



MODELO DE REGRESIÓN LOGÍSTICA PARA PREDICIR EL NIVEL DE RIESGO DE CONSUMO O ADICCIÓN A SUSTANCIAS PSICOACTIVAS ILEGALES EN POBLACIÓN COLOMBIANA

Logistic Regression Model to Predict the Risk Level Consumption or Addiction of Illegal Psychoactive Substances in the Colombian Population

Jefferson-Eduardo Campo-Yule 

Universidad del Cauca, Popayán, Colombia. 

jeffersoncy@unicauca.edu.co

Danny-Alberto Díaz-Magé 

Universidad del Cauca, Popayán, Colombia. 

dannyad@unicauca.edu.co

Hugo-Armando Ordoñez-Eraso 

Universidad del Cauca, Popayán, Colombia. 

hugoordonez@unicauca.edu.co

Fecha de recibido: 30-11-2024



Fecha de aceptado: 22-04-2025

RESUMEN

Una de las principales problemáticas sociales que se presenta en Colombia es el consumo de sustancias psicoactivas ilegales, entre las que la marihuana es la sustancia más consumida por las personas. Si bien se han creado diversos programas preventivos para reducir el consumo de estas sustancias, son pocos los estudios que usan la tecnología para abordar esta problemática. Por tanto, este estudio presenta un modelo de *machine learning* que predice el riesgo de consumo de sustancias psicoactivas ilegales en Colombia, con datos de la ENCSPA-2019. El modelo demostró un rendimiento óptimo y logró métricas superiores al 98 %, lo que permite identificar factores clave asociados al consumo que influyen en el nivel de riesgo de una persona y ofrece una herramienta útil para apoyar decisiones en salud pública y prevención.

Palabras clave: consumo; *machine learning*; modelo predictivo; sustancias psicoactivas ilegales.

ABSTRACT

One of the main social issues in Colombia is the consumption of illegal psychoactive substances, among which marijuana is the most consumed by individuals. Although various preventive programs have been created to reduce the consumption of these substances, few studies utilize technology to address this problem. Therefore, this study presents a machine learning model that predicts the risk of illegal psychoactive substance use in Colombia, using data from ENCSPA-2019. The model demonstrated optimal performance, achieving metrics above 98%, enabling the identification of key factors associated with consumption that influence an individual's risk level, and providing a useful tool to support public health decisions and prevention efforts.

Keywords: consumption; illegal psychoactive substances; *machine learning*; predictive model.

MODELO DE REGRESSÃO LOGÍSTICA PARA PREVER O NÍVEL DE RISCO DE CONSUMO OU DEPENDÊNCIA DE SUBSTÂNCIAS PSICOATIVAS ILEGAIS NA POPULAÇÃO COLOMBIANA

RESUMO

Um dos principais problemas sociais enfrentados na Colômbia é o consumo de substâncias psicoativas ilegais, sendo a maconha a substância mais consumida pela população. Embora diversos programas preventivos tenham sido criados para reduzir o consumo dessas substâncias, são poucos os estudos que utilizam a tecnologia para enfrentar essa problemática. Assim, este estudo apresenta um modelo de *machine learning* que prevê o risco de consumo de substâncias psicoativas ilegais na Colômbia, com dados da ENCSPA-2019. O modelo demonstrou um desempenho ótimo e alcançou métricas superiores a 98%, o que permite identificar fatores-chave associados ao consumo que influenciam o nível de risco de uma pessoa e oferece uma ferramenta útil para apoiar decisões em saúde pública e prevenção.

Palavras-chave: consumo; *machine learning*; modelo preditivo; substâncias psicoativas ilegais.

1. INTRODUCCIÓN

Según el Ministerio de Justicia y Derecho del Colombia, las sustancias psicoactivas ilegales (SPI), como marihuana, cocaína o bazuco, pueden traer consecuencias graves para la salud y el bienestar social [1]. Además, el Ministerio de Salud define el consumo de SPI como una alteración al sistema nervioso que puede causar desde problemas sociales hasta la muerte por sobredosis. Como respuesta a esta problemática, en Colombia se han creado programas como “Más mente, más prevención” y el “Plan Nacional de Promoción de la Salud, Prevención y Atención al Consumo de Sustancias Psicoactivas 2014-2021” para educar a la población sobre el consumo de estas sustancias y reducir sus efectos negativos, como problemas personales, laborales y sobredosis [2].

El consumo de SPI en Colombia ha aumentado notablemente, sobre todo entre los jóvenes, y son quienes más están expuestos a sustancias como marihuana, cocaína y bazuco. Esta problemática no solo afecta el consumo personal, sino que también impulsa el crecimiento del mercado ilegal. Aunque algunos logran dejar el consumo de estas sustancias, otros enfrentan graves consecuencias [1].

El 10,3 % de la población entre 12 y 65 años ha consumido alguna SPI. El consumo es mayor en edades entre 25 y 34 años, y aumenta con el nivel socioeconómico. Además, el estudio indica que el 50 % de las personas inicia el consumo entre los 15 y los 20 años [3]. La marihuana es la sustancia más consumida, con un 8,3 %. La mayor frecuencia de consumo de marihuana se da entre personas de 18 a 24 años y estratos altos. Le siguen la cocaína con un 2,1 %, esta es más común entre personas con un rango de edad de 18 a 24 años, y el bazuco con un 0,55 %, este tiene mayor consumo en estratos bajos.

Dado el avance de la tecnología, es fundamental crear una herramienta que apoye a entidades gubernamentales y de salud en la prevención del consumo de SPI en Colombia. El uso de la inteligencia artificial (IA), en especial de *machine learning* (ML), se presenta como una solución viable para desarrollar herramientas de modelos predictivos capaces de analizar datos y poder anticipar posibles riesgos en personas de diferentes edades. Por ejemplo, Gong *et al.* [4] hacen uso de Machine Learning y Deep Learning para desarrollar un modelo que identifica a pacientes con consumo indebido de SPI, mientras que Islam *et al.* [5] plantean un enfoque que predice la vulnerabilidad de una persona al uso y el abuso de SPI. Sin embargo, aunque ya existen estudios internacionales que aplican dichas técnicas, aún son escasos los trabajos que se enfocan en predecir el riesgo de consumo de estas sustancias en un contexto colombiano.

Una de las principales problemáticas en Colombia es el consumo de SPI, la cual ha presentado un aumento continuo a lo largo del tiempo. Por ello es necesario comprender y combatir esta situación identificando los diferentes estudios que se han desarrollado como solución para identificar a personas en riesgo de consumir SPI. Por otro lado, en los últimos años se realizaron diversos estudios que abarcan esta problemática, pero en países distintos a Colombia. Por ejemplo, Arif *et al.* [6] desarrollaron un modelo predictivo para predecir el riesgo de adicción a SPI, en esta investigación encontraron que el algoritmo de regresión logística superó una precisión del 97,91 % respecto a algoritmos como *K-nearest neighbors* (KNN), *support vector machine* (SVM), *Naïve bayes, classification, random forest* (RF), *regression trees, multilayer perception* (MLP), *adaptive boosting* (AdaBoost) y *gradient boosting machine* (GBM).

Así mismo, se planteó un marco que se basa en ML para identificar la vulnerabilidad individual hacia el abuso de SPI, este enfoque ayuda a prevenir la adicción en personas que consumen estas sustancias [7]. Además, otros estudios se han enfocado en el análisis del consumo de una única sustancia; es el caso del trabajo de Davis *et al.* [8], en el que se usan modelos como Random Forest y Regularized Cox regression para predecir el riesgo de recaída en el consumo de opioides. También está el trabajo de Han y Seo [9], cuyo objetivo principal es plantear un modelo para predecir el riesgo de consumo de marihuana mediante cigarrillos electrónicos. Para este estudio se enfocan en *least absolute shrinkage and selection operator* (LASSO) para predecir las variables de consumo de marihuana y el modelo *classification and regression tree* (CART) para predecir el número de personas que están en riesgo de consumo de marihuana mediante cigarrillos electrónicos.

Zoboroski *et al.* [10] se enfocaron en entidades como el ejército; en su estudio usan modelos o algoritmos como regresión logística, árboles de decisión y redes neuronales para predecir el riesgo de consumo de marihuana teniendo en cuenta rasgos de personalidad. Concluyeron que los jóvenes presentan un alto riesgo de consumo. Negriff *et al.* [11] utilizaron ML para examinar los factores de riesgo en pertenecientes y no pertenecientes al bienestar infantil; en este estudio usaron regresión logística, LASSO y SVM para crear el modelo que permitirá predecir qué personas están en riesgo de consumo de marihuana.

