



**ESCUELA SUPERIOR POLITÉCNICA DE CHIMBORAZO**  
**FACULTAD DE CIENCIAS**  
**CARRERA ESTADÍSTICA**

**COMPARACIÓN ENTRE LOS MODELOS DE REGRESIÓN  
LOGÍSTICA Y REDES NEURONALES ARTIFICIALES PARA  
DETERMINAR LOS FACTORES MATERNOS QUE INFLUYEN EN  
EL PESO DEL BEBÉ AL NACER EN INSTITUTO ECUATORIANO  
DE SEGURIDAD SOCIAL DE LA CIUDAD DE RIOBAMBA EN EL  
AÑO 2020 – 2021.**

**Trabajo de Titulación**

Tipo: Proyecto de Investigación

Presentado para optar el grado académico de:

**INGENIERA ESTADÍSTICA**

**AUTORAS:**

ALLAYCO ALVAREZ DIGNA ROCIO  
HUARACA GUASHPA MARÍA FERNANDA

**DIRECTORA:** ING. JOHANNA ENITH AGUILAR REYES, MGS.

Riobamba – Ecuador

2023

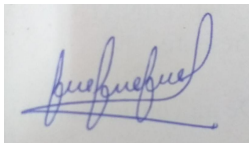
**©2023, Digna Rocio Allayco Alvarez y María Fernanda Huaraca Guashpa**

Autorizamos la reproducción total o parcial, con fines académicos, por cualquier medio o procedimiento, incluyendo la cita bibliográfica del documento, siempre y cuando se reconozca el Derecho de Autor.

Nosotras, Digna Rocio Allayco Alvarez y María Fernanda Huaraca Guashpa, declaramos que el presente Trabajo de Titulación es de nuestra autoría y los resultados del mismo son auténticos. Los textos en el documento que provienen de otras fuentes están debidamente citados y referenciados.

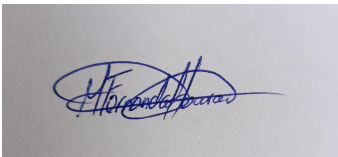
Como autores asumimos la responsabilidad legal y académica de los contenidos de este Trabajo de Titulación; el patrimonio intelectual pertenece a la Escuela Superior Politécnica de Chimborazo.

Riobamba, 29 de noviembre de 2023

A handwritten signature in blue ink, appearing to read 'Digna Rocio Allayco Alvarez', written over a horizontal line.

**Allayco Alvarez Digna Rocio**

**0606029783**

A handwritten signature in blue ink, appearing to read 'María Fernanda Huaraca Guashpa', written over a horizontal line.

**Huaraca Guashpa María Fernanda**

**0604875880**

**ESCUELA SUPERIOR POLITÉCNICA DE CHIMBORAZO**  
**FACULTAD DE CIENCIAS**  
**CARRERA ESTADÍSTICA**

El Tribunal del Trabajo de Titulación, certifica que: el Trabajo de Titulación; Tipo: Proyecto de Investigación. **COMPARACIÓN ENTRE LOS MODELOS DE REGRESIÓN LOGÍSTICA Y REDES NEURONALES ARTIFICIALES PARA DETERMINAR LOS FACTORES MATERNOS QUE INFLUYEN EN EL PESO DEL BEBÉ AL NACER EN INSTITUTO ECUATORIANO DE SEGURIDAD SOCIAL DE LA CIUDAD DE RIOBAMBA EN EL AÑO 2020-2021**, realizado por las señoritas: **ALLAYCO ALVAREZ DIGNA ROCIO** y **HUARACA GUASHPA MARÍA FERNANDA**, ha sido minuciosamente revisado por los Miembros del Tribunal del Trabajo de Titulación, el mismo que cumple con los requisitos científicos, técnicos, legales, en tal virtud el Tribunal autoriza su presentación.

|   | <b>FIRMA</b>   | <b>FECHA</b> |
|---|--|--------------|
| Dra. Jaqueline Elizabeth Balseca Castro<br><b>PRESIDENTA DEL TRIBUNAL</b>           |  | 2023-11-29   |
| Ing. Johanna Enith Aguilar Reyes, Mgs<br><b>DIRECTORA DEL TRABAJO DE TITULACIÓN</b> |  | 2023-11-29   |
| Msc. Tania Paulina Morocho Barrionuevo<br><b>ASESORA DEL TRABAJO DE TITULACIÓN</b>  |  | 2023-11-29   |

## **DEDICATORIA**

A Dios por ser mi fortaleza en los momentos más difíciles y adversos de mi corta vida, por darme una maravillosa y espléndida familia que me apoya de una forma incondicional a pesar de mis errores. A mis padres María Lucila Alvarez y Juan Manuel Allayco que siempre se mantuvieron a mi lado guiándome en cada paso, dándome la mano cuando me caía y estaba a punto de rendirme, no existen palabras para expresar mi infinito amor hacia ellos y darme la oportunidad de seguir mis sueños. A mis hermanos Liliana y Luis Allayco por el cariño y apoyo absoluto que me brindan a pesar de mis tropiezos.

**Digna**

Agradezco en primer lugar a Dios por ser quien guía mis pasos y no me abandona en mis momentos más difíciles. A mis queridos padres Juan Carlos Huaraca y Nancy Olivia Guashpa quienes me han inculcado valores y han moldeado a la persona en la que hoy me he convertido. Por ser el pilar fundamental en mi vida, por su paciencia, apoyo incondicional y su amor infinito, gracias a ellos he podido culminar esta etapa de mi vida. A mi hermano Jhonn Alexander Huaraca por el gran cariño y amor brindado, por ser un gran ejemplo en mi vida y un apoyo incondicional. Gracias familia por apoyarme y darme toda la fortaleza para poder cumplir todos mis sueños, sin ustedes nada de esto hubiera sido posible.

**María**

## **AGRADECIMIENTO**

Primeramente, agradecerle a Dios por permitirme culminar una etapa en mi vida, por darme la suficiente fuerza para seguir y jamás rendirme en mi travesía. A mis padres por darme la vida, brindarme su amor infinito y sus sabios consejos cuando todo lo veía perdido. A mi hermana Liliana por estar allí para mí, escuchándome siempre, por tu apoyo infinito, por ser la razón por la cual quiero ser una mejor persona cada día. A la Ing. Johanna Aguilar y la Ing. Paulina Morocho por haber impartido sus valiosos conocimientos para poder culminar con el presente trabajo de Investigación.

**Digna**

Mi más sincero agradecimiento a mi familia, a mi madre, padre y hermano que han sido mi inspiración para seguir adelante, gracias por todo el apoyo que me han brindado durante toda mi etapa estudiantil, ya que existió muchos momentos difíciles que sin su ayuda, consejos y amor no hubiera podido culminar ésta meta, este triunfo es más suyo que mío. Agradecer a la Escuela Superior Politécnica de Chimborazo que me abrió las puertas para construir lo que alguna vez fue un sueño, así mismo a la Ing. Johanna Aguilar y la Ing. Paulina Morocho quienes me han guiado y han compartido sus conocimientos en este trabajo de investigación.

**María**

## ÍNDICE DE CONTENIDO

|  |     |
|--|-----|
| ÍNDICE DE TABLAS . . . . .   | ix  |
| ÍNDICE DE ILUSTRACIONES . . . . .  | x   |
| RESUMEN . . . . .  | xi  |
| ABSTRACT . . . . .   | xii |
| INTRODUCCIÓN . . . . .   | 1   |
| <b>CAPÍTULO I</b>  |     |
| 1. PROBLEMA DE INVESTIGACIÓN . . . . .   | 3   |
| 1.1. Planteamiento del problema . . . . .  | 3   |
| 1.2. Limitaciones y delimitaciones . . . . .   | 3   |
| 1.3. Problema general de investigación . . . . .   | 3   |
| 1.4. Problemas específicos de investigación . . . . .                                      | 4   |
| 1.5. Objetivos . . . . .   | 4   |
| 1.5.1. <i>Objetivo general</i> . . . . .   | 4   |
| 1.5.2. <i>Objetivos específicos</i> . . . . .  | 4   |
| 1.6. Justificación . . . . .   | 4   |
| 1.6.1. <i>Justificación teórica</i> . . . . .  | 4   |
| 1.6.2. <i>Justificación práctica</i> . . . . .   | 5   |
| <b>CAPÍTULO II</b>   |     |
| 2. MARCO TEÓRICO . . . . .   | 6   |
| 2.1. Antecedentes de investigación . . . . .   | 6   |
| 2.2. Marco teórico . . . . .   | 7   |
| 2.2.1. <i>Modelo de regresión logística</i> . . . . .                                      | 7   |
| 2.2.2. <i>Estimación de los parámetros del modelo (Función de verosimilitud)</i> . . . . . | 8   |
| 2.2.3. <i>Hipótesis en regresión logística simple</i> . . . . .                            | 9   |
| 2.2.4. <i>Bondad de ajuste del modelo</i> . . . . .  | 10  |

|        |  |    |
|--------|--|----|
| 2.2.5. | <i>Redes neuronales artificiales</i>                     | 11 |
| 2.2.6. | <i>Curva característica operativa del receptor (ROC)</i> | 12 |
| 2.2.7. | <i>Área bajo la curva (AUC)</i>                          | 13 |
| 2.2.8. | <i>Comparar dos o más curvas ROC</i>                     | 14 |
| 2.3.   | <b>Marco conceptual</b>                                  | 14 |
| 2.3.1. | <i>Gestación</i>   | 14 |
| 2.3.2. | <i>Enfermedades durante el periodo de gestación</i>      | 14 |
| 2.3.3. | <i>Desarrollo fetal y el entorno materno</i>             | 15 |
| 2.3.4. | <i>Peso de un recién nacido</i>                          | 15 |
| 2.3.5. | <i>Bajo peso de un recién nacido</i>                     | 16 |
| 2.3.6. | <i>Muerte neonatal</i>                                   | 16 |

### CAPÍTULO III

|        |  |    |
|--------|--|----|
| 3.     | <b>MARCO METODOLÓGICO</b>                  | 17 |
| 3.1.   | <b>Tipo de investigación</b>               | 17 |
| 3.2.   | <b>Diseño de investigación</b>             | 17 |
| 3.3.   | <b>Localización del estudio</b>            | 17 |
| 3.4.   | <b>Población del estudio</b>               | 18 |
| 3.5.   | <b>Tamaño de la muestra</b>                | 18 |
| 3.6.   | <b>Método de muestreo</b>                  | 18 |
| 3.7.   | <b>Técnicas de recolección de datos</b>    | 18 |
| 3.8.   | <b>Modelo estadístico</b>                  | 18 |
| 3.9.   | <b>Identificación de variables</b>         | 18 |
| 3.9.1. | <i>Variable dependiente</i>                | 18 |
| 3.9.2. | <i>Variables independientes</i>            | 18 |
| 3.10.  | <b>Operacionalización de las variables</b> | 19 |

### CAPÍTULO IV

|        |   |    |
|--------|---|----|
| 4.     | <b>MARCO DE ANÁLISIS E INTERPRETACIÓN DE RESULTADOS</b> | 23 |
| 4.1.   | <b>Análisis descriptivo</b>                             | 23 |
| 4.2.   | <b>Regresión logística</b>                              | 31 |
| 4.2.1. | <i>Regresión logística binaria</i>                      | 31 |



|             |   |    |
|-------------|---|----|
| <b>4.3.</b> | <b>Redes neuronales artificiales</b>                            | 32 |
| <b>4.4.</b> | <b>Comparación del modelo de regresión con redes neuronales</b> | 35 |
|             | <b>CONCLUSIONES</b>   | 36 |
|             | <b>RECOMENDACIONES</b>  | 37 |
|             | <b>BIBLIOGRAFÍA</b>   |    |
|             | <b>ANEXOS</b>   |    |

## ÍNDICE DE TABLAS

|  |    |
|--|----|
| <b>Tabla 3–1</b> Operacionalización de las variables . . . . .                                       | 19 |
| <b>Tabla 4–1</b> Operacionalización de las variables . . . . .                                       | 23 |
| <b>Tabla 4–2</b> Frecuencias absolutas y relativas de la variable Hábitos tóxicos . . . . .          | 24 |
| <b>Tabla 4–3</b> Frecuencias absolutas y relativas de la variable Abortos anteriores . . . . .       | 25 |
| <b>Tabla 4–4</b> Frecuencias absolutas y relativas de la variable Antecedentes patológicos . . . . . | 26 |
| <b>Tabla 4–5</b> Frecuencias absolutas y relativas de la variable Antecedentes alérgicos . . . . .   | 27 |
| <b>Tabla 4–6</b> Frecuencias absolutas y relativas de la variable Bajo peso . . . . .                | 29 |
| <b>Tabla 4–7</b> Frecuencias absolutas y relativas de la variable Sexo del bebé . . . . .            | 30 |
| <b>Tabla 4–8</b> Frecuencias absolutas de los pesos del bebé en el entrenamiento . . . . .           | 32 |
| <b>Tabla 4–9</b> Frecuencias absolutas de los pesos del bebé para la validación . . . . .            | 32 |
| <b>Tabla 4–10</b> Porcentaje de exactitud de los modelos . . . . .                                   | 33 |
| <b>Tabla 4–11</b> Matriz de confusión . . . . .  | 34 |
| <b>Tabla 4–12</b> Áreas bajo la curva de los modelos de regresión logística y red neuronal . . . . . | 35 |

## ÍNDICE DE ILUSTRACIONES

|   |    |
|---|----|
| <b>Ilustración</b> : 2-1Diagrama de una neurona artificial . . . . .  | 12 |
| <b>Ilustración</b> : 2-2Curva Característica operativa del receptor . . . . .   | 13 |
| <b>Ilustración</b> : 4-1Base de datos de las madres en estado de gestación y los pesos de los recién nacidos . . . . .    | 23 |
| <b>Ilustración</b> : 4-2Diagrama de barras de las frecuencias absolutas de la variable Hábitos tóxicos . . . . .          | 25 |
| <b>Ilustración</b> : 4-3Diagrama de barras de las frecuencias absolutas de la variable Abortos anteriores . . . . .       | 26 |
| <b>Ilustración</b> : 4-4Diagrama de barras de las frecuencias absolutas de la variable Antecedentes patológicos . . . . . | 27 |
| <b>Ilustración</b> : 4-5Diagrama de barras de las frecuencias absolutas de la variable Antecedentes alérgicos . . . . .   | 28 |
| <b>Ilustración</b> : 4-6Diagrama de barras de las frecuencias absolutas de la variable Bajo peso                          | 29 |
| <b>Ilustración</b> : 4-7Diagrama de barras de las frecuencias absolutas de la variable Sexo del bebé                      | 30 |
| <b>Ilustración</b> : 4-8Codificación de las variables categóricas . . . . .   | 31 |
| <b>Ilustración</b> : 4-9Resumen del modelo de regresión logística . . . . .   | 31 |
| <b>Ilustración</b> : 4-10Diagrama de barras del porcentaje de exactitud de los modelos . . . . .                          | 33 |
| <b>Ilustración</b> : 4-11Red neuronal artificial del mejor modelo . . . . .   | 34 |
| <b>Ilustración</b> : 4-12Curvas de ROC para los modelos logísticos y red neuronal . . . . .                               | 35 |

## RESUMEN

El bajo peso al nacer se ha convertido en un problema de salud a nivel mundial, afectando de manera negativa a corto y largo plazo en el desarrollo del bebé. El presente trabajo tuvo como objetivo principal determinar los factores maternos que influyen en el peso del recién nacido en el IESS-Riobamba el año 2020-2021. La base de datos se obtuvo del área de estadística, con un total de 273 mujeres. Para el procesamiento de los modelos se utilizó el software R en donde se codificaron las variables cualitativas y posterior crear los modelos de las 2 técnicas. En el análisis estadístico descriptivo se obtuvo como resultado que la edad mínima de las madres fue de 15 años y la edad máxima de 46 años, se determinó que las enfermedades más comunes durante el embarazo fueron preeclampsia, diabetes gestacional, entre otros. Por último, se apreció que el 35.5 % de recién nacidos se encontraban con un bajo peso mientras que el 64.5 % contaban con un peso adecuado. En el modelo de regresión logística se obtuvo 7 variables significativas: hábitos tóxicos, semanas de gestación, peso de la madre, paridad, abortos, diagnóstico del embarazo e IMC, mientras que en el modelo de redes neuronales se obtuvo un modelo con 2 capas ocultas con 6 y 3 neuronas respectivamente. Finalmente, en las curvas ROC se obtuvo un área bajo la curva (AUC) de 91.8 % en la regresión logística y un 85.9 % en redes neuronales artificiales llegando a la conclusión que el modelo de regresión logística es el mejor modelo para determinar factores. Se recomienda a las instituciones de salud públicas y privadas que les interese la investigación, divulgar esta información a las madres que se encuentran en estado de gestación para que tomen medidas preventivas y su bebé no nazca con bajo peso.

