

# Modelos predictivos para la estimación de adolescentes con tendencia al alcoholismo

Fausto Salazar-Fierro<sup>1-2\*</sup>, Jorge Castañeda<sup>1-3</sup>, Marco Revelo-Aldás<sup>1-4-5</sup>

<sup>1</sup>Universidad Nacional Mayor San Marcos. Ecuador

<sup>2</sup>Universidad Técnica del Norte. Ecuador

<sup>3</sup>Universidad de Ciencias y Humanidades. Ecuador

<sup>4</sup>Instituto Técnico Superior Ibarra. Ecuador

<sup>5</sup>Universidad de Investigación de Tecnología Experimental Yachay. Ecuador

\*Autor para correspondencia: fausto.salazar@unmsm.edu.pe

Recibido: 2022/01/02

Aprobado: 2022/04/11

DOI: <https://doi.org/10.26621/ra.v1i26.779>

## RESUMEN

Una de las drogas más consumidas a nivel mundial, es sin lugar a duda el alcohol. Según algunas evidencias, los jóvenes suelen tomar contacto con el alcohol entre los 12 y 17 años, esto ha llevado a la realización de diferentes investigaciones con el fin de entender los patrones que pueden condicionar en los jóvenes el consumo de alcohol. El presente artículo tiene como objetivo analizar tres diferentes modelos predictivos basados en Machine Learning, con el fin de entender cuál de los modelos analizados responden de mejor manera al estudio de la predicción de la tendencia al alcoholismo en jóvenes. Para la realización del análisis, se ha tomado como base un dataset de 521 registros, obtenido de Kagle como modelo, el cual fue sometido al análisis de los tres modelos. Según las pruebas realizadas con los modelos predictivos, el modelo de Regresión Lineal posee una mayor precisión con un accuracy de 1.00 frente al 0.95 del modelo KNN y 0.98 del Árbol de Decisión. El estudio nos determina en las curvas ROC analizadas que el modelo de regresión lineal alcanza mejores resultados entre la sensibilidad de los verdaderos positivos y la especificidad de los falsos positivos. Por otro lado, debemos mencionar que según el dataset analizado, se tendría que los indicadores de predicción son: la zona donde vive el adolescente, el estado familiar en el que crece y la disposición de tiempo libre. Si bien el estudio no pretende ser concluyente, refleja la importancia de reconocer los factores psicosociales protectores en el diseño e implementación de programas de promoción y prevención asociados al consumo responsable de alcohol y de conducta no violenta con adolescentes desde una perspectiva salutogénica.

**Palabras clave:** Adolescencia, factores psicosociales, predicción, alcoholismo, aprendizaje automático

## ABSTRACT

One of the most consumed drugs worldwide is, without a doubt, alcohol. According to some evidence, young people usually come into contact with alcohol between the ages of 12 and 17, this has led to different investigations in order to understand what patterns can condition alcohol consumption in young people. The objective of this article is to analyze three different predictive models based on Machine Learning, in order to understand which of the analyzed models respond in the best way to the study of the prediction of the tendency to alcoholism in young people. To carry out the analysis, a data set of 521 records has been taken as a base, obtained from Kagle as a model, which was subjected to the analysis of the three models. According to the tests carried out with the predictive models, the Linear Regression model has greater precision with an accuracy of 1.00 compared to 0.95 for the KNN model and 0.98 for the Decision Tree. The study determines in the ROC curves analyzed that the linear regression model achieves better results between the sensitivity of true positives and the specificity of false positives. On the other hand, we must mention that according to the data set analyzed, the predictive indicators are the area where the adolescent lives, the family status in which he grows up, and the availability of free time. Although the study does not claim to be conclusive, it reflects the importance of recognizing protective psychosocial factors in the design and implementation of promotion and prevention programs associated with responsible alcohol consumption and non-violent behavior with adolescents from a salutogenic perspective.