Cabe destacar que, a pesar de la escasez de estudios sobre el consumo de SPI en Colombia, se han desarrollado aplicaciones de IA para analizar patrones de consumo de estas sustancias. Por ejemplo, Palomino *et al.* [12] emplearon redes neuronales profundas con el objetivo de identificar patrones de consumo y abuso de drogas; extrajeron las características principales de los datos de entrada, de los cuales obtuvieron que las características relacionadas con la economía familiar ayudan a explicar los patrones que se presentan para el consumo de las SPI.

Ordoñez *et al.* [13] propusieron un modelo basado en algoritmos de aprendizaje automático con el fin de encontrar tendencias en la adicción al consumo de sustancias psicoactivas en personas sin hogar. Sin embargo, como ya se mencionó, los estudios relacionados con el consumo de SPI en Colombia son muy escasos. Por ejemplo, se encontraron estudios sobre el uso de IA aplicados en otros ámbitos, como en el trabajo de Pinto Hidalgo [14], en el que se utilizó inteligencia geoespacial e IA con el fin de detectar lugares potenciales para la producción de pasta de coca. En este estudio se entrenó un modelo de aprendizaje profundo con diversas imágenes satelitales y se logró una precisión media de 90,07 %. Esto demuestra que la metodología propuesta, así como el modelo, pueden ser útiles para combatir el narcotráfico.

En conclusión, se encontraron diversos estudios relacionados con el consumo de SPI, tanto a nivel mundial como en Colombia, que mediante el uso de la IA y, en especial, con el uso de ML buscan realizar un análisis exhaustivo de datos de personas para la creación de modelos predictivos; sin embargo, en Colombia no se encontraron estudios que permitan analizar datos ni personales ni de consumo para predecir el nivel el riesgo de consumo de las personas según la información que brindan.

Cabe mencionar que las SPI contienen componentes sintéticos y naturales que son capaces de alterar el sistema nervioso, y afectan el comportamiento y el estado emocional de una persona [13]. A continuación, se muestran algunas de estas sustancias.

Marihuana: sustancia ilegal capaz de afectar directamente las funciones del cerebro, sobre todo las partes encargadas de memoria, aprendizaje, atención, entre otras.

Bazuco: sustancia tóxica que destruye el tejido cerebral y provoca una pérdida irreversible de la memoria; los efectos de esta sustancia son inmediatos y se vuelve adictiva debido a la rápida disolución en el torrente sanguíneo.

Cocaína: sustancia ilegal que al ser consumida afecta el sistema nervioso y el resto del cuerpo de forma inmediata; durante los primeros 30 minutos después de su consumo los efectos son la sensación de menor cansancio, hiperestimulación y mayor estado de alerta mental.

El consumo de este tipo de SPI es una problemática que afecta a la sociedad y se presenta en los diferentes tipos de edades: adolescentes, jóvenes, adultos y personas mayores. La [Figura 1](#) muestra la frecuencia de consumo de marihuana en relación con diferentes grupos de edad en la población de Colombia. Los datos están segmentados en cuatro grupos de edad: Adolescente (azul), Adulto (verde), Joven (amarillo) y Persona mayor (rojo).

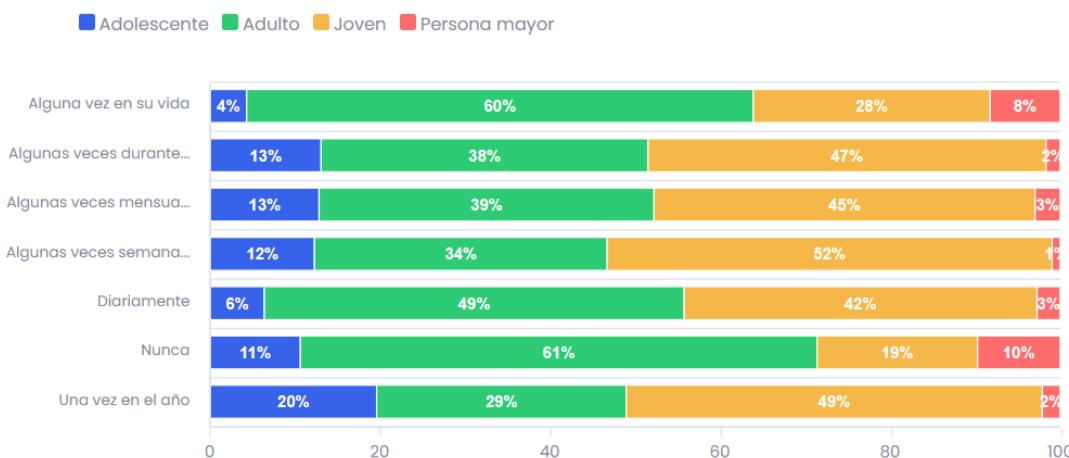


Figura 1. Frecuencia de consumo de marihuana vs. edad

Según la [Figura 1](#), se puede concluir que los jóvenes tienen la mayor frecuencia de consumo de marihuana en casi todas las categorías, sobre todo semanal y mensual. Los adultos también tienen una alta frecuencia de consumo, y se destacan en las categorías de consumo diario y alguna vez en su vida. En el grupo de adolescentes, aunque el consumo es menor en comparación con jóvenes y adultos, hay una notable proporción que consume alguna vez durante el año y alguna vez en su vida. El grupo de personas mayores tiene la menor frecuencia de consumo en todas las categorías.

De acuerdo con la [Figura 2](#), se puede concluir que los adultos tienen la mayor frecuencia de consumo de cocaína en casi todas las categorías, especialmente de manera diaria y semanal. Los jóvenes tienen una alta frecuencia de consumo mensual y varias veces durante el año. El grupo de adolescentes tiene una menor frecuencia de consumo en comparación con jóvenes y adultos; sin embargo, hay una notable proporción que consume cocaína alguna vez durante el año y alguna vez en su vida. En cuanto a las

personas mayores, este grupo tiene la menor frecuencia de consumo en todas las categorías.

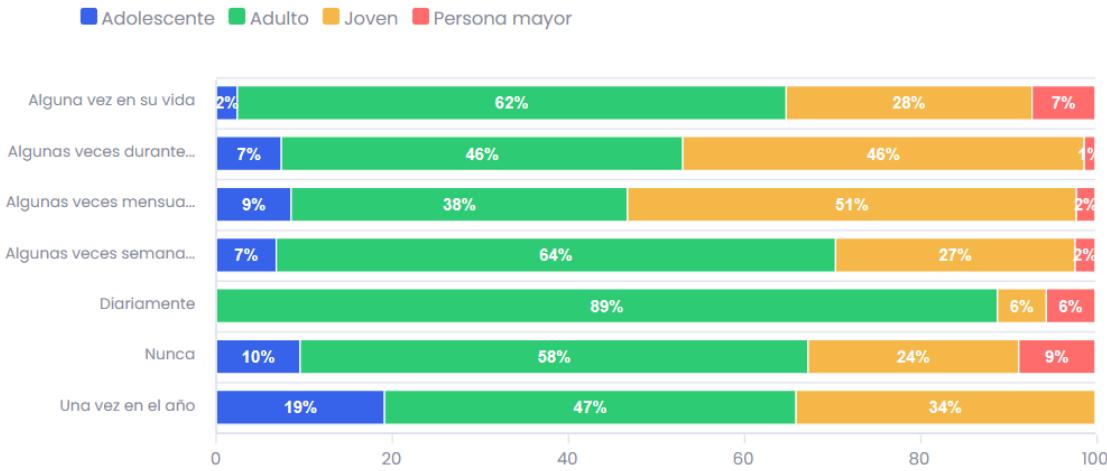


Figura 2. Frecuencia consumo de cocaína vs. edad

La Figura 3 muestra la frecuencia de consumo de bazuco en relación con diferentes grupos de edad en la población de Colombia. Los datos están segmentados en cuatro grupos de edad: Adolescente (azul), Adulto (verde), Joven (amarillo) y Persona mayor (rojo). Según lo anterior, se puede concluir que los adultos consumen con mayor frecuencia bazuco en casi todas las categorías, sobre todo semanal y mensual. Las personas mayores tienen una notable proporción de consumo alguna vez en su vida y durante el año. Los jóvenes tienen una frecuencia de consumo diaria y semanal, pero en menor proporción en comparación con adultos y personas mayores. En el grupo de adolescentes, aunque el consumo es menor en comparación con los otros grupos, hay una notable proporción que consume bazuco algunas veces mensualmente y alguna vez durante el año.

