Revista Innovación y Software
Vol. 3, No. 2, Mes Septiembre-Febrero, 2022
ISSN: 2708-0935
Pág. 26-39
https://revistas.ulasalle.edu.pe/innosoft

Tipo de artículo: Artículos originales
Temática: Inteligencia artificial
Recibido: 22/05/2022 | Aceptado: 14/06/2022 | Publicado: 30/09/2022

Identificadores persistentes:
ARK: ark:/42411/s9/a69
PURL: 42411/s9/a69

Aplicación de los árboles de decisión en el diagnóstico de Anemia en niños de la ciudad de Arequipa

Application of Decision Trees in the diagnosis of Anemia in children in the city of Arequipa

Indira Agramonte Mayhua [1], Alex Chaco Huamani [2], Alexander Valdiviezo Tovar[3] , Melody Ramos Challa [4*]

[1] Universidad Nacional de San Agustín. Arequipa, Perú. iagramontem@unsa.edu.pe
[2] Universidad Nacional de San Agustín. Arequipa, Perú. achacohu@unsa.edu.pe
[3] Universidad Nacional de San Agustín. Arequipa, Perú. avaldiviezot@unsa.edu.pe
[4] Universidad Nacional de San Agustín. Arequipa, Perú. mramoschal@unsa.edu.pe

[*] Autor para correspondencia: mramoschal@unsa.edu.pe

Resumen

Uno de los problemas más comunes en los niños que no son correctamente alimentados es la anemia. La deficiencia de hierro es perjudicial para los menores, pues impide que realicen sus actividades diarias por el cansancio extremo y fatiga. Debido a esta situación, el Estado peruano ha intentado disminuir el nivel de prevalencia de anemia a nivel nacional con campañas médicas en diferentes regiones, pese a ello, localidades como Caylloma en Arequipa aún mantienen un alto porcentaje de infantes anémicos, para ello se desarrolló una implementación mediante Árboles de Decisión con el lenguaje Python para poder determinar si un niño tiene anemia en base a los datos proporcionados.

Palabras clave: Árbol de decisión, anemia, inteligencia artificial.

Abstract

Anemia is one of the most common problems in children who are not adequately fed. Iron deficiency harms minors as it prevents them from carrying out their daily activities due to extreme tiredness and fatigue. Due to this situation, the Peruvian State has tried to reduce the level of prevalence of anemia at the national level with medical campaigns in different regions; despite this, localities such as Caylloma in Arequipa still maintain a high percentage of anemic infants, for which an implementation through Decision Trees with the python language to be able to determine if a child has anemia based on the data provided.

Keywords: *Decision tree, anemia, artificial intelligence.*

Introducción

Uno de los mayores problemas de salud pública en el mundo asociado al incremento de índices de morbilidad y mortalidad, es la anemia [1]; según la OMS (Organización Mundial de la Salud) en el año 2020 la anemia afectó en todo el mundo a 1620 millones de personas, lo que representaría al 24,8% de la población, mientras que, la máxima prevalencia se da en los niños en edad preescolar con 47,4% [2].

En nuestro país muchas mujeres gestantes y niños hasta los 11 años de edad mueren debido a tardíos diagnósticos por parte de los médicos siendo familias del sector rural los más afectados, esto debido a deficiencias nutricionales, bajos ingresos familiares, bajo nivel educativo, políticas de salud centralistas y al bajo financiamiento destinado por parte del estado peruano [3].

Realizar diagnósticos en niños no es una tarea fácil, no sólo basta con saber si un niño presenta anemia o no, sino también conocer el grado de afección; dato que podría ser determinante en la evaluación de futuras complicaciones en el paciente [4].

Por otro lado, existen diferentes indicadores para realizar tales diagnósticos los más comunes son talla, peso, niveles de hemoglobina, frecuencia cardíaca entre otros lo que permite evaluar la condición en la que se encuentra un individuo frente a esta enfermedad [5]. Otro factor importante a considerar también es la ubicación geográfica, en nuestro país la media de niveles de hemoglobina en regiones de la sierra es mayor que en las regiones de la costa.

Tomando en cuenta lo anteriormente descrito, en este artículo se pretende desarrollar un Árbol de decisión, basado en un sistema de información del estado nutricional de niños del Perú proporcionada por el Instituto Nacional de Salud con el fin de conocer el diagnóstico de anemia en niños pertenecientes a la región de Arequipa , de esta manera también anticipar el nivel de afección con una precisión aceptable [6].