**Palabras clave:** <PESO>, <REGRESIÓN LOGÍSTICA>, <REDES NEURONALES>, <ÁREA BAJO LA CURVA>, <CURVAS ROC>.



## SUMMARY/ABSTRACT

Low birth weight has become a worldwide health problem, negatively affecting the short- and long-term development of the baby. The main objective of this study was to determine the maternal factors that influence newborn birth weight at IESS-Riobamba in 2020-2021. The database was obtained from the statistics area, with a total of 273 women. For the processing of the models, the R software was used to code the qualitative variables and subsequently create the models of the 2 techniques. The descriptive statistical analysis showed that the minimum age of the mothers was 15 years and the maximum age was 46 years; it was determined that the most common diseases during pregnancy were preeclampsia, gestational diabetes, among others. Finally, it was found that 35.5% of the newborns were underweight while 64.5% were of adequate weight. In the logistic regression model, 7 significant variables were obtained: toxic habits, weeks of gestation, maternal weight, parity, miscarriages, pregnancy diagnosis and BMI, while in the neural network model a model with 2 hidden layers was obtained with 6 and 3 neurons respectively. Finally, in the ROC curves, an area under the curve (AUC) of 91.8% was obtained in logistic regression and 85.9% in artificial neural networks, concluding that the logistic regression model is the best model for determining factors. It is recommended that public and private health institutions interested in this research disseminate this information to pregnant mothers so that they can take preventive measures to prevent their babies from being born with low birth weight.

**Keywords:** <WEIGHT>, <LOGISTIC REGRESSION>, <NEURAL NETWORKS>, <AREA UNDER THE CURVE>, <ROC CURVES>.



Edgar Mesias Jaramillo Moyano  
0603497397

## INTRODUCCIÓN

Las redes neuronales artificiales y la regresión logística hoy en día se han convertido en técnicas muy habituales para realizar investigaciones, de esta manera (Moreno, 2002, p.7) menciona que las Redes Neuronales Artificiales son las que actualmente están causando un mayor impacto, debido a su extraordinaria aplicabilidad práctica. Recientemente esta tecnología ha captado la atención de los profesionales dedicados a la estadística y al análisis de datos, los cuales comienzan a incorporar las redes neuronales al conjunto de herramientas estadísticas orientadas a la clasificación de patrones y la estimación de variables continuas.

La regresión logística (RL) es uno de los instrumentos estadísticos más expresivos y versátiles de que se dispone para el análisis de datos en clínica y epidemiología. Su uso se universaliza y expande desde principios de los ochenta debido, especialmente, a las facilidades informáticas con que se cuenta desde entonces. En los últimos años se ha verificado una presencia muy marcada de esta técnica, tanto en la literatura orientada a tratar temas metodológicos como en los artículos científicos biomédicos. Fiel reflejo de esta tendencia es que el empleo de la RL suponía el 32 % de los artículos publicados por American Journal of Epidemiology de 1986 a 1990 y el 68 % de los que aparecieron en el mencionado quinquenio en New England Journal of Medicine, con lo cual quedó ubi-cada en el quinto puesto, solo superada por cuatro técnicas convencionales: t de Student, prueba Chi-cuadrado, análisis de la varianza y prueba de Fisher (OMS ,2022b: p.1).

Un primer estudio respecto a “Comparación de la red neuronal artificial y el modelo de regresión logística para los factores que afectan el peso al nacer” (Kirişci, 2019, p.3) que tuvo por objetivo comparar el análisis de redes neuronales artificiales (ANNs) y regresión logística para determinar los factores que afectan el peso al nacer. Este estudio incluyó a 223 recién nacidos. Se utilizaron los registros de bebés nacidos entre enero de 2017 y diciembre de 2017. Los datos se obtuvieron del distrito de Beykoz de Estambul. Se evaluó el análisis ANNs y regresivo logístico del método obtenido en base a estos registros. Se dio a conocer los resultados obtenidos indicando que, en el diagnóstico médico en términos de su capacidad de clasificación, estos dos métodos son aproximadamente iguales entre sí. Ninguno de los dos modelos puede cambiar al otro. Ambos modelos se pueden utilizar como complemento para ayudar con la toma de decisiones. Ambos modelos tienen el potencial de ayudar a los médicos con respecto a la comprensión de los factores de riesgo de peso del bebe y la estimación del riesgo.

En el estudio “Comparación entre regresión logística y redes neuronales para predecir cáncer de

piel en perros” (Chavez, 2019, p.2) aborda la comparación de éstas dos técnicas de redes neuronales artificiales y regresión logística para la predicción de cáncer de piel en perros. Las características que se han analizado son la edad, el sexo, raza, exposición al sol, albinismo y la aparición de dermatitis. Dichas características fueron validadas por el método de coeficiente de correlación y el análisis de componente principal. Los resultados obtenidos demostraron que la red neuronal artificial con validación cruzada supera al modelo de regresión logística. El valor de predicción generado por la red neuronal fue de 89.6 % mientras que la regresión logística obtuvo un 84 %.

Los trabajos de investigación y artículos mostrados anteriormente muestran la importancia y el aporte de la comparación de redes neuronales artificiales y regresión logística para determinar factores maternos que influyen en el peso del bebe al nacer, colaborando con guías de información para futuras investigaciones que consoliden las hipótesis investigativas establecidas.

## CAPÍTULO I

### 1. PROBLEMA DE INVESTIGACIÓN

#### 1.1. Planteamiento del problema

En los últimos años el bajo peso al nacer se ha convertido en un problema de salud pública a nivel mundial, siendo una de las principales causas de muerte fetal y neonatal en el transcurso de los primeros 5 años de vida, así como también en el retraso del desarrollo físico y cognitivo del recién nacido, afectando así de manera negativa a corto y largo plazo.

La UNICEF señala que cada año nacen alrededor de 20 millones de bebés con un peso inferior a los 2500g. Perú tiene el 11 % de niños recién nacidos con bajo peso, siendo así el porcentaje más alto en Sudamérica, seguido de Colombia y Venezuela con un 9% ocupando el segundo lugar. Según el Instituto Nacional de Estadísticas y Censo (OMS, 2022a: p.1), menciona que, en el año 2021, se cuenta con un porcentaje de 8,6% nacidos vivos con bajo peso al nacer en el Ecuador.

Se realizará una comparación mediante las técnicas de regresión logística y redes neuronales artificiales para determinar el modelo más confiable y efectivo que identifique las causas maternas: nutrición y enfermedades persistente durante el proceso de gestación de la madre, entendiendo así la magnitud y gravedad del problema por el que se encuentra actualmente atravesando.

#### 1.2. Limitaciones y delimitaciones

Una limitación respecto a los registros de las pacientes es que por cuestiones de confidencialidad el expediente médico es inaccesible sin la autorización. Dentro de las delimitaciones, se tomará como muestra a las pacientes en su último mes de gestación y que han tenido hijos registrados en el Hospital General Docente IESS desde el año 2020 hasta 2021.

#### 1.3. Problema general de investigación

¿Cuál de las dos técnicas Regresión logística y redes neuronales artificiales identifica de forma más verídica los factores maternos que influye en el peso del bebé al nacer en el IESS Riobamba en el periodo 2020-2021?



#### **1.4. Problemas específicos de investigación**

- ¿Cuáles son los posibles factores maternos asociados al bajo peso del bebé al nacer en el IESS?
- ¿Qué servicios públicos son los encargados de recolectar la información bebé nacidos con bajo en el IESS?
- ¿Qué técnicas se utilizará para determinar factores maternos que influyen en el bajo peso del bebé al nacer en el periodo 2020-2021?
- ¿Cuál modelo es el más adecuado que yo pueda utilizar para determinar los factores maternos asociados al bajo peso del recién nacido?

#### **1.5. Objetivos**

##### **1.5.1. *Objetivo general***

- Comparar los resultados de las técnicas: Regresión Logística y Redes Neuronales Artificiales, con el fin de establecer el modelo que identifique de manera más confiable los factores maternos que influyen en el peso del bebé al nacer en el IESS Riobamba en el periodo 2020-2021.

##### **1.5.2. *Objetivos específicos***

- Determinar mediante un marco teórico oportuno los posibles factores maternos asociados al bajo peso del bebé al nacer.
- Recolectar la información de los posibles factores maternos que influyen en el peso del bebé al nacer, usando las historias clínicas de las madres durante todo su periodo de gestación, atendidos en el IESS.
- Construir modelos mediante las técnicas: Regresión Logística y Redes Neuronales Artificiales.
- Realizar una comparativa del rendimiento de los resultados encontrados con los dos modelos mediante la curva ROC.

#### **1.6. Justificación**

##### **1.6.1. *Justificación teórica***

Un problema de salud muy importante y que es el causante de las derivaciones de muchas más enfermedades es el bajo peso al nacer del bebé, En Ecuador según el INEC en el año 2021 por cada

1000 habitantes se tiene 14.2 nacidos vivos y del cual el 8.6% de nacidos vivos tienen bajo peso al nacer. Para realizar el análisis se va a optar por; la aplicación de un modelo de regresión logística y redes neuronales artificiales con el estudio de las variables que intervienen en el bajo peso del bebé al nacer, permitiendo conocer los posibles factores maternos que originan el bajo peso en los bebés recién nacidos, para generar recomendaciones que permita reducir este acontecimiento mostrado en el periodo de estudio.

### **1.6.2. *Justificación práctica***

Con el presente trabajo de investigación mediante la adecuada revisión bibliográfica se determinará los posibles factores maternos asociados al bajo peso al nacer del bebé, al generar un modelo asociado de acuerdo con las técnicas estadísticas: Redes neuronales artificiales y Regresión logística respectivamente.

Una vez obtenido los modelos se evaluará y se determinará el “mejor” a través de las medidas de bondad de ajuste: el área bajo la curva ROC y así definir los factores maternos que influyen en el peso al nacer.

## CAPÍTULO II

### 2. MARCO TEÓRICO

#### 2.1. Antecedentes de investigación

Históricamente, el IESS nace de la caja de pensiones creada en 1928 para dar atención solo a los trabajadores públicos, el Instituto Ecuatoriano de Seguridad Social nace en 1942 con la Ley del Seguro Social Obligatorio que da cobertura también a los trabajadores del sector privado con prestaciones de enfermedad, maternidad, accidentes de trabajo y atención farmacológica, jubilación, montepío y mortuoria y posteriormente préstamos hipotecarios para la construcción de viviendas, para los afiliados (empleados y trabajadores con relación de dependencia casi exclusivamente urbanos).

En 1936 el IESS abrió sus puertas en atención de salud como Dispensario N.º 5 en dependencia de la caja de pensiones, en 1965 se suscribe el contrato para la construcción del Hospital del IESS para brindar los servicios de Consulta Externa, hospitalización, cirugía, odontología, laboratorio clínico e histopatológico.

En el año 2001 se aprueba la Ley de Seguridad Social. Su principal fuente de financiamiento son las aportaciones de empleados y empleadores del sector público y privado. El Hospital IESS de Riobamba cumplió el 11 de noviembre de 2014, 50 años de su creación, fundado como clínica, laboró por 13 años en las calles 10 de agosto y Pichincha; En 1969 se inaugura como hospital con una dotación de 56 camas. A partir de agosto de 1992 nuestro hospital utiliza las actuales instalaciones con 67 camas hospitalarias, con habitaciones que dan cabida a uno, dos o tres, quedando algunas áreas vacías (Chavez y Useche , 2021: pp.1231–1250)

Con lo antes mencionado cabe indicar que en este proyecto de investigación se establecen los principales factores maternos que intervienen en el peso del bebe al nacer a través de un marco teórico conceptual aplicando la terminología correcta misma que, se utilizó para la comparación entre las técnicas: regresión logística y redes neuronales artificiales para la categorización de recién nacidos a través de indicadores asociados al modelo matemático y en conjunto con la prueba para la validación del mismo: Curvas de ROC y se determinó como “el mejor” a aquel cuya efectividad para determinar factores maternos que influyen en el peso de bebe al nacer en el IESS de la ciudad de Riobamba fue mayor de esta manera, contribuimos con la institución de salud entregando un precedente investigativo que posteriormente puede ser replicado para posteriores avances médicos. Finalmente, la investigación comprende varios capítulos, el capítulo I engloba el problema de

investigación en conjunto con su respectiva justificación, objetivos y la hipótesis a comprobar al final de este trabajo investigativo. El capítulo II abarca el marco teórico referencial en donde con brevedad se describe teóricamente las técnicas empleadas: Regresión Logística y Redes Neuronales Artificiales y, respecto a las características que las madres tienen al momento de estar en gestación, el bajo peso del bebé al nacer y la mortalidad materna.