**Keywords:** Adolescence, psychosocial factors, prediction, alcoholism, Machine Learning

Fausto A Salazar-Fierro  [orcid.org/0000-0003-1558-6721](https://orcid.org/0000-0003-1558-6721)

Jorge E Castañeda-Albán  [orcid.org/0000-0001-7725-8702](https://orcid.org/0000-0001-7725-8702)

Marco D Revelo-Aldás  [orcid.org/0000-0001-9357-8861](https://orcid.org/0000-0001-9357-8861)



## INTRODUCCIÓN

La droga más consumida en el mundo es el alcohol, que constituye la toxicidad de mayor relevancia (De Las et al., s.f.), que generalmente se utiliza como brebaje, se considera como una enfermedad que se puede adquirir por varias vías causando dependencia (Gómez et al., 2019).

El alcoholismo define como una enfermedad crónica y progresiva producida por el abuso de consumo del alcohol que van ámbitos personales con secuelas de deterioro de salud, familiar, relaciones interpersonales entre otros, teniendo como mayores consumidores a los países más desarrollados donde ha ocurrido un cambio de patrón como es el caso de Europa donde el 70% admite haber ingerido alcohol (Méndez et al., 2018; Páez et al., 2021).

La adolescencia es una etapa de transición entre la infancia y la adultez entre los 10 y 19 años siendo una etapa muy difícil para el adolescente (Borrás, 2014). En donde tienen muchos problemas para manejar muchos cambios pudiendo desarrollar conductas de riesgo como el consumo de alcohol, drogas entre los 15 a 17 años por el interés de experimentar nuevos comportamientos debido a la inexperiencia del individuo (Ahumada et al., 2017).

El alcoholismo se ha convertido en el tercer factor de muerte prematura y de discapacidad con un 4% a nivel mundial, y el consumo excesivo de alcohol en jóvenes puede traer algunas consecuencias como relaciones sexuales riesgosas, accidentes de tránsito, enfermedades crónicas graves, entre otras (Martínez et al., 2017).

Entre las bebidas alcohólicas más conocidas en primer lugar se ubica la cerveza, seguida de los destilados, lastimosamente las políticas públicas de salubridad en América Latina son insuficientes para responder a este problema de salud mundial (Monteiro, 2013).

Las redes neuronales son redes formadas por nodos o neuronas interconectadas en varios tipos de orden jerárquico, el modelo de predicción de red neuronal se establecerá de acuerdo con la influencia de los factores (Hu et al., 2014).

La adolescencia es una etapa del ciclo vital especialmente vulnerable a la iniciación al consumo de alcohol puesto que muchos jóvenes no se muestran preocupados por la prevención de enfermedades en el futuro, sino que se plantean sobre todo vivir el presente (Mira et al., 2017).

Asimismo, puede interferir en otros aspectos cruciales como el rendimiento académico o la propensión a realizar comportamientos de riesgo tales como mantener relaciones sexuales desprotegidas, la conducción temeraria de vehículos o las conductas violentas o vandálicas (Motos et al., 2016).

Factores como el tabaco y el alcohol pueden actuar como facilitadores para una posterior experimentación con marihuana y una posible progresión hacia al consumo de otras drogas ilegales, tal y como propone Kandel (1975) en su modelo evolutivo.

De acuerdo con este modelo el uso de una droga legal es condición necesaria, pero no suficiente para el avance a una etapa posterior (consumo de otras drogas ilegales). No obstante, Kandel (1980) sugiere que la influencia social de la familia y los iguales tiene mayor relevancia en el consumo en las primeras etapas, mientras que los factores individuales pesan más en las últimas (Avellaneda et al., 2010).

Uno de los principales propósitos del análisis estadístico es hacer predicciones y proporcionar medidas adecuadas de la incertidumbre asociada con ellas.

Por lo tanto, los pronósticos deben ser de naturaleza probabilística, tomando la forma de distribuciones de probabilidad sobre cantidades y eventos futuros.