En ese sentido nos basamos en el Sistema de Información del Estado Nutricional (SIEN) que fue implementado por el Instituto Nacional de Salud - Centro Nacional de Alimentación y Nutrición (INS/CENAN) [7] con el fin de obtener un diagnóstico de anemia en niños menores a 11 años en la ciudad de Arequipa ya que dicho sistema realiza un proceso continuo y sistemático que registra, procesa, reporta y analiza información del estado nutricional de niños menores de

cinco años y mujeres gestantes que acuden a establecimientos de salud del primer nivel de atención del Ministerio de Salud [8].

El Árbol de decisión presentado es una estructura que nos ayudará a tomar como entrada un objeto o situación descrita por un conjunto de propiedades en este caso indicadores de anemia y proporcionará como salida una decisión de sí o no. En términos de capacidad, el Árbol de decisión es un método rápido y eficaz que va a permitir clasificar las entradas del conjunto de datos y proporcionar una buena capacidad de apoyo a las decisiones [9].

El sistema será de gran ayuda en el diagnóstico de anemia ya que tiene ciertos parámetros que pueden ser revisados utilizando una estructura de inteligencia artificial para conseguir datos más cercanos que nos permitan encontrar una relación entre el nivel de hemoglobina frente a un posible cuadro de anemia [10].

Trabajos Relacionados

Barrientos et al. [11] evalúan el desempeño de tres algoritmos basados en los árboles de decisión a partir de los resultados en su aplicación con el fin de determinar si la técnica puede llegar a ser una herramienta de soporte y ayuda eficaz en el tratamiento y diagnóstico médico correspondientes a la sintomatología que un médico especialista considera importante en el diagnóstico de cáncer de seno. Para realizar esto los autores utilizaron dos bases de datos las cuales contienen datos recopilados por expertos en patología con experiencia en la detección de cáncer de mama, estos datos permiten el diagnóstico para saber si el paciente tiene o no cáncer de mamá sin embargo es necesario una biopsia y una mamografía para indicar si el tumor es maligno. Esta investigación concluye que los árboles de decisión son posibles de construir a partir de datos médicos, sin embargo, los datos recopilados deben de pasar por un proceso de clasificación de esta manera se obtiene un margen de error mínimo.

R. Yen et al.[12] examinaron los factores de riesgo principales relacionados con las conductas alimentarias inadaptadas y la regulación de las emociones, y cómo sus interacciones afectan a la detección de los trastornos alimentarios. Para ello construyeron un modelo de árbol de decisión utilizando el aprendizaje automático sobre un conjunto de datos de 830 mujeres jóvenes chinas sin antecedentes con una edad media de 18,91 años. El conjunto de datos se dividió en datos de entrenamiento y de prueba con una proporción de 70% y 30%. Los resultados señalan que la inflexibilidad de la imagen corporal se identificó como el principal clasificador de las mujeres con alto riesgo de TCA, seguidas por la angustia psicológica y la insatisfacción corporal.

Pei et al.[13] establecieron un modelo predictivo basado en los factores de riesgo asociados a la diabetes mediante un árbol de decisión. En su investigación reclutaron a un total de 10.436 participantes que se sometieron a un examen de salud entre enero de 2017 y julio de 2017, de los cuales 3454 permanecieron en el conjunto final de datos. Para asignar el porcentaje de datos de entrenamiento y de prueba usaron el algoritmo J48, dando como resultado un 70% de datos para entrenamiento y 30% de datos de prueba. Su trabajo consta de 14 variables de entrada y 2 de salida, el modelo de árbol de decisión presentado identificó varios factores clave que están estrechamente relacionados con el desarrollo de la diabetes y que también son modificables. Además, el modelo alcanzó una exactitud de clasificación del 90,3% con una precisión del 89,7% y un recuerdo del 90,3%. Presentando como conclusión que su modelo de árbol de decisión estima el desarrollo de la diabetes en una población china adulta de alto riesgo con un gran potencial para la aplicación del control de la diabetes.

Bouza y Santiago [14] califican a los árboles de decisión como una herramienta muy potente, que permite la segmentación, clasificación, predicción, reducción, identificación y recodificación de los datos. En el campo de la medicina la toma de decisiones es fundamental, es por esto que algunas instituciones utilizan sistemas conocidos como Decision support systems, sistemas eficientes y fiables que ayudan a los médicos a obtener sus diagnósticos.