En el capítulo III con respecto al marco metodológico se indica el enfoque, nivel y diseño de investigación en donde se describe el proceso para la recolección y análisis de la información en base a la unidad de análisis y el tamaño muestral; la selección y descripción de variables que en conjunto con los algoritmos de clasificación empleados se determinara la factibilidad para la solución del problema.

En el capítulo IV se presenta el análisis de resultados obtenidos en la caracterización por regresión logística y redes neuronales artificiales para así finalmente presentar las conclusiones y recomendaciones.

## 2.2. Marco teórico

### 2.2.1. Modelo de regresión logística

Dada una variable respuesta categórica con dos niveles, la regresión logística modela la probabilidad de que,  $Y$  pertenezca a una categoría o nivel particular, dados los valores de un único predictor  $X$ . La clasificación depende del límite o threshold que se establezca (Gil, 2018).

$$\Pr(Y = k | X = x) \quad (2-1)$$

En regresión logística utilizamos la función logística:

$$p(X) = \frac{e^{\beta_0 + \beta_1 X}}{1 + e^{\beta_0 + \beta_1 X}} \quad (2-2)$$

que siempre producirá una curva en forma de S, comprendiéndose los valores de  $Y$  entre  $[0, 1]$ . La ecuación anterior puede reestructurarse como:

$$\frac{p(X)}{1 + p(X)} = e^{\beta_0 + \beta_1 X} \quad (2-3)$$

Donde  $\frac{p(X)}{1+p(X)}$  corresponde a los odds, pudiendo tomar cualquier valor entre 0 (muy baja probabilidad de éxito) e  $\infty$  (muy alta probabilidad de éxito). Este ratio, pues, indica cuanto más probable es el éxito que el fracaso. Introduciendo el logaritmo en ambos lados de la ecuación, obtenemos una función lineal.

$$\log\left(\frac{p(X)}{1-p(X)}\right) \quad (2-4)$$

La parte izquierda de la ecuación es lo que se conoce como logaritmo de odds (log-odds) o logit. En el ámbito de las ciencias de la salud, los modelos de regresión logística permiten el análisis de los resultados de manera explicativa o predictiva, además, se puede conocer la fuerza de asociación mediante los OR (odds ratio) de los factores de riesgo con el efecto estudiado de una manera independiente y conocer el valor predictivo de cada uno de ellos o bien del modelo (Pérez y Fiuza, et al, 2000, p.2).

Los odds ratio se presentan como una función exponencial de un conjunto de variables independientes.

Donde  $e$  es una constante,  $X_j$  es la  $j$ -ésima variable independientes,  $B_j$  es el coeficiente logístico de  $X_j$  y  $B_0$  es el coeficiente logístico constante.

Esta ecuación (2-1) indica que las probabilidades de ocurrencia de un evento están en función de  $e$  elevado a la potencia de una combinación lineal de las variables independientes. Cuando el coeficiente logístico  $B$  de una variable independiente  $X$  es positivo, indica que la variable independiente tiene probabilidades mayores de ocurrencia. Mientras que cuando el coeficiente logístico es negativo indica que la variable independiente hace que las probabilidades sean más bajas (Nayebi, 2020).

En un modelo de regresión logística todos los coeficientes logísticos de la variable independiente son distintos a cero y con un nivel de significancia inferior a 0,05 y según la medida de bondad de ajuste del modelo se indica qué tan bien se ajusta a los datos. Para el cálculo de los coeficientes logísticos se toman en cuenta el efecto de las variables independientes para maximizar la probabilidad de ocurrencia, la fuerza de la función de regresión logística se puede considerar como:

$$Z = B_0 + B_1X_1 + B_2X_2 + \dots + B_KX_K \quad (2-5)$$

La cual es una función lineal de variables independientes de  $X_1$  a  $X_k$ . Cuando existe un modelo de regresión logística ajustado se puede especificar el efecto de cada una de las variables independientes sobre las probabilidades de que ocurra o no un evento (Nayebi, 2020).

### **2.2.2. Estimación de los parámetros del modelo (Función de verosimilitud)**

Con el fin de estimar  $\beta = (\beta_1, \beta_2, \dots, \beta_k)$  y analizar el comportamiento del modelo estimado se toma una muestra aleatoria de tamaño  $n$  dada por  $(x_i, y_i)_{i=1, \dots, n}$  donde el valor de las variables independientes es  $x_i = (x_{i1}, x_{i2}, \dots, x_{ik}) = y_i \in [0, 1]$  es el valor observado de  $Y$  en el  $i$ -ésimo elemento

de la muestra (REGRESIÓN LOGÍSTICA, s. f.).

Como  $\left(\frac{Y}{X_1}, X_2, \dots, X_k\right) \in B[1, p(X_1, X_2, \dots, X_k; \beta)]$  la función de verosimilitud viene dada por:

$$L\left[\frac{\beta}{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)}\right] = \prod_{i=1}^n p_i^{y_i} (1 - p_i)^{1-y_i} \quad (2-6)$$

Donde:

$$p_i = p(x_i; \beta) = p[x_{i1}, x_{i2}, \dots, x_{ik}; \beta]_{i=1,2,\dots,n} \quad (2-7)$$

Debido a que la distribución de Y dado un conjunto de variables  $X = (x_1, x_2, \dots, x_p)$  no es normal y no existe homocedasticidad en los errores, la estimación del vector  $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_p)$  por el método de mínimos cuadrados no tiene propiedades óptimas en su lugar emplearemos el método de máxima verosimilitud para obtener los valores de los parámetros desconocidos que maximizan la probabilidad de obtener el conjunto observado de datos La función de verosimilitud adopta la forma:

$$l(\beta) = \prod_{i=1}^n (\pi(x_i)^{y_i} [1 - \pi(x_i)]^{1-y_i}) \quad (2-8)$$

Aplicando logaritmo neperiano, la expresión  $l(\beta)$  se define como:

$$L(\beta) = \ln[l(\beta)] = \sum_{i=1}^n \{y_i \ln[\pi(x_i)] + (1 - y_i) \ln[1 - \pi(x_i)]\} \quad (2-9)$$

Para encontrar el valor de  $\beta$  se deriva de  $L(\beta)$  con respecto a  $\beta_0, \beta_1, \dots, \beta_p$  y se iguala al valor cero obteniéndose:

$$\sum_{i=1}^n [y_i - \pi(x_i)] = 0 \quad (2-10)$$

Para encontrar la solución de este conjunto de ecuaciones se utiliza el método iterativo de Newton Raphson que se lo puede realizar mediante algún paquete estadístico.

### 2.2.3. Hipótesis en regresión logística simple

La hipótesis nula que se desea contrastar es si la variable independiente no influye significativamente sobre la variable dependiente. La hipótesis alternativa es: la variable independiente influye significativamente sobre la variable dependiente (Dianderas, 2007,p.2). Entre los procedimientos para contrastar las hipótesis anteriores están:

**Evaluación estadística del coeficiente:** En donde se usa como estadístico de prueba a

$$\frac{\widehat{\beta}_1 - \beta}{ee(\widehat{\beta}_1)} \quad (2-11)$$

el cual corresponde a una variable aleatoria que tiene distribución t con  $n - 2$  grados de libertad. (En el caso de la regresión logística múltiple el número de grados de libertad es  $n - K - 1$  grados de libertad, en donde K es el número de variables independientes).

**Evaluación estadística de WALD:** en donde se usa como estadístico de prueba a

$$Wald = \frac{(\widehat{\beta}_1)^2}{(ee(\widehat{\beta}_1))^2} \quad (2-12)$$

#### 2.2.4. Bondad de ajuste del modelo

Se utilizan dos tipos de contrastes: (a) Contrastos que analizan la bondad de ajuste desde un punto de vista global. (b) Contrastos que analizan la bondad de ajuste paso a paso (REGRESIÓN LOGÍSTICA, s. f.).

**Contraste de bondad de ajuste global de Hosmer-Lemeshow** El índice de bondad de ajuste:

$$z^2 = \sum_{i=1}^n \frac{(y_i - \hat{p}_i)^2}{\hat{p}_i(1 - \hat{p}_i)} \quad (2-13)$$

Donde  $\hat{p}_i = p(x_{i1}, x_{i2}, \dots, x_{ik}; \hat{\beta})_{i=1, \dots, n}$ ,  $z^2 \approx X_{n-k}^2$  si el modelo ajustado es cierto

#### Contraste de bondad de ajuste de Hosmer-Lemeshow

Evalúa la bondad de ajuste del modelo construyendo una tabla de contingencia a la que aplica un contraste tipo chi-cuadrado.

Calcula los deciles de las probabilidades estimadas  $(\hat{p}_i)_{i=1, \dots, n}$ ,  $(D_1, D_2, \dots, D_9)$ , dividiendo los datos observados en diez categorías dadas por:  $A_j = \{(\hat{p}_i)_{i=1, \dots, n} \in (D_{j-1}, D_j)_{j=1, \dots, 10}\}$  donde  $D_0 = 0$  y  $D_{10} = 1$

El estadístico de contraste:

$$T = \sum_{j=1}^{10} \frac{(e_j - n_j \bar{p}_j)^2}{n_j \bar{p}_j (1 - \bar{p}_j)} \quad (2-14)$$

Donde  $\begin{cases} n_j \equiv \text{numero } A_j (j = 1, \dots, 10) \\ e_j \equiv y_{i=1}^{n} e_n A_j (j = 1, \dots, 10) \end{cases}$   
 $\bar{p}_j = \sum_{i \in A_j} \bar{p}_i$

p-valor del contraste:  $p[X_g^2 \geq T_{\text{observado}}]$

#### AIC

El criterio de información de Akaike es una herramienta estadística útil para elegir el número de

retrasos de p y q basándose en la suma de los cuadrados de los errores que busca minimizarlos a partir de varias combinaciones de los parámetros.

$$AIC = \ln(\hat{\Theta})^2 + \frac{2}{n}r \quad (2-15)$$

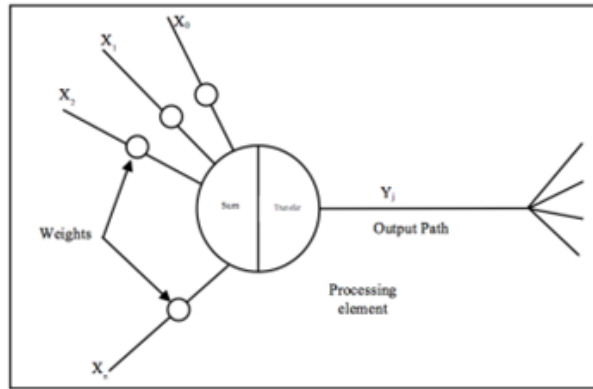
Donde:  $\ln$  : Logaritmo neperiano  $\hat{\Theta}^2$  Suma residual de cuadrados dividida entre el número de observaciones  $n$  : Número de observaciones  $r$  : Número total de parámetros (incluyendo el término constante)

### **2.2.5. Redes neuronales artificiales**

La primera neurona artificial fue inventada por W. McCulloch y W. Pitts, de la Universidad de Chicago en 1943. Un modelo binario cuyo estado es 1- Activo o 0- Inactivo. Calculaba la suma de sus entradas con el valor de cada entrada modulado por la eficacia sináptica correspondiente y toma una decisión comparando esta suma con un cierto umbral ya prefijado. Si la suma es superior al umbral, la neurona se activa; caso contrario se mantiene inactiva. Por lo tanto, cada neurona toma una decisión teniendo en cuenta la evolución del estado global de la red (Gómez, et al., 1994, p.169-179). Las Redes Neuronales Artificiales, ANN (Artificial Neural Networks) fueron creadas en base a las redes neuronales biológicas del cerebro humano. La creación de las redes neuronales artificiales busca producir una estructura de procesamiento, crear conexiones en base a la experiencia para generar un proceso de aprendizaje con información básica debido a que el aprendizaje de las ANN está limitado. Para la selección de una red se debe realizar en función de las características del problema a resolver. Éstos se pueden clasificar en aplicaciones de predicción, clasificación, asociación, conceptualización, filtrado y optimización (Olabe, s. f., p.4).

En las redes neuronales existe un elemento procesador o PE (Process Element) equivalente a una neurona, posee varias entradas que son combinadas para posterior cambiar los valores dependiendo de una función en la transferencia y se envía a la salida del procesador (Olabe, s.f., p.2). Esta salida se puede transferir a otros formando sinapsis entre varias neuronas como se muestra en la ilustración 1





**Ilustración 2-1:** Diagrama de una neurona artificial

**Fuente:** (Rivas Asanza, Mazón Olivo, 2018).

Esta técnica aplica un conjunto de entradas a la neurona y cada una representa una salida. Cada entrada es multiplicada por un peso o ponderación que corresponde al grado de conexión de la sinapsis. Las entradas ponderadas se suman y se determina el nivel de activación de la neurona artificial y se indica por la siguiente ecuación.

$$NET = X * W \quad (2-16)$$

Donde NET es la salida, X el vector de entrada y W el vector de pesos. Para obtener el valor de NET, se procesa mediante una función de activación F, es decir, producir una señal de salida a la neurona (OUT), esta función puede ser lineal o no lineal que simule las características de transferencias de las neuronas biológicas. Las funciones F más utilizadas son la función Sigmoid:

$$OUT = \frac{1}{1 + e^{-NET}} \quad (2-17)$$

y Tangente hiperbólica:

$$OUT = \tanh(NET) \quad (2-18)$$

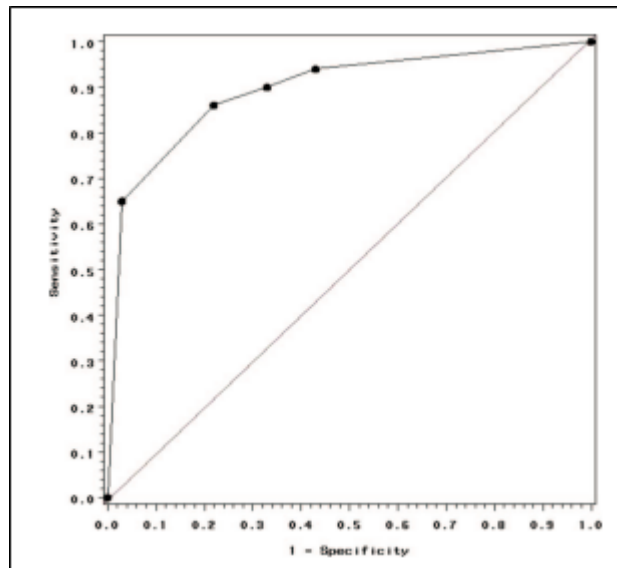
Una de las desventajas de esta técnica es que no toma en cuenta muchas de las características de las neuronas biológicas, como por ejemplo la omisión de retardos o de sincronismo en la generación de la salida (Olabe, s.f., p.5).