En recientes estudios psiquiátricos se aun utilizados modelos de aprendizaje autónomo para predecir trastornos psiquiátricos con mucha exactitud lo que permite elaborar sistemas para tomar decisiones clínicas identificando variables influyentes (Jin Park et al., s.f.).

En la medicina también se han realizado trabajos como es la predicción de tuberculosis en adultos usando modelos de predicción empleando datos que fueron sintetizados con revisión narrativa con tablas y cifras (Peetluk et al., 2021).

Actualmente la aplicación de redes neuronales en un enfoque paramétrico para el pronóstico usando algoritmos para mejorar la demanda de diversos productos (Herrera-granda & Chicaiza-ipeales, s.f.).

El uso de técnicas de minería de datos bayesiana y árboles de decisión permiten identificar patrones como etnias, género, lugar de residencia, etc. (Mendoza, 2020).

Para la presente investigación, se plantea la siguiente pregunta:

¿Cuáles son los factores que inciden para que un adolescente tenga tendencia al consumo de alcohol?

Los datos se descargaron de un repositorio de dataset Kaggle que se considera que contiene las variables necesarias para poder realizar la predicción de la pregunta planteada.

## MÉTODOS

El dataset analizado es un archivo de texto separado por comas (CSV) con un total de 521 registros y 17 variables, que tiene la siguiente estructura.

**Tabla 1.** Tipos de datos de las variables

Variable	Alias	Tipo de dato
Edad	Edad	Entero
Género	Género	Cadena
Urbano Rural	Urb_Rur	Booleano
Estado Familiar	Fam_Estado	Booleano
Es Romántico	Romantico	Booleano
La madre trabaja	Madre_Trabaja	Booleano
El padre trabaja	Padre_Trabaja	Booleano
Tiene tiempo libre	T_Libre	Booleano
Salidas sociales	Salidas	Booleano
Estado abstinencia	Abstinencia	Booleano
Es irritable	Irritabilidad	Booleano
Tiene tutor	Tutor	Booleano
Escuela privada	Sec_Privada	Booleano
Es sedentario	Sedentario	Booleano
Posee internet	Internet	Booleano
Tiene depresión	Depresión	Booleano
Resultado	class	Booleano

Fuente: <https://www.kaggle.com/>

Para el modelo de predicción se realizaron las siguientes condiciones:

Si la edad es menor a 15 años se minimiza el consumo de alcohol. Si es mayor a 15 años, no tiene abstinencia, el lugar de residencia es urbano, es romántico, tiene tiempo libre y presenta cuadros depresivos, los modelos de predicción nos muestran que existe una tendencia a consumir alcohol; siendo estas variables con mayor relevancia en el análisis.

Adicionalmente, si el hogar es disfuncional, tiene salidas constantes, los padres trabajan, estudia en una escuela privada, no realiza ejercicio frecuentemente, posee internet y no tiene ninguna persona que se responsabilice de sus acciones, los modelos nos muestran valores de incidencia al consumo de licor; tomando en cuenta que estas variables son de menor relevancia para el modelo.

Otras combinaciones que se aplicaron como el género, la irritabilidad de las personas, no tuvieron relevancia en los modelos de acuerdo con las condiciones aplicadas.

## RESULTADOS Y DISCUSIÓN

Las pruebas se realizaron con el software en línea Colab de Google y su lenguaje de programación Python, usando instrucciones que muestran el resultado del factor K del vecino más próximo (KNN), regresión lineal y árbol de decisión, mediante el dataset.

Para los tres modelos se consideró el método de partición de datos, donde los modelos se entrenaron con el 30% de la muestra y se probó la precisión de los modelos con el restante 70%.

Se realizó un balance de datos, tomándose en cuenta la variable de salida (class), permitiendo que se nivelen, el balanceo se realizó con el fin de estimar su distribución en la población, antes de la tabulación de las otras variables.

Una vez realizado el tratamiento de datos se obtuvo una media de edad de 17 años, con una desviación estándar de 1.95.

