

Determinación del mejor modelo para predecir la Parasitosis intestinal en niños de 5 a 9 años en el Hospital pediátrico Alfonso Villagómez Román

Determination of the best model to predict intestinal parasitosis in children from 5 to 9 years of age at the Alfonso Villagómez Román Pediatric Hospital.

Johanna Enith Aguilar Reyes

Ingeniera en Estadística Informática
Master en gestión y liderazgo educacional. Docente investigador en el grupo MOODELING de la Escuela Superior Politécnica de Chimborazo.
Universidad Politécnica Estatal del Carchi
johanna.aguilar@upec.edu.ec
<https://orcid.org/0000-0002-1230-2503>

Manuel Antonio Meneses Freire

Doctor en matemática
Máster en técnicas estadísticas
Doctor dentro del programa de doctorado en estadística e investigación operativa
Universidad Politécnica Estatal del Carchi
ameneses@unach.edu.ec
<https://orcid.org/0000-0001-8182-3153>

Resumen

Al comparar dos técnicas de clasificación: Árboles de clasificación y Regresión Logística para predecir la parasitosis intestinal en niños de 5 a 9 años atendidos en el Hospital Pediátrico Alfonso Villagómez Román de la ciudad de



Imaginario Social
Entidad editora
REDICME (reg-red-18-0061)

e-ISSN: 2737-6362
mayo especial 2023 Vol. 6-2-2023
<http://revista-imaginariosocial.com/index.php/es/index>

Recepción: 20 de marzo de 2023
Aceptación: 20 de abril de 2023

156- 183

Atribución/Reconocimiento-NoComercial- CompartirIgual 4.0 Licencia Pública Internacional — CC

BY-NC-SA 4.0
<https://creativecommons.org/licenses/by-nc-sa/4.0/legalcode.es>

Riobamba, se establece el mejor modelo a través de la significancia de los factores asociados para la variable en estudio. La base de datos para el análisis fue proporcionada por el Área de estadística de la casa de salud a través de los repositorios existentes se tomó información del periodo 2019 -2021.

Los modelos encontrados se obtuvieron usando el software SPSS y su modelador SPSS MODELER empleado para evaluar la capacidad predictiva a través de las medidas de bondad de ajuste: Tasa de error y Curva Roc; con este análisis el modelo por árboles de clasificación fue el mejor por tener una precisión general del 65,7% y una tasa de error del 24,49%, los factores asociados al diagnóstico son: Triage, Saturación de oxígeno, Temperatura axilar, Talla, Peso, Frecuencia cardiaca mínima y Tipo de seguro de salud. Cabe indicar que la diferencia en la eficiencia de predicción para cada técnica fue mínima por lo que sería factible probar otras técnicas ya sea siguiendo la línea de árboles de decisión o modelos de regresión.

Palabras clave: Regresión logística, Árboles de clasificación, Estadística, Parasitosis intestinal, Pronósticos.

Abstract

By comparing two classification techniques: Classification Trees and Logistic Regression to predict intestinal parasitosis in children from 5 to 9 years old attended at the Alfonso Villagómez Román Pediatric Hospital in the city of Riobamba, the best model is established through the significance of the associated factors for the variable under study. The database for the analysis was provided by the Statistics Area of the health center through the existing repositories, information was taken from the period 2019 -2021.

The models found were obtained using SPSS software and its modeler SPSS MODELER used to evaluate the predictive capacity through the measures of goodness of fit: error rate and Roc Curve; with this analysis the model by classification trees was the best for having an overall accuracy of 65.7% and an error rate of 24.49%, the factors associated with the diagnosis are: Triage, Oxygen saturation, Axillary temperature, Height, Weight, Minimum heart rate and Type of health insurance. It should be noted that the difference in predictive efficiency for each technique was minimal, so it would

be feasible to test other techniques either along the lines of decision trees or regression models.

Key words: Logistic regression, Classification trees, Statistics, Intestinal parasitosis, Predictors.

Introducción

La estadística es un área aplicable a distintos campos del conocimiento en los que el ser humano se desenvuelve, uno de ellos la medicina, en la cual, a través de la técnica para la recolección, análisis y el procesamiento de la información generada por los establecimientos de salud tiene como objetivo otorgar posibles soluciones a diversos problemas que enfrentan, contribuyendo a la correcta toma de decisiones, satisfaciendo las necesidades propias y de los usuarios.

Las enfermedades ocasionadas por parásitos intestinales está considerada como uno de los problemas más comunes en la salud humana, estos patógenos afectan a poblaciones de todas las edades, pero con mayor frecuencia a los niños quienes se encuentran vulnerables al contagio por distintos factores como el entorno social, alimentación, higiene y recientemente según estudios se ha incluido la migración como una causa relevante ya que la gran afluencia de personas de distintos lugares en el mundo provocó un incremento notorio en los diagnósticos, sobre todo ha alertado acerca de nuevas especies de parásitos no tan habituales en la región, siendo esto un problema para las diferentes casas de salud quienes se enfrentan a un panorama fuera de lo común y en ocasiones no cuenta con el personal y el equipo adecuado para sobrellevar el tratamiento de la enfermedad.

Antecedentes investigativos

Las naciones unidas en su artículo “OMS alerta sobre infección de parásitos intestinales en países en desarrollo” menciona que, gran parte por no mencionar la mayoría de los niños residentes en países en vías de desarrollo se han visto afectados por gusanos intestinales contribuyendo a la malnutrición de estos y,

disminuye las probabilidades de crecimiento, desarrollo y aprendizaje (OMS, 2008).

(Batista Rojas & Álvarez Hernández, 2013) en su artículo “Parasitismo intestinal en niñas y niños mayores de 5 años de Ciudad Bolívar” se aplicó un estudio descriptivo y transversal a 320 pacientes mayores de 5 años atendidos en este recinto durante julio 2011- marzo 2012 con la finalidad de caracterizarlos según algunas variables de interés: tipos de parásitos, particularidades del abastecimiento y tratamiento del agua de consumo, lugar de deposición, hábitos higiénico-sanitarios y síntomas más frecuentes se concluyó que, los pacientes presentaban poliparitismo cuyo síntoma relevante es el dolor abdominal y se atribuye a la, ausencia de sanidad en el entorno puesto que los habitantes se abastecen de agua almacenada en pipas o tanques además de, la ausencia de baterías sanitarias y el hábito de andar.

(Montero Perez & Huilca Peralta, 2018) en su trabajo de investigación “Parasitosis intestinal, estado nutricional y diagnóstico bacteriológico en manos de niños de un jardín de la zona rural de Huancayo”, consideró la investigación de tipo aplicada y diseño cuantitativo - no experimental descriptivo, se realizó un examen parasitológico, así como también un diagnóstico bacteriológico donde se hizo la siembra en agar Macconkey, y por último se realizó un cuestionario para saber el estado nutricional de los niños concluyendo que el factor estado nutricional con relación a su edad, peso y talla fue baja.

En Ecuador los diversos síntomas y manifestaciones de la parasitosis constituyen las diez primeras causas de consulta pediátrica. Durante la infancia es frecuente la anemia en niños parasitados que, a largo plazo se convierte en alteraciones del desarrollo ponderal, psicomotriz e intelectual (Menacho Chávez, 2022).