### 2.2.6. Curva característica operativa del receptor (ROC)

El análisis de las curvas ROC surgió a principios de los años cincuenta para el análisis de la detección de las señales de radar. Las curvas ROC son gráficos en los cuales se representa la

sensibilidad en función de los falsos positivos (1-especificidad) de la prueba diagnóstica, donde cada punto de la curva representa un par correspondiente a un nivel de decisión determinado (Armesto, 2021,p.3).

Para crear una curva ROC, se debe tabular las sensibilidades y especificaciones para diferentes de prueba continua. Se obtiene una lista de diferentes valores de prueba y la sensibilidad y especificación. En el eje y se traza la sensibilidad, es decir, tasa de verdaderos positivos y en el eje x la especificidad 1 o tasa de falsos positivos para los valores tabulados.



**Ilustración 2-2:** Curva Característica operativa del receptor

**Fuente:** (Mandrekar, 2010).

El gráfico que se va generando es una curva escalonada, de modo que cuando se obtiene un verdadero positivo la curva se desplazará verticalmente y en caso de que se obtengan falsos positivos la curva se desplazará horizontalmente.

### 2.2.7. Área bajo la curva (AUC)

El índice de precisión global de la prueba de diagnóstico viene dado por el valor del área bajo la curva, este valor está comprendido entre 0.5 (azar) y 1 (perfecta discriminación). Swets, clasifica la exactitud de la prueba del siguiente modo: si el valor del área está comprendido entre 0.5 y 0.7 entonces la exactitud es baja, si está comprendido entre 0.7 y 0.9 la exactitud es regular-alta (dependiendo de lo que estemos estudiando) y si es superior a 0.9 la exactitud de la prueba es alta. Por lo tanto, el valor del área bajo la curva resume la curva ROC en su conjunto que la utilización de este valor nos permite hacer comparaciones de puntos de dos curvas que tengan igual sensibilidad o especificidad y un enfoque global de comparación de la exactitud de las pruebas comparando sus

respectivas áreas bajo la curva. Gráficamente tendrá mayor precisión aquella curva que esté situada más arriba y la izquierda (Armesto y España, 2011, p. 3).

### **2.2.8. Comparar dos o más curvas ROC**

Se puede comparar dos o más pruebas para evaluar la capacidad predictiva de dos o más biomarcadores para el mismo evento. La prueba con el AUC más alto puede considerarse mejor. Sin embargo, cuando los valores específicos de sensibilidad y especificidad son relevantes en la comparación, se realiza con los AUC parciales.

Cuando un mismo individuo se somete a dos o más pruebas diagnósticas diferentes de interés se consideran ROC correlacionadas. Por otro lado, cuando el individuo se somete a dos pruebas de diagnóstico diferente se denomina curvas ROC no correlacionadas. Comparar curvas ROC no correlacionadas es algo simple y se basa en una forma de estadística Z que utiliza la diferencia en el área bajo las curvas y la SD de cada AUC (Pérez y Martín, 2023).

## **2.3. Marco conceptual**

### **2.3.1. Gestación**

Es el período que transcurre entre la implantación en el útero del óvulo fecundado y el momento del parto. Durante la gestación tiene lugar los procesos fisiológicos de crecimiento y desarrollo del feto en el interior del útero materno. El embarazo humano dura unas 40 semanas desde el primer día de la última menstruación o 38 desde la fecundación (aproximadamente 9 meses) (Peris y Sant, 2021).

### **2.3.2. Enfermedades durante el periodo de gestación**

Durante el embarazo la madre pasa por varios cambios fisiológicos que por lo general no ponen en riesgos la vida y desarrollo de la madre y del bebé. Sin embargo, durante este periodo las infecciones de diversa índole pueden poner en riesgo el curso normal del mismo e incluso comprometer la salud del neonato, provocando los partos prematuros y el bajo peso al nacimiento. (Alonso, et al, 2010,p.85). Según, (Gómez, 2000,p.99) Se conocen 2 tipos principales de hipertensión durante el embarazo.

- La hipertensión que aparece por primera vez durante la gestación a partir de las 20 semanas, en el trabajo de parto o el puerperio y en la que la presión se normaliza después de terminar la gestación.
- El embarazo que acaece en mujeres con hipertensión crónica esencial o secundaria a otra enfermedad.

Se han hecho algunos estudios que describen los cambios de la presión arterial durante todo el embarazo y se acepta de una forma general que la presión arterial comienza a disminuir a finales del primer trimestre, alcanza su nivel mínimo en el segundo trimestre y aumenta de nuevo al comenzar el tercero, para alcanzar su punto máximo al término del embarazo (Gómez, 2000,p.99).

Por otro lado, en un estudio realizado (Rosell, 2006,p.53) indica que las edades extremas de la vida reproductiva de la mujer han estado asociadas estadísticamente con la enfermedad hipertensiva además, el hábito de fumar es causa contribuyente para la aparición de la preeclampsia y la obesidad expone a la gestante que la padece a numerosas complicaciones durante el embarazo, entre ellas, diabetes mellitus, hipertensión arterial crónica y trabajo de parto laborioso.

### **2.3.3. *Desarrollo fetal y el entorno materno***

Los cuatro factores más pertinentes en la vida fetal son: a) el retraso del crecimiento intrauterino (RCIU); b) el parto prematuro de un feto con crecimiento normal para la edad gestacional; c) la sobrenutrición in útero; y d) los factores intergeneracionales. Hay muchos datos, obtenidos principalmente en los países desarrollados, que demuestran que el RCIU se asocia a un mayor riesgo de cardiopatía coronaria, accidente cerebrovascular, diabetes e hipertensión arterial. Tal vez sea la pauta de crecimiento, es decir, un crecimiento fetal restringido seguido de un crecimiento posnatal muy rápido de recuperación, lo que realmente imprima los mecanismos subyacentes de las enfermedades. Por otro lado, un tamaño grande al nacer (macrosomía) también se acompaña de un mayor riesgo de diabetes y enfermedades cardiovasculares. Entre la población adulta de la India, se observó una relación entre la disminución de la tolerancia a la glucosa y un índice ponderal alto (exceso de peso) al nacer. En los indios pima se halló una relación en forma de U con el peso al nacer, mientras que entre los estadounidenses de origen mexicano no se daba esa relación. Un peso de nacimiento elevado también se asocia a un mayor riesgo de cáncer de mama y de otro tipo de tumores malignos (OMS, 2018).

Esto nos da a entender que la talla y el peso ideal de un recién nacido se calcula teniendo presente tanto la morbilidad y la mortalidad inmediatas, de igual manera las consecuencias que estas traen en el futuro, como es la susceptibilidad a las enfermedades graves.

### **2.3.4. *Peso de un recién nacido***

Según la Clasificación Internacional de Enfermedades (CIE) perteneciente a la OMS, un neonato normal se define teniendo en cuenta el peso al nacer de 2500 a 4000 gramos, peso para edad gestacional entre los percentiles 10 y 90, y edad gestacional de 37 a 42 semanas. Así, la edad gestacional modifica la interpretación del peso al nacer, como sucede con el prematuro. Dichas

variables implican diferentes condiciones del estado de salud intrauterino, y por lo tanto del tamaño renal (Bianchi et al., 2016).

La determinación del crecimiento fetal normal se basa en la comparación de las medidas antropométricas del neonato con los estándares obtenidos de neonatos sanos por provenir de embarazos sin patología detectada. Sin embargo, cuanto más prematuro es un recién nacido, aún sin patología evidente, más sospecha debiéramos tener sobre su crecimiento normal, ya que el parto prematuro es consecuencia de una enfermedad fetal (Pacora, 2005, p.212).

### **2.3.5. *Bajo peso de un recién nacido***

La Organización Mundial de la Salud (OMS) define «bajo peso al nacer» como un peso al nacer inferior a 2500 g. El bajo peso al nacer sigue siendo un problema significativo de salud pública en todo el mundo y está asociado a una serie de consecuencias a corto y largo plazo (OMS, 2017).

### **2.3.6. *Muerte neonatal***

Instituciones internacionales señalan que cada año se suceden más de 10 millones de muertes infantiles en niños menores de 5 años; 3,9 millones ocurren en el período neonatal, la mayoría por causas prevenibles y casi todas en países pobres. Estas cifras varían según las tasas globales de mortalidad. En países con alta mortalidad infantil, un 20% de las muertes ocurren en el período neonatal, mientras que en países con baja mortalidad más de un 50% de las muertes ocurre en recién nacidos. No existe mucha información acerca de las causas directas de muertes neonatales en comunidades de bajos recursos, pero se estima que el 24% es debido a infecciones respiratorias, 29% a asfixia en el parto, 24% por complicaciones de la prematuridad y 7% por tétanos. Un mejor conocimiento de la epidemiología materno-neonatal puede contribuir a una mejor atención y tomar medidas preventivas (Faneite et al., 2004).

En Venezuela diversos autores se han preocupado por estudiar la mortalidad neonatal, y en general señalan sus hallazgos la conveniencia de prestarle atención permanente a este importante indicador de salud pública. Se ha estudiado la mortalidad perinatal durante el lapso 1993-1997, con cifras promedios elevadas, con tasas de 45,46 por mil nacidos vivos, con tendencia al incremento. Su componente de muerte fetal fue 22,98 y el neonatal 23,11. Esta situación nos ha motivado a realizar un seguimiento a este difícil problema, nos propusimos determinar la incidencia de la mortalidad neonatal, revelar las características maternas, determinar las condiciones obstétricas, mostrar los factores relacionados, y finalmente hacer las recomendaciones pertinentes (Faneite et al., 2004).

## **CAPÍTULO III**

### **3. MARCO METODOLÓGICO**

#### **3.1. Tipo de investigación**

Por el método de investigación será cuantitativa debido a que se va a realizar una comparación del modelo de regresión logística y redes neuronales artificiales para determinar los factores maternos que influyen en el peso del bebe al nacer, según el objetivo de estudio será una investigación aplicada ya que utiliza conocimientos basados en los diferentes modelos a aplicarse para posterior a ellos ser empleada, según el nivel de profundización en el objeto de estudio será exploratoria y descriptiva por lo que se pretende determinar los principales factores maternos que influyen en el peso del recién nacido mediante el uso del modelo de regresión logística y redes neuronales artificiales, según la manipulación de variables es no experimental debido a que no se manipula las variables en estudio dado que la matriz de información proviene de una fuente primaria, según el tipo de inferencia será inductiva ya que se identifica los factores maternos asociados al bajo peso del recién nacido a través de un análisis de regresión logística y redes neuronales artificiales, según el periodo temporal es longitudinal ya que se observa un evento (periodo de gestación) durante un tiempo determinado con el propósito de observar sus cambios en el estado nutricional y patologías de la madre.

#### **3.2. Diseño de investigación**

Se utiliza el método de investigación cuantitativa y según la manipulación de las variables es un diseño no experimental, durante el desarrollo del proyecto se trabajará con algoritmos para la obtención de modelos matemáticos de regresión logística y redes neuronales artificiales que mediante las medidas de bondad de ajuste se determinará el “mejor” modelo cuya capacidad de asociación de variables es mayor al utilizar la tasa de error y el área bajo la curva ROC a través del AUC.

#### **3.3. Localización del estudio**

La investigación planteada se llevará a cabo en el Instituto Ecuatoriano de Seguridad Social (IESS), ubicada en la calle Chile 3555, en la ciudad de Riobamba.

### **3.4. Población del estudio**

La población de estudio pertenece a los pacientes registrados en el Instituto Ecuatoriano de Seguridad Social (IESS).

### **3.5. Tamaño de la muestra**

La muestra pertenece a todas las madres en estado de gestación registradas en el Instituto Ecuatoriano de Seguridad Social (IESS) durante el periodo 2020-2021.

### **3.6. Método de muestreo**

No se aplicará un método de muestreo ya que los datos serán proporcionados por el Instituto Ecuatoriano de Seguridad Social (IESS).

### **3.7. 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.

### **3.8. Modelo estadístico**

La comparativa entre modelos de regresión y redes neuronales artificiales estará sujeta a la revisión previa del historial clínico de las pacientes que asisten al IESS, de quienes se obtendrá los datos para las variables independientes, las que serán utilizadas para modelar los factores asociados al bajo peso al nacer del bebé a través de las técnicas planteadas.

### **3.9. Identificación de variables**

#### **3.9.1. *Variable dependiente***

- Peso del recién nacido

#### **3.9.2. *Variables independientes***

- Edad materna
- Peso (kg) de la madre
- Talla (cm) de la madre

- IMC
- Diagnóstico durante el embarazo
- Semanas de gestación
- Paridad
- Hábitos tóxicos
- Abortos anteriores
- Antecedentes patológicos personales y alergias
- Sexo del bebé

### 3.10. Operacionalización de las variables

**Tabla 3–1:** Operacionalización de las variables

| Nombre de la variable      | Descripción   | Tipo de variable | Escala de medición | Categoría o intervalo |
|----------------------------|---|------------------|--------------------|-----------------------|
| Peso(Kg) del recién nacido | El primer peso del bebé, tomado justo después de nacer.               | Cuantitativa     | Razón              | [2 ,4] kg             |
| Edad materna               | Edad en la que se encuentra la madre gestante.                        | Cuantitativa     | Ordinal            | [15 ,46] años         |
| Peso (kg) de la madre      | Peso de la madre en la última semana de gestación.                    | Cuantitativa     | Razón              | [34 ,95] kg           |
| Talla (cm) de la madre     | Talla de la madre en la última semana de gestación.                   | Cuantitativa     | Razón              | [128 ,165] cm         |
| IMC de la madre            | Índice de masa corporal de la madre en la última semana de gestación. | Cuantitativa     | Razón              | [70 ,228] IMC         |

*Continúa en la siguiente página*



**Tabla 3-1** : *Continuación de la página anterior*

| <b>Nombre de la variable</b> | <b>Descripción</b>                                   | <b>Tipo de variable</b> | <b>Escala de medición</b> | <b>Categoría o intervalo</b>   |
|------------------------------|--|-------------------------|---------------------------|--|
| Tipo de parto                | Tipo de parto que da termino al estado de gestación. | Cualitativa             | Nominal                   | Parto múltiple<br>Parto múltiple por cesárea<br>Parto cesárea con histerectomía<br>Parto cesárea de emergencia<br>Parto único espontaneo |