La distribución de los cuartiles y sus datos de edad se muestran a continuación en la figura 1.

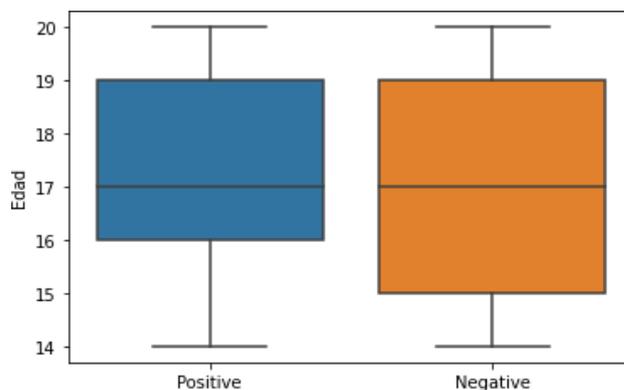


Figura 1. Caja y bigotes variable edad

Para el tratamiento de datos se ha ejecutado el proceso para cambiar los valores nominales una categorización 0 para negativos (200) y 1 para los positivos (320) de la variable class, que determina el valor de la predicción. La matriz de confusión resultante del modelo predictivo KNN, se muestra en la figura 2.

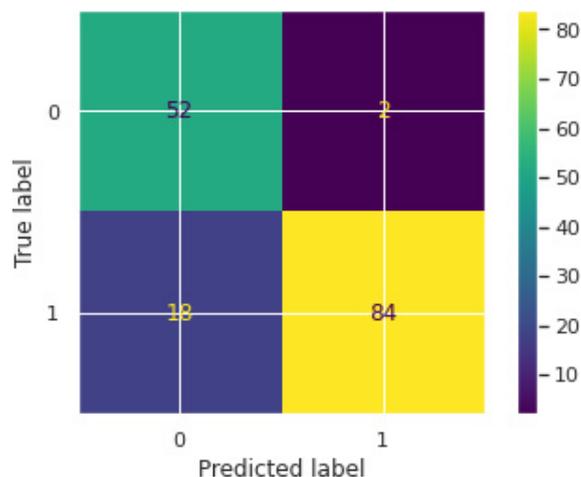


Figura 2. Matriz de confusión modelo KNN

El índice de Gini mide el grado de pureza de un nodo, mide la probabilidad de no sacar dos registros de la misma clase del nodo. A mayor índice de Gini menor pureza, por lo que se selecciona la variable con menor Gini ponderado con la que empieza nuestro modelo de árbol de decisión como se despliega en la figura 3.

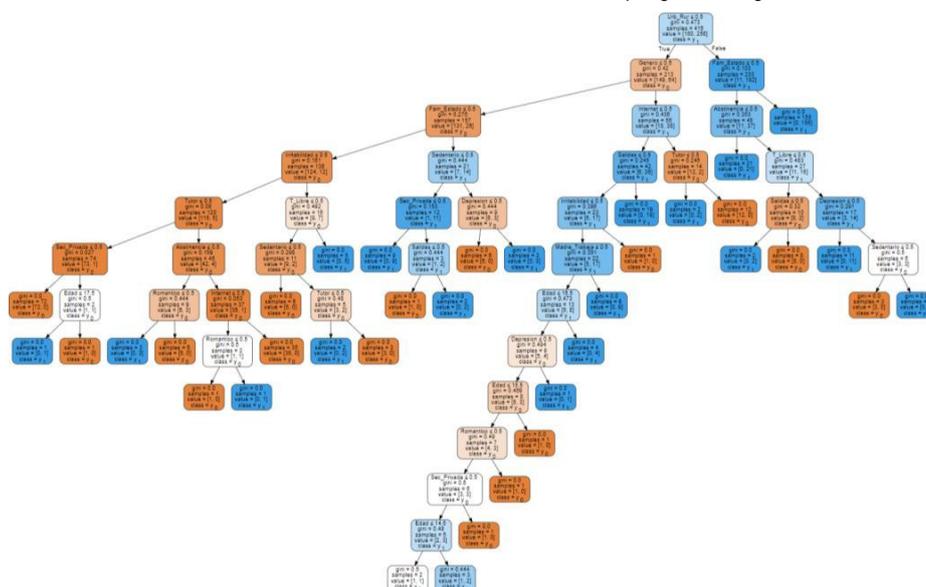


Figura 3. Modelo Árbol de decisión.