Tomando en cuenta las investigaciones anteriormente descritas, se plantea realizar un árbol de decisión que pueda obtener un diagnóstico de anemia en base a un conjunto de datos que pueden ser revisados utilizando una estructura de inteligencia artificial.

Marco Teórico

A. Árboles de Decisión

Según Solarte y Soto [15] los árboles de decisión es una técnica de aprendizaje inductivo supervisado no paramétrico, se utiliza en la predicción y se emplea en el campo de la inteligencia artificial. Por otro lado, Landín y Romero [16] nos indican que los árboles de decisión posibilitan la uniformidad del diagnóstico y tratamiento en pacientes. Se basa en la aplicación de un conjunto de reglas y el uso de funciones lógicas, un aporte del uso de esta técnica es el plantear un problema de forma concisa y que a partir del mismo problema analizar todas las opciones posibles [17].

B. Data reduction (python)

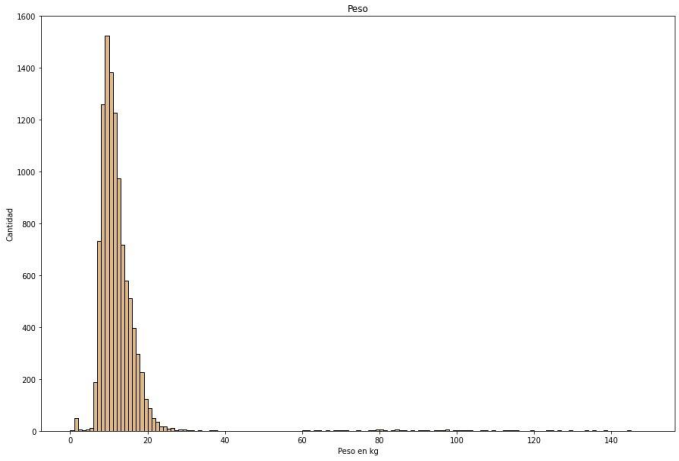
Data reduction es una técnica para reducir la cantidad de variables en un conjunto de datos[18][19]. En tareas de Machine Learning como la regresión o la clasificación, a menudo hay demasiadas variables, conocidas también como características, con las que trabajar. Cuanto mayor sea el número de características, más difícil será modelarlas. Además, algunas de estas, pueden ser bastante redundantes, añadiendo ruido al conjunto de datos y no tiene sentido tenerlas en los datos de entrenamiento. Es aquí donde se tiene que reducir la cantidad de variables[19].

El proceso de data reduction transforma los datos de un conjunto de características de alta dimensión a un conjunto de características de baja dimensión a través de dos componentes: la selección de características y la extracción de características. En la selección de características, se eligen subconjuntos más pequeños de características de un conjunto de datos de muchas dimensiones para representar el modelo mediante el filtrado, la envoltura o la incrustación. La extracción de características reduce el número de dimensiones de un conjunto de datos para modelar las variables y realizar el análisis de componentes. También es importante que las propiedades significativas presentes en los datos no se pierdan durante la transformación [18][19].

C. Análisis de datos

a. Histogramas:

En la presente sección se mostrará los histogramas pertenecientes a algunas variables a utilizar en el presente estudio. Estos diagramas ayudan a comprender la distribución de los datos y comprobar valores extremos o atípicos.



Revista Innovación y Software
Vol. 3, No. 2, Mes Septiembre-Febrero, 2022
ISSN: 2708-0935
Pág. 26-39
https://revistas.ulasalle.edu.pe/innosoft

Figura 1. Cantidad de niños/niñas divididos por peso.

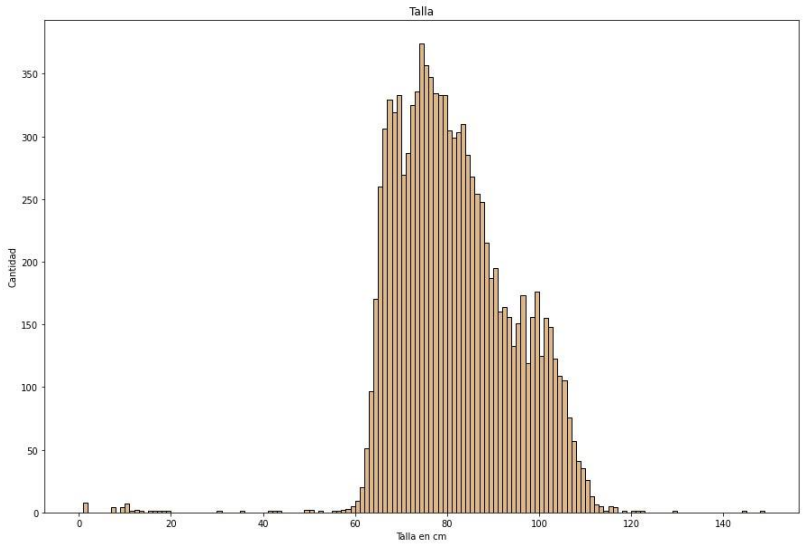


Figura 2. Cantidad de niños/niñas agrupados por la estatura.

