



UNIVERSIDAD TÉCNICA PARTICULAR DE LOJA

*La Universidad Católica de Loja*

ÁREA TÉCNICA

TÍTULO DE INGENIERO EN INFORMÁTICA

Técnicas de predicción aplicados a criaderos del mosquito transmisor de enfermedades como Dengue, Zika y Chikungunya en la ciudad de Machala

TRABAJO DE TITULACIÓN

AUTOR: Adrian Delgado, Jefferson Eduardo

DIRECTOR: Valdiviezo Díaz, Priscila Marisela, Mgtr.

CENTRO UNIVERSITARIO MACHALA

2018



*Esta versión digital, ha sido acreditada bajo la licencia Creative Commons 4.0, CC BY-NC-SA: Reconocimiento-No comercial-Compartir igual; la cual permite copiar, distribuir y comunicar públicamente la obra, mientras se reconozca la autoría original, no se utilice con fines comerciales y se permiten obras derivadas, siempre que mantenga la misma licencia al ser divulgada. <http://creativecommons.org/licenses/by-nc-sa/4.0/deed.es>*

*Loja, septiembre del 2018*

## APROBACIÓN DE LA DIRECTORA DEL TRABAJO DE TITULACIÓN

Magister.

Priscila Marisela Valdiviezo Díaz.

DOCENTE DE LA TITULACIÓN

De mi consideración:

El presente trabajo de titulación: Técnicas de predicción aplicados a criaderos del mosquito transmisor de enfermedades como Dengue, Zika y Chikungunya en la ciudad de Machala, realizado por Adrian Delgado Jefferson Eduardo, ha sido orientado y revisado durante su ejecución, por cuanto se aprueba la presentación del mismo.

Loja, septiembre de 2018

f) .....

## DECLARACIÓN DE AUTORÍA Y CESIÓN DE DERECHOS

“Yo, Adrian Delgado Jefferson Eduardo declaro ser autor del presente trabajo de titulación: Técnicas de predicción aplicados a criaderos del mosquito transmisor de enfermedades como Dengue, Zika y Chikungunya en la ciudad de Machala, de la Titulación de Informática, siendo Priscila Marisela Valdiviezo Díaz directora del presente trabajo; y eximo expresamente a la Universidad Técnica Particular de Loja y a sus representantes legales de posible reclamos o acciones legales. Además, certifico que las ideas, conceptos, procedimientos y resultados vertidos en el presente trabajo investigativo, son de mi exclusiva responsabilidad.

Adicionalmente declaro conocer y aceptar la disposición del Art. 88 del Estatuto Orgánico de la Universidad Técnica Particular de Loja que en su parte pertinente textualmente dice: “Forman parte del patrimonio de la Universidad la propiedad intelectual de investigaciones, trabajos científicos o técnicos y tesis de grado o trabajos de titulación que se realicen con el apoyo financiero, académico o institucional (operativo) de la Universidad”

f) .....

Autor: Jefferson Eduardo Adrian Delgado

Cédula: 0703183756

## **DEDICATORIA**

A mis hijos Joel y Francis, que son mi motivo constante de superación, a mi esposa Norma, por su comprensión y motivación permanente en cada etapa de mi vida.

A mis padres Luis y Blanca por su apoyo incondicional, a mis hermanos Vladimir y Liliana, a toda mi familia por su ánimo a seguir superándome.

## **AGRADECIMIENTO**

Un agradecimiento a la Universidad Técnica Particular de Loja, en especial a la Mgtr. Priscila Valdiviezo, Mgtr. Guido Riofrío, Mgtr. Rodrigo Barba, por el conocimiento y apoyo que me brindaron, a la Mgtr. Gladys Tenesaca, por su bondad y ánimo constante.

A la Universidad Técnica de Machala, en especial al Mgtr. Efraín Beltrán y PhD. Iván Ramírez, por su ayuda incondicional y desinteresada.

Al Departamento de Control Vectorial del Distrito de Salud 07D02 Machala, especialmente a la Mgtr. Tania Ordóñez, por su apoyo y colaboración.

A mis amigos y todas las personas que me ayudaron a la consecución de este trabajo de titulación.

## INDICE DE CONTENIDOS

<b>APROBACIÓN DE LA DIRECTORA DEL TRABAJO DE TITULACIÓN</b> .....	ii
<b>DECLARACIÓN DE AUTORÍA Y CESIÓN DE DERECHOS</b> .....	iii
<b>DEDICATORIA</b> .....	iv
<b>AGRADECIMIENTO</b> .....	v
<b>INDICE DE CONTENIDOS</b> .....	vi
<b>ÍNDICE DE FIGURAS</b> .....	viii
<b>ÍNDICE DE TABLAS</b> .....	x
<b>RESUMEN</b> .....	1
<b>ABSTRACT</b> .....	2
<b>INTRODUCCIÓN</b> .....	3
<b>CAPÍTULO I. PLANTEAMIENTO</b> .....	4
1.1 Descripción del problema .....	5
1.2 Propuesta de solución.....	5
1.3 Metodología .....	6
<b>CAPÍTULO II. MARCO TEÓRICO</b> .....	8
2.1 Minería de Datos.....	9
2.2 Árboles de decisión .....	12
2.3 Series Temporales.....	13
2.4 Red Bayesiana .....	14
2.4.1. Concepto.....	14
2.4.2. Ingeniería del conocimiento .....	15
2.5 Red Neuronal .....	16
2.5.1. Concepto.....	16
2.5.2. Ventajas .....	16
2.5.3. Estructura .....	17
2.5.4. Elementos.....	20
2.5.5. Fases.....	20
2.5.5.1. Fase de aprendizaje o entrenamiento.....	20
2.5.5.2. Fase de operación o ejecución.....	21
2.5.6. Métricas de evaluación .....	22
2.6 Trabajos relacionados .....	22
2.6.1. Sistema de predicción de ruido urbano mediante redes neuronales. ....	22