*Continúa en la siguiente página*

**Tabla 3-1 :** *Continuación de la página anterior*

| Nombre de la variable           | Descripción  | Tipo de variable | Escala de medición | Categoría o intervalo  |
|---------------------------------|--|------------------|--------------------|--|
| Diagnóstico durante el embarazo | La enfermedad o infección que tuvo durante el periodo de gestación | Cualitativa      | Nominal            | Ninguno<br>Amenaza de aborto<br>Anemia<br>Diabetes gestacional<br>Diabetes, HTA<br>Enfermedades endocrinas<br>Epilepsia<br>Hipertensión gestacional<br>Hipotiroidismo<br>Infección de partes de las vías urinarias<br>Infección de tracto urinario<br>Infección del riñón<br>Infección genital<br>Preeclampsia leve a moderada<br>Preeclampsia severa<br>Síndrome HELLP<br>Trabajo de parto prematuro sin parto<br>Vómitos |
| Semanas de gestación            | Número de semanas que tiene la madre gestante                      | Cuantitativa     | Ordinal            | [23 ,41] semanas   |
| Paridad                         | Número de embarazos que ha tenido                                  | Cuantitativa     | Nominal            | [0 ,5] hijos   |

*Continúa en la siguiente página*

**Tabla 3-1 :** *Continuación de la página anterior*

| <b>Nombre de la variable</b>        | <b>Descripción</b>   | <b>Tipo de variable</b> | <b>Escala de medición</b> | <b>Categoría o intervalo</b>   |
|-------------------------------------|--|-------------------------|---------------------------|--|
| Hábitos tóxicos                     | Consumo de sustancias que afectan el desarrollo del bebé                           | Cualitativa             | Nominal                   | Ninguno<br>Alcohol<br>Tabaco<br>Otras drogas   |
| Abortos anteriores                  | Número de abortos que ha tenido anterior al último embarazo                        | Cuantitativa            | Razón                     | [0 ,2] abortos   |
| Antecedentes patológicos personales | Enfermedades que ha tenido antes del embarazo que complican el desarrollo del bebé | Cualitativa             | Nominal                   | Ninguno<br>Cáncer de seno<br>Cáncer en cráneo, estómago y seno<br>Cardiopatía<br>Colelitiasis<br>Discapacidad intelectual<br>Gastritis<br>Hipertensión<br>Hipotiroidismo<br>Parálisis facial<br>Queratocono<br>Sobrepeso |
| Género del bebé                     | Información correspondiente al sexo-biológico del recién nacido                    | Cualitativa             | Nominal                   | Masculino<br>Femenino  |

**Fuente:** Base de datos de las madres gestantes en el periodo 2020-2021 atendidas en el IESS.

**Realizado por:** Allayco, Digna y Huaraca, Maria, 2023.



**Tabla 4-1 :** Continuación de la página anterior

|              | <b>Edad</b> | <b>Peso</b> | <b>Talla</b> | <b>Semanas G</b> | <b>Paridad</b> | <b>Peso-bebé</b> | <b>IMC</b> |
|--------------|-------------|-------------|--------------|------------------|----------------|------------------|------------|
| Máximo       | 45.5        | 95          | 156.0        | 41.4             | 5.0            | 4.0              | 228.2      |
| Percentil 25 | 25.9        | 58.4        | 147.0        | 37.5             | 0.0            | 2.4              | 126.2      |
| Percentil 50 | 30.5        | 63.0        | 152.0        | 38.7             | 1.0            | 2.6              | 143.1      |
| Percentil 75 | 34.7        | 68.4        | 156.0        | 39.6             | 2.0            | 3.0              | 164.7      |

**Fuente:** Base de datos madres gestantes y peso del bebé al nacer en el IESS Riobamba periodo 2020-2021.

**Realizado por:** Allayco, Digna y Huaraca, María, 2023.

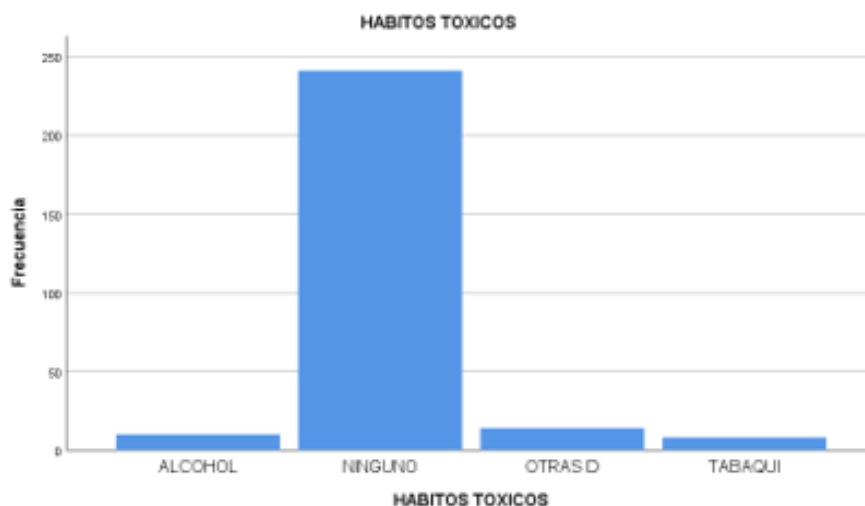
En la tabla 4-1 Estadísticos descriptivos de las variables cuantitativas, se obtiene que la edad mínima de las mujeres embarazadas es de 15 años y la edad máxima de 45, con una media de 30 años y una desviación estándar de 6.8 en el periodo 2020-2021. En cuanto a las semanas de gestación se tiene un mínimo de 23.2 y un máximo de 41.4 semanas, con una media de 38.4 y desviación estándar de 2.1, lo que indica que en su mayoría las madres gestantes superan las 37 semanas que son mínimas para que un bebé no tenga complicaciones en su desarrollo. Así mismo, durante este periodo se observa que se tiene una media de 1 en cuanto a la paridad, es decir, muchas de las mujeres reciben a su primer hijo. Y por último, se observa que en el peso del bebé se tienen un mínimo de 1.5 y un máximo de 5.0 kg, con una media de 2.7, lo cual llama mucho la atención debido a que si un bebé no supera los 2.5 kg se considera como bajo peso del recién nacido.

**Tabla 4-2:** Frecuencias absolutas y relativas de la variable Hábitos tóxicos

| <b>Clase</b> | <b>F. Absoluta</b> | <b>F. Relativa</b> |
|--------------|--------------------|--------------------|
| Alcohol      | 10                 | 3.7                |
| Ninguno      | 241                | 88.3               |
| Otras Drogas | 14                 | 5.1                |
| Tabaquismo   | 8                  | 2.9                |
| <b>TOTAL</b> | <b>273</b>         | <b>100.0</b>       |

**Fuente:** Base de datos madres gestantes y peso del bebé al nacer en el IESS Riobamba periodo 2020-2021.

**Realizado por:** Allayco, Digna y Huaraca, María, 2023.



**Ilustración 4-2:** Diagrama de barras de las frecuencias absolutas de la variable Hábitos tóxicos

**Fuente:** Base de datos madres gestantes y peso del bebé al nacer en el IESS Riobamba periodo 2020-2021.

**Realizado por:** Allayco, Digna y Huaraca, María, 2023.

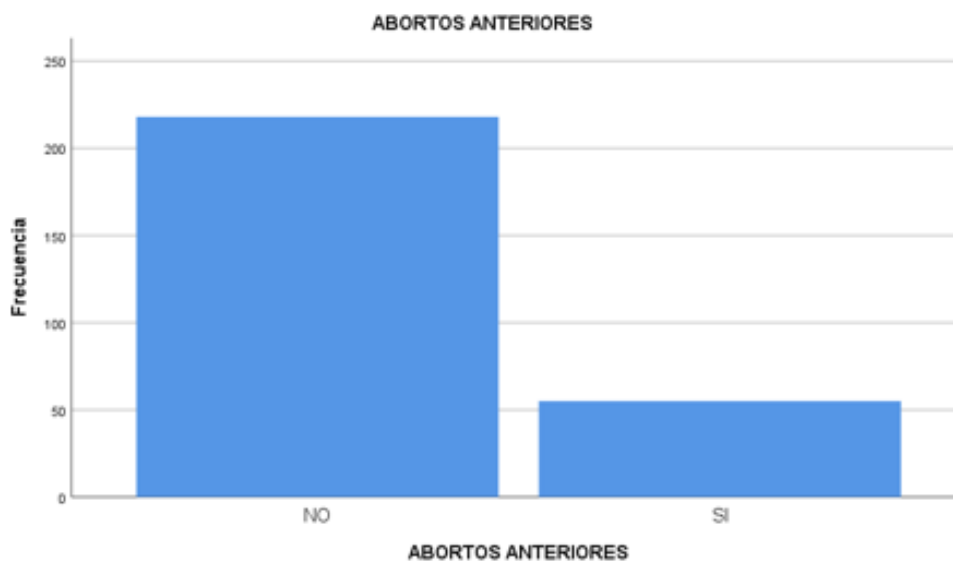
En la tabla 4-2 Frecuencias absolutas de la variable Hábitos tóxicos y en la Figura 4-2 Diagrama de barras de las frecuencias absolutas de la variable hábitos tóxicos, se puede observar que existe mayor frecuencia en la clase "ninguno", es decir, la mayoría de las mujeres no consumen alcohol, tabaco o drogas durante su embarazo con un total de 241 mujeres. Sin embargo, se observa que si existe una mínima frecuencia de consumo de otras drogas y alcohol en el período de enero 2020 a diciembre 2021.

**Tabla 4-3:** Frecuencias absolutas y relativas de la variable Abortos anteriores

| Clase        | F. Absoluta | F. Relativa |
|--------------|-------------|-------------|
| No           | 218         | 79.9        |
| Si           | 55          | 20.1        |
| <b>TOTAL</b> | 273         | 100.0       |

**Fuente:** Base de datos madres gestantes y peso del bebé al nacer en el IESS Riobamba periodo 2020-2021.

**Realizado por:** Allayco, Digna y Huaraca, María, 2023.



**Ilustración 4–3:** Diagrama de barras de las frecuencias absolutas de la variable Abortos anteriores

**Fuente:** Base de datos madres gestantes y peso del bebé al nacer en el IESS Riobamba periodo 2020-2021.

**Realizado por:** Allayco, Digna y Huaraca, María, 2023.

En la tabla 4–3 Frecuencias absolutas de la variable Abortos anteriores y en la Figura 4–3 Diagrama de barras de las frecuencias absolutas de la variable Abortos anteriores, se observa que existe una mayor frecuencia de mujeres que no han tenido abortos anteriores con un total de 218 embarazadas durante el periodo enero 2020 a diciembre 2021.

**Tabla 4–4:** Frecuencias absolutas y relativas de la variable Antecedentes patológicos

| Clase            | F. Absoluta | F. Relativa |
|------------------|-------------|-------------|
| Cáncer de seno   | 2           | 0.7         |
| Cáncer en cráneo | 1           | 0.4         |
| Cardiopatía      | 2           | 0.7         |
| Colelitiasis     | 6           | 2.2         |
| Gastritis        | 5           | 1.8         |
| Hipertensión     | 14          | 5.1         |
| Hipotiroidismo   | 21          | 7.7         |
| HTA              | 2           | 0.7         |
| Ninguno          | 216         | 79.1        |
| Queratocono      | 1           | 0.4         |

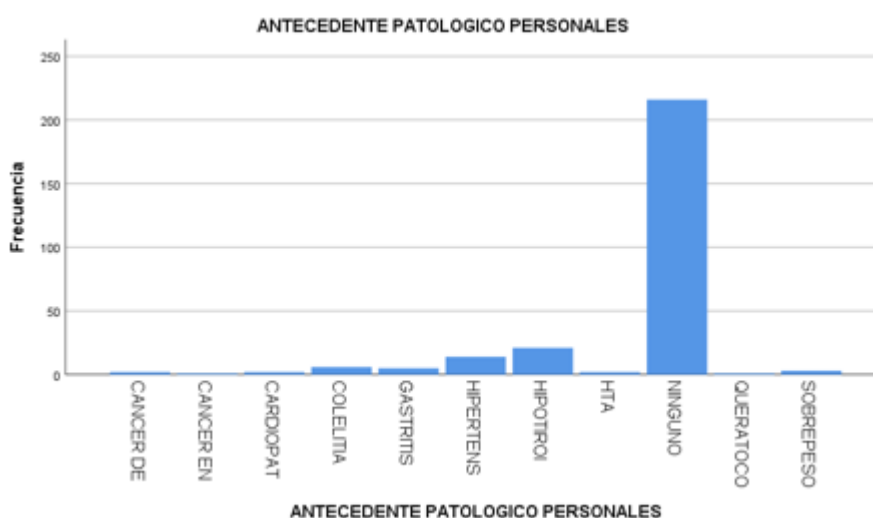
*Continúa en la siguiente página*

**Tabla 4-4 :** Continuación de la página anterior

| Clase        | F. Absoluta | F. Relativa  |
|--------------|-------------|--------------|
| Sobrepeso    | 3           | 1.1          |
| <b>TOTAL</b> | <b>273</b>  | <b>100.0</b> |

**Fuente:** Base de datos madres gestantes y peso del bebé al nacer en el IESS Riobamba periodo 2020-2021.

**Realizado por:** Allayco, Digna y Huaraca, María, 2023.



**Ilustración 4-4:** Diagrama de barras de las frecuencias absolutas de la variable Antecedentes patológicos

**Fuente:** Base de datos madres gestantes y peso del bebé al nacer en el IESS Riobamba periodo 2020-2021.

**Realizado por:** Allayco, Digna y Huaraca, María, 2023.

En la tabla 4-4 Frecuencias absolutas de la variable Antecedentes patológicos y en la Figura 4-4 Diagrama de barras de las frecuencias absolutas de la variable Antecedentes patológicos, se observa que existe una mayor frecuencia en la clase ninguno, es decir, en total 216 mujeres en estado de gestación no han tenido antecedentes de enfermedades antes del embarazo durante el periodo enero 2020 a diciembre 2021.