Para el modelo de regresión lineal se obtuvieron los valores de correlación entre las variables del dataset como se muestra en la Tabla 2.

**Tabla 2.** Valores de matriz de correlación

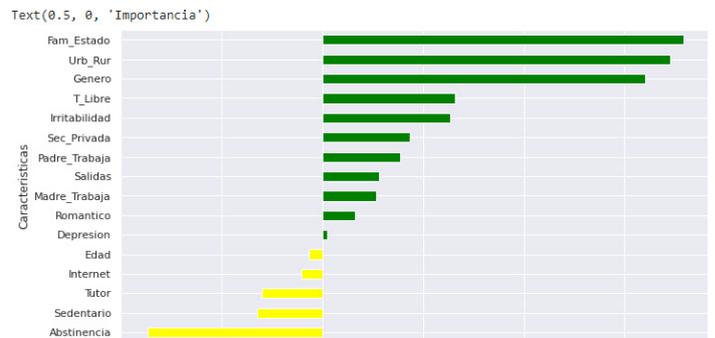
	Edad	Genero	Urb_Rur	Funestado	Romantico	M_Trabajo	P_Trabajo	T_Libre	Salidas	Abstinencia	Irritabilidad	Tutor	Sec_Privada	Sedentario	Internet	Depresión	Clases
Edad	1	-0,037	0,055	0,114	-0,026	0,118	0,031	-0,028	0,138	0,117	0,102	0,064	0,11	0,123	0,036	0,073	0,034
Genero	-0,037	1	0,269	0,312	0,282	0,124	0,22	-0,209	0,208	0,052	0,014	0,102	0,332	0,091	-0,328	0,005	0,449
Urb_Rur	0,055	0,269	1	0,599	0,447	0,263	0,374	0,087	0,235	0,088	0,238	0,15	0,442	0,153	-0,144	0,127	0,666
Fam_Estado	0,114	0,312	0,599	1	0,406	0,332	0,317	0,028	0,331	0,129	0,203	0,116	0,442	0,181	-0,311	0,099	0,649
Romantico	-0,026	0,282	0,447	0,406	1	0,283	0,244	0,09	0,069	-0,005	0,14	0,088	0,264	0,11	-0,203	0,169	0,437
Madre_Trabaja	0,118	0,124	0,263	0,332	0,283	1	0,18	0,028	0,301	0,309	0,147	0,336	0,273	0,263	0,09	0,046	0,243
Padre_Trabaja	0,031	0,22	0,374	0,317	0,244	0,18	1	-0,064	0,294	0,144	0,239	0,264	0,374	0,32	-0,053	0,03	0,343
T_Libre	-0,028	-0,209	0,087	0,028	0,09	0,028	-0,064	1	-0,148	0,125	0,161	0,136	-0,196	-0,1	0,205	0,054	0,11
Salidas	0,138	0,208	0,235	0,331	0,069	0,301	0,294	-0,148	1	0,291	0,077	0,178	0,364	0,412	0,015	0,109	0,251
Abstinencia	0,117	0,052	0,088	0,129	-0,005	0,309	0,144	0,125	0,291	1	0,114	0,453	0,117	0,216	0,267	0,002	-0,013
Irritabilidad	0,102	0,014	0,238	0,203	0,14	0,147	0,239	0,161	0,077	0,114	1	0,127	0,152	0,202	0,044	0,128	0,299
Tutor	0,064	0,102	0,15	0,116	0,088	0,336	0,264	0,136	0,178	0,453	0,127	1	0,187	0,25	0,29	-0,066	0,047
Sec_Privada	0,11	0,332	0,442	0,442	0,264	0,273	0,374	-0,196	0,364	0,117	0,152	0,187	1	0,233	-0,222	-0,009	0,432
Sedentario	0,123	0,091	0,153	0,181	0,11	0,263	0,32	-0,1	0,412	0,216	0,202	0,25	0,233	1	0,041	0,159	0,122
Internet	0,036	-0,328	-0,144	-0,311	-0,203	0,09	-0,053	0,205	0,015	0,267	0,044	0,29	-0,222	0,041	1	0,029	-0,268
Depresion	0,073	0,005	0,127	0,099	0,169	0,046	0,03	0,054	0,109	0,002	0,128	-0,066	-0,009	0,159	0,029	1	0,072
class	0,034	0,449	0,666	0,649	0,437	0,243	0,343	0,11	0,251	-0,013	0,299	0,047	0,432	0,122	-0,268	0,072	1