2.6.2. Entrenamiento de redes neuronales basado en algoritmos evolutivos.....	23
2.6.3. Modelamiento matemático y por redes neuronales artificiales del crecimiento de Spirulina sp. en fotobiorreactor con fuente de luz fluorescente e iluminación en estado sólido.....	23
<b>CAPÍTULO III. ANÁLISIS Y PREPARACIÓN DE DATOS .....</b>	<b>25</b>
3.1 Obtención de los datos.....	26
3.1.1. Datos de Control Vectorial. ....	26
3.1.2. Datos de Clima. ....	28
3.1.3. Datos socioeconómicos.....	28
3.2 Preparación de los datos .....	29
<b>CAPÍTULO IV. SELECCIÓN DE TÉCNICA DE PREDICCIÓN .....</b>	<b>32</b>
4.1 Selección de los datos .....	33
4.1.1. Ensayo 1.....	34
4.1.2. Ensayo 2.....	35
4.1.3. Ensayo 3.....	36
4.2 Series Temporales.....	36
4.3 Selección de la herramienta .....	38
4.4 Crear Red Neuronal en Matlab .....	38
4.5 Entrenamiento de la Red Neuronal .....	41
<b>CAPÍTULO V. IMPLEMENTACIÓN DE LA TÉCNICA .....</b>	<b>43</b>
5.1 Creación del scrip de la Red Neuronal Entrenada .....	44
5.2 Ingreso de datos y simulación en la Red Neuronal Entrenada.....	45
5.3 Obtención de los resultados .....	48
<b>CAPÍTULO VI. EVALUACIÓN DE LA TÉCNICA.....</b>	<b>50</b>
6.1 Métricas de evaluación.....	51
6.2 Mapa temático.....	54
<b>CONCLUSIONES.....</b>	<b>59</b>
<b>RECOMENDACIONES .....</b>	<b>61</b>
<b>BIBLIOGRAFÍA.....</b>	<b>62</b>
<b>ANEXOS.....</b>	<b>64</b>
<b>Anexo 1 .....</b>	<b>65</b>
<b>Anexo 2.....</b>	<b>66</b>
<b>Anexo 3.....</b>	<b>67</b>
<b>Anexo 4.....</b>	<b>68</b>

<b>Anexo 5</b> .....	69
<b>Anexo 6</b> .....	72
<b>Anexo 7</b> .....	75
<b>Anexo 8</b> .....	77
<b>Anexo 9</b> .....	78

## ÍNDICE DE FIGURAS

<b>Figura 1.</b> Diagrama de la solución planteada. ....	6
<b>Figura 2.</b> Proceso de descubrimiento de conocimiento. ....	10
<b>Figura 3.</b> Clasificación de las técnicas de Data Mining. ....	11
<b>Figura 4.</b> Ejemplo de árbol de decisión para decidir si esperar por una mesa. ....	13
<b>Figura 5.</b> Ejemplo de Serie de Tiempo. ....	14
<b>Figura 6.</b> Ejemplo de red bayesiana. ....	15
<b>Figura 7.</b> Esquema de una red neuronal. ....	17
<b>Figura 8.</b> Ejemplo de Red Neuronal Feedforward de Capa Simple. ....	18
<b>Figura 9.</b> Ejemplo de Red Neuronal Feedforward Multi - Capa. ....	19
<b>Figura 10.</b> Ejemplo de Red Neuronal Recurrente. ....	19
<b>Figura 11.</b> Matriz de correlación. ....	33
<b>Figura 12.</b> Influencia de la variable Precipitación frente a Casas Positivas con presencia del mosquito Aedes aegypti. ....	37
<b>Figura 13.</b> Influencia de las variables Temperatura máxima y mínima frente a Casas Positivas con presencia del mosquito Aedes aegypti. ....	37
<b>Figura 14.</b> Arquitectura de la red neuronal – Ensayo 2. ....	40
<b>Figura 15.</b> Mejor rendimiento del grupo de datos de entrenamiento de la red neuronal – Ensayo 2. ....	42
<b>Figura 16.</b> Generación del scrip de la red neuronal entrenada. ....	44
<b>Figura 17.</b> Resumen del scrip de la red neuronal entrenada. ....	45
<b>Figura 18.</b> Resultado del entrenamiento de la red neuronal artificial – Ensayo 1. ....	51
<b>Figura 19.</b> Resultado del entrenamiento de la red neuronal artificial – Ensayo 2. ....	51
<b>Figura 20.</b> Resultado del entrenamiento de la red neuronal artificial – Ensayo 3. ....	52
<b>Figura 21.</b> Análisis de regresión de la red neuronal – Ensayo 2. ....	53
<b>Figura 22.</b> Mapa temático de la ciudad de Machala con las casas positivas, trabajadas por el personal del Departamento de Control Vectorial del Distrito de Salud 07D02, semana 49 del mes de diciembre de 2017. ....	55
<b>Figura 23.</b> Mapa temático de la ciudad de Machala con las casas positivas utilizando red neuronal artificial, semana 49 del mes de diciembre de 2017. ....	56
<b>Figura 24.</b> Mapa temático de la ciudad de Machala con las casas positivas trabajadas por el personal del Departamento de Control Vectorial del Distrito de Salud 07D02, semana 20 del mes de mayo de 2017. ....	57

**Figura 25.** Mapa temático de la ciudad de Machala con las casas positivas utilizando red neuronal artificial, semana 20 del mes de mayo de 2017. ....57

## ÍNDICE DE TABLAS

<b>Tabla 1.</b> Datos de visitas domiciliarias.....	26
<b>Tabla 2.</b> Datos climáticos.....	28
<b>Tabla 3.</b> Datos Socioeconómicos.....	28
<b>Tabla 4.</b> Obtención de nuevas variables.....	30
<b>Tabla 5.</b> Descripción de los valores de correlación.....	34
<b>Tabla 6.</b> Variables seleccionadas de acuerdo al grado de correlación, Ensayo 1.....	35
<b>Tabla 7.</b> Variables seleccionadas de acuerdo al grado de correlación, Ensayo 2.....	35
<b>Tabla 8.</b> Variables seleccionadas de acuerdo al grado de correlación, Ensayo 3.....	36
<b>Tabla 9.</b> Parámetros de entrenamiento de la Red Neuronal Ajustada.....	42
<b>Tabla 10.</b> Datos para el ingreso en la red neuronal entrenada para la predicción – Ensayo 1. .....	46
<b>Tabla 11.</b> Datos para el ingreso en la red neuronal entrenada para la predicción – Ensayo 2. .....	47
<b>Tabla 12.</b> Datos para el ingreso en la red neuronal entrenada para la predicción – Ensayo 3. .....	47
<b>Tabla 13.</b> Comparación de datos trabajados frente a datos de predicción – Ensayo 1, 2 y 3. .....	49
<b>Tabla 14.</b> Resumen de resultados de las métricas de evaluación.....	53
<b>Tabla 15.</b> Estratificación de Riesgo Entomológico para Índices aélicos.....	55
<b>Tabla 16.</b> Comparación de resultados de predicción - semana 49 del año 2017.....	58

## RESUMEN

El objetivo del presente trabajo de titulación es el desarrollo de una herramienta de predicción de criaderos de mosquitos *Aedes aegypti* en la ciudad de Machala, este vector transmite enfermedades como Dengue, Zika y Chikungunya.

La técnica de predicción utilizada en esta investigación es la red neuronal artificial desarrollada con el software propietario Matlab, el resultado de predicción obtenido se lo comparará con los resultados obtenidos por el Departamento de Control Vectorial del Distrito de Salud 07D02 Machala, para medir la eficacia de la predicción.