Metodología y Desarrollo

A. Recursos

Para realizar esta investigación fue necesaria la búsqueda de datos que nos permitieran tener acceso a información verídica. Por esta razón el dataset utilizado fue extraído del Repositorio de Datos del Instituto Nacional de Salud, la investigación que lo acompaña es Sistema de Información del Estado Nutricional de niños y gestantes Perú - INS/CENAN. Las herramientas que se utilizaron para realizar este proyecto fueron: Visual Studio Code, Google Colab que nos permiten crear código para data reducción y crear el árbol de decisión a partir de un data set.

B. Limpieza de datos

Data Reduction, es el proceso de reducir la cantidad de capacidad requerida para almacenar datos. La reducción de datos puede aumentar la eficiencia del almacenamiento y reducir los costos. Los proveedores de almacenamiento a menudo describirán la capacidad de almacenamiento en términos de capacidad bruta y capacidad efectiva, que se refiere a los datos después de la reducción. En esta

ocasión lo que se busca es reducir la cantidad de atributos(columnas) del dataset Niños Arequipa, esta reducción es debido a que mucha de esta información es irrelevante y en ocasiones inutil.

```
if row['Peso'] == '#¡NULO!' or row['Talla'] == '#¡NULO!':
            continue
```

Con la anterior sentencia se elimina o se evita aquellas filas con datos incompletos, es decir, aquellas filas que tengan como contenido !#NULO! . Esto se hace para que esta información no se tome en cuenta por que esta información se vería reflejada en el árbol de decisión que se realizará a posterior.

```
            Diresa = row['Diresa']
            DxAnemia = row['Dx_Anemia']
            Sexo = row['Sexo']
            EdadMeses = row['EdadMeses']
            Peso = row['Peso']
            Talla = row['Talla'] Hemoglobina
            = row['Hemoglobina']
            HBC = row['HBC']
            ProvinciaREN = row['ProvinciaREN']
            DistritoREN = row['DistritoREN']
```

Para la siguiente parte, el código nos muestra un filtro donde solo se escogen los atributos que son importantes. En otras palabras las atributos con una mayor relevancia son las elegidas para que continúen en un nuevo archivo csv, además DxAnemia es nuestro dato de salida por lo que también se le toma en cuenta.

```
with open('Niños_AREQUIPA_V2.csv', 'w', newline='') as csvfile:
             fieldnames   =   ['Diresa',   'DxAnemia','Sexo',   'EdadMeses',
'Peso','Talla','Hemoglobina','HBC','ProvinciaREN','DistritoREN']
    writer = csv.DictWriter(csvfile, fieldnames=fieldnames)
    writer.writeheader() writer.writerows(new_data)
```

Revista Innovación y Software
Vol. 3, No. 2, Mes Septiembre-Febrero, 2022
ISSN: 2708-0935
Pág. 26-39
https://revistas.ulasalle.edu.pe/innosoft