La Organización Panamericana de la Salud, en el año 2016, realizó un estudio con el nombre de “Prevalencia de parasitismo intestinal en niños quechuas de zonas rurales montañosas del Ecuador”, se analizó la relación entre algunas variables sanitarias (el uso de letrinas, la disponibilidad de métodos adecuados de almacenamiento y tratamiento de agua dentro de esto también se estudió la aplicación de proyectos comunitarios para proteger las fuentes de agua potable) además de, las enfermedades

y la prevalencia de parasitismo infantil. Se buscó una gran variedad de parásitos, la mayoría de ellos patógenos, que se transmiten por diversas vías, como el agua, los alimentos, el suelo y las heces fecales. Se realizó exámenes coproparasitario a 149 niños de comunidades rurales quechuas que viven en las montañas de la provincia de Chimborazo, dando como resultado el 85,7% de las muestras presentaban al menos uno de los 10 parásitos estudiados y 63,4% contenían dos o más tipos de parásitos. No se encontraron diferencias significativas entre el número de casos informados de enfermedades como: diarrea, fiebre, infecciones respiratorias, vómitos y otras (Revista Panamericana de la Salud Pública, 2008)

(Pazmiño Gómez, y otros, 2018) en su artículo “ Parasitosis intestinal y estado nutricional en niños de 1-3 años de un centro infantil del Cantón Milagro”, corresponde un estudio cuantitativo de carácter descriptivo, apoyado por una investigación de campo que permitió la recolección de datos antropométricos y las muestras fecales para la realización de exámenes coprológicos a 38 niños que asisten frecuentemente al CIBV, dando como resultados lo siguiente: 23 (60,5%) de los niños y niñas presentaron parasitosis intestinal, quienes se ven afectados con bajo peso debido a la presencia de parásitos, en la encuesta a los padres de familia como parte de los instrumentos de investigación, más del 65% de los hogares no tienen una adecuada norma de cuidado para prevenir la infección por parásitos intestinales, el consumo de agua sin hervir así como la ingesta de frutas y legumbres sin lavado previo. Concluyendo que la carencia de servicios básicos, hábitos de higiene alimentaria, el desconocimiento de los riesgos son determinantes para la presencia de parásitos intestinales en los infantes y por ende afecta el estado nutricional.

Por otra parte, en el artículo titulado “Árboles de Clasificación vs. Regresión Logística en el desarrollo de habilidades genéricas en ingeniería” desde un enfoque experimental evalúa el desempeño de ambos modelos en el contexto de dos habilidades genéricas de ingeniería (razonamiento cuantitativo y comprensión lectora), incorporando dos escenarios predictores separados (solo indicadores y

solo construcciones derivadas del análisis de componentes principales “ACP”), dentro del cual los dos métodos presentan un desempeño similar respecto a indicadores mientras que en el escenario de construcciones la regresión logística destaca significativamente (Perez Rave & Gonzalez Echeverria, 2018).

En la tesis titulada “Comparación de modelos de clasificación: regresión logística y árboles de clasificación para evaluar el rendimiento académico” se comparan los modelos de regresión logística binaria y árboles de clasificación (CHAID) para evaluar el rendimiento académico. El comportamiento de estos modelos fue medido por cuatro indicadores: Sensibilidad, Curva ROC, Índice de GINI e Índice de Kappa, determinado que el árbol de clasificación es modelo más adecuado para clasificación y predicción (Lizares Castillo, 2017).

Todos los estudios descritos previamente son referencias significativas para el desarrollo de la presente publicación puesto que, cada investigación tiene un enfoque, metodología y resultados distintos, pero llegan a coincidir en que la infección por parásitos intestinales es una enfermedad común alrededor del mundo, siendo los niños el grupo más propenso a contraer estas bacterias debido a que por la falta de conocimiento y cuidado influyen en la correcta evolución de los menores; inclinándonos a indagar minuciosamente sobre la problemática y brindarle a la unidad de salud un antecedente correctamente argumentado que contribuya a la toma de decisiones para controlar o erradicar dicha patología.

Materiales y métodos

Tipo de investigación

La investigación es de tipo cuantitativo, debido que se analizó el número de pacientes diagnosticados con parasitosis intestinal en el HPAVR, según el objeto de estudio es aplicado ya que se emplea diferentes análisis para comparar cuál de los dos métodos es mejor para realizar pronósticos, según el nivel de profundización en el objeto de estudio es exploratorio e inferencial ya que este estudio no se lo ha realizado antes, según la manipulación de las variables es no experimental pues los datos se obtuvieron mediante las historias clínicas de los pacientes pediátricos, según la inferencia es inductiva ya que se analiza las metodologías de árboles de clasificación

y regresión para especificar que método es mejor y según el periodo temporal es longitudinal donde se analizó pacientes pediátricos internados en diferentes años calendario.

Población de estudio

La población de estudio pertenece a los pacientes atendidos en el periodo 2019-2021 en el Hospital Pediátrico Alfonso Villagómez Román.

Tamaño de la muestra

La muestra pertenece a todos los pacientes de 5 a 9 años diagnosticados con parasitosis intestinal en el Hospital Pediátrico Alfonso Villagómez Román.

Técnicas de recolección de datos

No se aplicó una técnica de recolección de datos ya que la información con los pacientes registrados en la entidad de salud fue entregada por el área de estadística de la institución.

Modelo Estadístico

La comparativa entre modelos de regresión y árboles de clasificación estará sujeta a la revisión previa del historial clínico de los pacientes que asisten a la casa de salud, de quienes obtendremos los datos para las variables independientes, las que serán utilizadas para modelar la parasitosis intestinal a través de las técnicas planteadas.

Identificación de variables

Variable dependiente

- Diagnóstico de alta (Parasitosis intestinal y Otro caso)

Variables independientes

- Cantón
- Grupo Cultural
- Edad
- Género

- Frecuencia cardiaca mínima
- Frecuencia respiratoria mínima
- Triage
- Temperatura axilar (°C)
- Peso (Kg)
- Talla (cm)
- Saturación de oxígeno
- Tipo de seguro

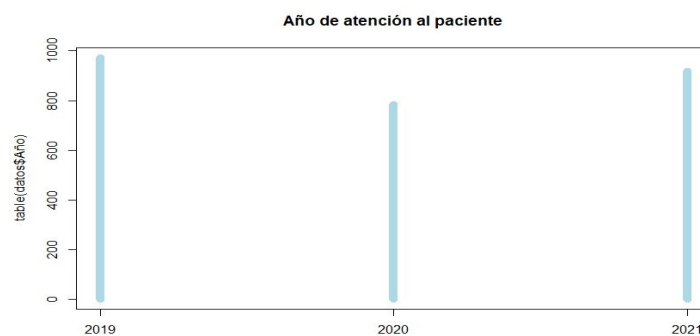
Resultados y discusión

El análisis estadístico toma en consideración un colectivo de 2675 individuos o pacientes pediátricos de 5 a 9 años que fueron atendidos en el servicio de emergencia del Hospital Pediátrico Alfonso Villagómez Román (HPAVR) de Riobamba en el periodo 2019 – 2021, la matriz de información contiene 14 variables, 7 de tipo cuantitativo y cualitativo respectivamente.