A continuación, se muestra el gráfico de la matriz de correlación en base a datos obtenidos (Figura 4).

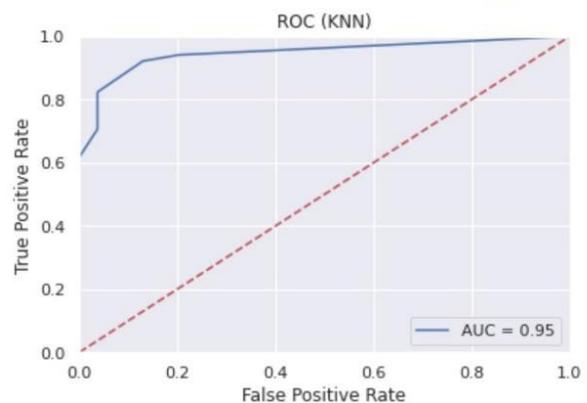


**Figura 4.** Gráfico de la matriz de correlación

Los recuadros con tendencia a color negro en la Figura 4 representan las variables de menor relevancia de acuerdo con los valores del dataset. A partir de la matriz de correlación se desprende la importancia de las variables respecto a la predicción (Figura 5); donde las variables en verde inciden para que los valores resultantes sean positivos, así mismo, las variables en amarillo inciden a los negativos.



**Figura 5.** Gráfico de la matriz de correlación



**Figura 6.** ROC (KNN)

Curvas ROC de los modelos predictivos a continuación Figura 6, se muestran las curvas ROC de los modelos predictivos, donde se puede observar la sensibilidad frente a la especificidad de los verdaderos positivos ante los verdaderos negativos.

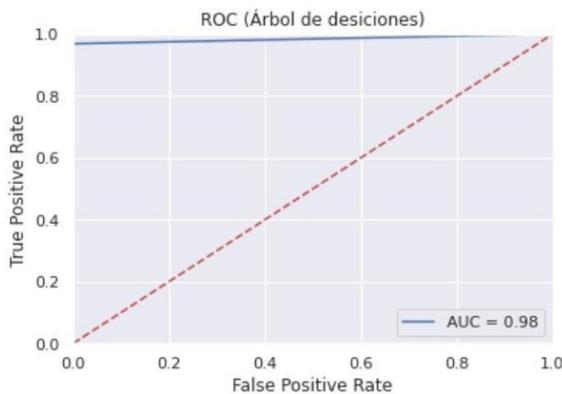


Figura 7. ROC (Árbol de decisiones)

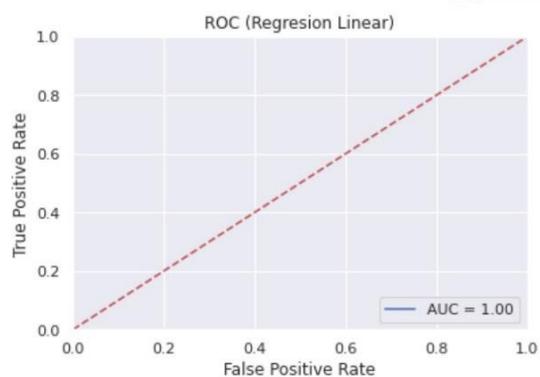


Figura 8. ROC (Regresión Lineal)

Según las Figuras 6, 7 y 8 de las curvas ROC mostradas, se puede apreciar que la regresión lineal tiene una mayor precisión.