Los resultados se presentarán en mapas temáticos de la ciudad de Machala elaborados en el software QGIS con licencia GNU GPL, clasificados de acuerdo a la escala de riesgo entomológico utilizado por el Distrito de Salud 07D02 Machala.

**PALABRAS CLAVES:** Predicción, *Aedes aegypti*, criaderos de mosquitos, red neuronal artificial, mapas temáticos.

## ABSTRACT

The objective of the present titration work is the development of a prediction tool for mosquito breeding *Aedes aegypti* in the city of Machala, this vector transmits diseases such as Dengue, Zika and Chikungunya.

The prediction technique used in this research is the artificial neural network developed with proprietary software Matlab, the prediction result obtained will be compared with the results obtained by the Department of Vector Control of the Health District 07D02 Machala, to measure the effectiveness of the prediction.

The results will be presented in thematic maps of the city of Machala prepared in the software QGIS with GNU GPL license, classified according to the scale of entomological risk used by the Health District 07D02 Machala.

**KEYWORDS:** Prediction, *Aedes aegypti*, mosquito breeding sites, artificial neural network, thematic maps.

## INTRODUCCIÓN

Una de las mayores preocupaciones del ser humano es la conservación de la salud, especialmente si esta puede ser mantenida con ciertos cuidados y educación sanitaria para prevenir enfermedades, ciudades costeras como Machala están expuestas a enfermedades tropicales como Dengue, Zika y Chikungunya, todas estas enfermedades son transmitidas a través de la picadura del mosquito hembra infectado del género *Aedes Aegypti*, por tal motivo la eliminación o control de sus criaderos de desarrollo se constituyen en una actividad importante para contrarrestar la presencia de este mosquito.

Por tal motivo se desarrolla una herramienta de predicción de criaderos de mosquitos *Aedes aegypti*, con la combinación de datos de control vectorial, clima y socioeconómicos, utilizando técnicas de predicción como redes neuronales artificiales, cuyos resultados de predicción son graficados en mapas temáticos.

El trabajo de titulación tiene la siguiente estructura:

Capítulo I. Planteamiento: Descripción del problema y propuesta de solución.

Capítulo II. Marco Teórico: Fundamento teórico para la realización del presente trabajo de investigación.

Capítulo III. Análisis y preparación de datos: Descripción, recopilación y preparación de las variables que conforman el dataset del trabajo de investigación.

Capítulo IV. Selección de técnica de predicción: Elección de Redes Neuronales Artificiales como técnica de predicción.

Capítulo V. Implementación de la técnica: Simulación de la red neuronal con datos que no forman parte del dataset.

Capítulo VI. Evaluación de la técnica: Aplicación de métricas de evaluación para establecer el grado de acierto de los resultados de la red neuronal.

**CAPÍTULO I.**  
**PLANEAMIENTO**

## **1.1 Descripción del problema**

Enfermedades tropicales como el Dengue, Zika, Chikungunya son endémicas en la parte baja de la provincia de El Oro, especialmente la ciudad de Machala capital de la provincia. Estas enfermedades son transmitidas por el mosquito hembra infectado del género *Aedes aegypti*, y la principal amenaza es la presencia de sus criaderos, por lo general estos mosquitos se desarrollan en aguas limpias, lo cual constituye un gran problema, porque la transmisión de estas enfermedades provoca en la población ausentismo laboral, pérdidas económicas y aumento de la atención de salud pública y privada, con lo que retrasa el progreso productivo y económico de la ciudad.

Varios son los factores que inciden en el incremento de los criaderos del mosquito *Aedes aegypti*, por ejemplo: la acumulación de agua para consumo humano, presencia de lluvias, aumento de la temperatura ambiental, mala práctica de educación sanitaria de la población, entre otras.

A nivel de la provincia de El Oro, según datos consultados en la página oficial de la Dirección Nacional de Vigilancia Epidemiológica del Ministerio de Salud del Ecuador, en el periodo 2011 al 2016, el promedio de registros anuales de la enfermedad del dengue supera los 3000 casos, alrededor de 600 casos de la enfermedad de Fiebre Chikungunya en el período 2015 – 2016 y 7 casos de Zika en el año 2016, cabe mencionar que estas dos últimas enfermedades reportaron casos en los años 2015 y 2016 respectivamente.

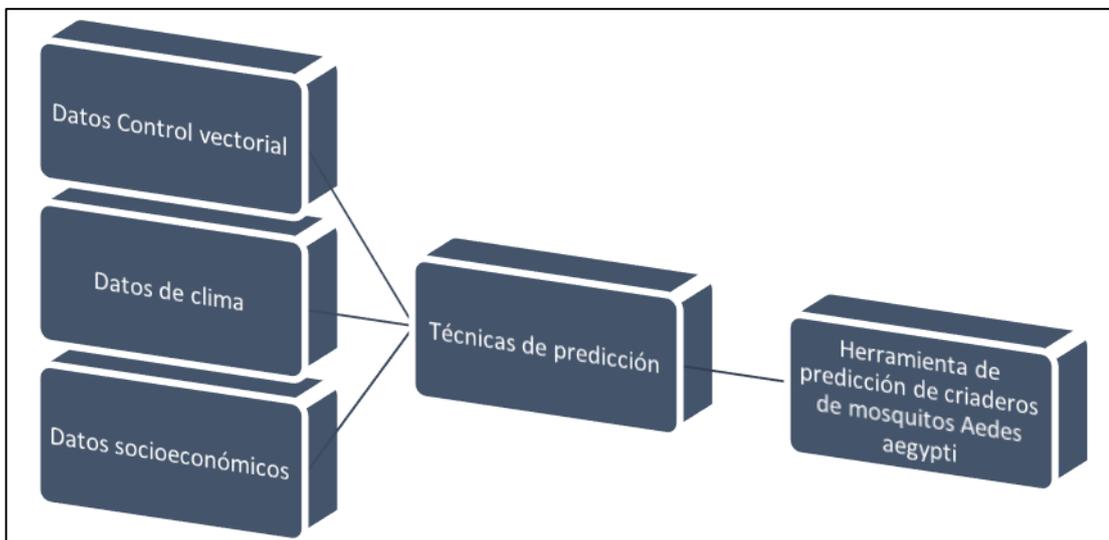
El Departamento de Control Vectorial del Distrito de Salud 07D02 del Ministerio de Salud, es el responsable de realizar actividades de control y prevención de enfermedades en la ciudad de Machala, pero no cuenta con una herramienta que le permita pronosticar la presencia de criaderos de mosquitos, especialmente *Aedes aegypti*.

## **1.2 Propuesta de solución**

El objetivo principal de la propuesta es crear una herramienta que permita identificar los lugares o zonas en la ciudad de Machala donde se podría presentar un incremento de los criaderos de mosquitos, brindando a las autoridades de salud encargadas de realizar el control de estos vectores, una herramienta de predicción para detectar prematuramente la presencia de criaderos y tomar medidas necesarias para evitar la transmisión de enfermedades tropicales como Dengue, Zika o Chikungunya.