Análisis exploratorio univariado

Para el presente trabajo de investigación en primer lugar se realizó un análisis descriptivo de las variables mediante representaciones gráficas de las frecuencias de cada una de ellas, así como también un análisis estadístico básico para las variables cuantitativas seleccionadas. Los resultados encontrados se detallan a continuación:

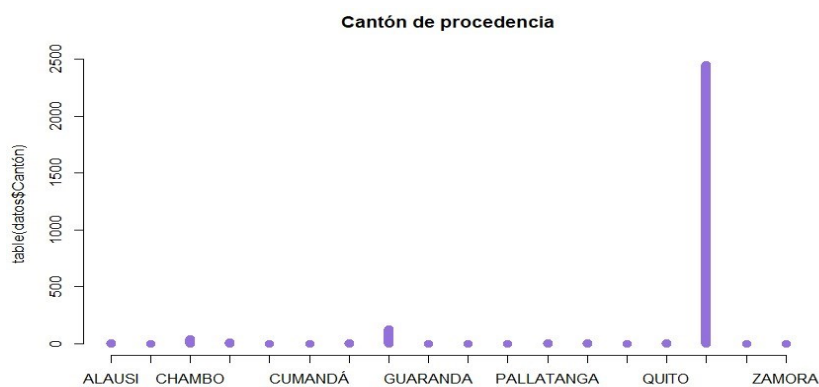
Figura 1: Distribución de la variable “Año”



Fuente: Elaboración propia

En la gráfica se representa la afluencia de pacientes que ingresaron al hospital en los 3 años de estudio donde, observamos que el año con mayor ingreso fue en el 2019, esta variación podemos decir que se origina a raíz de la pandemia por COVID-19 que afectó a la población y provocó una crisis sanitaria en el sistema de salud. En 2020 y 2021 se observa un descenso y un leve incremento en los ingresos hospitalarios respectivamente.

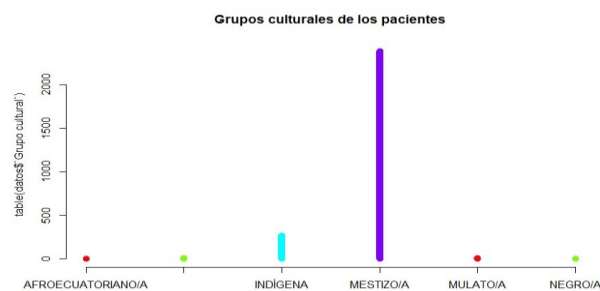
Figura 2: Distribución de la variable Cantón



Fuente: Elaboración propia

En lo que respecta al cantón de residencia según el gráfico indica que la mayoría de los pacientes que acudieron a la casa de salud son locales es decir oriundos del cantón Riobamba, sin embargo, también se registra pacientes de otros cantones del país en menor proporción, esto último puede estar ligado a que al ser un hospital de tercer nivel y específicamente para pacientes pediátricos quienes llegan al lugar posiblemente son transferidos de otra entidad de salud en su cantón natal.

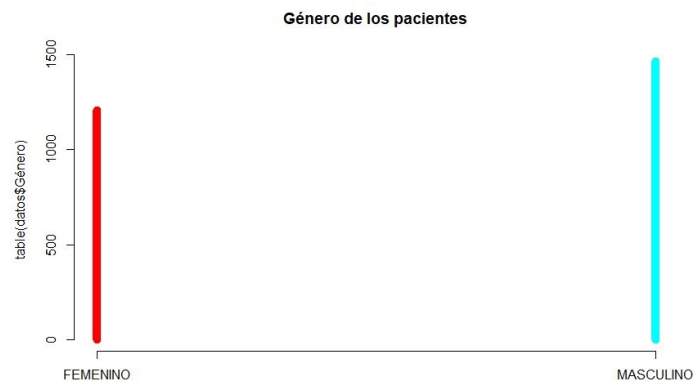
Figura 3: Distribución de la variable "Grupo cultural"



Fuente: Elaboración propia

Con respecto al grupo cultural con el cual se identifican los pacientes según el gráfico hay un gran grupo de individuos que se identifican como Mestizo/a, seguido de aquellos que se identifican como Indígenas y en menor proporción se encuentran los grupos blanco, mulato y negro.

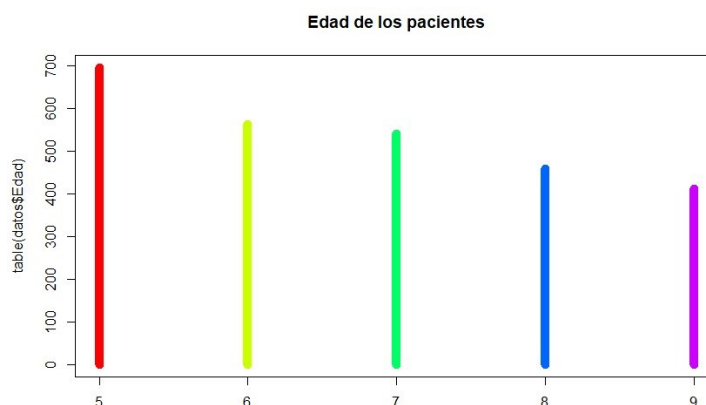
Figura 4: Distribución de la variable "Género"



Fuente: Elaboración propia

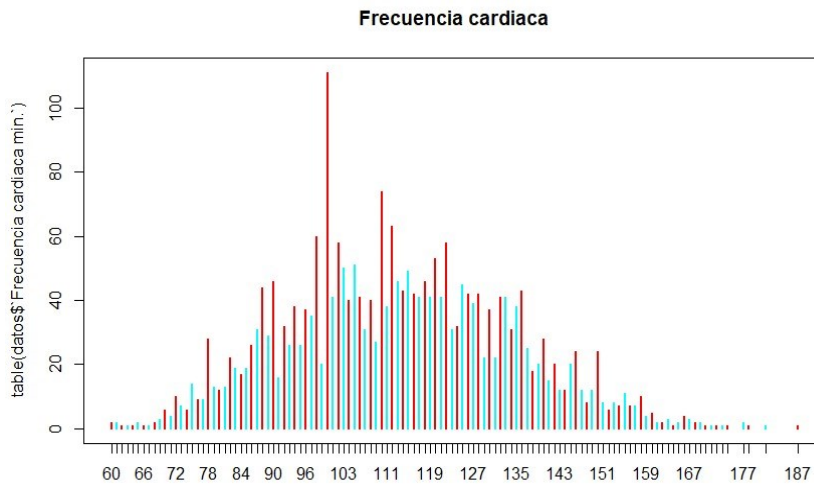
Según los registros de la casa de salud en los tres años de estudio en el hospital se registró un gran conglomerado de pacientes masculinos y en leve proporción pacientes femeninas.

Figura 5: Distribución de la variable "Edad"



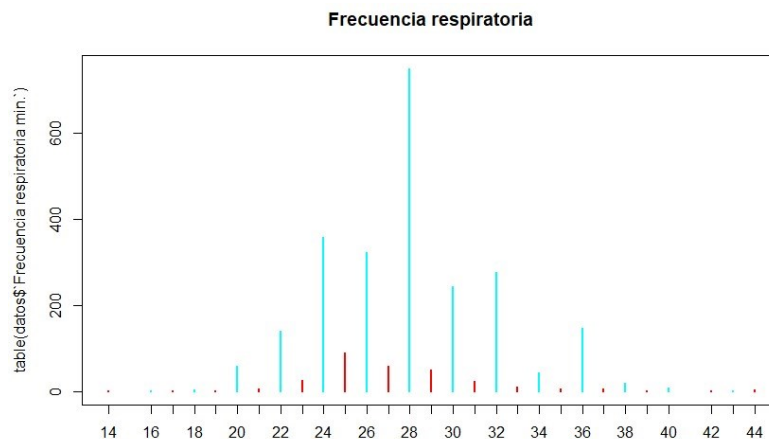
Fuente: Elaboración propia

Con respecto a la edad cabe recalcar que para el estudio se limitó a pacientes pediátricos entre 5-9 años por lo tanto podemos observar que la edad con mayor frecuencia es 5 años, además observamos que el resto de las edades presenta un descenso en sus proporciones.