Finalmente se crea un nuevo archivo csv con la información separada. Este nuevo archivo csv llamado "Niños_AREQUIPA_V2.csv" contiene una cabezera con los siguientes datos: 'Diresa', 'DxAnemia', 'Sexo', 'EdadMeses', 'Peso', 'Talla', 'Hemoglobina', 'HBC', 'ProvinciaREN', 'DistritoREN'. y es acompañado por la información que está separado en un array new_data.

```
            DxAnemia Sexo  EdadMeses   Peso  Talla  Hemoglobina    HBC  \
Diresa
AREQUIPA  Anemia Leve    F         52  16.90  109.0         11.8  10.67
AREQUIPA       Normal    M         14  11.09   82.1         13.8  13.47
AREQUIPA       Normal    M         10  10.67   73.3         11.8  11.47
AREQUIPA       Normal    F         13   9.22   74.8         12.3  11.97
AREQUIPA       Normal    F         14   8.51   73.7         11.5  11.49
...               ...  ...        ...    ...    ...          ...    ...
AREQUIPA       Normal    F         48  19.60  109.8         13.5  12.37
AREQUIPA       Normal    F          6   7.20   72.0         13.8  12.67
AREQUIPA       Normal    M          6   7.50   67.0         12.9  11.77
AREQUIPA       Normal    M         24  12.90   85.0         14.7  13.64
AREQUIPA       Normal    M         46  19.00  102.5         15.2  14.14

         ProvinciaREN     DistritoREN
Diresa
AREQUIPA     AREQUIPA      PAUCARPATA
AREQUIPA     CAYLLOMA           MAJES
AREQUIPA     CAYLLOMA           MAJES
AREQUIPA     CAYLLOMA           MAJES
AREQUIPA   CONDESUYOS      RIO GRANDE
...               ...             ...
AREQUIPA     AREQUIPA      PAUCARPATA
AREQUIPA     AREQUIPA      PAUCARPATA
AREQUIPA     AREQUIPA  CERRO COLORADO
AREQUIPA     AREQUIPA  MARIANO MELGAR
AREQUIPA     AREQUIPA  MARIANO MELGAR

[10553 rows x 9 columns]
```

Figura 3. Limpieza de datos

C. Librerías a utilizar

Para elaborar Árbol de Decisión en python se utilizaron las siguientes librerías:

A. Pandas: Pandas es una librería de Python especializada en el manejo y análisis de estructuras de datos. Las principales características de esta librería son: Define nuevas estructuras de datos basadas en los arrays de la librería NumPy pero con nuevas funcionalidades. Permite leer y escribir fácilmente ficheros en formato CSV, Excel y bases de datos SQL. Permite acceder a los datos mediante índices o nombres para filas y columnas. Ofrece métodos para ordenar, dividir y combinar conjuntos de datos. Permite trabajar con series temporales.

B. DecisicionTreeClassifier : La biblioteca Scikit Learn tiene una función de módulo DecisionTreeClassifier() para implementar el clasificador de árboles de decisión con bastante facilidad.

C. Confusion_matrix : La matriz de confusión sirve para evaluar la precisión de una clasificación. Por definición, una matriz de confusiones tal que es igual al número de observaciones que se sabe que están en el grupo y se predijo que estén en el grupo.

D. StringIO : Proporciona un medio conveniente para trabajar con texto en memoria utilizando la interfaz de programación de archivo (read() , write() , etc.). Utilizamos StringIO para construir cadenas grandes puede ofrecer ahorros en el rendimiento sobre algunas otras técnicas de concatenación de cuerdas en algunos casos.

E. Pydotplus : Se pueden utilizar tanto para resolver problemas de clasificación como de regresión. Una de sus principales ventajas es la facilidad con la que se puede interpretar los resultados en base a reglas. Permitiendo no solo obtener un resultado, sino inspeccionar los motivos por los que se llega a una predicción dada.

F. Numpy : El principal beneficio de NumPy es que permite una generación y manejo de datos extremadamente rápido. NumPy tiene su propia estructura de datos incorporada llamado arreglo que es similar a la lista normal de Python, pero puede almacenar y operar con datos de manera mucho más eficiente.

G. Seaborn : Es una librería de visualización de datos para Python desarrollada sobre matplotlib . Ofrece una interfaz de alto nivel para la creación de atractivas gráficas.

H. matplotlib : Con esta librería es posible crear trazados, histogramas, diagramas de barra y cualquier tipo de gráfica con ayuda de algunas líneas de código. Se trata de una herramienta muy completa, que permite generar visualizaciones de datos muy detalladas.

I. Tree : Los árboles de decisión son representaciones gráficas de posibles soluciones a una decisión basadas en ciertas condiciones, es uno de los algoritmos de aprendizaje supervisado más utilizados en machine learning y pueden realizar tareas de clasificación o regresión.

D. Importación de datos

Una de las dificultades que se presenta al realizar el desarrollo de un árbol de decisión es escoger el atributo apropiado, ya que este atributo debe ubicarse como raíz del árbol de decisión para que los demás atributos se generen como descendientes. Este proceso se realiza de forma recursiva en cada nodo. Antes es necesario realizar data reduction y limpieza de datos. Se obtuvieron 10553 datos divididos en 9 columnas. Los datos a importar deben de haber pasado por un proceso de reducción y limpieza, con el fin de detectar registros corruptos, registros imprecisos, datos duplicados o mal formateados.

```
dataset = pd.read_csv("/content/drive/My Drive/IA-2022/ninos_arequipa_V2.csv", index_col=0, encoding='latin-1')
#veamos cuantas dimensiones y registros contiene
dataset.shape
print(dataset)
```