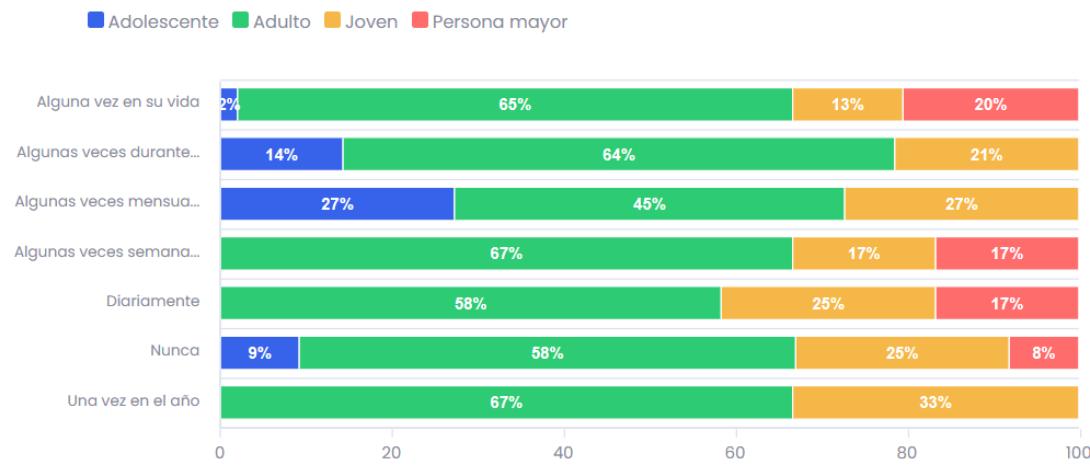


Figura 3. Frecuencia consumo de bazuco vs. edad

Por lo tanto, con base en los resultados obtenidos sobre el uso de ML en la predicción del riesgo de consumo de SPI y teniendo en cuenta los gráficos presentados, se propone desarrollar un modelo de ML

que ayude a entidades gubernamentales y de salud en Colombia en la toma de decisiones para prevenir o reducir el consumo de estas sustancias. Esta problemática afecta a la población de manera cotidiana, y el objetivo principal de esta investigación y creación del modelo es servir como base para futuras investigaciones que deseen abordar esta problemática específica en Colombia. El uso de herramientas tecnológicas como ML y el constante avance de la tecnología permitirán el desarrollo de nuevos modelos que puedan servir como herramientas para identificar y ayudar a las personas en riesgo de consumir SPI en Colombia.

Es importante considerar que aún siguen existiendo diversas complejidades, como la disparidad en la cantidad de datos que hay respecto al consumo de ciertas sustancias. La diferencia en la disponibilidad de datos entre personas que consumen marihuana, cocaína y bazuco es muy notoria. Asimismo, es relevante tener en cuenta el año en que se recopilaron los datos, ya que para este estudio se utilizaron datos del año 2019, datos que pueden variar con el paso del tiempo. Estas variaciones temporales pueden influir en los resultados y en la efectividad del modelo propuesto.

Con base en lo anterior, esta investigación busca proporcionar una herramienta que identifique y apoye a personas que están en riesgo; educar y concientizar al público sobre los riesgos asociados al consumo de SPI, y contribuir al análisis de datos e investigación en el campo de la IA. La herramienta usa un modelo de ML que permite predecir el nivel de riesgo de una persona a partir del análisis de datos personales y datos relacionados con SPI. La solución presentada será de gran ayuda para los entes gubernamentales y de salud, y les permitirá identificar a aquellas personas que se encuentran en riesgo de consumir estas sustancias.

2. METODOLOGÍA

En la [Figura 4](#), se presenta el diseño para la creación del modelo y la herramienta que usa el modelo para predecir el riesgo de consumo de SPI en Colombia. El proceso para la creación se realizó con base en las siguientes subsecciones.

2.1 Conjunto de datos

El conjunto de datos seleccionado se tomó a partir de la Encuesta Nacional de Consumo de Sustancias Psicoactivas en Población General (ENCSPA-2019), realizada por el Departamento Administrativo Nacional de Estadística (DANE) y cuyo objetivo fue conocer el estado sobre el consumo de las diferentes drogas, tanto sustancias legales como ilegales para el año 2019. De este conjunto de datos solo se tuvo en cuenta la información relacionada con el consumo de SPI como marihuana, bazuco y cocaína. El conjunto de datos se divide en diferentes capítulos: desde información general de las personas encuestadas hasta información sobre el consumo de las diferentes sustancias identificadas en la encuesta; además, cabe resaltar que el conjunto de datos seleccionado es de acceso abierto y gratuito (<https://microdatos.dane.gov.co/index.php/catalog/680/study-description>).

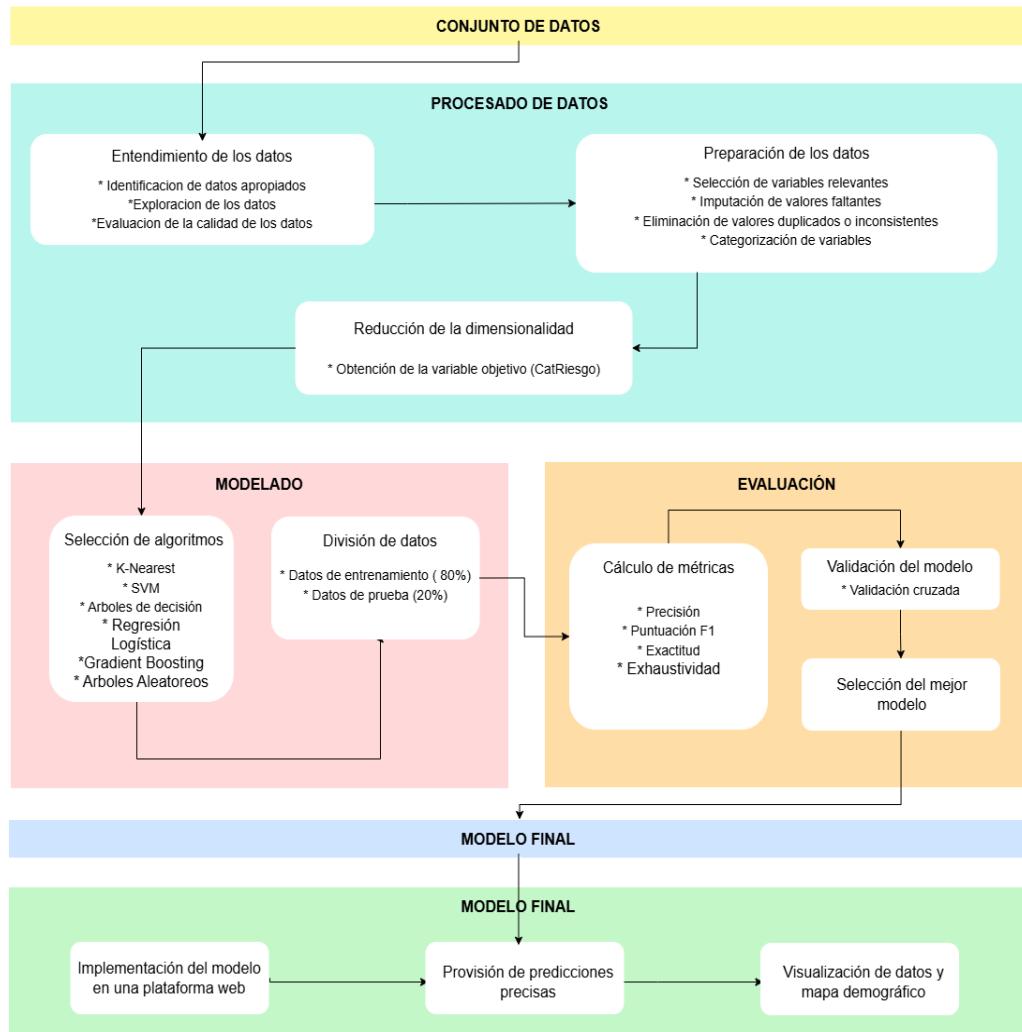


Figura 4. Modelo propuesto

2.2 Procesamiento de datos

1) Comprensión de datos. El desarrollo de estas actividades y los procesos definidos para el proyecto se iniciaron desde la comprensión de los datos, lo que implicó identificar los datos apropiados de acuerdo con la temática seleccionada. En este caso se seleccionó el conjunto de datos registrado por el DANE acerca del nivel de consumo de SPI en Colombia ([Tabla 1](#)).