La investigación será apoyada por el Centro de Investigación de Salud Global de la Universidad Upstate de Nueva York y Universidad Técnica de Machala, y posteriormente se coordinará con el Ministerio de Salud Pública del Ecuador para implementar un plan piloto en el Distrito de Salud de Machala para realizar actividades de prevención y control de criaderos antes que estos se conviertan en un potencial peligro para la salud de la población.

El desarrollo de la investigación se basará en técnicas de predicción donde tendrá como datos de entrada: Datos de control vectorial, Datos de clima y Datos Socioeconómicos, cuyos detalles de los atributos que los conforman se describen en la sección Análisis y preparación de los datos, los mismos que deberá predecir la presencia de criaderos del mosquito *Aedes aegypti* representado en un mapa temático de la ciudad de Machala.



**Figura 1.** Diagrama de la solución planteada.

**Fuente:** El autor

**Elaborado por:** El autor

### 1.3 Metodología

Las técnicas de predicción serán aplicadas a los datos obtenidos de las zonas o barrios de la ciudad de Machala, los datos de control vectorial serán proporcionados por el Departamento de Control Vectorial del Distrito de Salud de Machala del Ministerio de Salud Pública, los datos de clima serán facilitados por el Instituto Nacional de Meteorología e Hidrología – INAMHI y los datos socioeconómicos será obtenidos de la página web del Instituto Nacional de Estadísticas y Censos – INEC. Para el modelo se analizará varias alternativas como Redes Neuronales, Series Temporales, Redes Bayesianas, Árboles de decisión.

Las fases de desarrollo serían: obtener los datos mencionados, prepararlos para que sirvan como fuente de entrada, crear la herramienta utilizando técnicas de predicción como las redes neuronales, entrenarla utilizando el algoritmo de regulación bayesiana y evaluar la herramienta utilizando métricas como el valor de regresión y el error cuadrático medio, posteriormente representar en un mapa temático de la ciudad de Machala los resultados obtenidos, utilizando software GIS (Sistema de Información Geográfica).

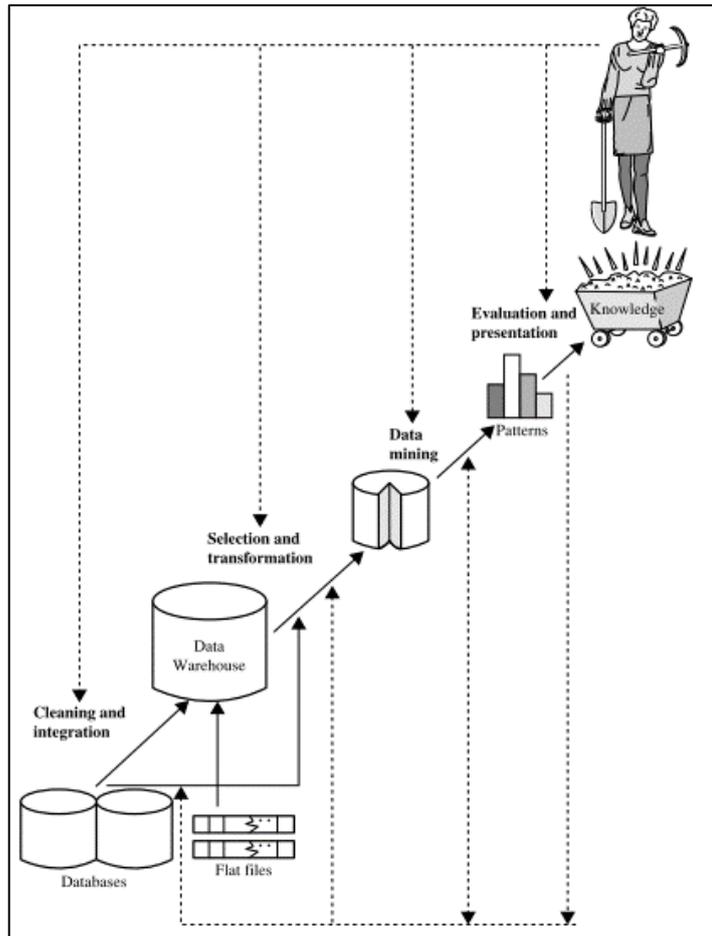
**CAPÍTULO II.**  
**MARCO TEÓRICO.**

## 2.1 Minería de Datos

Como menciona (Dueñas-Reyes, 2009), es “un proceso no trivial de identificación válida, novedosa, potencialmente útil y entendible de patrones comprensibles que se encuentran ocultos en los datos” (p. 143). Un concepto más adecuado debería ser “minería de conocimiento de datos” (Jiawei, Kamber, Han, Kamber, & Pei, 2012, p. 6).

Para descubrir el conocimiento se debe seguir los siguientes pasos como lo mencionan (Jiawei et al., 2012, p. 6):

1. Limpieza de datos (para eliminar el ruido y datos inconsistentes)
2. Integración de datos (donde se pueden combinar múltiples fuentes de datos)
3. Selección de datos (donde los datos relevantes para la tarea de análisis se recuperan de la base de datos)
4. Transformación de datos (donde los datos se transforman y consolidan en formularios apropiados para la minería realizando operaciones de resumen o agregación)
5. La minería de datos (un proceso esencial en el que se aplican métodos inteligentes para extraer patrones de datos)
6. Evaluación de patrones (para identificar los patrones verdaderamente interesantes que representan el conocimiento basado en medidas de interés)
7. Presentación del conocimiento (donde la visualización y las técnicas de representación del conocimiento se utilizan para presentar el conocimiento minado a los usuarios)



**Figura 2.** Proceso de descubrimiento de conocimiento.  
**Fuente:** (Jiawei et al., 2012)  
**Elaborado por:** El autor

“En general, cualquiera que sea el problema a resolver, no existe una única técnica para solucionarlo, sino que puede ser abordado siguiendo aproximaciones distintas” (Aluja, 2001, p. 484).

La minería de datos cuenta con tres técnicas: predictivas, descriptivas y auxiliares.

Las técnicas predictivas tienen fundamento teórico, este es puesto a prueba y comparado posteriormente para dar como válido el modelo.