**Tabla 4-5:** Frecuencias absolutas y relativas de la variable Antecedentes alérgicos

| Clase       | F. Absoluta | F. Relativa |
|-------------|-------------|-------------|
| Complejo B  | 2           | 0.7         |
| Diclofenaco | 1           | 0.4         |

*Continúa en la siguiente página*



**Tabla 4-5 :** Continuación de la página anterior

| Clase          | F. Absoluta | F. Relativa  |
|----------------|-------------|--------------|
| Hidrocortizona | 2           | 0.7          |
| Metoclopramida | 2           | 0.7          |
| Ninguno        | 262         | 96.0         |
| Penicilina     | 4           | 1.5          |
| <b>TOTAL</b>   | <b>273</b>  | <b>100.0</b> |

**Fuente:** Base de datos madres gestantes y peso del bebé al nacer en el IESS Riobamba periodo 2020-2021.

**Realizado por:** Allayco, Digna y Huaraca, María, 2023.



**Ilustración 4-5:** Diagrama de barras de las frecuencias absolutas de la variable Antecedentes alérgicos

**Fuente:** Base de datos madres gestantes y peso del bebé al nacer en el IESS Riobamba periodo 2020-2021.

**Realizado por:** Allayco, Digna y Huaraca, María, 2023.

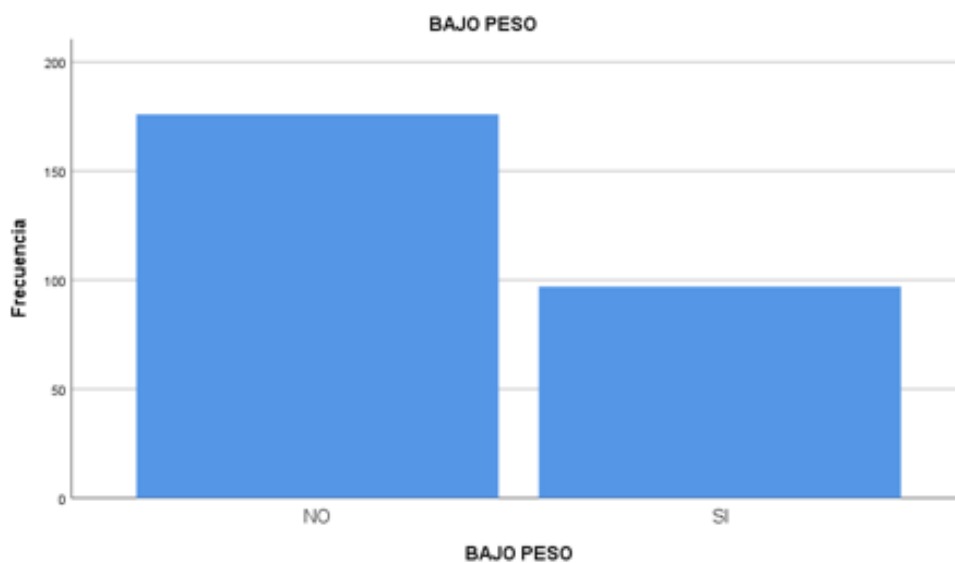
En la tabla 4-5 Frecuencias absolutas de la variable Antecedentes alérgicos y en la Figura 4-5 Diagrama de barras de las frecuencias absolutas de la variable Antecedentes alérgicos, se observa que existe una mayor frecuencia en la clase ninguno con un total 262 mujeres en estado de gestación no tienen antecedentes alérgicos durante el periodo enero 2020 a diciembre 2021.

**Tabla 4-6:** Frecuencias absolutas y relativas de la variable Bajo peso

| Clase        | F. Absoluta | F. Relativa |
|--------------|-------------|-------------|
| No           | 176         | 64.5        |
| Si           | 97          | 35.5        |
| <b>TOTAL</b> | 273         | 100.0       |

**Fuente:** Base de datos madres gestantes y peso del bebé al nacer en el IESS Riobamba periodo 2020-2021.

**Realizado por:** Allayco, Digna y Huaraca, María, 2023.



**Ilustración 4-6:** Diagrama de barras de las frecuencias absolutas de la variable Bajo peso

**Fuente:** Base de datos madres gestantes y peso del bebé al nacer en el IESS Riobamba periodo 2020-2021.

**Realizado por:** Allayco, Digna y Huaraca, María, 2023.

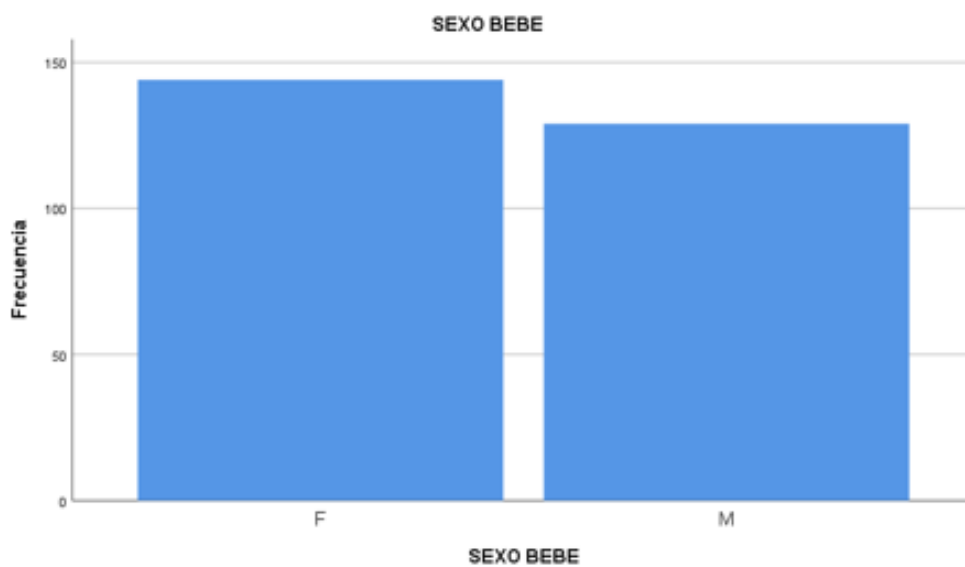
En la tabla 4-6 Frecuencias absolutas de la variable Bajo peso y en la Figura 4-6 Diagrama de barras de las frecuencias absolutas de la variable Bajo peso, se observa que existe una mayor frecuencia en la clase NO, lo que indica que con un total de 176 bebés nacieron con un peso adecuado durante el periodo enero 2020 a diciembre 2021.

**Tabla 4-7:** Frecuencias absolutas y relativas de la variable Sexo del bebé

| Clase        | F. Absoluta | F. Relativa |
|--------------|-------------|-------------|
| F            | 144         | 52.7        |
| M            | 129         | 47.3        |
| <b>TOTAL</b> | 273         | 100.0       |

**Fuente:** Base de datos madres gestantes y peso del bebé al nacer en el IESS Riobamba periodo 2020-2021.

**Realizado por:** Allayco, Digna y Huaraca, María, 2023.



**Ilustración 4-7:** Diagrama de barras de las frecuencias absolutas de la variable Sexo del bebé

**Fuente:** Base de datos madres gestantes y peso del bebé al nacer en el IESS Riobamba periodo 2020-2021.

**Realizado por:** Allayco, Digna y Huaraca, María, 2023.

En la tabla 4-7 Frecuencias absolutas de la variable Sexo del bebé y en la Figura 4-7 Diagrama de barras de las frecuencias absolutas de la variable Sexo del bebé, se observa que existe una mayor frecuencia de bebés con sexo femenino con un total de 144 recién nacidos durante el periodo enero 2020 a diciembre 2021.

## 4.2. Regresión logística

|   | EDAD  | Peso, kg. | TALLA | TIPO_DE_PARTO | DIAGNOSTICO_D_EMBARAZO | SEMANAS_DE_GESTACION | HABITOS_TOXICOS | ABORTOS_ANTERIORES |
|---|-------|-----------|-------|---------------|------------------------|----------------------|-----------------|--------------------|
| 1 | 40.00 | 57.5      | 140   | 2             | 0                      | 35.6                 | 0               | 1                  |
| 2 | 39.26 | 68.3      | 158   | 5             | 0                      | 38.2                 | 0               | 1                  |
| 3 | 36.46 | 62.2      | 148   | 4             | 9                      | 39.6                 | 0               | 0                  |
| 4 | 21.13 | 40.1      | 142   | 4             | 0                      | 40.2                 | 0               | 0                  |
| 5 | 27.75 | 59.3      | 147   | 1             | 17                     | 36.9                 | 0               | 0                  |
| 6 | 19.21 | 65.2      | 143   | 5             | 0                      | 37.7                 | 1               | 0                  |

|   | PARIDAD | ANTECEDENTE_PAT_PERSONALES | ANTECEDENTES_ALERGICOS | PESO_BEBE_NACER, Kg. | T_PESO | SEXO | IMC    |
|---|---------|----------------------------|------------------------|----------------------|--------|------|--------|
| 1 | 1       | 6                          | 0                      | 2.40                 | 1      | 0    | 112.70 |
| 2 | 2       | 0                          | 0                      | 2.98                 | 0      | 1    | 170.50 |
| 3 | 2       | 0                          | 0                      | 3.60                 | 0      | 1    | 136.24 |
| 4 | 0       | 0                          | 0                      | 2.55                 | 0      | 1    | 80.86  |
| 5 | 1       | 8                          | 0                      | 2.46                 | 1      | 1    | 128.14 |
| 6 | 0       | 0                          | 0                      | 3.00                 | 0      | 0    | 133.33 |

**Ilustración 4–8:** Codificación de las variables categóricas

**Fuente:** Base de datos madres gestantes y peso del bebé al nacer en el IESS Riobamba periodo 2020-2021.

**Realizado por:** Allayco, Digna y Huaraca, María, 2023.

En la figura 4–8 Codificación de las variables categóricas se observa que las variables de tipo categóricas como diagnóstico del embarazo, hábitos tóxicos, abortos anteriores, antecedentes patológicos personales, antecedentes alérgicos y sexo del bebe, las cuales se codifican según el número de clases de cada variable para posterior encontrar el modelo de regresión logística binaria, con la finalidad de identificar de una mejor manera los factores que influyen en el bajo peso del bebe al nacer. Además, se estableció como factor (Dummy) nuestra variable dependiente bajo peso.

### 4.2.1. Regresión logística binaria

| Coefficients:              |          |            |         |          |     |  |
|----------------------------|----------|------------|---------|----------|-----|--|
|                            | Estimate | Std. Error | z value | Pr(> z ) |     |  |
| (Intercept)                | 24.03186 | 5.37583    | 4.470   | 7.81e-06 | *** |  |
| HABITOS_TOXICOS            | 2.92026  | 1.14235    | 2.556   | 0.010578 | *   |  |
| SEMANAS_DE_GESTACION       | -0.43080 | 0.13033    | -3.305  | 0.000949 | *** |  |
| Peso, kg.                  | -0.22744 | 0.06189    | -3.675  | 0.000238 | *** |  |
| SEXO                       | -1.33408 | 0.40205    | -3.318  | 0.000906 | *** |  |
| PARIDAD                    | -0.73051 | 0.19452    | -3.756  | 0.000173 | *** |  |
| ABORTOS_ANTERIORES         | 1.16247  | 0.38815    | 2.995   | 0.002745 | **  |  |
| DIAGNOSTICO_D_EMBARAZO     | 0.09375  | 0.03741    | 2.506   | 0.012206 | *   |  |
| IMC                        | 0.04199  | 0.01800    | 2.333   | 0.019662 | *   |  |
| ANTECEDENTE_PAT_PERSONALES | 0.11001  | 0.06392    | 1.721   | 0.085244 | .   |  |

**Ilustración 4–9:** Resumen del modelo de regresión logística

**Fuente:** Base de datos madres gestantes y peso del bebé al nacer en el IESS Riobamba periodo 2020-2021.

**Realizado por:** Allayco, Digna y Huaraca, María, 2023.

En la figura 4–9 Resumen del modelo de regresión logística binaria, se observa que para la creación del modelo se utilizan 9 variables ya categorizadas y se obtiene que de éstas, 8 son significativas en el modelo esto debido a que su p-valor es menor al nivel de significancia de 0.05, es decir, estas variables influyen significativamente en el modelo de regresión logística. Obteniendo así la

siguiente ecuación:

$$T\_peso = 24.031 + 2.9202 * h\_toxicos - 0.4308 * SA\_gestacion - 0.2274 * Peso - 1.33340 * sexo - 0.7305 * paridad + 1.1624 * abortos + 0.0937 * Diag\_embarazo + 0.0419 * IMC$$

Además, se calcula el criterio de información de Akaike, el cual nos indica que cuanto menor sea el valor mejor es el ajuste del modelo, al realizar el cálculo dio como resultado un número respectivamente bajo igual a 200.71, esto nos indica que tiene un buen ajuste. Por otro lado, se calcula también el R2 de McFadden debido a que no existe el valor  $R^2$  para la regresión logística. Es por ello que se calcula una métrica conocida como  $R^2$  de McFadden, que varía de 0 a poco menos de 1. En este criterio se indica que si el valor es superior a 0.40 se indica que el modelo se ajusta muy bien a los datos. Obteniendo así un valor de 0.49, lo que nos demuestra que el modelo se ajusta adecuadamente a los datos.

#### 4.3. Redes neuronales artificiales

Antes de realizar la creación de un modelo de redes neuronales artificiales en R, se introduce la base de datos que tenga las variables categorizadas. En la aplicación de las redes neuronales artificiales se debe rescalas las variables de entrada para que estén en el mismo rango.

**Tabla 4–8:** Frecuencias absolutas de los pesos del bebé en el entrenamiento

| Adecuado | Bajo Peso |
|----------|-----------|
| 151      | 6         |

**Fuente:** Base de datos madres gestantes y peso del bebé al nacer en el IESS Riobamba periodo 2020-2021.

**Realizado por:** Allayco, Digna y Huaraca, María, 2023.

En la tabla 4–8 Frecuencias absolutas del peso del bebé para entrenamiento de la red neuronal se observa que del 80% de la base utilizada en el entrenamiento, 151 bebé nacieron con el peso adecuado mientras que los 6 restantes con bajo peso.

**Tabla 4–9:** Frecuencias absolutas de los pesos del bebé para la validación

| Adecuado | Bajo Peso |
|----------|-----------|
| 38       | 17        |

**Fuente:** Base de datos madres gestantes y peso del bebé al nacer en el IESS Riobamba periodo 2020-2021.

**Realizado por:** Allayco, Digna y Huaraca, María, 2023.

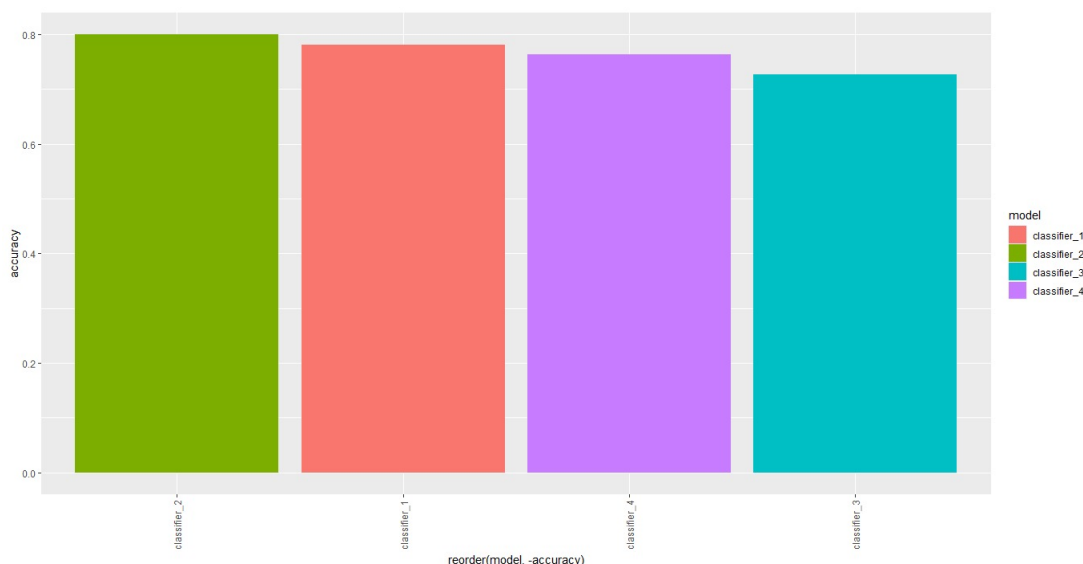
Por otro lado, en la tabla 4–9 Frecuencias absolutas del peso del bebé para la validación de la red neuronal se observa que en el 20% de la base utilizada para la validación del modelo, 38 bebés nacieron con el peso adecuado mientras que los 17 restantes con bajo peso.