Con el conjunto de datos seleccionado ([Tabla 1](#)), se realizó una exploración de datos con el fin de comprender el contenido y estructura, incluida la identificación de tipos de variables, distribución de los datos mediante el análisis de tendencias, patrones, gráficas y estadísticas para comprender mejor los datos. Finalmente, se evaluó la calidad de los datos, se identificaron variables o datos faltantes, datos atípicos y datos en los cuales se identificaron inconsistencias, entre otros problemas.

Tabla 1. Descripción del conjunto de datos

Variable	Descripción
Departamento	Departamento en el que vive una persona.
residentes_hogar	Número de personas que residen en el hogar de una persona.
d2_04_num_hijos	Número de hijos vivos que tiene una persona.
per_edad_tipo	Etapa de desarrollo o edad en la que se encuentra una persona.
estrato_tipo	Estrato social en el que se encuentra una persona.
situacion_tipo	Actividad en la que ocupa mayormente el tiempo una persona.
frecuencia_consumo_marihuana_tipo	Frecuencia con la que consume marihuana una persona.
frecuencia_consumo_cocaina_tipo	Frecuencia con la que consume cocaína una persona.
frecuencia_consumo_bazuco_tipo	Frecuencia con la que consume bazuco una persona.
per_sexo_tipo	Sexo con el que se identifica (hombre o mujer) una persona.
vive_padre_hogar_tipo	En el hogar de una persona vive o no el padre.
vive_madre_hogar_tipo	En el hogar de una persona vive o no la madre.
vivienda_tipo	Tipo de vivienda en la que vive una persona.
d_01_aporta_dinero_hogar_tipo	Aporta o no económicamente al hogar una persona.
d_08_estado_salud_tipo	Estado de salud con el que cuenta una persona.
d_09_deprimido_tipo	Se ha sentido deprimida una persona.
d_10_poco_interes_tipo	Poco interés al realizar las cosas que le gusta hacer a una persona.
d_11_h_conocimiento_riesgo_fumar_marihuana_frecuentemente_tipo	Conocimiento sobre el riesgo de consumir marihuana de forma frecuente.
d_11_k_conocimiento_riesgo_cocaina_frecuentemente_tipo	Conocimiento sobre el riesgo de consumir cocaína de forma frecuente.
d_11_n_conocimiento_riesgo_fumar_bazuco_frecuentemente_tipo	Conocimiento sobre el riesgo de consumir bazuco de forma frecuente.
d_12_b_presenta_problema_consumo_sp_barrio_tipo	En el barrio de una persona se presenta o no se presenta el consumo de sustancias psicoactivas.
d_12_c_presenta_problema_expendio_sp_barrio_tipo	En el barrio de una persona se presenta o no se presenta el expendio de sustancias psicoactivas.
d2_01_etnia_tipo	Etnia a la que pertenece una persona.
d2_03_estado_civil_tipo	Estado civil en el que se encuentra una persona.
d2_05_nivel_educativo_tipo	Mayor nivel educativo en el que se encuentra o que ya cursó una persona.
g_01_familiares_consumen_sp_tipo	Los familiares de una persona consumen o no consumen sustancias psicoactivas.
g_02_amigos_consumen_sp_tipo	Los amigos de una persona consumen o no consumen sustancias psicoactivas.
g_03_curiosidad_probar_sp_tipo	Una persona ha sentido o no curiosidad por probar sustancias psicoactivas.
g_04_probaria_sp_tipo	Decisión por probar sustancias psicoactivas.
g_05_posibilidad_probar_sp_tipo	Se le ha presentado la posibilidad o no de probar sustancias psicoactivas a una persona.
g_06_a_posibilidad_conseguir_marihuana_tipo	Facilidad o posibilidad para poder conseguir marihuana.
g_06_b_posibilidad_conseguir_cocaina_tipo	Facilidad o posibilidad para poder conseguir cocaína.
g_06_c_posibilidad_conseguir_bazuco_tipo	Facilidad o posibilidad para poder conseguir bazuco.
g_07_alguien_ofrecio_comprar_probar_sp_tipo	A una persona le han ofrecido comprar o probar sustancias psicoactivas.
g_01_a_num_familiares_consumen_sp_imp_tipo	Número de familiares que consumen sustancias psicoactivas.
g_02_a_num_amigos_consumen_sp_imp_tipo	Número de amigos que consumen sustancias psicoactivas.
g_08_a_ofrecieron_marihuana_imp_tipo	A una persona le han ofrecido y hace cuánto tiempo le ofrecieron marihuana.
g_08_b_ofrecieron_cocaina_imp_tipo	A una persona le han ofrecido y hace cuánto tiempo le ofrecieron cocaína.
g_08_c_ofrecieron_bazuco_imp_tipo	A una persona le han ofrecido y hace cuánto tiempo le ofrecieron bazuco.
CatRiesgo (variable objetivo)	Nivel de riesgo en el que se encuentra la persona.

2) Preparación de datos. En la preparación y transformación de los datos, se seleccionaron las variables más relevantes para el modelo, pues muchas variables originales no tenían suficiente relevancia. Se aplicaron técnicas de imputación de datos en valores faltantes o nulos, se eliminaron algunos valores duplicados y valores que presentaban inconsistencias. También se aplicaron técnicas de categorización en algunas variables para mejorar el rendimiento y la simplificación del modelo.

3) Reducción de la dimensionalidad. Por último, se creó la variable objetivo a partir de las variables seleccionadas. Para esto, se aplicó la técnica análisis de componentes principales (PCA), la cual asegura que la información de las variables más significativas se concentre en esta nueva variable y pueda orientar al modelo predictivo hacia las características más importantes y relevantes. La variable objetivo resultante consta de cinco clases: 1) riesgo muy bajo, 2) riesgo bajo, 3) riesgo medio, 4) riesgo alto y 5) riesgo muy alto. A continuación, se detallan estas categorías.

1. Riesgo muy bajo: esta clase incluye individuos con características y factores de riesgo mínimos asociados al consumo de SPI. Por lo general, estas personas cuentan con un entorno familiar estable, acceso a educación y empleo, y una red de apoyo social robusta. En el contexto colombiano, se pueden incluir personas que viven en sitios con baja incidencia de violencia y crimen, además de estar comprometidas en la promoción de estilos de vida saludables. Estas personas tienen oportunidades de desarrollo personal y profesional. Tales factores contribuyen a crear un entorno protector que minimiza los riesgos de consumo de SPI [15].
2. Riesgo bajo: en esta clase, se encuentran personas con algunos factores de riesgo presentes, pues pueden estar expuestas ocasionalmente a entornos o situaciones de riesgo, pero cuentan con mecanismos de protección y apoyo que les ayudan a mantenerse alejadas del consumo de SPI. En el contexto colombiano, podría incluir a jóvenes en situaciones vulnerables que aún tienen acceso a programas educativos y comunitarios de prevención. La disposición de las entidades gubernamentales en implementar programas de prevención puede hacer una diferencia significativa en reducir el riesgo en estos casos y alcanzar un impacto positivo frente a las consecuencias producidas por el consumo de SPI [15].
3. Riesgo medio: en esta clase, se encuentran personas con una combinación equilibrada entre factores de riesgo y protección. Aunque no se encuentran en la situación más vulnerable, tienen una mayor probabilidad de estar expuestas a situaciones que pueden llevar al consumo de SPI. En el contexto colombiano, se podría incluir a personas en comunidades con altos índices de desempleo y violencia que, de cierta manera, tienen algunos recursos y apoyo social. La efectividad en el seguimiento de procesos sociales y la planificación conjunta entre organizaciones y entidades gubernamentales son cruciales para abordar estos riesgos de manera integral [15].
4. Riesgo alto: en esta clase, las personas presentan múltiples factores de riesgo que aumentan considerablemente la probabilidad de consumir SPI. Estos factores suelen incluir entornos familiares inestables, falta o nula disponibilidad de acceso a educación y empleo, y exposición a entornos con presencia de alta violencia y criminalidad. En Colombia, esta categoría podría representar a personas en zonas de conflicto o comunidades marginadas, en las que las opciones de desarrollo son limitadas y la exposición a las SPI son altas. Por ejemplo, en contextos como el del departamento de La Guajira, se ha identificado un aumento en las situaciones conflictivas como también la falta de conciencia social sobre la problemática, la mala utilización del tiempo libre y los altos índices de desempleo; por tanto, han incrementado los riesgos como problemas sociales, conductas delictivas, pandillas y otros problemas relacionados con el consumo de SPI [16].