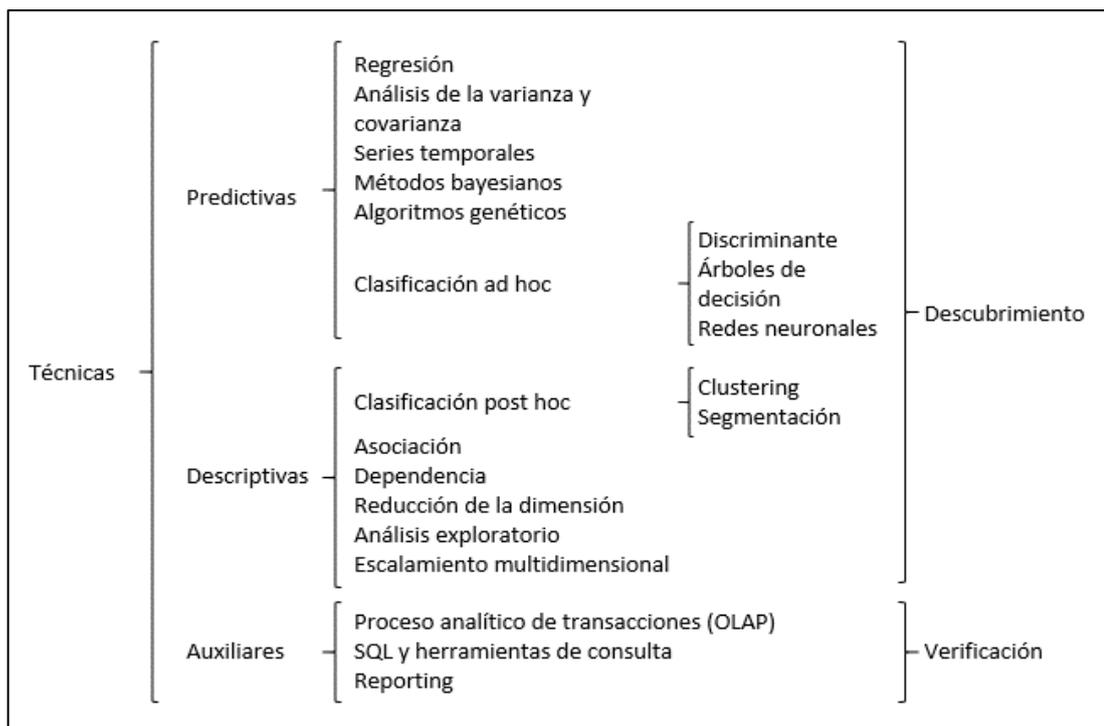
Los modelos de las técnicas predictivas deben seguir las siguientes etapas: a) identificación, selección del modelo para los datos de la investigación; b) estimación, preparación de datos y adaptación del modelo; c) análisis, verificación del modelo; y d) predicción, corresponde al uso del modelo para predecir los valores de las variables objetivos.

En estas técnicas se encuentran las regresiones, series temporales, árboles de decisión, redes neuronales, algoritmos genéticos, análisis discriminante y redes bayesianas. (López, 2007, p. 8).

Para las técnicas descriptivas no se cuentan con variables objetivos ni predictoras, los modelos son resultado del reconocimiento de patrones en los datos de la investigación.

En las técnicas descriptivas encontramos las técnicas de asociación y dependencia, las técnicas de análisis exploratorio de datos, las técnicas de reducción de la dimensión y de escalamiento multidimensional. (López, 2007, p. 8).

Las técnicas auxiliares se establecen a partir de las técnicas de estadística descriptiva y cualquier otra estrategia que permita verificar su resultado; son técnicas limitadas y de soporte.



**Figura 3.** Clasificación de las técnicas de Data Mining.

**Fuente:** (López, 2007)

**Elaborado por:** El autor

A continuación, se analizan técnicas de minería de datos predictivas como: Árboles de decisión, Redes Neuronales, Series Temporales y Redes Bayesianas, con el objetivo de seleccionar la más indicada para la consecución del objetivo planteado.

## 2.2 Árboles de decisión

Un árbol de decisión es una representación visual basado en reglas de decisiones para la solución de un determinado problema.

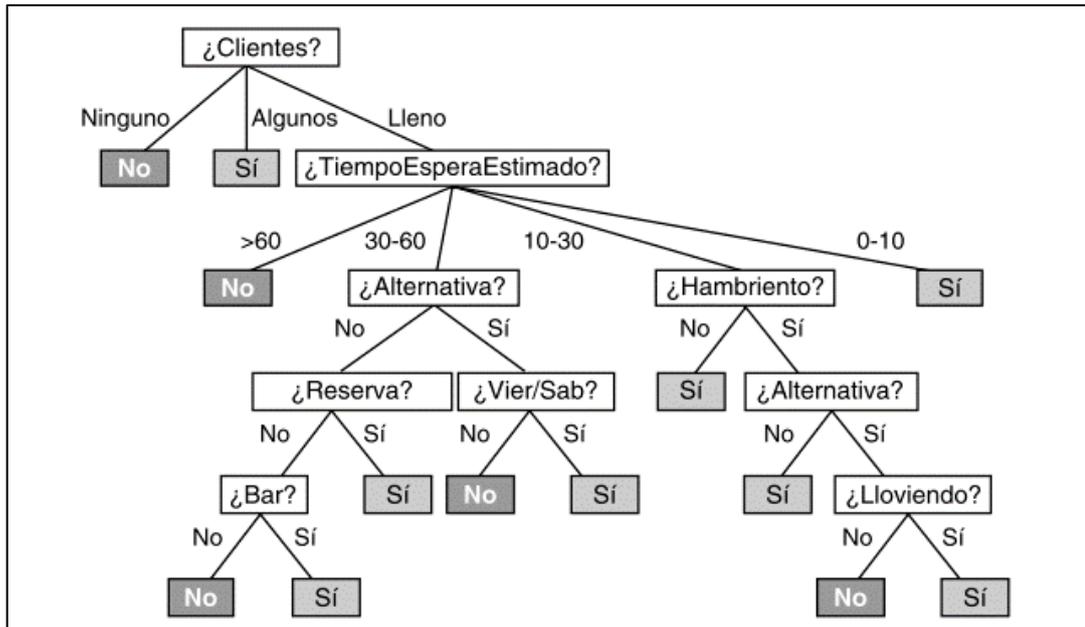
Como lo menciona (Erandi et al., 2009) “ Un árbol de decisión es un modelo de predicción cuyo objetivo principal es el aprendizaje inductivo a partir de observaciones y construcciones lógicas” (p. 20).

(Rubio & Rubio, 2012) afirma que “Los árboles de decisión constituyen representaciones gráficas en forma de diagramas ordenados de secuencias temporales lógicas y de estructura arborescente” (p. 4).

En su investigación Moreno, Miguel, García, & Polo (2008) indican que, los árboles de decisión consisten en la creación de un modelo de clasificación que parten de datos de entrenamiento y de un incitador. El modelo consta de una secuencia de patrones que permiten la identificación de clases.

Un árbol de decisión realiza una serie de pruebas para poder tomar una decisión, los nodos internos del árbol son el resultado de la decisión tomada a partir del valor de una de sus características. (Norvig, P., & Russell, 2014).

En la siguiente figura se indica un ejemplo de un árbol de decisión.