**Tabla 4–10:** Porcentaje de exactitud de los modelos

| Modelo | Porcentaje de exactitud |
|--------|-------------------------|
| 1      | 0.781                   |
| 2      | 0.8                     |
| 3      | 0.727                   |
| 4      | 0.763                   |

**Fuente:** Base de datos madres gestantes y peso del bebé al nacer en el IESS Riobamba periodo 2020-2021.

**Realizado por:** Allayco, Digna y Huaraca, María, 2023.

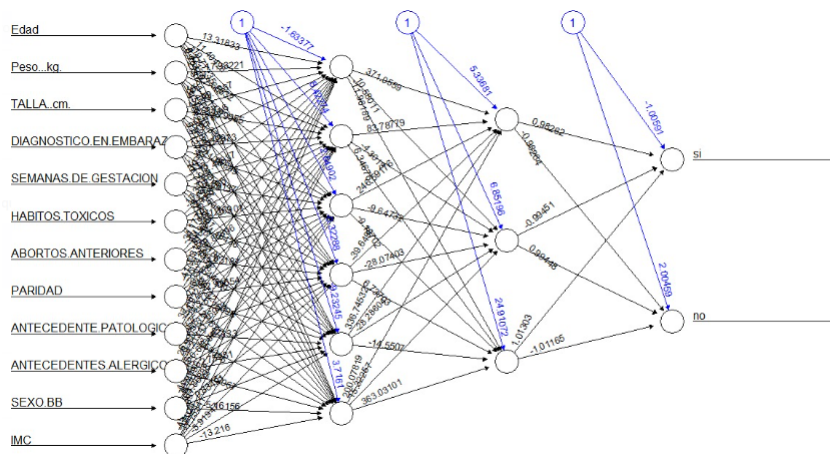


**Ilustración 4–10:** Diagrama de barras del porcentaje de exactitud de los modelos

**Fuente:** Base de datos madres gestantes y peso del bebé al nacer en el IESS Riobamba periodo 2020-2021.

**Realizado por:** Allayco, Digna y Huaraca, María, 2023.

Por otro lado, en la tabla 4–10 Porcentaje de exactitud de los modelos y en la ilustración 4–10 Diagrama de barras del porcentaje de exactitud de los modelos se observa que el segundo modelo obtiene un porcentaje de exactitud del 80% por lo cual, se selecciona el modelo 2.



**Ilustración 4–11:** Red neuronal artificial del mejor modelo

**Fuente:** Base de datos madres gestantes y peso del bebé al nacer en el IESS Riobamba periodo 2020-2021.

**Realizado por:** Allayco, Digna y Huaraca, María, 2023.

En la ilustración 4–11 se observa que en la capa de entrada se encuentran todas las variables de la base de datos y en la capa de salida se encuentra la variable dependiente peso del bebé, mientras que, en las capas ocultas se encuentran dos, una capa oculta con 6 neuronas y la otra capa con 3 neuronas.

**Tabla 4–11:** Matriz de confusión

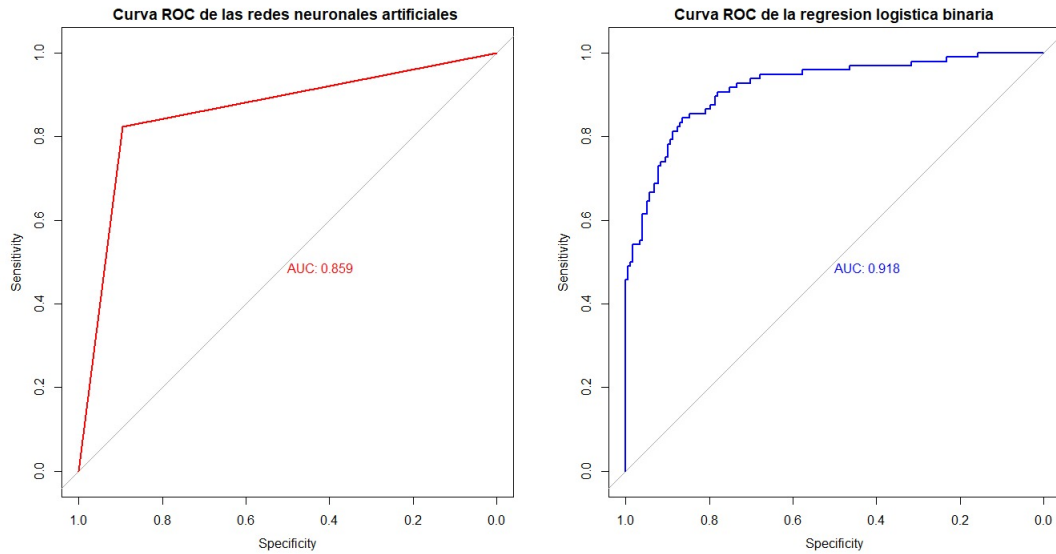
|           | P. Verdadera | P. Verdadera |
|-----------|--------------|--------------|
| Predichos | 0            | 1            |
| 0         | 37           | 1            |
| 1         | 3            | 14           |

**Fuente:** Base de datos madres gestantes y peso del bebé al nacer en el IESS Riobamba periodo 2020-2021.

**Realizado por:** Allayco, Digna y Huaraca, María, 2023.

En la tabla 4–11 Matriz de confusión se puede observar que de los 55 pesos de los bebés, 37 se clasifican correctamente en la clase 0, y 14 se clasifican en la clase 1. Sin embargo, 3 pesos se clasifican como bajo peso pero pertenecen a la muestra que tiene el peso adecuado y 1 peso se clasifica como peso adecuado siendo bajo peso. Obteniendo un porcentaje de clasificación errónea del 7.27%, es decir, el porcentaje de clasificar adecuadamente es del 92.72%.

#### 4.4. Comparación del modelo de regresión con redes neuronales



**Ilustración 4–12:** Curvas de ROC para los modelos logísticos y red neuronal

**Fuente:** Base de datos madres gestantes y peso del bebé al nacer en el IESS Riobamba periodo 2020-2021.

**Realizado por:** Allayco, Digna y Huaraca, María, 2023.

**Tabla 4–12:** Áreas bajo la curva de los modelos de regresión logística y red neuronal

| AUC M. logístico | AUC M. red neuronal |
|------------------|---------------------|
| 0.92             | 0.86                |

**Fuente:** Base de datos madres gestantes y peso del bebé al nacer en el IESS Riobamba periodo 2020-2021.

**Realizado por:** Allayco, Digna y Huaraca, María, 2023.

En la Figura 4–12 Curvas de ROC para los modelos logísticos y red neuronal y en la tabla 4–12 Áreas bajo la curva de los modelos de regresión logística y red neuronal se observa que la mayor área presenta el modelo del regresión logística binaria de 0.92, lo que nos indica que dicho modelo nos ayuda con mayor precisión a determinar los factores maternos que influyen en el peso del bebé al nacer, sin embargo no podemos descartar el modelo de red neuronal ya que presenta una área bajo la curva de 0.86



## CONCLUSIONES

- Mediante el análisis teórico se ha determinado los siguientes factores maternos que tienen alguna influencia en el peso del recién nacido: sexo del bebé, edad materna, Índice de Masa Corporal (IMC), paridad, hábitos nutricionales, edad gestacional, peso materno en el embarazo, intervalo de embarazo, educación de la madre, atención prenatal, grávida de las cuales algunas de estas se utilizan como variables en el estudio para determinar los modelos de regresión logística y redes neuronales.
- La información brindada por parte del Hospital IESS, en donde se detallaba la información e historial patológico de las madres en el periodo enero 2020-diciembre 2021, ayudó a conocer los factores maternos que tienen mayor relevancia en las historias clínicas de la madre en estado de gestación, entre ellas están el peso y talla de la madre, las enfermedades adquiridas durante el embarazo, los hábitos tóxicos, las semanas de gestación, paridad, abortos anteriores, IMC de la madre y sexo del recién nacido

- Al utilizar la técnica de regresión logística binaria se obtuvo la siguiente ecuación:

$$T\_peso = 24.031 + 2.9202 * h\_toxicos - 0.4308 * SA\_gestacion - 0.2274 * Peso - 1.33340 * sexo - 0.7305 * paridad + 1.1624 * abortos + 0.0937 * Diag\_embarazo + 0.0419 * IMC$$

En la cual las variables hábitos tóxicos, semanas de gestación, peso de la madre, sexo del bebé, paridad, abortos anteriores, diagnósticos durante el embarazo e IMC de la madre son las variables significativas, con un valor de  $R^2$  de McFadden de 0.49 lo que indica que es un buen modelo y se ajusta a la base de datos.

En cuanto a la red neuronal se obtiene un modelo con 2 capas ocultas, la primera con 6 neuronas y la segunda con 3, con 100 veces que se pasarán los datos de entrenamiento dando un porcentaje de exactitud del 80 % del modelo Construir modelos mediante las técnicas: Regresión Logística y Redes Neuronales Artificiales.

- Al realizar las curvas de ROC de los 2 modelos planteados, se obtuvo que el área bajo la curva para el modelo de regresión logística es de 0.92 y el AUC para la red neuronal artificial es de 0.86, lo que nos indica que el mejor modelo al comparar mediante las curvas ROC es el modelo logístico.

## **RECOMENDACIONES**

- Realizar una investigación bibliográfica más profunda sobre los posibles factores maternos que influyen en el desarrollo del peso del bebé al nacer.
- Ampliar la base de datos de las madres gestantes y los pesos de los bebés al nacer con información proporcionada por el Instituto Ecuatoriano Seguridad Social (IESS) u otras instituciones de salud pública o privada.
- Aplicar nuevas técnicas estadísticas que permitan continuar con la investigación y determinar con mayor exactitud los factores maternos más influyentes en el peso del recién nacido.
- Se recomienda a las instituciones de salud públicas y privadas divulgar esta información a las madres que se encuentran en estado de gestación para que tomen medidas preventivas y su bebé no nazca con bajo peso.

## BIBLIOGRAFÍA

**ALONSO, A., et al** *Enfermedades periodontales durante el embarazo y su asociación con parto prematuro y bajo peso al nacimiento: Evidencias y controversias*. Avances en Periodoncia e Implantología Oral [en línea], 2010, vol. 22, no 2, p. 85-89. [Consulta: 15 junio 2023]. Disponible en: [https://scielo.isciii.es/scielo.php?pid=S1699-65852010000200004&script=sci\\_arttext&tlng=pt](https://scielo.isciii.es/scielo.php?pid=S1699-65852010000200004&script=sci_arttext&tlng=pt)

**ARMESTO, D; ESPAÑA, B.** *Pruebas diagnósticas: curvas ROC*. *Electron J Biomed*. [En línea] ESPAÑA, 2011, vol. 1, p. 77-82. [Consulta: 26 julio 2023]. Disponible en: <https://www.biomed.uninet.edu/2011/n1/armesto.pdf>

**BIANCHI, M., et al** *DETERMINACIÓN DEL TAMAÑO Y VOLUMEN RENAL EN RECIÉN NACIDOS CONSIDERADOS NORMALES SEGÚN CRITERIOS DE LA OMS MEDIANTE ECOGRAFÍA. ESTUDIO DESCRIPTIVO, OBSERVACIONAL Y TRANSVERSAL CON REVISIÓN DE LA BIBLIOGRAFÍA*. [En línea] Revista de nefrología, diálisis y trasplante. [Consulta: 15 junio 2023]. Disponible en: [http://www.scielo.org.ar/scielo.php?pid=S2346-85482016000400003&script=sci\\_arttext](http://www.scielo.org.ar/scielo.php?pid=S2346-85482016000400003&script=sci_arttext)

**CHÁVEZ, R.** *Comparación entre regresión logística y redes neuronales para predecir cáncer de piel en perros*. [En línea] (Trabajo de investigación). (Ingeniería) Universidad de Lima, PERÚ, 2019. pp. 1-14. [Consulta: 20 junio 2023]. Disponible en: <https://repositorio.ulima.edu.pe/handle/20.500.12724/8401>

**DIANDERAS, E.** *MODELO DE REGRESIÓN LOGÍSTICA*. [En línea] LIMA, 2007, Vol. 10, N° 20. pp. 1-3. [Consulta: 17 julio 2023]. Disponible en: [https://repositorio.ulima.edu.pe/bitstream/handle/20.500.12724/8401/Chavez\\_Martinez\\_Renato.pdf?sequence=3&isAllowed=y](https://repositorio.ulima.edu.pe/bitstream/handle/20.500.12724/8401/Chavez_Martinez_Renato.pdf?sequence=3&isAllowed=y)

**FANEITE, P., et al** . [En línea] Departamento de Obstetricia y Ginecología. Hospital "Dr. Adolfo Prince Lara", Puerto Cabello, Estado Carabobo, Venezuela, 2004. [Consulta: 10 junio 2023]. Disponible en: [http://ve.scielo.org/scielo.php?pid=S0048-77322004000300002&script=sci\\_arttext](http://ve.scielo.org/scielo.php?pid=S0048-77322004000300002&script=sci_arttext)

**FIUZA, M. y RODRIGUEZ, J.** *La regresión logística: una herramienta versátil*. [En línea] Artículo de revista académica, ECUADOR, 2010, Vol. 20, pp. 477-565.