5. Riesgo muy alto: esta clase incluye a personas con la mayor probabilidad de consumir SPI, debido a la presencia de múltiples y severos factores de riesgo. Estos individuos suelen vivir en entornos extremadamente desfavorecidos, con escaso acceso a servicios de salud y educación, y con alta exposición a violencia y actividades delictivas. En Colombia, se presentan estas situaciones en personas que viven en la calle, en zonas de conflicto armado, o en comunidades donde el narcotráfico y la delincuencia organizada tienen fuerte presencia. Estos problemas sociales concurrentes afectan el comportamiento humano y comprometen el desarrollo integral de la persona y su familia [16].

2.3 Modelado

En cuanto al modelado, se seleccionaron varios algoritmos considerando los que mejor se adecuaron al problema ([Tabla 2](#)), como también al nuevo conjunto de datos una vez realizado el procesado, para ello se dividieron los datos en conjuntos de entrenamiento con un 80 % y prueba un 20 %. Se realizó la optimización de hiperparámetros a cada uno de los modelos evaluados utilizando técnicas como GridSearch y RandomizedSearch para encontrar los mejores hiperparámetros para todos los modelos seleccionados. Por una parte, GridSearch hace referencia a una búsqueda exhaustiva de los hiperparámetros seleccionados [17], en los que se hace una búsqueda en cuadricula para obtener el mejor resultado. Cabe resaltar que, al agregar muchos hiperparámetros para la optimización de un modelo, esta optimización puede tomar mucho tiempo. Por otro lado, RandomizedSearch, como su nombre lo indica, es una búsqueda aleatoria de los valores de los parámetros seleccionados; para este caso no se usan todos los posibles valores de cada hiperparámetro, sino que mediante el parámetro n_iter se indica el número de iteraciones que se realizarán para la optimización del modelo [18].

Tabla 2. Algoritmos testeados

Modelo	Descripción
Support vector machine	SVM es una de las técnicas de ML usadas para problemas de clasificación en <i>big data</i> y ciencia de datos; sin embargo, es matemáticamente complejo y muy costoso computacionalmente, dado que correlaciona datos con un espacio de características de grandes dimensiones [19].
K-nearest	KNN es uno de los algoritmos más populares aplicados en la minería de datos, se basa en buscar la mayor proximidad para realizar las predicciones o clasificaciones sobre la agrupación de un grupo de datos individual. Es de resaltar que si bien este modelo se puede aplicar a problemas de regresión y clasificación, normalmente se aplica a problemas de clasificación [20].
Random forest	En primer lugar, es necesario conocer un nuevo concepto sobre los árboles de decisión, debido a que el algoritmo RF está compuesto por múltiples árboles de decisión. La lógica de los árboles de decisión se basa en responder una pregunta con Sí o No para tomar un camino u otro, con el fin de obtener una predicción [21].
Xgboost	XGBoost es una implantación de boosting que busca usar múltiples modelos de aprendizaje para crear un nuevo modelo más robusto. XGBoost fue diseñado para mejorar la velocidad y escalabilidad computacionales [22].
Logistic regression	La regresión logística se utiliza como proceso para modelar la probabilidad de un resultado. Es uno de los algoritmos más comunes usados para la creación de modelos predictivos, en los que se usan tanto para escenarios binarios como para problemas de clasificación [23].
Decision tree	El árbol de decisión es un algoritmo supervisado que se usa para resolver problemas de regresión y clasificación; este algoritmo se basa en nodos en diferentes ramas que ayudan a obtener un resultado para una predicción [24].

2.4 Evaluación

Para la evaluación de los modelos y posterior selección del mejor, se compararon los rendimientos de acuerdo con los resultados de métricas como *Precision*, *Accuracy*, *Sensitivity* y *F1-Score* ([Tabla 3](#)); también se realizó validación cruzada para asegurar que el modelo se adapta ante datos nuevos y finalmente se revisó que el modelo cumpliera con los requisitos y expectativas que se plantearon al inicio del proyecto.

Tabla 3. Métricas de evaluación

Modelo	Descripción
	Se entiende como la capacidad del modelo clasificador para no etiquetar como positiva una muestra que es negativa [25].
<i>Precision</i>	$Precision = tp / (tp + fp)$ tp Número de verdaderos positivos fp Número de falsos positivos
<i>Accuracy</i>	Muestra la fracción o recuento de las predicciones que fueron correctas; por ejemplo, en clasificación la función de precisión devuelve 1,0 si todo el conjunto de etiquetas predichas coincide con el verdadero conjunto de etiquetas, de lo contrario ese valor puede bajar hasta cero, lo que significa que no hubo coincidencia alguna [26].
<i>F1-Score</i>	$Accuracy = cp / tp$ cp Número de predicciones correctas tp Número total de predicciones Se interpreta como un promedio ponderado entre la precisión y la exhaustividad, en la que F1-Score alcanza su mejor valor en 1 y su peor valor en 0 [27]. $F1 = 2 * (precision * recall) / (precision + recall)$ Se define como la capacidad que tiene el modelo para identificar todas las etiquetas positivas en un conjunto de datos, en el que el mejor valor es 1 y el peor valor es cero [28].
<i>Recall</i>	$Recall = tp / (tp + fn)$ tp Número de verdaderos positivos fn Número de falsos negativos

2.5 Despliegue

El modelo se implementó en una plataforma web a través de un servidor en la nube, con el objetivo de que el modelo predictivo sea accesible y utilizable para los usuarios finales. Se aseguró que, mediante las entradas del modelo proporcionadas por el usuario a través de un formulario, el modelo haga predicciones precisas y brinde información detallada. Esto les permite a los usuarios comprender y concientizarse sobre el riesgo de consumo de SPI. Además, se incluyeron herramientas para visualizar los datos mediante gráficas estadísticas, que facilitan la comparación y comprensión de las diferentes variables utilizadas en el modelo. Se complementó la plataforma con un mapa demográfico de Colombia, que ayuda a entender la distribución general de la población en riesgo de consumir SPI en regiones o áreas específicas.

3. RESULTADOS

En el análisis de datos, la comparación del rendimiento de distintos modelos de clasificación es muy importante para determinar cuál es el modelo más adecuado para garantizar el éxito en tareas de clasificación. Por ello, en esta investigación se evaluaron los algoritmos SVM, KNN, RF, XGBoost, *decision tree* y *logistic regression*. Además, utilizando las métricas de desempeño exactitud (*Accuracy*), precisión

(Precisión), recuperación (Recall) y la puntuación F1 (F1-score), se obtuvo una perspectiva más clara sobre cuál de estos modelos se debe seleccionar. Los datos presentados en la [Tabla 4](#) se obtuvieron del test_score, y reflejan el rendimiento de cada modelo en el conjunto de datos de prueba.

Tabla 4. Métricas de desempeño sin optimizar

Modelo	Accuracy	Precision	Recall	F1-Score
SVM	0,984100	0,978580	0,976901	0,977038
KNN	0,909500	0,893786	0,872923	0,877132
RF	0,938900	0,940112	0,940700	0,940600
XGBoost	0,955300	0,941037	0,941058	0,939251
Decision tree	0,924300	0,924034	0,923900	0,921167
Logistic regression	0,980400	0,981103	0,980400	0,980343

Los datos obtenidos muestran que los modelos de clasificación más avanzados, como SVM, XGBoost y *logistic regression*, tienen un mejor desempeño en comparación con métodos más simples como KNN y *decision tree*. En cuanto a los algoritmos que destacan por su alta precisión y F1-Score, se encuentran *logistic regression* y SVM, esto indica que tienen una excelente capacidad para clasificar correctamente las instancias.

Otro aspecto importante es optimizar los hiperparámetros para mejorar el rendimiento de los modelos, lo que se logra a través de una correcta elección de hiperparámetros y técnicas como GridSearchCV y RandomizedSearchCV. Los datos presentados en la [Tabla 5](#) se obtuvieron del test_score y reflejan el rendimiento de cada modelo optimizado en el conjunto de datos de prueba.

Tabla 5. Métricas de desempeño optimizadas

Modelo	Accuracy	Precision	Recall	F1-Score
SVM	0,985500	0,978681	0,976905	0,977154
KNN	0,913600	0,900797	0,881983	0,885219
RF	0,946000	0,948153	0,946700	0,945900
XGBoost	0,967300	0,953189	0,953022	0,951497
Decision tree	0,934900	0,937064	0,936500	0,933783
Logistic regression	0,986500	0,987063	0,986500	0,986478

Con base en las métricas obtenidas después de la optimización de hiperparámetros, se puede observar que el modelo SVM muestra un rendimiento muy alto en todas sus métricas. XGBoost mejora significativamente sus métricas después de la optimización. Pero —con base en el análisis exhaustivo de las métricas, datos obtenidos y después del proceso de optimización de parámetros mediante las técnicas RandomizedSearch y GridSearch— se decidió seleccionar *logistic regression* como el mejor modelo para esta tarea de clasificación, pues su desempeño es superior en todas las métricas evaluadas y garantiza una excelente capacidad predictiva, lo que lo hace ideal para el problema abordado en este estudio.