En relación con el trabajo de Gómez y Londoño (2013), en el cual realizan un análisis de un comportamiento no violento en adolescentes que consumen alcohol basándose en modelos predictores bajo ecuaciones estructurales, concluyendo que el consumo de alcohol y la actitud no violenta en adolescentes se relacionan de forma directa, a diferencia de que, en este trabajo, se aplica el análisis con particionamiento de árbol de decisión, algoritmo basado en distancia y red neuronal.

## CONCLUSIONES

Entre los tres modelos analizados, el modelo de regresión lineal posee mayor precisión entre la sensibilidad de los verdaderos positivos y la especificidad de los falsos positivos, según el análisis de las curvas ROC obtenidas, por consiguiente, tiene una mayor probabilidad de pronosticar la tendencia al alcohol en adolescentes.

De conformidad con la muestra analizada, se observa que los indicadores psicosociales más relevantes para determinar la posibilidad de la tendencia al alcoholismo son el estado familiar, si vive en zona urbana o rural, el género y el tiempo libre. Por otro lado, los indicadores psicosociales que determinan una tendencia negativa al consumo de alcohol son la edad, el uso del internet, el tipo de tutor y el sedentarismo.

**Agradecimientos:** De consideración especial un agradecimiento a la Universidad Nacional San Marcos, Universidad Técnica del Norte y al Instituto Técnico Superior Ibarra.

**Contribución de los autores:** Todos los autores contribuyeron al diseño del estudio. Todos los autores leyeron y aprobaron el manuscrito final.

**Fuente de financiamiento:** Esta investigación no recibió financiamiento externo.

**Conflicto de intereses:** Los autores declaran no tener ningún conflicto de intereses. Los patrocinadores no tuvieron ningún papel en el diseño del estudio, recopilación, análisis o interpretación de datos, redacción del documento, o en la decisión de publicar los resultados.