[Consulta: 4 junio 2023]. Disponible en: <https://www.revistanefrologia.com/es-la-regresion-logistica-una-herramienta-articulo-X0211699500035664>

**GÓMEZ, et al** *Aprendizaje con redes neuronales artificiales*. [En línea] Ensayos: Revista de la Facultad de Educación de Albacete, ESPAÑA, 2000, no 9, p. 169-180. [Consulta: 8 mayo 2023]. Disponible en: <https://redined.educacion.gob.es/xmlui/handle/11162/228254>

**GÓMEZ, E.** *Trastornos hipertensivos durante el embarazo*. *Revista cubana de obstetricia y ginecología*. [En línea] Revista Cubana de Obstetricia y Ginecología, 2010, 26(2), P. 99-114 [Consulta: 13 marzo]. Disponible en: [http://scielo.sld.cu/scielo.php?pid=S0138-600X2000000200006&script=sci\\_arttext&tlng=en](http://scielo.sld.cu/scielo.php?pid=S0138-600X2000000200006&script=sci_arttext&tlng=en)

**KIRIŞCI, M.** *Comparison of artificial neural network and logistic regression model for factors affecting birth weight*. [En línea] Research Article, Switzerland, 2019. vol. 1, no 4, p. 378.[Consulta: 21 julio 2023]. Disponible en: <https://link.springer.com/article/10.1007/s42452-019-0391-x>

**MANYA, P.** *PLAN MÉDICO FUNCIONAL HOSPITAL IESS RIOBAMBA*. [En línea] Artículo de revista académica, ECUADOR, 2020. vol. 1, p. 38.[Consulta: 24 julio 2023]. Disponible en: <https://www.iess.gob.ec/documents/10162/3321614/PMF+HOSPITAL+RIOBAMBA.pdf>

**MORENO, J, et al.** *Redes neuronales artificiales aplicadas al análisis de datos*. [En línea] Doctorado Psicología, Facultad de Psicología, Tesis doctoral de la Facultad de Psicología, Universitat de Les Illes Balears, Palma de Mallorca, España, 2002.[Consulta: 25 julio 2023]. Disponible en: <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=eae427113918d9780e856aed7cba91fb8d414209>

**NAYEBI, H.** *Advanced statistics for testing assumed casual relationships*. [En línea] Springer International Publishing, 2020. vol. 1, p. 378.[Consulta: 29 junio 2023]. Disponible en: <https://link.springer.com/book/10.1007/978-3-030-54754-7>

**OLABE, X.** *Redes neuronales artificiales y sus aplicaciones*. [En línea] Publicaciones de la Escuela de Ingenieros, 1998, Escuela Superior de Ingeniería de Bilbao, EHU. [Consulta: 19 julio 2023]. Disponible en: [https://ocw.ehu.eus/pluginfile.php/40137/mod\\_resource/content/1/redes\\_neuro/contenidos/pdf/libro-del-curso.pdf](https://ocw.ehu.eus/pluginfile.php/40137/mod_resource/content/1/redes_neuro/contenidos/pdf/libro-del-curso.pdf)

**OMS** *Metas mundiales de nutrición 2025: documento normativo sobre bajo peso al nacer.* [En línea] 2022. [Consulta: 2 junio 2023]. Disponible en: <https://www.who.int/es/publications/i/item/WHO-NMH-NHD-14.5#:~:text=El%20objetivo%20para%202025%20es,millones%20a%20unos%2014%20millones.>

**PÉREZ, J. y PÉREZ, M.** *La curva ROC.* [En línea] Research Article, Medicina de Familia, SEMERGEN, 2023, vol. 49, no 1, p. 10-21.[Consulta: 3 julio 2023]. Disponible en: <https://www.sciencedirect.com/science/article/abs/pii/S1138359322001952>

**PERIS, M, et al.** *Embarazo.* [En línea] FMC-Formación Médica Continuada en Atención Primaria, 2021, vol. 28, no 2, p. 112-117.[Consulta: 9 julio 2023]. Disponible en: <https://www.sciencedirect.com/science/article/abs/pii/S113420722030178X>

**RIVAS, W. y MAZÓN, B.** *Redes neuronales artificiales aplicadas al reconocimiento de patrones.* [En línea] En IV Workshop de Investigadores en Ciencias de la Computación. ECUADOR, 2018. [Consulta: 18 julio 2023]. Disponible en: <http://repositorio.utmachala.edu.ec/handle/48000/12499>

**ROSELL, E, et al.** *Factores de riesgo de la enfermedad hipertensiva del embarazo.* [En línea] Research Article, 2006 [Consulta: 9 julio 2023]. Disponible en: [http://scielo.sld.cu/scielo.php?pid=S1025-02552006000500007&script=sci\\_arttext](http://scielo.sld.cu/scielo.php?pid=S1025-02552006000500007&script=sci_arttext)



## ANEXOS

```
1 #####regresion logistica binaria #####
2
3 #1. llamado de nuestra base de datos
4
5 base<-read.csv("BASE.csv", header = TRUE, sep = ";", dec = ".")
6 base
7 # 2. reconocer nuestra base de datos
8 ##### Nombre de nuestra base de datos#####
9 #Estructura de nuestra base de datos#####
10 names(base)
11 str(base)
12 head(base)
13 #INTERPRETACION##
14 ### Se trata de una base de datos que en su mayoria de sus variables
15 ### son de tipo categorico salvo las variables edad,talla,IMC,peso,
16 ### paridad,abortos anteriores.
17
18 # 3. Volver factor(Dummy) nuestra variable dependiente o
19 # nuestra variable objetivo
20 install.packages("car")
21 install.packages("cardata")
22 library(car) ## para recodificar
23 table(base$T_PESO)
24 # nos indica que existen ----casos de bebes que nacieron con
25 #bajos pero indicado por el valor 1 y ----casos de bebes que
26 #no nacieron con bajo peso indicado por el valor 0
27
28 base$T_PESO<-as.factor(base$T_PESO)
29 levels(base$T_PESO) #niveles de la variable dependiente
30
31 ##### 4. calcular el modelo
32
33 #funcion glm: es una funcion que nos permite general modelos
34 #lineales generalizados que es una derivacion del modelo
35 #lineal general pero que esta especialmente
36 #adecuado para tener una variable dependiente dicotomica
```

```

37 #de tipo factor.
38 log1<-glm(T_PESO~EDAD+Peso.kg.+TALLA+DIAGNOSTICO_D_EMBARAZO+
39           SEMANAS_DE_GESTACION+HABITOS_TOXICOS+
40           ABORTOS_ANTERIORES+PARIDAD+
41           ANTECEDENTE_PAT_PERSONALES+
42           ANTECEDENTES_ALERGICOS+IMC+SEXO, data=base,
43           family = binomial(link = "logit"))
44 log1
45 # 1. Modelo sin predictores
46 null_model <- glm(T_PESO~1, data =base, family = "binomial")
47 # 2. modelo completo
48 full_model <- glm(T_PESO~., data = base, family = "binomial")
49 # 3. funcion step ()
50 step_model<-step(null_model, scope = list(lower = null_model,
51           upper = full_model), direction = "forward")
52
53 ##Análisis descriptivo del modelo
54
55 summary(step_model)
56 ##Criterio de arkaike del modelo
57
58 AIC(step_model)
59 ##R^2 de McFadden
60 library(DescTools)
61 PseudoR2(step_model, which = "McFadden")
62 # estimamos la probabilidad
63 step_prob <- predict(step_model, type = "response")
64
65 # Pintamos ROC of the stepwise model
66
67 library(pROC)
68 ROC<-roc(base$T_PESO, step_prob)
69 plot(ROC, col = "blue", print.auc=T)
70 title("Curva ROC de la regresion logistica binaria",
71       line = 2.5)
72 ##Area bajo la curva del modelo
73 auc(ROC)
74
75 ###Redes neuronales artificiales ####

```

```

76
77 #Conjunto de Datos
78 dataset<- read.csv("BASER.csv", header = TRUE,
79                   sep = ";",dec = ",")
80 dataset
81 #Haciendo uso de la funci n str, podemos explorar la
82 #estructura del dataframe que contiene el conjunto de datos:
83 str(dataset)
84 colnames(dataset)[13] <- "PESO"
85 dataset$PESO <-factor(dataset$PESO, levels = c("0", "1"),
86                   labels = c("no", "si"))
87 summary(dataset)
88
89 #En este sentido, la aplicaci n de las redes neuronales artificiales
90 #exige que las variables de entrada est n todas
91 #rescaladas en el mismo rango, por lo que haremos uso de la
92 #funci n scale para cumplir con esta condici n:
93 dataset[, c(1:12)] <- scale(dataset[, c(1:12)])
94 summary(dataset)
95
96 #Por ltimo , vamos a crear un conjunto de entrenamiento (80%) y
97 #un conjunto de validaci n (20%) a partir de nuestro
98 #dataset original, esto con la intenci n de validar la calidad
99 #de la clasificaci n producida por la red neuronal.
100 #Para esto, hagamos uso de la funci n sample.split del
101 #paquete caTools:
102 library(caTools)
103 set.seed(1234)
104 split <- sample.split(dataset$PESO, SplitRatio = 0.80)
105 training_set <- subset(dataset, split == TRUE)
106 test_set <- subset(dataset, split == FALSE)
107
108
109 ##GRAFICA DE LA NEURONA #####
110
111 model.net = neuralnet(
112   formula = PESO~.,
113   data = dataset,
114   hidden = c(6,3),

```



```

115 stepmax = 1e+5,
116 linear.output = TRUE,
117 act.fct = "logistic",)
118
119 plot(model.net, rep="best")
120 ##nombres de las variables del modelo
121
122 names(model.net)
123 unlist(model.net$weights)
124
125 #A fin de asegurarnos de que la proporción de estrellas pulsares
126 #y no pulsares es aproximadamente la misma en ambos
127 #conjuntos de datos, veamos la distribución de los tipos de estrella
128 #en cada uno:
129 table(training_set$PESO)
130 table(test_set$PESO)
131
132 #La implementación de la red neuronal la haremos haciendo uso
133 #del paquete H2O, el cual es una librería para análisis
134 #predictivo y machine learning que incorpora funciones para crear
135 #redes neuronales artificiales así como modelos de
136 #deep learning. En primer lugar, es necesario entonces cargar la
137 #librería e inicializarla:
138 #install.packages("h2o")
139 library(h2o)
140 h2o.init(nthreads = -1)
141
142 ###MODELOS ###
143
144
145 classifier_1 = h2o.deeplearning(y = 'PESO',
146                               training_frame = as.h2o(training_set),
147                               activation = 'Rectifier',
148                               hidden = c(10,4,5),
149                               epochs = 100,
150                               train_samples_per_iteration = -2,
151                               seed=1234)
152
153 classifier_2 = h2o.deeplearning(y = 'PESO',

```

```

154         training_frame = as.h2o(training_set),
155         activation = 'Rectifier',
156         hidden = c(6,10),
157         epochs = 100,
158         train_samples_per_iteration = -2,
159         seed=1234)
160
161 classifier_3 = h2o.deeplearning(y = 'PESO',
162         training_frame = as.h2o(training_set),
163         activation = 'Rectifier',
164         hidden = c(30,6,9),
165         epochs = 1e+5,
166         train_samples_per_iteration = -2,
167         seed=1234)
168
169 classifier_4 = h2o.deeplearning(y = 'PESO',
170         training_frame = as.h2o(training_set),
171         activation = 'Rectifier',
172         hidden = c(24,50),
173         epochs = 1e+5,
174         train_samples_per_iteration = -2,
175         seed=1234)
176
177 ###predicciones de los modelos
178
179 predict_test_1 <- h2o.predict(object = classifier_1,
180         newdata = as.h2o(test_set))
181 accuracy_1_new <- mean(predict_test_1["predict"] ==
182         as.h2o(test_set)$PESO)
183
184 predict_test_2 <- h2o.predict(object = classifier_2,
185         newdata = as.h2o(test_set))
186 accuracy_2_new <- mean(predict_test_2["predict"] ==
187         as.h2o(test_set)$PESO)
188
189 predict_test_3 <- h2o.predict(object = classifier_3,
190         newdata = as.h2o(test_set))
191 accuracy_3_new <- mean(predict_test_3["predict"] ==
192         as.h2o(test_set)$PESO)

```

```

193
194 predict_test_4 <- h2o.predict(object = classifier_4,
195                               newdata = as.h2o(test_set))
196 accuracy_4_new <- mean(predict_test_4["predict"] ==
197                          as.h2o(test_set)$PESO)
198
199 #install.packages("glue")
200 library(glue)
201 glue("Accuracy del Modelo 1 Nuevo: {accuracy_1_new} \n
202       Accuracy del Modelo 2 Nuevo: {accuracy_2_new} \n
203       Accuracy del Modelo 3 Nuevo: {accuracy_3_new} \n
204       Accuracy del Modelo 4 Nuevo : {accuracy_4_new} \n
205       " )
206
207 ###Grafica de los modelos ####
208
209 y_pred <- as.vector(ifelse(predict_test_2$predict == 'no', 0, 1))
210 y_test_set <-ifelse(test_set$PESO == 'no', 0, 1)
211
212 ##matriz de confucion
213
214 cm <- table(y_test_set, y_pred)
215 cm
216 ##vector del porcentaje de exactitud
217
218 accuracy <- c(accuracy_1_new, accuracy_2_new,
219               accuracy_3_new, accuracy_4_new)
220 model<- c("classifier_1", "classifier_2",
221           "classifier_3", "classifier_4")
222
223 #install.packages("ggplot2")
224 library(ggplot2)
225
226 data<- data.frame(accuracy, model)
227 ggplot(data, aes(x= reorder(model, -accuracy),
228                  y=accuracy, fill=model ),
229         title("Accuracy de Modelos Planteados "))+
230 geom_bar(stat ="identity")+
231 theme(axis.text.x = element_text(angle = 90,

```

```
232         vjust = 0.5, hjust=1))
233
234
235
236 ##CURVAS DE ROC#####
237
238 par(mfrow=c(1,2))
239 ###redes neuronales
240
241 curva_roc <- roc(y_test_set, y_pred)
242 p1<-plot(curva_roc,col ="red",print.auc=T);p1
243 title("Curva ROC de las redes neuronales artificiales ", line = 2.5)
244 #Area bajo la curva del modelo
245 auc(curva_roc)
246 ##regresion logistica binaria
247
248 ROC<-roc(base$T_PESO, step_prob)
249 p2<-plot(ROC,col ="blue",print.auc=T);p2
250 title("Curva ROC de la regresion logistica binaria", line = 2.5)
251 #Area bajo la curva del modelo
252 auc(ROC)
```



epoch

Dirección de Bibliotecas y  
Recursos del Aprendizaje

UNIDAD DE PROCESOS TÉCNICOS Y ANÁLISIS BIBLIOGRÁFICO Y  
DOCUMENTAL

REVISIÓN DE NORMAS TÉCNICAS, RESUMEN Y BIBLIOGRAFÍA

Fecha de entrega: 10 / 01 / 2024

|  |
|--|
| <b>INFORMACIÓN DEL AUTOR/A (S)</b>   |
| Digna Rocio Allayco Alvarez<br>Maria Fernanda Huaraca Guashpa                |
| <b>INFORMACIÓN INSTITUCIONAL</b>   |
| <b>Facultad:</b> Ciencias  |
| <b>Carrera:</b> Estadística  |
| <b>Título a optar:</b> Ingeniera Estadística                                 |
| <b>f. Analista de Biblioteca responsable:</b> Ing. Rafael Inty Salto Hidalgo |

2152-DBRA-UPT-2023