Después de seleccionar el modelo que obtuvo mejor resultados en las métricas frente a los otros modelos evaluados, y después de optimizar los hiperparámetros para mejorar su rendimiento, se generaron diversas gráficas. Estas gráficas ayudan a entender el comportamiento del modelo, así como su eficacia y precisión al momento de realizar nuevas predicciones.

Es importante aclarar que, al tener datos desbalanceados en el conjunto de datos, la precisión para cada clase se puede ver afectada con el aumento o disminución según la cantidad de datos que existan para cada clase. Por ejemplo, en la clase 4 se puede evidenciar una notable diferencia en la precisión respecto a las otras clases; en esta clase el desbalance en los datos es notoria.

Por su parte, se presenta la importancia de las características para cada clase de la variable objetivo. Estos datos fueron obtenidos a partir de la propiedad `coef` del modelo de regresión logística, que calcula la importancia de cada característica según los parámetros configurados para el modelo. Cada clase puede verse influenciada de manera distinta por estas características, lo que se refleja en las precisiones obtenidas.

En la [Figura 5](#) se presenta la importancia de las características para la clase 0 (riesgo muy bajo) y 1 (riesgo bajo) de la variable objetivo. En la [Tabla 6](#) se pueden observar las características más influyentes y menos influyentes para las clases evaluadas en la [Figura 3](#).

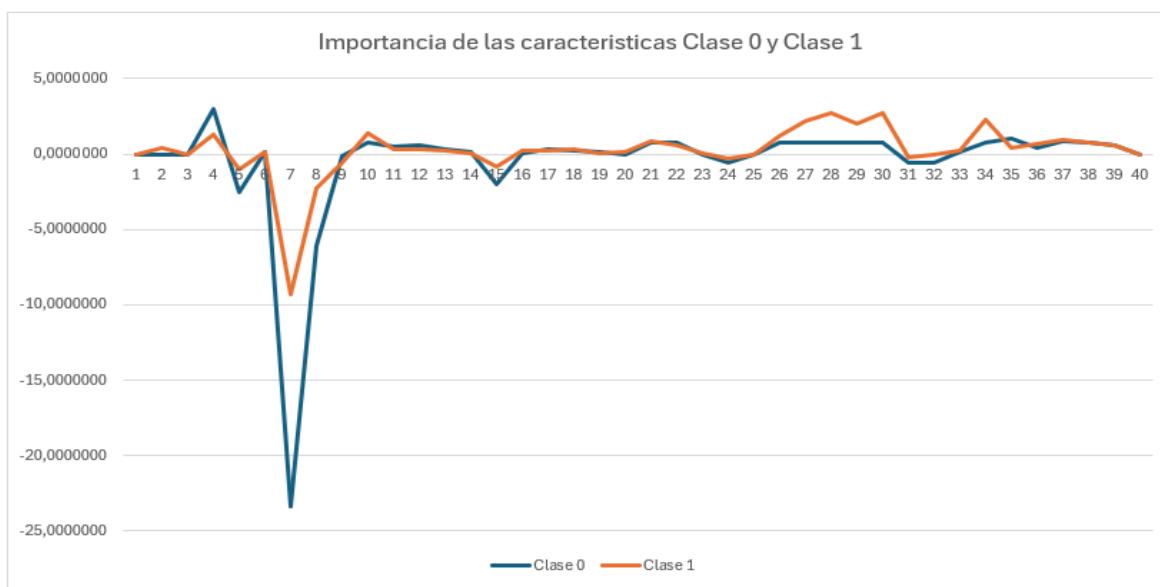


Figura 5. Importancia de las características para la clase 0 (riesgo muy bajo) y clase 1 (riesgo bajo)

En la [Tabla 6](#), se puede observar que tanto para la clase 0 como para la clase 1 de la variable objetivo las características con menor influencia son iguales para las dos clases; sin embargo, las clases con mayor influencia difieren.

Tabla 6. Características con mayor y menor influencia para la clase 0 y 1

Clase	Características más influyentes	Características menos influyentes
0 (Riesgo muy bajo)	5. per_edad_tipo	8. frecuencia_consumo_marihuana_tipo
	35.g_01_a_num_familiares_consumen_sp_imp_tipo	9. frecuencia_consumo_cocaina_tipo
	37.g_08_a_ofrecieron_marihuana_imp_tipo	6. estrato_tipo
	38.g_08_b_ofrecieron_cocaina_imp_tipo	15. d_08_estado_salud_tipo
	28.g_03_curiosidad_probar_sp_tipo	8. frecuencia_consumo_marihuana_tipo
1 (Riesgo bajo)	30.g_05_posibilidad_probar_sp_tipo	9. frecuencia_consumo_cocaina_tipo
	34.g_07_alguien_ofrecio_comprar_probar_sp_tipo	6. estrato_tipo
	27.g_02_amigos_consumen_sp_tipo	15. d_08_estado_salud_tipo

En la [Figura 6](#), se presenta la importancia de las características para la clase 2 (riesgo medio) de la variable objetivo. En la [Tabla 7](#) se presentan las características más influyentes y menos influyentes para la clase 2.

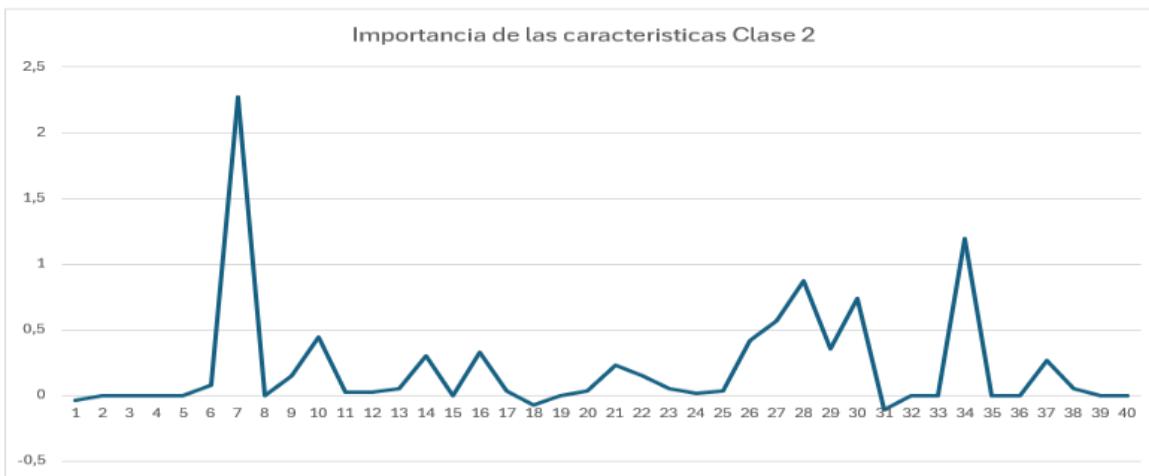


Figura 6. Importancia de las características para la clase 2 (riesgo medio).

Tabla 7. Características con mayor y menor influencia para la clase 2

Clase	Características más influyentes	Características menos influyentes
2 (Riesgo medio)	8. frecuencia_consumo_marihuana_tipo 34.g_07_alguien_ofrecio_comprar_probar_sp_tipo 28.g_03_curirosidad_probar_sp_tipo	31.g_06_a_posibilidad_conseguir_marihuana_tipo 18.d_11_h_conocimiento_riesgo_fumar_marihuana_frecuentemente_tipo 1. departamento

Como se mencionó, las características para cada clase difieren, pues para esta clase (clase 2) las características con menor influencia son diferentes a las de la clase 0 y 1.

Respecto a la [Figura 7](#), se presenta la importancia de las características para las clases 3 y 4 de la variable objetivo; además, cabe destacar que las características más influyentes y menos influyentes se presentan en la [Tabla 8](#), en la que se logra evidenciar la diferencia entre las características para las clases mencionadas.

Por otra parte, para usar el modelo final propuesto en conjunto con los datos procesados, se desarrolló una aplicación web (<https://preventapp-89091.web.app>). Esta aplicación consta de diversas funcionalidades, como herramientas para graficar y comparar los datos entre las diferentes variables del conjunto de datos, que incluyen gráficos de barras, gráficos de pastel y un mapa de calor por departamentos, entre otras opciones.