Figura 4. Importación de datos

Posteriormente se verifica la información del dataset.

```
print('Información en el dataset:')
print(dataset.keys())
dataset.info()
```

Figura 5. Verificación del Dataset

Preparación de los datos.

```
#variables predictoras
X = dataset.iloc[:,2:7]

#variables a predecir
Y = dataset.iloc[:,0]

#Separo los datos de "train"
#en entrenamiento y prueba para probar los algoritmos
X.head()
```

Figura 6. Preparación de datos

Creación del modelo, se crearon las variables “X” y “Y” para el entrenamiento; las variables “x” y “y” para realizar la prueba.

```
#X_train y Y_train para entrenamiento
# X_test y Y_test para prueba
X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size=0.75, random_state=0)
X_train.info()
```

Figura 7. Creación del modelo

Validación del entrenamiento.

```
algoritmo = DecisionTreeClassifier(max_depth= 8)
arbol_anemia = algoritmo.fit(X_train,Y_train)
```

```
fig =plt.figure(figsize=(25,20))  #dimension
tree.plot_tree(arbol_anemia,feature_names=list(X.columns.values),class_names=list(Y.values), filled=True)
plt.show()
```

Revista Innovación y Software
Vol. 3, No. 2, Mes Septiembre-Febrero, 2022
ISSN: 2708-0935
Pág. 26-39
https://revistas.ulasalle.edu.pe/innosoft

Figura 8. Algoritmo para el entrenamiento

Análisis de Resultados

La matriz de confusión nos permite tener un resumen del rendimiento del algoritmo.

```
Matriz_C=confusion_matrix(Y_test, Y_pred)
Matriz_C
```

```
array([[1254,    0,    0,   23],
       [   0,  601,    1,    0],
       [   0,    0,   10,    0],
       [  23,    6,    0, 5997]])
```

Figura 9. Matriz de confusión

```
Precision_G = np.sum(Matriz_C.diagonal())/np.sum(Matriz_C)
Precision_G
```

```
0.9933038534428301
```

Figura 10. Precisión global del modelo

```
Precision_anemia_leve = ((Matriz_C[0,0]))/sum(Matriz_C[0,])
Precision_anemia_leve
```

```
0.9819890368050117
```

Figura 11. Precisión - Anemia Leve

```
Precision_anemia_moderada = ((Matriz_C[1,1]))/sum(Matriz_C[1,])
Precision_anemia_moderada
```

```
0.9983388704318937
```

Figura 12. Precisión - Anemia Moderada

Revista Innovación y Software
Vol. 3, No. 2, Mes Septiembre-Febrero, 2022
ISSN: 2708-0935
Pág. 26-39
https://revistas.ulasalle.edu.pe/innosoft

```
Precision_anemia_severa = ((Matriz_C[2,2]))/sum(Matriz_C[2,])
Precision_anemia_severa
```

```
1.0
```

Figura 13. Precisión - Anemia Severa

```
Precision_normal = ((Matriz_C[3,3]))/sum(Matriz_C[3,])
Precision_normal
```

```
0.9951875207434451
```