Figura 6: Distribución de la variable "Frecuencia cardiaca"

Fuente: Elaboración propia

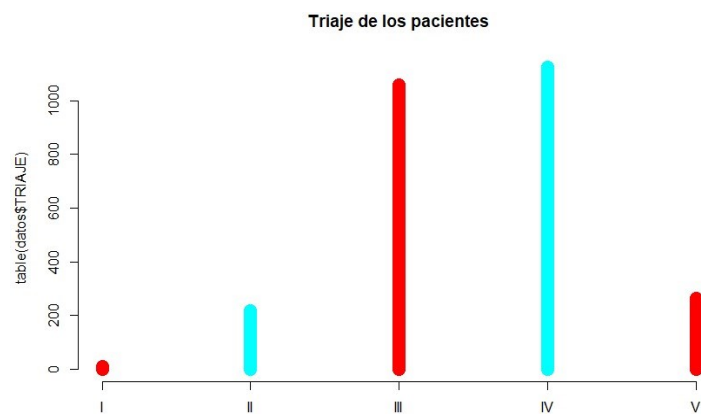
La frecuencia cardiaca registra el número de pulsaciones del paciente por unidad de tiempo, en el grafico podemos observar que los valores se encuentran entre 60 y 187 pulsaciones por unidad de tiempo siendo 100 el valor que mayor frecuencia se presenta en los pacientes. Además, cabe indicar que algunos valores estarán por encima o debajo de los rangos normales de frecuencia cardiaca en los niños de 5-9 años y esta variación pudo ser a otros factores o condiciones en las que ingreso el paciente al hospital.

Figura 7: Distribución de la variable "Frecuencia respiratoria"

Fuente: Elaboración propia

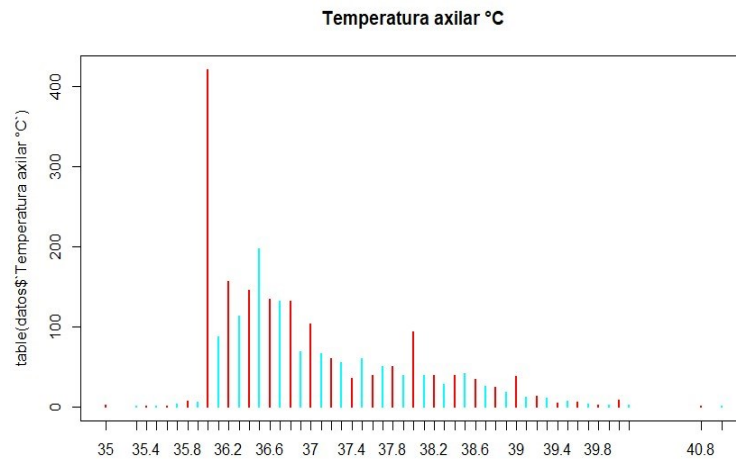
La frecuencia de respiración mínima contabiliza el número de respiraciones del paciente en un periodo específico, respecto a los registros del hospital y como se observa en el gráfico la variable se encuentra en un intervalo de 14 a 44 respiraciones, y hay una gran proporción de niños entre 5-9 años cuya frecuencia respiratoria es de 28 mientras que los demás se encuentran por encima o debajo de este valor y, como se indicó anteriormente los valores registrados varían con respecto a los valores normales para estas edades debido a factores que inciden para que esta se eleve o disminuya respectivamente.

Figura 8: Distribución de la variable "Triage"



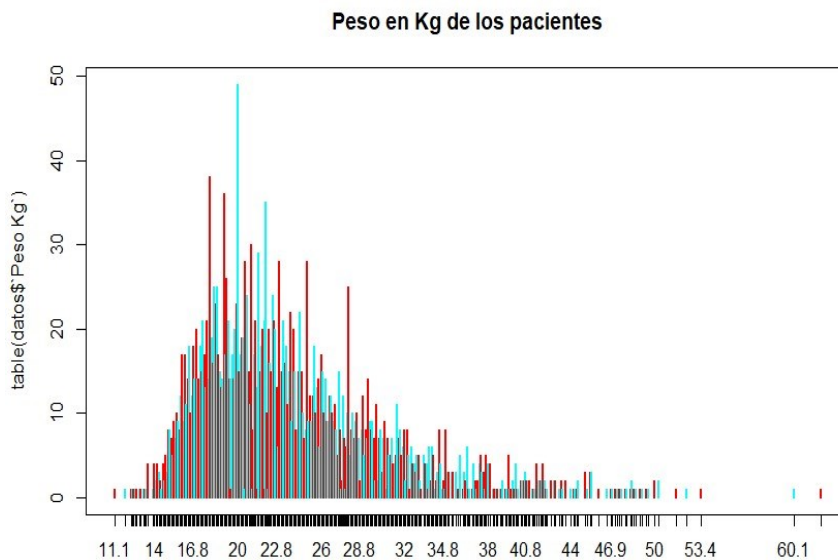
Fuente: Elaboración propia

El triage hace referencia al nivel de gravedad con la que un paciente ingresa al servicio de emergencia del hospital pediátrico, siendo el Nivel I el más crítico y el Nivel V el menos urgente a ser atendido, para el presente estudio se obtuvo que los pacientes ingresaban en su mayoría con un nivel de emergencia III y IV, a pesar que se presentaron pacientes con un Nivel I de emergencia, fueron pocos, lo cual es gratificante ya que la vida de la mayoría de pacientes no se encontraba en un riesgo crítico.

Figura 9: Distribución de la variable "Temperatura axilar"

Fuente: Elaboración propia

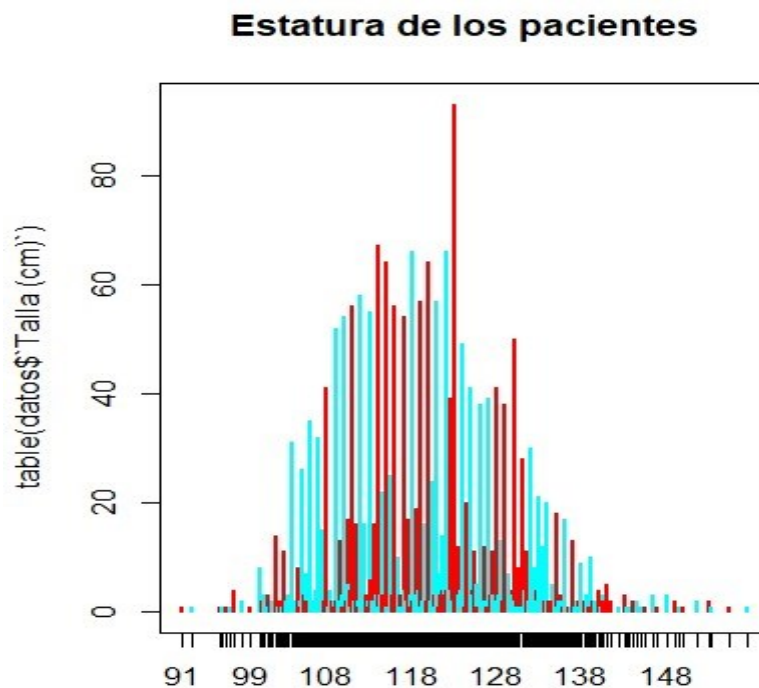
En el análisis de la temperatura axilar podemos observar que el mayor número de pacientes ronda entre los 36 °C, temperatura considerada dentro de los parámetros normales, pero existen casos que llegan a preocupar debido a que presentaban temperaturas de 35 °C y 41 °C estando estas consideradas dentro de los parámetros de hipotermia e hiperpirexia (fiebre) respectivamente, entendiendo que los pacientes acudieron a la casa de salud en condiciones críticas de salud.