**Figura 4.** Ejemplo de árbol de decisión para decidir si esperar por una mesa.

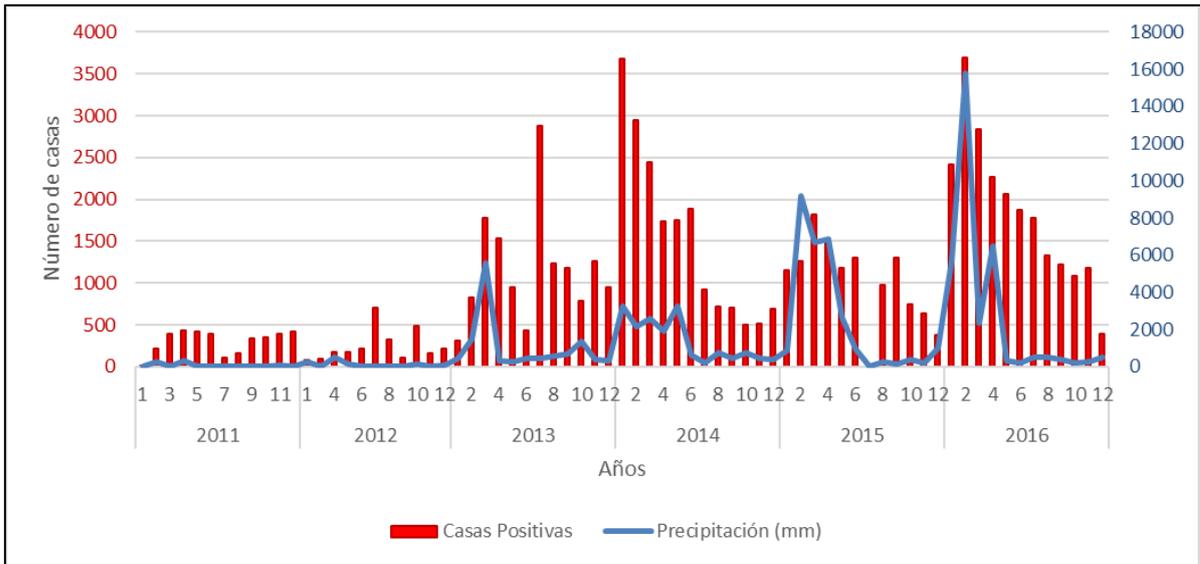
**Fuente:** (Norvig, P., & Russell, 2014)

**Elaborado por:** El autor

### 2.3 Series Temporales

Una serie temporal es un concepto matemático o sucesión ordenada, de variables a través del tiempo, solo se tiene un resultado u observación por cada proceso de las variables que forman parte de la serie temporal, estos valores a su vez, varían con el transcurso del tiempo conforme a las leyes de probabilidad.(Gras, 2001).

Las series de tiempo de acuerdo a (Martínez F.J, Ordieres J., Pernía A., 2005) citado por (Dueñas-Reyes, 2009) “permiten buscar patrones a partir de grandes cantidades de datos. Algunas de sus variables están en función del tiempo. Esta técnica se utiliza a partir del comportamiento histórico de los datos, que permite modelar los componentes básicos de la serie”.



**Figura 5.** Ejemplo de Serie de Tiempo.

**Fuente:** El autor

**Elaborado por:** El autor

De acuerdo a la evolución histórica de la serie temporal, permite identificar tendencias, ciclos y estacionalidad de los datos a representar, con lo que permite realizar predicciones futuras. (Aluja, 2001).

Las series temporales son modelos que describen el proceso que generan los datos observados mediante el análisis, los modelos que se representan en el dominio del tiempo se deducen de la función de autocorrelación, y los modelos que se representan en el dominio de la frecuencia se argumentan de la función de densidad espectral. (Gras, 2001).

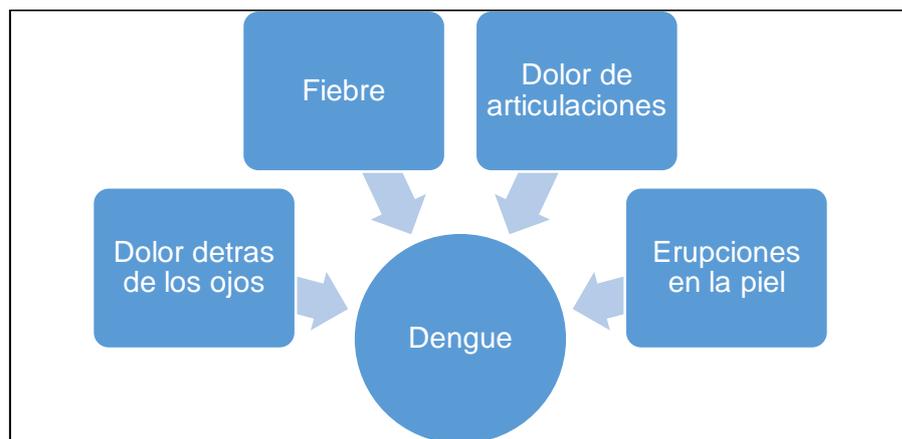
## 2.4 Red Bayesiana

### 2.4.1. Concepto.

Las redes bayesianas son modelos de probabilidad de causa y efecto, son grafos dirigidos acíclicos, de valores aleatorios representados por nodos y arcos, que representan a las variables y las relaciones respectivamente.

La topología de las redes bayesianas, permiten establecer si existe relación entre las variables, y luego de un proceso de probabilidad son capaces de resolver problemas de predicción.

Las redes bayesianas pueden realizar tareas de clasificación que se define con una sola de las variables objetivo de la base de datos, mientras que todas las demás son los datos del tema a clasificar; en el conjunto de datos pueden existir muchas variables, pudiendo existir alguna relación directa con la variable objetivo, aunque otras variables pueden tener una influencia directa sobre la clase. (Felgaer, 2004).



**Figura 6.** Ejemplo de red bayesiana.  
**Fuente:** El autor  
**Elaborado por:** El autor

#### 2.4.2. Ingeniería del conocimiento.

Consiste en definir el grafo y las tablas de probabilidad de la red bayesiana, por lo general es necesario ejecutar varias instancias con ayuda de algoritmos de minería de datos. (Rodríguez & Dolado, 2007).

Los pasos para la construcción de la ingeniería del conocimiento son los siguientes:

- Seleccionar la variable objetivo.
- Definir de la estructura de las redes bayesiana. Las redes generadas son más densas por medio de las relaciones causales, extendiendo las autonomías condicionales eliminando arcos innecesarios.
- Definir las tablas de probabilidad para los nodos.

- Evaluar y validar su funcionamiento, a través del análisis de sensibilidad, para establecer si los cambios en los valores de las variables introducidas tienen algún efecto en el resultado.

La Red Bayesiana cuenta con clasificadores siendo estos: Clasificador Bayesiano Simple (NBC) adquiriendo los valores de predicción posterior y, Redes Bayesianas Dinámicas (RBD) mostrando la situación de las variables en determinado estado de tiempo. (Corso & Gibellini, 2011).