Figura 14. Precisión - Sin Anemia

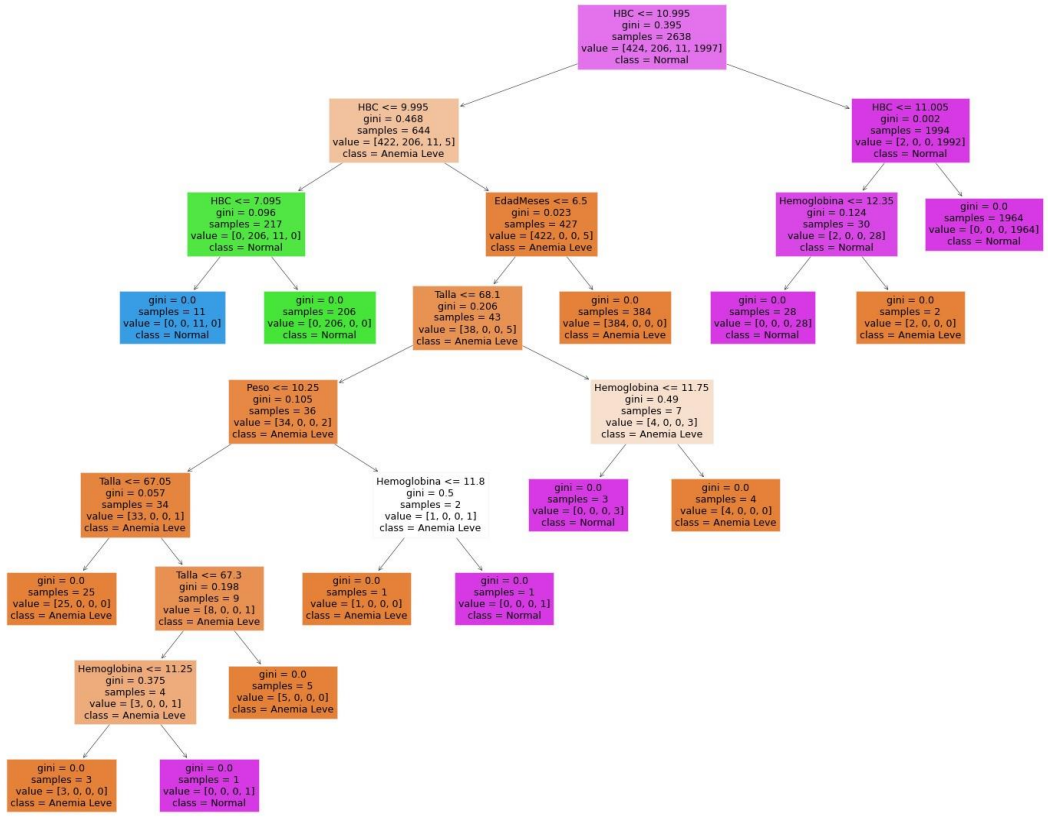


Figura 15. Árbol generado

Conclusiones y Trabajos Futuros

Los resultados obtenidos con el modelo de clasificación por árboles de decisión indican que es capaz de generar modelos con los datos que se encuentran en el dataset. Durante este trabajo, existen diversas áreas que quedan abiertas y en las que es posible trabajar. Algunas de estas investigaciones están relacionadas directamente al tema de este trabajo y son resultado de la investigación durante el desarrollo. Otras investigaciones no están relacionadas directamente sin embargo pueden ser retomadas posteriormente o como una opción de exploración a otros investigadores.

A continuación, se presentan algunos trabajos futuros que pueden desarrollarse. Además, se sugieren algunos temas específicos para apoyar y mejorar el trabajo de inteligencia artificial propuesto. Entre los posibles trabajos futuros se destacan:

- Aplicación de Árboles de Decisión en el diagnóstico de Anemia en niños en el Perú.
- Árboles de Decisión en el diagnóstico de Cáncer.
- Calidad de predicción del diagnóstico de anemia.
- Tratamiento de la matriz de confusión para disminuir la tasa de falsos positivos y generar un mejor diagnóstico de la anemia.

Referencias

[1] Jara, Hambre, desnutrición y anemia: una grave situación de salud pública. Revista Gerencia Y Políticas de Salud, 7(15), 7–10. [Online] Available at: http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S1657-70272008000200001.

[2] Al-kassab-Córdova, A., Méndez-Guerra, C., & Robles-Valcarcel, P. Factores sociodemográficos y nutricionales asociados a anemia en niños de 1 a 5 años en Perú. Revista Chilena de Nutrición, 47(6), 925–932. https://doi.org/10.4067/s0717-75182020000600925.

[3] O. Munares “Niveles de hemoglobina en gestantes atendidas en establecimientos del Ministerio de Salud del Perú” [online document], 2011. Available: Oxford Reference Online, https://www.scielosp.org/article/rpmesp/2012.v29n3/329-336/es/ [Accessed: Jun 21, 2022].

[4] J. Mendoza, “Sistema experto para alertar y brindar alternativa de tratamiento para la anemia en niños de la provincia de Jaén,” M.S. tesis, Facultad de Ingeniería, Universidad Católica Santo Toribio de Mogrovejo, Chiclayo, Perú, 2021. [Online]. Available: https://tesis.usat.edu.pe/handle/20.500.12423/3689

[5] F. Andrade, “Modelo de regresión Dirichlet bayesiano: aplicación para estimar la prevalencia del nivel de anemia infantil en centros poblados del Perú,” M.S. tesis, Escuela de Postgrado, Pontificia Universidad Católica del Perú, Lima, Perú, 2020. [Online]. Available: https://tesis.pucp.edu.pe/repositorio/handle/20.500.12404/18683