Figura 10: Distribución de la variable "Peso"

Fuente: Elaboración propia

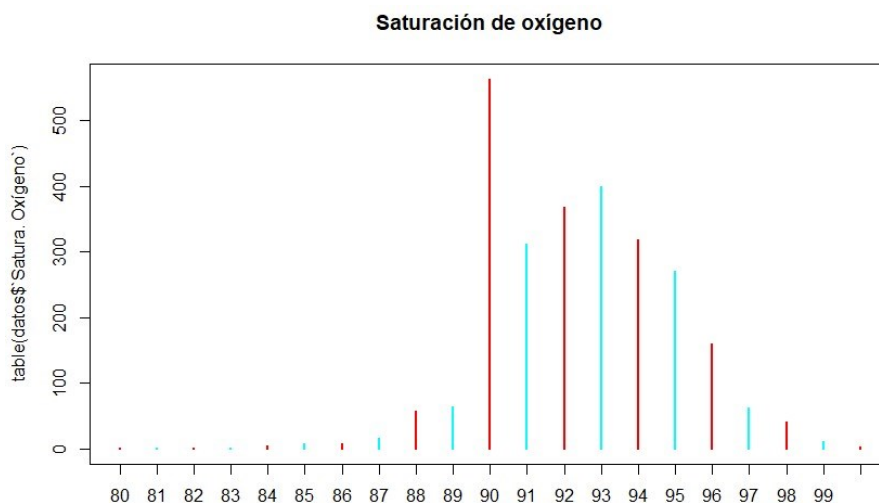
Con respecto al peso, podemos observar que la media es equivalente a 23.91 Kg dándonos a entender que la gran mayoría de pacientes registrados mantienen un peso adecuado para la edad que presentan, lo que si llama la atención es que existieron pacientes con tendencia a la desnutrición, ya que sus pesos fueron un tanto bajos con valores mínimos de 11.1 Kg.

Figura 11: Distribución de la variable "Talla"



Fuente: Elaboración propia

Con respecto al análisis de la estatura o talla, podemos decir que la media de los pacientes estudiados ronda en una media de 119.8 cm de estatura, esto puede ser debido a que la estatura promedio de los ecuatorianos no es muy alta, pero se registró que la talla más baja es equivalente a 91.0 cm dándonos a entender que se trata de un paciente de 5 años, siendo la edad más corta establecida para este estudio.

Figura 12: Distribución de la variable “Saturación de oxígeno”

Fuente: Elaboración propia

La saturación de oxígeno hace referencia a las cantidades de oxígeno que circulan por el torrente sanguíneo, siendo así, que un valor de 80 se encuentra fuera de los parámetros normales indicando que los pacientes se les debió otorgar servicios de salud de manera urgente y prioritaria para salvaguardar sus vidas.

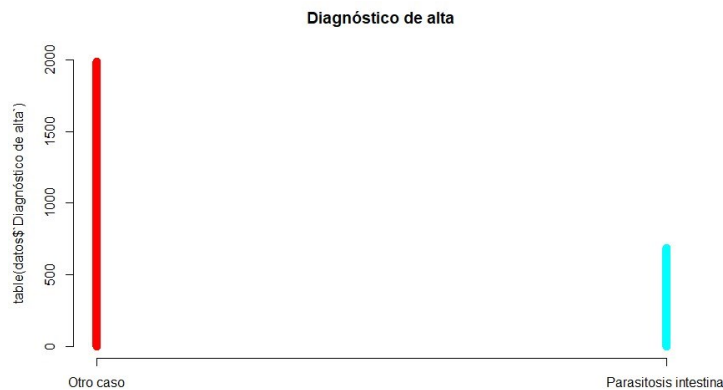
Figura 13: Distribución de la variable “Peso”

Fuente: Elaboración propia

En el análisis se observó que la gran mayoría de pacientes que acuden a esta casa de salud no poseen un seguro social, por lo cual los gastos de salud se deben cubrir por parte del estado, esto puede deberse a la situación tan crítica en la que se encuentra nuestro país actualmente los ecuatorianos, para obtener un seguro de

salud deben hacer aportaciones mensuales, las cuales solo una minoría de pacientes contaba con algún seguro público o privado.

Figura 14: Distribución de la variable “Diagnóstico de alta”



Fuente: Elaboración propia

Se observa que, en la ciudad de Riobamba y sus alrededores, presentan casos de parasitosis intestinal, y al estar esta casa de salud al servicio para transferencia de pacientes en estado crítico, asumimos que los casos atendidos de parasitosis fueron de carácter severo, lo cual puede estar atentando seriamente la salud de los niños e inclusive atentando contra la vida de estos.

Técnicas de modelado Modelo de clasificación: Regresión Logística Binaria

La clasificación mediante regresión logística tiene como objetivo estimar, clasificar e identificar las variables influyentes en la parasitosis intestinal; para la construcción del modelo mediante esta técnica se utilizó todo el conjunto de estudio.

La selección de las variables para el modelo se lo realizó con el método de escalonado hacia adelante que inicia solo con el término constante y va agregando al mismo modelo aquellas variables independientes cuyo nivel de asociación sea significativo estadísticamente respecto a la variable dependiente, obteniendo el modelo que a continuación se presenta:

Figura 15: Variables del Modelo de Regresión por el método de pasos hacia adelante

	B	Error estándar	Wald	gl	Sig.	Exp(B)	
Paso 1 ^a	Triage	-,254	,056	20,439	1	,000	,776
	Constante	-,179	,199	,806	1	,369	,836
Paso 2 ^b	Triage	-,298	,057	27,207	1	,000	,742
	Satur.oxig	,091	,018	24,580	1	,000	1,096
	Constante	-8,471	1,686	25,243	1	,000	,000
Paso 3 ^c	Triage	-,354	,063	31,265	1	,000	,702
	Temp.axilar	-,112	,055	4,157	1	,041	,894
	Satur.oxig	,082	,019	18,775	1	,000	1,085
	Constante	-3,257	3,052	1,139	1	,286	,038
Paso 4 ^d	Frec.car.min	,008	,003	10,237	1	,001	1,008
	Triage	-,347	,064	29,814	1	,000	,707
	Temp.axilar	-,198	,062	10,401	1	,001	,820
	Satur.oxig	,090	,019	22,049	1	,000	1,094
	Constante	-1,758	3,090	,324	1	,569	,172
Paso 5 ^e	Frec.car.min	,010	,003	13,565	1	,000	1,010
	Frec.resp.min	-,029	,012	5,588	1	,018	,971
	Triage	-,362	,064	32,096	1	,000	,696
	Temp.axilar	-,188	,062	9,300	1	,002	,828
	Satur.oxig	,085	,019	19,426	1	,000	1,089
	Constante	-,993	3,117	,101	1	,750	,370

Fuente: Elaboración propia

En el cuadro presentando anteriormente se observa que las variables en estudio: Triage, Temperatura axilar, Saturación de oxígeno, Frecuencia respiratoria mínima y Frecuencia cardiaca mínima son significativas (p -valor < 0.05) por lo tanto están incluidas al modelo y concluimos que influyen sobre el diagnostico de los pacientes atendidos en el HPAVR con un nivel de confianza del 95%. El modelo obtenido mediante regresión logística es el siguiente:

1

$$p_i = 1 + e^z$$

Donde

$$Z = -0,993 + 0,85(\text{Saturación oxígeno}) - 0,188(\text{Temperatura axilar}) - 0,362(\text{Triage}) + 0,29(\text{Frecuencia respiratoria mínima}) + 0,01(\text{Frecuencia cardíaca mínima})$$

Del modelo presentado se observa que el signo de los coeficientes en algunas variables es positivo y significa que la variable aumenta la probabilidad del suceso

estudiado, o lo que es lo mismo decir que aumenta la probabilidad de que un paciente sea diagnosticado con parasitosis intestinal.