Destaca especialmente la sección “Formularios > Consulte su nivel de riesgo”, en la que los usuarios pueden predecir su nivel de riesgo de consumo o adicción a sustancias psicoactivas mediante un formulario interactivo. Esta funcionalidad principal proporciona una herramienta útil para que las personas evalúen y comprendan su propio riesgo en relación con el consumo de SPI.

En la [Figura 8](#) se presentan las preguntas relacionadas con el conocimiento y el consumo de sustancias psicoactivas, así como el estado de salud de las personas. Estas preguntas son claves para predecir el riesgo de consumo o adicción de sustancias psicoactivas ilegales.

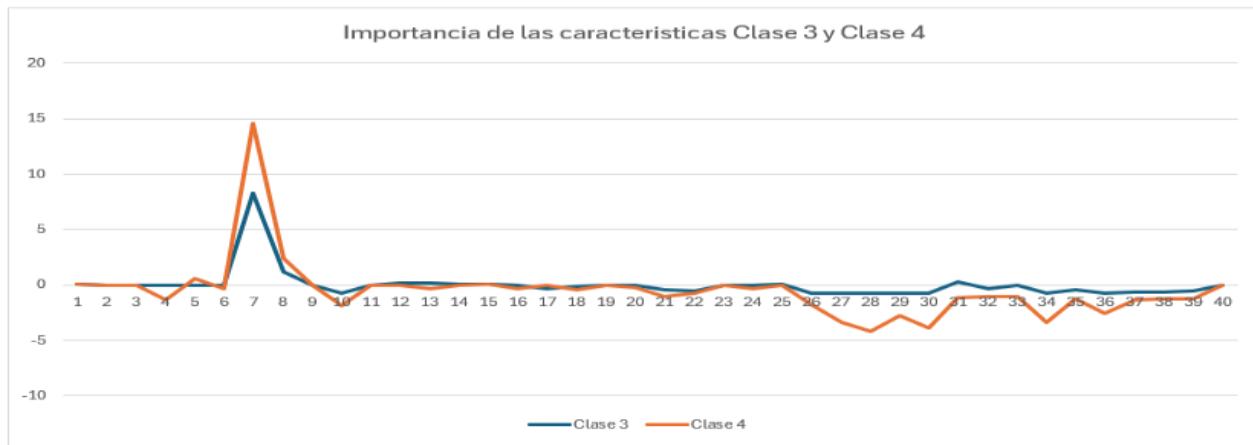


Figura 7. Importancia de las características para la clase 3 (riesgo alto) y la clase 4 (riesgo muy alto)

Tabla 8. Características con mayor y menor influencia para las clases 3 y 4

Clase	Características más influyentes	Características menos influyentes
3 (Riesgo alto)	8. frecuencia_consumo_marihuana_tipo	36.g_02_a_num_amigos_consumen_sp_imp_tipo
	9. frecuencia_consumo_cocaina_tipo	27. g_02_amigos_consumen_sp_tipo
	31.g_06_a_posibilidad_conseguir_marihuana_tipo	28. g_03_curirosidad_probar_sp_tipo
	13. vivienda_tipo	30. g_05_posibilidad_probar_sp_tipo
4 (Riesgo muy alto)	8. frecuencia_consumo_marihuana_tipo	28. g_03_curirosidad_probar_sp_tipo
	9. frecuencia_consumo_cocaina_tipo	30. g_05_posibilidad_probar_sp_tipo
	6. estrato_tipo	27. g_02_amigos_consumen_sp_tipo
	15. d_08_estado_salud_tipo	34.g_07_alguien_ofrecio_comprar_probar_sp_tipo

Figura 8. Preguntas de información sobre consumo para el formulario de la aplicación web

Una vez que el formulario ha sido completado correctamente, se generará una predicción sobre el nivel de riesgo de la persona. En este caso, se presenta el nivel de riesgo que predijo el modelo, acompañado de una breve explicación (Figura 9) sobre el significado de ese nivel. Además, se proporcionan detalles

sobre las características que más influyeron en la obtención de este nivel de riesgo, junto con una gráfica interactiva ([Figura 10](#)) que presenta las diferentes características que tuvieron mayor influencia en el resultado de la predicción.



Figura 9. Resultado de la predicción

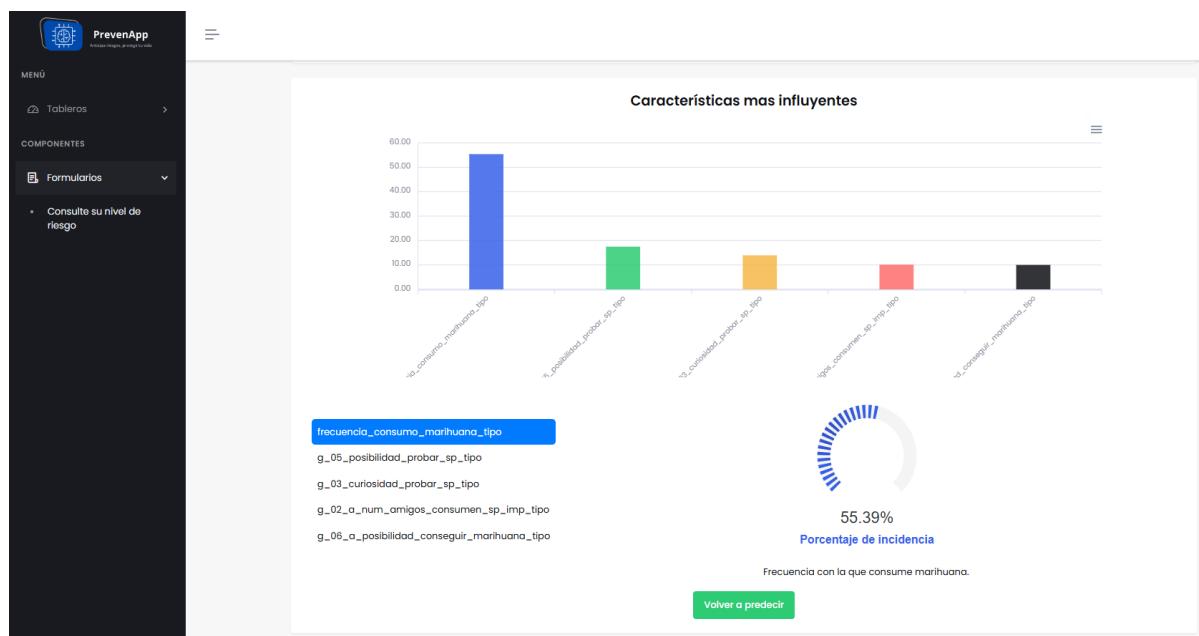


Figura 10. Importancia de las características de la predicción

4. CONCLUSIONES

El desarrollo y evaluación de un modelo de ML para predecir el nivel de riesgo de consumo de SPI en la población colombiana ha demostrado ser una herramienta valiosa para la toma de decisiones en programas de prevención de consumo y adicciones. Los objetivos específicos se centraron en definir un modelo predictivo, construir una herramienta web para visualizar las predicciones y evaluar el rendimiento del modelo a través de diversas métricas.

Los factores de riesgo a nivel individual, familiar y social contribuyen en gran medida a la probabilidad de consumir estas sustancias. Para abordar este problema, se han implementado programas de prevención como “Mas mente, más prevención” o el “Plan Nacional de Promoción de la Salud, Prevención y Atención al Consumo de Sustancias Psicoactivas 2014-2021”, pero no cuentan con herramientas tecnológicas avanzadas para analizar los datos de manera rápida y efectiva [2]. El modelo desarrollado en esta investigación puede servir como una base crucial para la toma de decisiones de las entidades gubernamentales, entidades de salud y sectores que aborden este problema, ya que actualmente no existe un programa que brinde un apoyo tan integral y avanzado en este ámbito. Esta herramienta permitirá una evaluación más precisa y ágil del riesgo, y facilitará la implementación de intervenciones más eficaces y personalizadas.

La metodología aplicada incluyó la selección del conjunto de datos adecuado, el procesamiento y reducción de la dimensionalidad, el entrenamiento y la evaluación de los modelos, la optimización de hiperparámetros y el despliegue de una herramienta web. Para los modelos se decidió utilizar algoritmos como SVM, KNN, RF, XGBoost, *decision tree* y *logistic regression*; el rendimiento de estos algoritmos se evaluó a través de métricas de desempeño como Accuracy, Precision, Recall y F1-Score.

Los resultados indicaron que modelos sofisticados como SVM y *logistic regression* tienen un mejor desempeño en comparación con métodos más simples. En el caso de la optimización de hiperparámetros, *logistic regression* fue el que obtuvo mejor rendimiento con una exactitud del 98,65 %, recuperación del 98,70 % y F1-score del 98,64 %. Este modelo se implementó en la plataforma web para hacer las nuevas predicciones y poder visualizar los datos relevantes para garantizar una mejor comprensión del riesgo de consumo de SPI.