## **2.5 Red Neuronal**

### **2.5.1. Concepto.**

Una red neuronal artificial se basa en las neuronas del cerebro humano, las cuales se encargan de transmitir mensajes a través de impulsos eléctricos para comunicarse entre sí.

Como lo mencionan Norvig, P., & Russell (2014) “los investigadores de inteligencia artificial y de estadística se han interesado en las propiedades más abstractas de las redes neuronales, tales como su habilidad para desarrollar computación distribuida, para tolerar el ruido en la entrada, y para el aprendizaje”. (p. 838).

Otras técnicas de aprendizaje como las redes bayesianas también poseen estas propiedades, pero las redes neuronales continúan siendo una de las formas más difundidas y seguras de crear sistemas de aprendizaje.

### **2.5.2. Ventajas.**

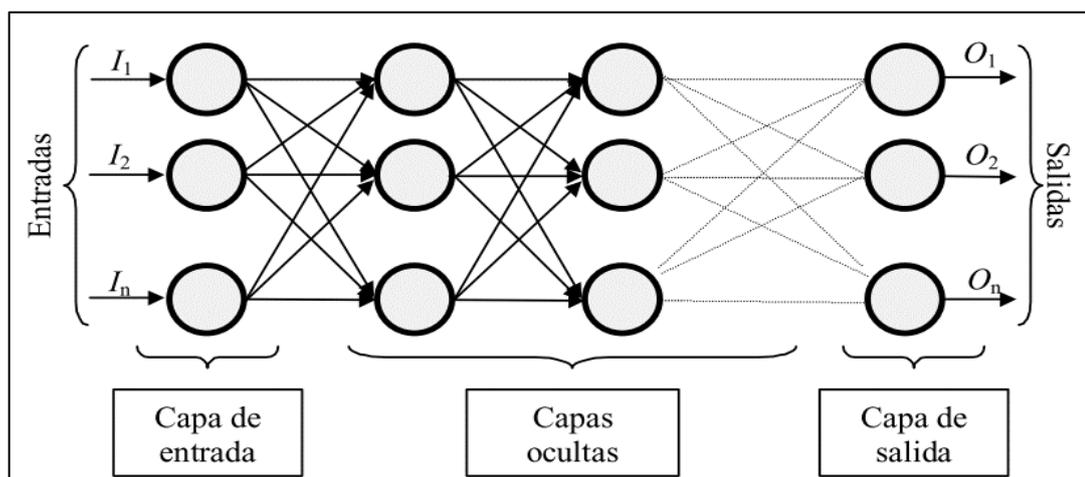
Las ventajas de las redes neuronales según indica Matich (2001) “son capaces de aprender de la experiencia, de generalizar de casos anteriores a nuevos casos, de abstraer características esenciales a partir de entradas que representan información irrelevante, etc.” (p. 8)

Algunas ventajas de las redes neuronales son las siguientes:

- **Aprendizaje Adaptativo.** Las redes neuronales están en capacidad de aprender a ejecutar operaciones establecidas en un entrenamiento o en una actividad de inicio.
- **Auto-organización.** Es la capacidad de modificación total de la red neuronal, para la resolución de un determinado objetivo a través del aprendizaje.
- **Tolerancia a fallos.** Si una parte de la red neuronal sufre algún daño, esta puede seguir operando sin que colapse por completo la red neuronal.
- **Operación en tiempo real.** La operación de las redes neuronales se realiza en forma paralela, lo que proporciona una rapidez de respuesta.
- **Fácil inserción dentro de la tecnología existente.** Las redes neuronales pueden adaptarse fácilmente por medio de chips especializados.

### 2.5.3. Estructura.

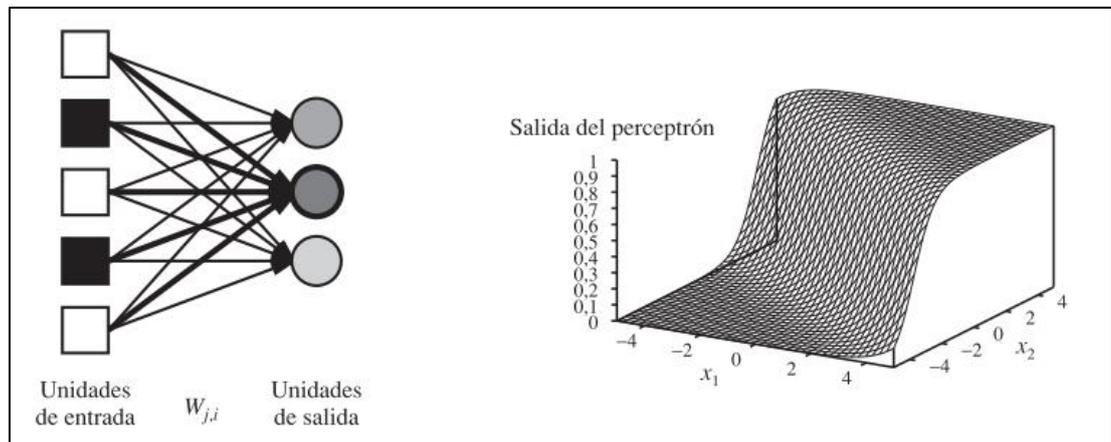
La estructura de las redes neuronales consta de tres secciones: entradas, capas ocultas y salidas. Las entradas constituyen a las variables de ingreso de los datos, estas se interconectan con la o las capas ocultas, y posteriormente obtener el resultado o la salida.



**Figura 7.** Esquema de una red neuronal.  
**Fuente:** (Matich, 2001)  
**Elaborado por:** El autor

En referencia a la estructura de las Redes Neuronales artificiales, existen 3 categorías: Redes Feedforward de Capa Simple, Redes Feedforward Multi-Capas y Redes Recurrentes. (Jirón Araya, 2015).