[6] G. Canaza, "Modelo predictivo de riesgo asociado a la anemia en niños menores de 5 años en la Microred Yauri provincia de espinar – Cusco, 2019," M.S. tesis, Facultad de Ingenieria, Estadistica e Informatica, Universidad Nacional del Altiplano, Puno, Perú, 2021. [Online]. Available: http://repositorio.unap.edu.pe/handle/UNAP/15335

[7] Sistema de información del Estado Nutricional de niños y gestantes Perú - INS/CENAN (Instituto Nacional de Salud Centro Nacional de Alimentación y Nutrición). Instituto Nacional de Salud - Centro Nacional de Alimentación y Nutrición. Nov, 2021. Available: https://datos.ins.gob.pe/sv/dataset/sistema-de-informacion-del-estado-nutricional-de-ninos-y-gestantes-peru-ins-cenan-2 019-2020

[8] Sistema de información del Estado Nutricional de niños y gestantes Perú - INS/CENAN (Instituto Nacional de Salud Centro Nacional de Alimentación y Nutrición) - Repositorio de Datos - Instituto Nacional de Salud. Ins.gob.pe. [Online] Available at: https://datos.ins.gob.pe/sv/dataset/sistema-de-informacion-del-estado-nutricional-de-ninos-y-gestantes-peru-ins-cenan-2 019-2020.

[9] Russel, S., & Norvig, P. (2012). Artificial intelligence—a modern approach 3rd Edition. In The Knowledge Engineering Review. https://doi.org/10.1017/S0269888900007724

[10]Dogan, S., & Turkoglu, I. (2008). Iron-deficiency anemia detection from hematology parameters by using decision trees. International Journal of Science & Technology, 3(1), 85-92.

[11]Martínez, R. E. B., Ramírez, N. C., Mesa, H. G. A., Suárez, I. R., Trejo, M. D. C. G., León, P. P., & Morales, S. L. B. (2009). Árboles de decisión como herramienta en el diagnóstico médico. Revista médica de la Universidad Veracruzana, 9(2), 19-24.

[12]Ren, Y., Lu, C., Yang, H., Ma, Q., Barnhart, W. R., Zhou, J., & He, J. (2022). Using machine learning to explore core risk factors associated with the risk of eating disorders among non-clinical young women in China: A decision-tree classification analysis. *Journal of Eating Disorders*, *10*(1). https://doi.org/10.1186/s40337-022-00545-6

[13]Pei, D., Yang, T., & Zhang, C. Estimation of diabetes in a high-risk adult chinese population using j48 decision tree model. Diabetes, Metabolic Syndrome and Obesity: Targets and Therapy 13 ,2020. https://doi.org/10.2147/DMSO.S279329

[14]Bouza, C., & Santiago, A. La minería de datos: árboles de decisión y su aplicación en estudios médicos. Modelación Matemática de Fenómenos del Medio Ambiente y la Salud, 2, 64-78. 2012.

[15]Martínez, G. R. S., & Mejía, J. A. S. Árboles de decisiones en el diagnóstico de enfermedades cardiovasculares. Scientia et technica, 16(49), 104-109, 2011.

[16]Landín Sorí, M., & Romero Sánchez, R. E. Árboles de decisiones para el diagnóstico y tratamiento de pacientes con glaucoma neovascular. Revista Archivo Médico de Camagüey, 16(4), 514-527, 2012.

[17]Zuniga, C., & Abgar, N. Breve aproximación a la técnica de árbol de decisiones. Recuperado de https://niefcz. files. wordpress. com/2011/07/breve-aproximacion-a-la-tecnica-de-arbol-de-decisiones. pd, 2011.

[18]Tech Target "Dimensionality reduction". Recuperado de https://www.techtarget.com/whatis/definition/dimensionality-reduction

[19]Barla, N. Dimensionality Reduction for Machine Learning. Recuperado de https://neptune.ai/blog/dimensionality-reduction, 2022

Artículos originales

Indira Agramonte Mayhua, Alex Chaco Huamani , Alexander Valdiviezo Tovar, Melody Ramos Challa

Aplicación de los árboles de decisión en el diagnóstico de Anemia en niños de la ciudad de Arequipa
Application of Decision Trees in the diagnosis of Anemia in children in the city of Arequipa

Innovación y Software
, p. 26
vol. 3, no. 2 39
2022
Universidad La Salle, Perú
facin.innosoft@ulasalle.edu.pe

ISSN: 2708-0927 / ISSN-E: 2708-0935