización cultural
Pares y disponibilidad	Normas del grupo y facilidad de acceso percibida	Habilidades de resistencia y control de venta a menores
Trabajo juvenil y estrés	Responsabilidades, tiempo disponible y bienestar	Apoyo psicosocial, permanencia y protección escolar
Vínculo escolar	Sentido de pertenencia y acompañamiento educativo	Tutorías y participación estudiantil

Doi: https://doi.org/10.63688/3wfr8616

Nota. Síntesis interpretativa de categorías del manuscrito institucional y del marco de variables definido para los modelos.

El uso de variables culturalmente sensibles exige una interpretación situada. No se clasificó la identidad lingüística, el origen migratorio ni la pertenencia comunitaria como factores de riesgo por sí mismos. Tampoco se incluyó la hoja de coca como sustituto del desenlace relacionado con alcohol. La operacionalización se concentró en hechos observables, prácticas de supervisión y percepciones de exposición, para reducir el peligro de reproducir prejuicios sociales en los resultados de un algoritmo.

4. DISCUSIÓN

El cambio de un abordaje predominantemente descriptivo a un enfoque asistido por aprendizaje automático amplía las posibilidades de organizar factores familiares y culturales, pero también modifica las exigencias de evidencia. Los hallazgos cualitativos del contexto de Integración del Norte permiten construir dimensiones pertinentes; el ejercicio computacional, en cambio, ilustra técnicas de clasificación sin estimar prevalencia ni riesgo real de la población escolar. Esta separación es fundamental para interpretar los resultados a la luz de Luwedde y Peltzer (2023), cuyo trabajo nacional describe asociaciones poblacionales de Bolivia con una base de datos efectivamente observada.

Los factores familiares adquieren especial relevancia por su naturaleza potencialmente modificable. Las revisiones de Ryan et al. (2010) y Yap et al. (2017) sostienen que las prácticas parentales pueden relacionarse con diferentes comportamientos de consumo adolescente; Pizarro et al. (2017) aporta evidencia de la importancia del contexto familiar en un vecindario peruano. La caracterización escolar coincide temáticamente con estas líneas al describir acompañamiento desigual, responsabilidades familiares y modelos adultos. Sin embargo, una puntuación de clasificación no debería interpretarse como juicio sobre las familias ni como relación causal automática.

La incorporación de variables culturales constituye uno de los aportes del enfoque. MacArthur et al. (2020) mostraron que las normas de consumo adquieren sentido en redes sociales e interacciones locales, lo cual invita a entender la exposición a festividades o labores familiares como un contexto que necesita interpretación. En entornos bolivianos, el reconocimiento de la ch’alla, los sistemas de reciprocidad y el papel tradicional de la hoja de

coca evita lecturas simplificadoras. El foco analítico se mantiene en la disponibilidad de alcohol, la edad de los participantes, las oportunidades de consumo y las prácticas de orientación, no en atribuir propiedades de riesgo a una cultura específica.

Los factores de pares y medios digitales también son adecuados para un diseño multivariable. Ivaniushina y Titkova (2021) reunieron evidencia de procesos de influencia y selección de amistades, mientras que Eisenberg et al. (2014) analizaron la formación de normas sociales y Nesi et al. (2017) estudiaron la exposición digital a contenidos relacionados con alcohol. Desde el punto de vista de modelación, esto respalda incorporar preguntas sobre creencias normativas, aceptación grupal y contactos con contenidos, siempre que los instrumentos midan cada constructo con claridad y eviten interpretaciones moralizantes de la conectividad juvenil.

La literatura predictiva muestra la utilidad de comparar distintos modelos antes de proponer aplicaciones educativas. Afzali et al. (2019) demostraron que los algoritmos pueden evaluarse entre estudios y contextos culturales, aunque la capacidad de generalización depende del conjunto de variables y de la población. Squeglia et al. (2017) exploraron predictores del inicio del consumo con información multidimensional, mientras que Lee (2025) aplicó métodos de aprendizaje automático y profundo a adolescentes coreanos. Kim et al. (2025) amplió el análisis a cohortes internacionales y mostró la relevancia de contrastar modelos en poblaciones diferenciadas. Estos antecedentes no autorizan extrapolar valores de desempeño obtenidos fuera de Bolivia a una unidad educativa concreta.