## REFERENCIAS

- Ahumada-Cortez, J. G., Gámez-Medina, M. E., & Valdez-Montero, C. (2017). El consumo de alcohol como un problema de salud pública. *Revista Cubana de Higiene y Epidemiología*, 50(3), 425–426. <https://doi.org/10.35197/rx.13.02.2017.01.ja>
- Avellaneda, S., Pérez, A., Gras, M. E., & Font-Mayolas, S. (2010). Patrones de consumo de alcohol en adolescentes. *Psicothema*, 22(2), 189–195. [www.psicothema.com](http://www.psicothema.com)
- Borrás, T. (2014). Adolescencia: definición, vulnerabilidad y oportunidad. *Correo Científico Médico*, 18(1), 05–07.
- De Las, M., Olite, M. P., González Aguiar, B., Guevara, A. M., Fernández López, O., & Luaces, Y. T. (n.d.). El alcoholismo y su repercusión: un enfoque desde la psicología de la salud Alcoholism and its Effects: an Approach Based on Health Psychology. <http://www.revfinlay.sld.cu/index.php/finlay/article/view/314>
- Gómez, O., Gómez, Y., Ricardo, M., Pérez, J., Hernández, E., & Wilber, R. (2019). Comportamiento del alcoholismo en adolescentes. *Acta Médica del Centro*, 13(3), 13. <https://www.medigraphic.com/pdfs/medicadelcentro/mec-2019/mec193d.pdf>
- Herrera-granda, I. D., & Chicaiza-ipeales, J. A. (n.d.). *Artificial Neural Networks for Bottled Water Demand Forecasting: A Small Business Case Study*. 1, 362–373. <https://doi.org/10.1007/978-3-030-20518-8>
- Hu, Y. G., Sun, S., & Wen, J. Q. (2014). Agricultural machinery spare parts demand forecast based on BP neural network. *Applied Mechanics and Materials*, 635–637, 1822–1825. <https://doi.org/10.4028/www.scientific.net/AMM.635-637.1822>. <https://doi.org/10.1371/journal.pone.0255626>
- Jin Park, S., Jung Lee, S., Kim, H., Kwon Kin, J., Ji-Won, C., Soo-Jung, L., Kppk Lee, H., Jin Kum, D., & Young Chol, I. (n.d.). Machine learning prediction of dropping out of outpatients with alcohol use disorders. <https://doi.org/10.1371/journal.pone.0255626>
- Martínez-Cardona, M. del C., Muñoz-Borja, P., & Barberena-Borja, N. (2017). Consumo de alcohol. *Revista de Estudios e Investigación En Psicología y Educación*, 220–224. <https://doi.org/10.17979/reipe.2017.0.14.2932>
- Mendez-Ruiz, M. D., Ortiz-Moreno, G. A., Eligio-Tejada, I. A., Yáñez-Castillo, B. G., & Villegas-Pantoja, M. Á. (2018). Risk perception and alcohol consumption in high school adolescents. *Aquichan*, 18(4), 438–448. <https://doi.org/10.5294/aqui.2018.18.4.6>
- Mendoza, L. (2020). *Estudio Comparativo de Naive Bayes, Árbol de Decisión, Redes Neuronales y K-NN*. Research Gate, June. [https://www.researchgate.net/publication/342452229\\_Estudio\\_Comparativo\\_de\\_Naive\\_Bayes\\_Arbol\\_de\\_Decision\\_Redets\\_Neuronales\\_y\\_K-NN](https://www.researchgate.net/publication/342452229_Estudio_Comparativo_de_Naive_Bayes_Arbol_de_Decision_Redets_Neuronales_y_K-NN)

- Navalón, Mira, A., & Ruiz-Callado, R. (2017). Consumo de sustancias psicoactivas y rendimiento académico. Una investigación en estudiantes de educación secundaria obligatoria. *Health and Addictions/Salud y Drogas*, 17, 1 (ene. 2017), 45–52. DOI:<https://doi.org/10.21134/haaj.v17i1.278>.
- Monteiro, M. G. (2013). Alcohol y Salud Pública en América Latina: ¿cómo impedir un desastre sanitario? *Adicciones*, 25(2), 99–105. <https://doi.org/10.20882/ADICCIONES.56>
- Motos Sellés, P., Cortés Tomás, M. T., & Giménez Costa, J. A. (2016). Age of Onset, Motives and Amount of Alcohol Consumed in the Determination of Consequences in College Binge Drinkers. *Universitas Psychologica*, 15(2), 1–12. <https://doi.org/10.11144/Javeriana.upsy15-2.edcm>
- Páez-Rodríguez, Y., Páez-Rodríguez, Y., Lago-Díaz, Y., Valle-Pimienta, T., Sosa-Hernández, J., Coro-Carrasco, Z., & Báez-Delgado, M. (2021). Consumo de alcohol en adolescentes de la escuela preuniversitaria Federico Engels de Pinar del Río. *Archivo Médico Camagüey*, 25(2), 243–254. <http://revistaamc.sld.cu/index.php/amc/article/view/8072>
- Peetluk, L. S., Ridolfi, F. M., Rebeiro, P. F., Liu, D., Rolla, V. C., & Sterling, T. R. (2021). Systematic review of prediction models for pulmonary tuberculosis treatment outcomes in adults. *BMJ Open*, 11(3). <https://doi.org/10.1136/bmjopen-2020-044687>