Prueba ómnibus para la significancia del modelo

$H_0 =$ El modelo no es significativo

$H_1 =$ El modelo es significativo

Figura 16: Prueba ómnibus de coeficientes del modelo

		Chi-cuadrado	gl	Sig.
Paso 1	Escalón	20,609	1	,000
	Bloque	20,609	1	,000
	Modelo	20,609	1	,000
Paso 2	Escalón	24,921	1	,000
	Bloque	45,530	2	,000
	Modelo	45,530	2	,000
Paso 3	Escalón	4,196	1	,041
	Bloque	49,726	3	,000
	Modelo	49,726	3	,000
Paso 4	Escalón	10,283	1	,001
	Bloque	60,009	4	,000
	Modelo	60,009	4	,000
Paso 5	Escalón	5,646	1	,017
	Bloque	65,656	5	,000
	Modelo	65,656	5	,000

Fuente: Elaboración propia

El contraste sobre el nivel de significancia global del modelo se contrastó con el estadístico Razón de Verosimilitud (Prueba Ómnibus). El cuadro anterior nos muestra un valor de Chicuadrado de 65,656 con un p-valor (0.00) menor al nivel establecido 0.05, por lo que concluimos que hay una relación significativa entre las variables independientes y su predictora entonces el modelo es significativo.

Tabla de clasificación de Regresión Logística

El modelo estimado a través de regresión logística binaria presentó un error de predicción de 25,607%, la matriz de confusión asociada al modelo es la siguiente:

Tabla 1: Matriz de Confusión

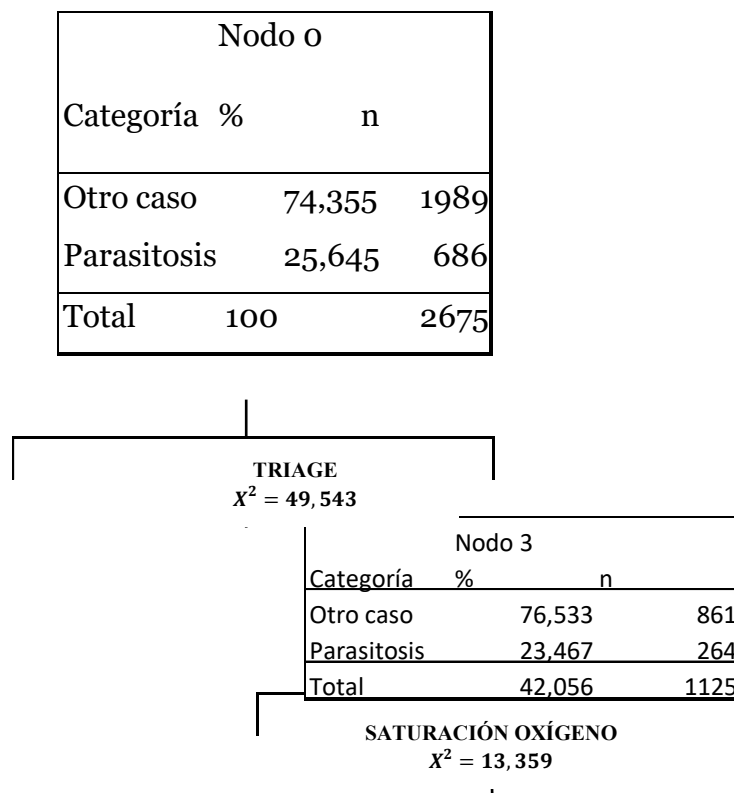
		Diagnóstico	
		Otro caso	Parasitismo
Diagnóstico	Otro caso	1991	0
	Parasitismo	684	0

Fuente: Elaboración propia

Modelo de clasificación: Árboles de Clasificación

La selección de las variables para el modelo por árboles de clasificación se utilizó el método de crecimiento CHAID, con un número de casos mínimo aceptable en un nodo padre de 100 y nodo hijo de 50 con profundidad 4 obtuvimos el siguiente resultado.

Figura 15: Árbol de clasificación



III;IV

Nodo 8			
Categoría	%	n	
Otro caso	67,871	169	
Parasitosis	32,129	80	
Total	9,308	249	

PESO	
-------------	--

$X^2 = 22,238$

< = 23.900

Nodo 18		
Categoría	%	n
Otro caso	55,97	75
Parasitosis	44,03	59
Total	5,009	134

SATURACION CARGENO	
---------------------------	--

$X^2 = 4,634$

< = 95.00

Nodo 25		
Categoría	%	n
Otro caso	47,143	33
Parasitosis	52,857	37
Total	2,617	70

Fuente: Elaboración propia

- a) El nodo 0 describe la variable dependiente: diagnóstico de alta, especificando el porcentaje de los pacientes registrados con parasitosis intestinal u otra enfermedad.
- b) La variable dependiente se ramifica en 4 nodos derivados de la variable “Triage” considerando esta como variable principal predictora.

- c) Fijamos atención en el nodo 2 puesto el valor de Chi-cuadrado es mayor a los otros, como nuestro objetivo de interés es conocer los factores que modelan la parasitosis intestinal nos concentramos en los nodos cuya ramificación se extienda hasta alcanzar el objeto de estudio.
- d) Los nodos 2, 3, 4 ramifican a los nodos 5,6,7,8,9,10 según la temperatura axilar, saturación de oxígeno y frecuencia cardiaca mínima respectivamente.
- e) Los nodos ramificados anteriormente se vuelven a abrir y se incluyen las variables Talla, Peso y Tipo de seguro de salud, llegando al penúltimo nivel del árbol las clasificaciones se han realizado adecuadamente para el diagnóstico 0 = “otro caso”, pues si observamos la selección de los grupos con respecto a los porcentajes a una profundidad de 3 la mayoría de los pacientes registro otra enfermedad diferente a la parasitosis intestinal f) A una profundidad de 4 el nodo terminal 25 partiendo del nodo 3 con saturación de oxígeno > 94 el 67.87% no son pacientes con parasitosis, este nodo se ramifica de acuerdo con el Peso el nodo 18 agrupa el 55.97% de pacientes con otro diagnóstico si el peso es inferior a 23.9. Continuando la ruta para pacientes con saturación de oxígeno ≤ 95 visualizamos el objetivo y en este nodo el 52,86% está clasificado con diagnóstico de parasitosis intestinal.
- g) Finalmente, en síntesis, los nodos que definen el diagnóstico de los pacientes con parasitosis intestinal Frecuencia cardiaca min, Triage, Temperatura axilar, Peso, Talla, Saturación de oxígeno, Tipo seguro de salud.