La comparación de regresión logística, bosque aleatorio y potenciación de gradiente permite discutir interpretabilidad y flexibilidad. La primera puede presentar efectos condicionales bajo hipótesis funcionales explícitas; el segundo combina árboles con variación de predictores como planteó Breiman (2001); y los sistemas de boosting, descritos por Chen y Guestrin (2016), pueden capturar interacciones complejas. El uso de implementaciones disponibles en scikit-learn, según Pedregosa et al. (2011), facilita la reproducibilidad. Con cien registros, sin embargo, la estabilidad de los parámetros y la evaluación de subgrupos son especialmente sensibles al azar y al desequilibrio de clases.

En el componente de interpretabilidad, las metodologías de explicación local constituyen una alternativa para inspeccionar predicciones individuales. Lundberg y Lee (2017) introdujeron el marco SHAP, capaz de estimar contribuciones de variables a una predicción. Su

ISSN: 3103-1307
Vol. 4 Núm. 1 (2026)

utilidad depende de que el modelo tenga validez empírica y de que las explicaciones no sean confundidas con mecanismos causales. Para una escuela, ninguna atribución algorítmica debe sustituir entrevistas, conocimiento pedagógico ni deliberación con familias y estudiantes; la interpretación responsable debe situar las variables dentro de relaciones sociales concretas.

Los controles metodológicos resultan centrales. Varma y Simon (2006) advirtieron que la selección y validación sobre las mismas particiones conduce a estimaciones optimistas. Collins et al. (2024), mediante TRIPOD+AI, enfatizaron el reporte transparente de población, desenlaces, predictores, datos faltantes y desempeño; Moons et al. (2025) establecieron criterios para juzgar sesgos y aplicabilidad en estudios de predicción. En la investigación educativa ello exige explicitar cómo se dividen los datos, cuáles decisiones fueron tomadas antes del análisis y qué procedimientos impedirán que información posterior al desenlace se introduzca como predictor.

El componente computacional aquí presentado cumple una función metodológica de exploración. Los patrones programados en una matriz sintética no pueden emplearse para afirmar que la supervisión, el estrés, las celebraciones o las amistades poseen determinada magnitud de asociación en adolescentes reales. Tampoco es posible estimar sensibilidad, AUC o calibración para una institución cuando sus respuestas individuales no están incorporadas. En un análisis empírico, estas estimaciones deben basarse en observaciones reales, respetar un protocolo de evaluación y acompañarse de intervalos de incertidumbre y análisis de equidad cuando el tamaño de los subgrupos lo permita.

Las implicaciones preventivas no dependen únicamente de obtener una puntuación predictiva. Hurley et al. (2019), al revisar programas de prevención con participación de padres, mostraron que el trabajo con familias representa una línea de intervención importante. El contexto educativo analizado sugiere combinar comunicación intrafamiliar, tutorías, prevención del acceso comercial de menores y participación comunitaria. Una plataforma predictiva, de utilizarse, debería orientar decisiones agregadas y voluntarias de apoyo, con prohibición de usos disciplinarios o discriminatorios. La dimensión tecnológica solo añade valor si fortalece intervenciones humanas culturalmente respetuosas.

Finalmente, el diálogo entre investigación cualitativa y analítica de datos permite plantear preguntas sobre la heterogeneidad del consumo entre estudiantes. Las categorías

Doi: https://doi.org/10.63688/3wfr8616

contextuales ayudan a decidir qué variables merecen medición; los modelos cuantitativos pueden examinar combinaciones y priorizar hipótesis; y la interpretación sociocultural evalúa si esas relaciones son plausibles y útiles en el territorio. La contribución científica reside en conectar estos planos sin sustituir el razonamiento educativo por un resultado matemático ni atribuir a los algoritmos una neutralidad social que no poseen.