La principal limitación de este estudio fue la presencia de clases desbalanceadas en el conjunto de datos, ya que esto afectó directamente la precisión del modelo, lo que requirió el uso de técnicas adicionales para solucionar dicho problema. Además, aunque se seleccionó un modelo final basado en su rendimiento óptimo, es importante aclarar que estos modelos no son perfectos, por lo que se pueden presentar predicciones en las que el modelo no generalice bien nuevos datos que se ingresen, lo que da como resultado predicciones inexactas en algunos casos.

En cuanto a las recomendaciones para futuros trabajos, estos podrían enfocarse en la obtención y utilización de conjuntos de datos más balanceados para aumentar la precisión del modelo. Además, se sugiere profundizar en otros algoritmos de aprendizaje automático como redes neuronales profundas (CNN) que pueden capturar patrones más complejos en los datos o modelos ensamblados que permitan combinar múltiples modelos y mejorar tanto la predicción como la robustez del modelo final. Otra sugerencia para tener en cuenta es emplear técnicas más avanzadas de optimización de hiperparámetros que permitan continuar mejorando el rendimiento predictivo y ser más eficientes que GridSearch y RandomizedSearch. Por último, este modelo se podría extender para predecir el riesgo de otras SPI sin limitarse únicamente a marihuana, bazuco y cocaína, o incluir la predicción de múltiples adicciones simultáneamente para dar una respuesta de predicción más detallada.

REFERENCIAS

- [1] Ministerio de Justicia y Derecho, *Situación del Consumo en Colombia*, 2024.
<https://www.minjusticia.gov.co/programas-co/ODC/Paginas/Situaci%C3%B3n-consumo.aspx>
- [2] Ministerio de Salud, *Prevención del consumo de Sustancias Psico-Activas (SPA)* 2024. <https://www.minsalud.gov.co/salud/publica/SMental/Paginas/convivencia-desarrollo-humano-sustancias-psicoactivas.aspx>
- [3] W. Ruiz Orejuela *et al.*, *Estudio nacional de consumo de sustancias psicoactivas Colombia*, 2019.
- [4] H. Gong, C. Xie, C. Yu, N. Sun, H. Lu, Y. Xie, "Psychosocial factors predict the level of substance craving of people with drug addiction: A machine learning approach," *International Journal of Environmental Research and Public Health*, vol. 18, no. 22, e12175, Nov. 2021.
<https://doi.org/10.3390/ijerph182212175>
- [5] U. I. Islam, I. H. Sarker, E. Haque, M. M. Hoque, "Predicting Individual Substance Abuse Vulnerability using Machine Learning Techniques," Arxiv, Dec. 2020. <http://arxiv.org/abs/2101.03184>
- [6] M. A. I. Arif, S. I. Sany, F. Sharmin, M. S. Rahman, M. T. Habib, "Prediction of addiction to drugs and alcohol using machine learning: A case study on Bangladeshi population," *International Journal of Electrical and Computer Engineering*, vol. 11, no. 5, pp. 4471-4480, Oct. 2021.
<https://doi.org/10.11591/ijece.v11i5.pp4471-4480>
- [7] U. I. Islam, E. Haque, D. Alsalmán, M. N. Islam, M. A. Moni, I. H. Sarker, "A Machine Learning Model for Predicting Individual Substance Abuse with Associated Risk-Factors," *Annals of Data Science*, vol. 10, no. 6, pp. 1607–1634, 2022. <https://doi.org/10.1007/s40745-022-00381-0>
- [8] J. P. Davis *et al.*, "Identifying individual and environmental predictors of opioid and psychostimulant use among adolescents and young adults following outpatient treatment," *Drug Alcohol Depend*, vol. 233, e109359, Apr. 2022. <https://doi.org/10.1016/j.drugalcdep.2022.109359>
- [9] D. H. Han, D. C. Seo, "Identifying risk profiles for marijuana vaping among U.S. young adults by recreational marijuana legalization status: A machine learning approach," *Drug Alcohol Depend*, vol. 232, e109330, Mar. 2022. <https://doi.org/10.1016/j.drugalcdep.2022.109330>
- [10] L. Zoboroski, T. Wagner, B. Langhals, "Classical and neural network machine learning to determine the risk of marijuana use," *International Journal of Environmental Research and Public Health*, vol. 18, no. 14, e7466, Jul. 2021. <https://doi.org/10.3390/ijerph18147466>
- [11] S. Negriff, B. Dilkin, L. Matai, E. Rice, "Using machine learning to determine the shared and unique risk factors for marijuana use among child-welfare versus community adolescents," *PLoS One*, vol. 17, no. 9, e0274998, Sep. 2022. <https://doi.org/10.1371/journal.pone.0274998>
- [12] K. Palomino, C. R. Berdugo, J. I. Vélez, "Leading consumption patterns of psychoactive substances in Colombia: A deep neural network-based clustering-oriented embedding approach," *PLoS One*, vol. 18, no. 8, e0290098, Aug. 2023. <https://doi.org/10.1371/journal.pone.0290098>
- [13] H. Ordoñez, R. Timarán-Pereira, J.-S. González-Sanabria, "Trends in Addiction to Psychoactive Substances Among Homeless People in Colombia Using Artificial Intelligence," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 9, n.º 3, pp. 52-60 2024.
<https://doi.org/10.9781/ijimai.2024.02.002>

- [14] J. J. Pinto Hidalgo, J. A. Silva Centeno, "Geospatial Intelligence and Artificial Intelligence for Detecting Potential Coca Paste Production Infrastructure in the Border Region of Venezuela and Colombia," *Journal of Applied Security Research*, vol. 18, no. 4, pp. 1000-1050, 2023.
<https://doi.org/10.1080/19361610.2022.2111184>
- [15] C. Brito Carrillo, "Diagnóstico prevención integral y evaluación de factores individuales, familiares y sociales asociados al uso y abuso de sustancias," 2020, <https://www.minjusticia.gov.co/programas-co/ODC/Publicaciones/Publicaciones/DIAGNOSTICO%20DE%20LA%20PROBLEMATICA%20DEL%20CONSUMO%20DE%20SUSTANCIAS%20PSICOACTIVAS%20DEPARTAMENTO%20DE%20LA%20GUAJIRA.pdf>
- [16] C. J. Brito Carrillo, E. E. Pinto Aragon, "Comprehensive attention in vulnerable sectors victims of the use and abuse of psychoactive substances in the juvenile penal responsibility processes law 1098 of 2006, in the department of La Guajira. Colombia," *Cultura, Educación y Sociedad*, vol. 7, no. 1, pp. 104-116, Jan. 2016. <https://doi.org/10.17981/cultedusoc.07.1.2016.8>
- [17] Scikit-Learn, *Grid Search CV*, 2024.
https://scikit-learn.org/0.18/modules/generated/sklearn.model_selection.GridSearchCV.html
- [18] Scikit-Learn, *Randomized Search CV*, 2024. https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.RandomizedSearchCV.html
- [19] IBM, *Funcionamiento de SVM*, 2024.
<https://www.ibm.com/docs/es/spss-modeler/saas?topic=models-how-svm-works>
- [20] IBM, *¿Qué es KNN?*, 2024. <https://www.ibm.com/mx-es/topics/knn>
- [21] IBM, *¿Qué es un bosque aleatorio?*, 2024. <https://www.ibm.com/mx-es/topics/random-forest>
- [22] IBM, *¿Qué es boosting?*, 2024. <https://www.ibm.com/es-es/topics/boosting>
- [23] IBM, *¿Qué es la regresión logística?*, 2024. <https://www.ibm.com/mx-es/topics/logistic-regression>
- [24] IBM, *¿Qué es un árbol de decisión?*, 2024. <https://www.ibm.com/es-es/topics/decision-trees>
- [25] Scikit-Learn, *Precision Score*, 2024.
https://scikit-learn.org/stable/modules/generated/sklearn.metrics.precision_score.html
- [26] Scikit-Learn, *Metrics and scoring: quantifying the quality of predictions*, 2024.
https://scikit-learn.org/stable/modules/model_evaluation.html#accuracy-score
- [27] Scikit-Learn, *F1 Score*, 2024.
https://scikit-learn.org/0.15/modules/generated/sklearn.metrics.f1_score.html
- [28] Scikit-Learn, *Recall Score*, 2024.
https://scikit-learn.org/stable/modules/generated/sklearn.metrics.recall_score.html