Tabla de clasificación: Algoritmo CHAID

El modelo estimado a través de regresión logística binaria presento un error de predicción de 24,495%, la matriz de confusión asociada al modelo es la siguiente:

Tabla 2: Matriz de confusión árbol clasificación

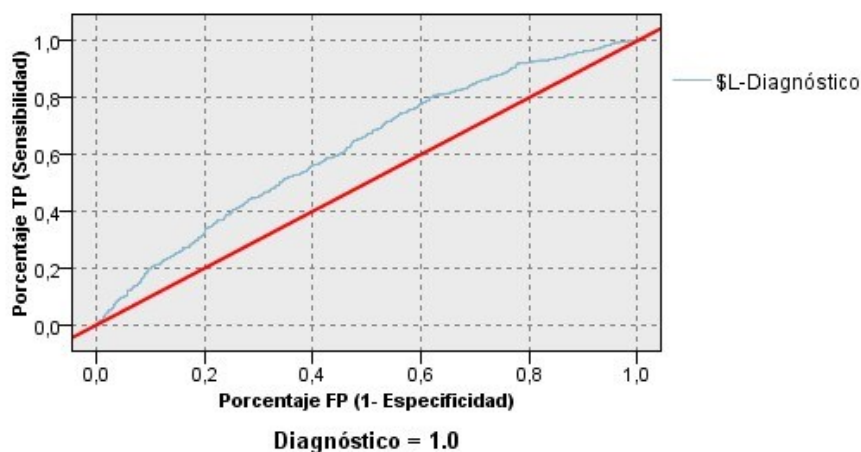
		Diagnóstico	
		Otro caso	Parasitismo
Diagnóstico	Otro caso	1993	0
	Parasitismo	682	0

Fuente: Elaboración propia

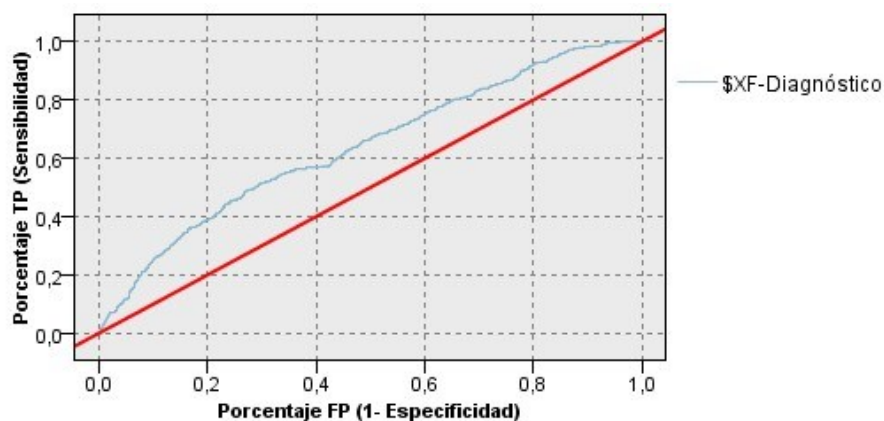
Evaluación de las técnicas de clasificación Comparación de los modelos: Regresión Logística y Árboles de Clasificación

Considerando la tasa de error mediante la matriz de confusión de cada modelo encontrado, (Tabla 3-1, Tabla 3-2) se analizó la capacidad predictiva de estos mediante el AUC de la curva ROC y sus graficas respectivas, a continuación, se presenta los resultados:

Figura 17: Curva ROC: Regresión Logística



Fuente: Elaboración propia

Figura 18: Curva ROC: Árbol de Clasificación

Diagnóstico = 1.0

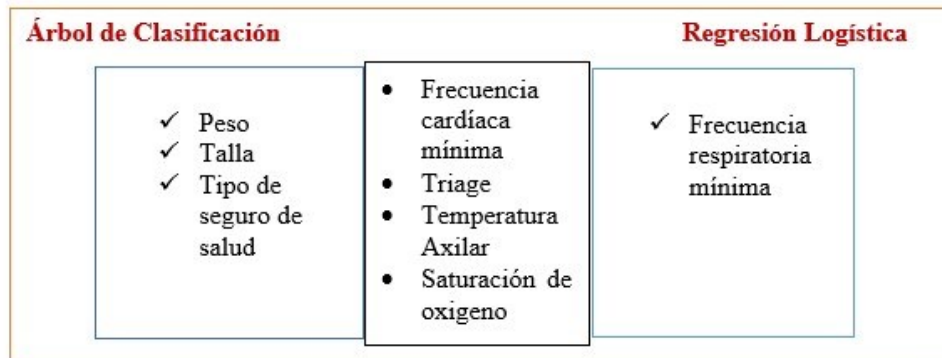
Fuente: Elaboración propia

Tabla 3: Áreas bajo la curva (AUC)

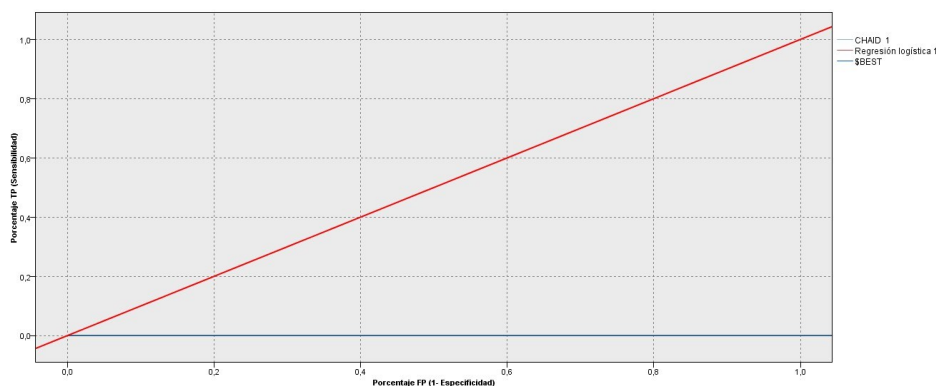
	Área bajo la curva (AUC)
Regresión Logística	62,8%
Árbol de clasificación	65,7%

Fuente: Elaboración propia

Una vez analizado las representaciones graficas de la curva ROC y el área bajo la curva (AUC) de los modelos encontrados, se observa que el modelo encontrado mediante Árbol de clasificación presenta un mayor AUC = 65,7% por lo tanto tiene mayor capacidad predictiva a comparación de la regresión logística cuyo AUC = 62,8%, y esto tiene sentido ya que al comparar con la matriz de confusión según la tasa del error, los árboles de clasificación, clasificaría el diagnostico de los pacientes con un error mínimo.

Figura 19: Comparativa entre los factores asociados mediante RI y AC

Fuente: Elaboración propia

Figura 3-20: Curva ROC: Regresión Logística vs. Árbol de Clasificación

Fuente: Elaboración propia

Conclusiones

Conforme a los datos obtenidos, se demostró que los árboles de clasificación, con un área bajo la curva ROC equivalente al 65.7%, son más eficientes que los modelos de regresión logística, los cuales presentan un área bajo la curva ROC del 62.8%.

Los posibles factores asociados a la parasitosis intestinal fueron determinados mediante un análisis teórico dándonos como resultado los siguientes: Temperatura axilar, peso, talla, edad, saturación de oxígeno, presión arterial, grupos culturales, frecuencia respiratoria, frecuencia cardíaca, y las variables anexadas al estudio fueron determinadas como posibles factores influyentes en el sector de estudio.

El análisis de las historias clínicas brindó las variables para este estudio, y a su vez permitió consolidar la base de datos que fue utilizada, aplicando las respectivas restricciones como el grupo etario, análisis discriminante de datos para poder obviar a los pacientes con signos vitales erróneamente establecidos, etc. mejorando la calidad del trabajo presentado.