5. CONCLUSIÓN

El estudio identificó una estructura de análisis integrada por comunicación y supervisión familiar, observación de modelos adultos, exposición en espacios productivos y comunitarios, influencia de pares, bienestar emocional y vinculación escolar. Estas dimensiones permiten comprender el consumo adolescente como un fenómeno contextual y constituyen una base de operacionalización para su estudio mediante aprendizaje automático. La comparación computacional de regresión logística, bosque aleatorio y potenciación de gradiente mostró cómo evaluar discriminación, equilibrio de clasificación y calidad probabilística mediante validación cruzada. Por haberse realizado sobre registros sintéticos, este ejercicio demuestra un procedimiento y no determina asociaciones empíricas ni capacidad predictiva para los estudiantes de la institución.

El enfoque propuesto requiere combinar herramientas computacionales con comprensión intercultural, resguardo ético y criterios preventivos. La contribución principal consiste en establecer una ruta metodológica que permite pasar de descripciones sociocomunitarias a variables medibles, sin confundir prácticas tradicionales con conductas de consumo ni emplear predicciones para etiquetar a adolescentes. La validación con registros originales es indispensable antes de formular recomendaciones basadas en niveles individuales de riesgo.

REFERENCIAS BIBLIOGRÁFICAS

Afzali, M. H., Sunderland, M., Stewart, S., Masse, B., Seguin, J., Newton, N., Teesson, M., & Conrod, P. (2019). Machine-learning prediction of adolescent alcohol use: A cross-study, cross-cultural validation. *Addiction, 114*(4), 662–671. https://doi.org/10.1111/add.14504

Breiman, L. (2001). Random forests. *Machine Learning, 45*, 5–32. https://doi.org/10.1023/A:1010933404324

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). Association for Computing Machinery. https://doi.org/10.1145/2939672.2939785

Collins, G. S., Moons, K. G. M., Dhiman, P., Riley, R. D., Beam, A. L., Van Calster, B., Ghassemi, M., Liu, X., Reitsma, J. B., van Smeden, M., Boulesteix, A.-L., Camaradou, J. C., Celi, L. A., Denaxas, S., Denniston, A. K., Glocker, B., Golub, R. M., Harvey, H., Heinze, G., . . . Logullo, P. (2024). TRIPOD+AI statement: Updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ, 385*, e078378. https://doi.org/10.1136/bmj-2023-078378

Eisenberg, M. E., Toumbourou, J. W., Catalano, R. F., & Hemphill, S. A. (2014). Social norms in the development of adolescent substance use: A longitudinal analysis of the International Youth Development Study. *Journal of Youth and Adolescence, 43*(9), 1486–1497. https://doi.org/10.1007/s10964-014-0111-1

Hurley, E., Dietrich, T., & Rundle-Thiele, S. (2019). A systematic review of parent based programs to prevent or reduce alcohol consumption in adolescents. *BMC Public Health, 19*, 1451. https://doi.org/10.1186/s12889-019-7733-x

Ivaniushina, V., & Titkova, V. (2021). Peer influence in adolescent drinking behavior: A meta-analysis of stochastic actor-based modeling studies. *PLOS ONE, 16*(4), e0250169. https://doi.org/10.1371/journal.pone.0250169

Kim, S., Kim, H., Kim, S., Lee, H., Hammoodi, A., Choi, Y., Kim, H. J., Smith, L., Kim, M. S., Fond, G., Boyer, L., Baik, S. W., Lee, H., Park, J., Kwon, R., Woo, S., & Yon, D. K. (2025). Machine learning–based prediction of substance use in adolescents in three independent worldwide cohorts: Algorithm development and validation study. *Journal of Medical Internet Research, 27*, e62805. https://doi.org/10.2196/62805

Lee, S. (2025). Development of deep learning auto-encoder algorithms for predicting alcohol use in Korean adolescents based on cross-sectional data. *Social Science & Medicine, 367*, 117690. https://doi.org/10.1016/j.socscimed.2025.117690