Para realizar predicciones sobre parasitosis intestinal en el Hospital Pediátrico Alfonso Villagómez Román podemos utilizar los árboles de clasificación, si bien no tienen una capacidad predictora muy elevada, funcionan de mejor manera que la regresión logística.

Los árboles de clasificación presentan un 2.9% más eficiencia que el modelo de regresión logística obtenido. Una causa de este suceso es que los árboles de clasificación presentan porcentajes de discriminación dejando en claro cuáles son los límites de aceptación en los valores de las variables estudiadas.

Referencias

Abellana Sangra, R., & Farran Codina, A. (2015). Identificación, impacto y tratamiento de datos perdidos y atípicos en epidemiología nutricional. Recuperado el 1 de 01 de 2023, de https://www.renc.es/imagenes/auxiliar/files/NUTR.%20COMUN.%20SUPL.%2012015_Tratamiento%20atipicos.pdf

Apt Baruch, W. L. (2013). Parasitología Humana (1ra ed.). M.G.H. Recuperado el 12 de 11 de 2022

Batista Rojas, O., & Álvarez Hernández, Z. (2013). Parasitismo intestinal en niñas mayores de 5 años de Ciudad Bolívar. *SciELO*, 17(4). Recuperado el 2 de 2 de 2023, de http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1029-30192013000400004

Bonilla Pulgar, G. E., & Bonilla Nina, G. E. (2020). *Revista de Historia, Patrimonio, Arqueología y Antropología Americana*. Recuperado el 20 de 10 de 2022, de <http://www.rehpa.net/ojs/index.php/rehpa/article/view/27/57>

- Congacha Ortega, G. N. (11 de 12 de 2020). DSpace ESPOCH. Recuperado el 24 de 7 de 2022, de <http://dspace.esPOCH.edu.ec/handle/123456789/14551>
- De'ath, G., & Fabricius, K. (2000). Classification and Regression Trees: A Powerful Yet Simple Technique for Ecological Data Analysis. *Ecology*. Recuperado el 9 de Junio de 2022, de <https://cybertesis.unmsm.edu.pe/handle/20.500.12672/7122>
- Global Health, Division of Parasitic Diseases and Malaria. (4 de enero de 2022). CDC. Recuperado el 11 de 11 de 2022, de Acerca de los parásitos: <https://www.cdc.gov/parasites/es/about.html>
- Gujarati, D. N., & Porter, D. C. (2010). *Econometría* (Quinta ed.). (Interamericana, Ed.) México DF, DF, Mexico: Mc Graw-Hill.
- Lema Punín, D. C., & Inga Miguitama, M. A. (11 de noviembre de 2018). FRECUENCIA DE PARASITOSIS INTESTINAL POR MICROSCOPIA DIRECTA EN LOS ESTUDIANTES DE LAS ESCUELAS RURALES DE LA PARROQUIA SAN BARTOLOMÉ-2017. Recuperado el 12 de 11 de 2022, de <http://dspace.ucuenca.edu.ec/handle/123456789/30073>
- Lizares Castillo, M. (2017). Comparación de modelos de clasificación: regresión logística y árboles de clasificación para evaluar el rendimiento académico. CYBERTESIS. Obtenido de <https://hdl.handle.net/20.500.12672/7122>
- Lojano Collaguazo, R. I., & Lojano Punin, M. A. (11 de Noviembre de 2018). PREVALENCIA DE ENTEROPARASITOSIS Y FACTORES DE RIESGO EN ESCOLARES DE LA UNIDAD EDUCATIVA CHIQUINTAD, 2017. Recuperado el 14 de Noviembre de 2022, de <http://dspace.ucuenca.edu.ec/handle/123456789/30073>
- Menacho Chávez, C. M. (2022). Repositorio Universidad Técnica del Norte. Recuperado el 2 de 2 de 2023, de FACTORES DE RIESGO DE PARASITOSIS INTESTINAL EN MENORES DE 5 AÑOS CENTRO DE SALUD IMANTAG, 2022:

- <http://repositorio.utn.edu.ec/bitstream/123456789/12737/2/06%20ENF%201308%20TRA%20BAJO%20DE%20GRADO.pdf>
- Montero Perez, A. P., & Huilca Peralta, G. D. (2018). Repositorio Institucional Continental. Recuperado el 2 de 2 de 2023, de <https://hdl.handle.net/20.500.12394/8796>
- OMS. (2008). Noticias ONU. Recuperado el 18 de 10 de 2022, de <https://news.un.org/es/story/2008/08/1140951>
- Paho. (s.f.). Paho.org. Recuperado el 12 de 11 de 2022, de <https://www3.paho.org/spanish/ad/dpc/cd/psit-programpage.htm#:~:text=La%20OPS%2FOMS%2ocalcula%2oque,95%25%20en%20algunas%20otribus%20ind%C3%ADgenas>.
- Pazmiño Gómez, B. J., Pérez Ayol, L., López Orozco, L., Vinueza Freire, W., Cadena Alvarado, J., Rodas Pazmiño, J., . . . Rodas Neira, E. (2018). Parasitosis intestinal y estado nutricional en niños de 1-3 años de un centro infantil del cantón Milagro. *Ciencia Unemi*, 11(26), 143-149. doi: <https://doi.org/10.29076/issn.2528-7737vol6iss10.2013pp9-18p>
- Peña, D. (2002). Datos Atípicos. En *Análisis de Datos Multivariantes*.
- Perez Rave, J., & Gonzalez Echeverria, F. (2018). Árboles de clasificación vs regresión logística en el desarrollo de competencias genéricas en ingeniería. *Scielo*, 22(4), 1519-1541. doi:<https://doi.org/10.13053/cys-22-4-2804>.
- Quimica.es. (s.f.). Huesped (biología). Recuperado el 12 de 11 de 2022, de https://www.quimica.es/enciclopedia/Hu%C3%A9sped_%28biolog%C3%A9da%29.html
- Revista Panamericana de la Salud Pública. (2008). Prevalencia de parasitismo intestinal en niños quechuas de zonas rurales montañosas de Ecuador. *Scielo*. Recuperado el 2 de 2 de 2023, de <https://scielosp.org/article/rpsp/2008.v23n2/125-125/>
- Reynolds, L. (2021). eHOW En Español. ¿Hervir el agua la hace destilada? Recuperado el 16 de 11 de 2022, de https://www.ehowenespanol.com/hervir-agua-destiladacomo_173637/

-
- Serna Pineda, S. C. (2009). Comparación de árboles de regresión y clasificación y regresión logística. Recuperado el 2 de 2 de 2023, de <https://repositorio.unal.edu.co/handle/unal/2421>
- Srivastava, T. (2019). 11 Important Model Evaluation Metrics for Machine Learning Everyone should know. Analytics Vidhya.
- Timofeev, R. (2004). Classification and Regression Trees (CART) Theory and Applications.
Recuperado el 2023 de 2 de 2, de https://www.academia.edu/13700196/Classification_and_Regression_Trees_CART_Theory_and_Applications
- Vallejo Samaniego, D. G. (2010). Wordpress. Recuperado el 20 de 10 de 2022, de <https://digvas.wordpress.com/personajes/alfonso-villagomez-roman/#:~:text=Este%20m%C3%A9dico%20oriobambe%C3%B1o%20naci%C3%B3n%20e%20l,y%20a%20aliviar%20los%20dolores%20humanos>.
- Viada, C., Carlos, B., Ballesteros, J., Fors, M., Robaina, M., & Uranga, R. (2016). REVISIÓN SISTEMÁTICA DE LOS MÉTODOS DE IMPUTACIÓN DE DATOS FALTANTES.