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems* (Vol. 30). https://proceedings.neurips.cc/paper_files/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html

Luwedde, E., & Peltzer, K. (2023). Levels and pattern of alcohol consumption among adolescents in Bolivia: A national cross-sectional survey in 2018. *Adolescents, 3*(1), 1–9. https://doi.org/10.3390/adolescents3010001

MacArthur, G. J., Hickman, M., & Campbell, R. (2020). Qualitative exploration of the intersection between social influences and cultural norms in relation to the development of alcohol use behaviour during adolescence. *BMJ Open, 10*(3), e030556. https://doi.org/10.1136/bmjopen-2019-030556

Moons, K. G. M., Damen, J. A. A., Kaul, T., Hooft, L., Andaur Navarro, C., Dhiman, P., Beam, A. L., Van Calster, B., Celi, L. A., Denaxas, S., Denniston, A. K., Ghassemi, M., Heinze, G., Kengne, A. P., Maier-Hein, L., Liu, X., Logullo, P., McCradden, M. D., Liu, N., . . . van Smeden, M. (2025). PROBAST+AI: An updated quality, risk of bias, and applicability assessment tool for prediction models using regression or artificial intelligence methods. *BMJ, 388*, e082505. https://doi.org/10.1136/bmj-2024-082505

Nesi, J., Rothenberg, W. A., Hussong, A. M., & Jackson, K. M. (2017). Friends' alcohol-related social networking site activity predicts escalations in adolescent drinking: Mediation by peer norms. *Journal of Adolescent Health, 60*(6), 641–647. https://doi.org/10.1016/j.jadohealth.2017.01.009

Organización Mundial de la Salud. (2024). *Alcohol*. https://www.who.int/news-room/fact-sheets/detail/alcohol/

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12*, 2825–2830. https://www.jmlr.org/papers/v12/pedregosa11a.html

Pizarro, K. W., Bustamante, I. V., & Surkan, P. J. (2017). Family factors and adolescent problem drinking in a high-risk urban Peruvian neighborhood. *Substance Use & Misuse, 52*(2), 194–202. https://doi.org/10.1080/10826084.2016.1223137

Ryan, S. M., Jorm, A. F., & Lubman, D. I. (2010). Parenting factors associated with reduced adolescent alcohol use: A systematic review of longitudinal studies. *Australian & New Zealand Journal of Psychiatry, 44*(9), 774–783. https://doi.org/10.1080/00048674.2010.501759

ISSN: 3103-1307

Vol. 4 Núm. 1 (2026)

Squeglia, L. M., Ball, T. M., Jacobus, J., Brumback, T., McKenna, B. S., Nguyen-Louie, T. T., Sorg, S. F., Paulus, M. P., & Tapert, S. F. (2017). Neural predictors of initiating alcohol use during adolescence. *American Journal of Psychiatry, 174*(2), 172–185. https://doi.org/10.1176/appi.ajp.2016.15121587

Varma, S., & Simon, R. (2006). Bias in error estimation when using cross-validation for model selection. *BMC Bioinformatics, 7*, 91. https://doi.org/10.1186/1471-2105-7-91

Yap, M. B. H., Cheong, T. W. K., Zaravinos-Tsakos, F., Lubman, D. I., & Jorm, A. F. (2017). Modifiable parenting factors associated with adolescent alcohol misuse: A systematic review and meta-analysis of longitudinal studies. *Addiction, 112*(7), 1142–1162. https://doi.org/10.1111/add.13785

CONFLICTO DE INTERESES

No existe.

FINANCIAMIENTO

No aplica.

CONTRIBUCIÓN DE AUTORÍA

Conceptualización: Wilder Mel Mamani Chaca.
Metodología: Wilder Mel Mamani Chaca.
Investigación: Wilder Mel Mamani Chaca.
Redacción – borrador original: Wilder Mel Mamani Chaca.
Redacción – revisión y edición: Wilder Mel Mamani Chaca.