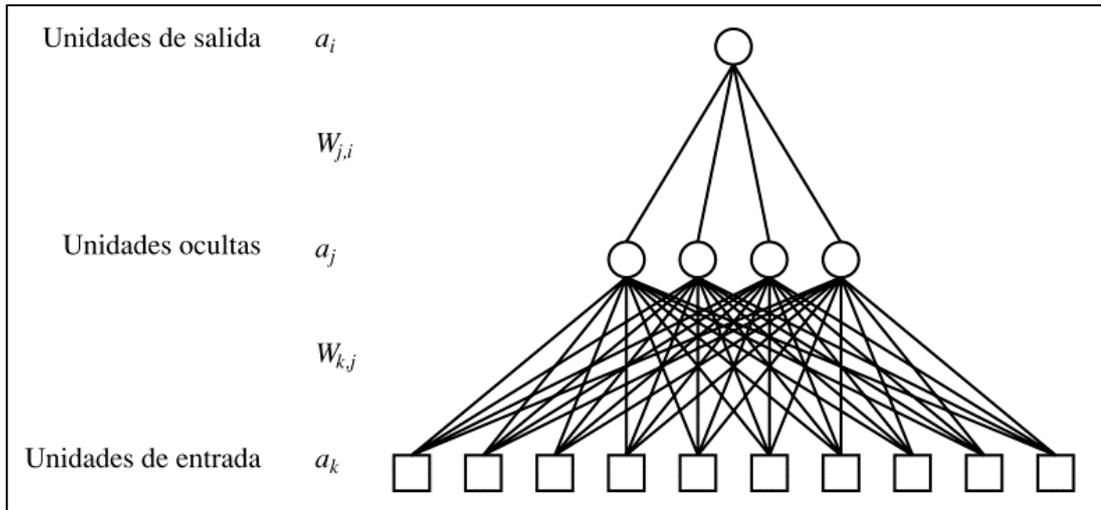
**Redes Feedforward de Capa Simple:** Es una red donde todas las entradas están conectadas directamente a las salidas se denomina red de capa simple, o red perceptrón; cada salida es independiente de las otras, el peso afecta únicamente a una de las salidas. (Norvig, P., & Russell, 2014).



**Figura 8.** Ejemplo de Red Neuronal Feedforward de Capa Simple.

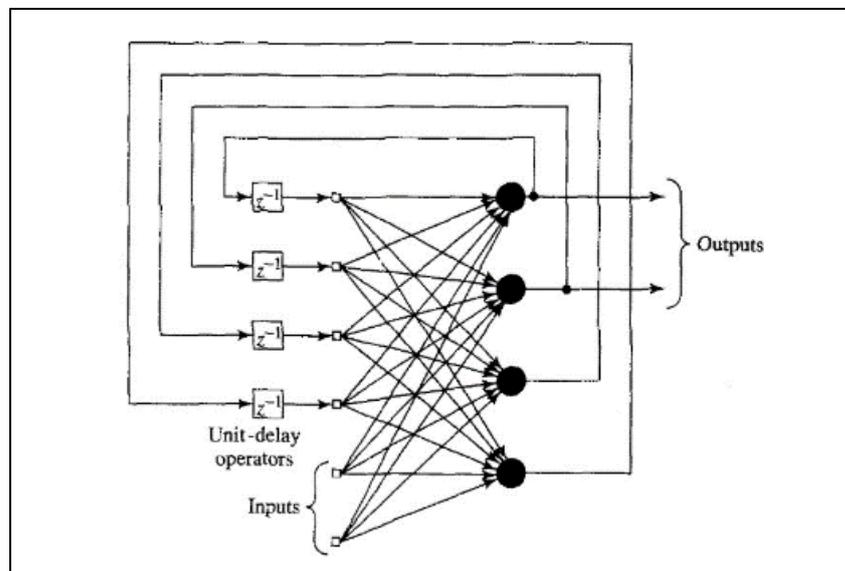
**Fuente:** (Norvig, P., & Russell, 2014)

**Redes Feedforward Multi-Capas (Multilayer Feedforward Networks):** Estas redes se diferencian por estar diseñadas por una o más capas o cultas, los nodos internos de cálculo tienen por nombre neuronas ocultas o unidades ocultas; la tarea de las neuronas ocultas es realizar los cálculos entre la capa de entrada externa y la salida de la red. Los nodos de origen en la capa de entrada proporcionan los elementos del patrón de activación, los mismos que conforman las señales de entrada atribuidas a los nodos de cálculo en la primera capa oculta. La respuesta de la red de acuerdo al patrón de activación proporcionado por los nodos fuentes en la capa de entrada son las señales de salida de las neuronas ocultas en la capa de salida. (Jirón Araya, 2015).



**Figura 9.** Ejemplo de Red Neuronal Feedforward Multi - Capa.  
**Fuente:** (Norvig, P., & Russell, 2014)  
**Elaborado por:** El autor

**Redes Recurrentes (Recurrent Networks):** Este tipo de redes se caracterizan por tener al menos un loop de retroalimentación; el aprendizaje y el rendimiento de la red se ven influenciadas por la estructura de recurrencia. Suponiendo que la red neuronal contiene unidades no-lineales, las unidades de retardo provocan una conducta dinámica no lineal. (Jirón Araya, 2015).



**Figura 10.** Ejemplo de Red Neuronal Recurrente.  
**Fuente:** (Jirón Araya, 2015)  
**Elaborado por:** El autor

#### 2.5.4. Elementos.

Los elementos que conforman una red neuronal artificial son los siguientes:

- **Entradas:** Son los valores de las  $n$  variables que toma como ingreso la red neuronal.
- **Pesos Sinápticos:** La red neuronal asigna pesos a las múltiples variables que recibe como entrada, esto es lo que le da importancia a una variable de otra, estos pesos pueden tener valores positivos o excitadores, negativos o inhibidores, puede también llegar a tener el valor de cero, con lo que la comunicación entre esas neuronas se pierde.
- **Función de activación:** “La función de activación determina el estado de activación actual de la neurona en base al potencial resultante y al estado de activación anterior de la neurona”. (Bertona, 2005, p. 9).
- **Función de salida:** Es valor resultante de la red neuronal, este valor dependerá de los pesos y la interacción de las neuronas en las capas ocultas.

#### 2.5.5. Fases.

Las redes neuronales artificiales cuentan con dos fases: fase de entrenamiento que consiste en utilizar datos para entrenamiento y la fase de operación donde utiliza datos que no fueron utilizados en el entrenamiento, esto con el fin de realizar la validación.

##### 2.5.5.1. *Fase de entrenamiento.*

El entrenamiento de una red neuronal se basa en la asignación de los pesos sinápticos, que es lo que determina la eficiencia en la solución de un determinado problema.

“Una vez introducidos todos los ejemplos se comprueba si se ha cumplido cierto criterio de convergencia; de no ser así se repite el proceso”. (Isasi Viñuela & Galván León, 2000, p. 10).

“La finalización del periodo de aprendizaje se puede determinar: Mediante un número fijo de ciclos, Cuando el error descienda por debajo de una cantidad preestablecida, Cuando la modificación de los pesos sea irrelevante”. (Isasi Viñuela & Galván León, 2000, p. 11).

El aprendizaje de las redes neuronales artificiales se lo puede clasificar en lo siguiente:

- **Aprendizaje supervisado:** Para este tipo de aprendizaje, el conjunto de datos de aprendizaje tiene dos tipos de atributos: los propios datos y alguna información referente a la respuesta del problema; para verificar si el comportamiento adecuado de este aprendizaje se comenta que existe alguna persona externa encargado de realizar esta comprobación, modificando adecuadamente los pesos mediante la comparación entre la salida producida y la esperada. (Isasi Viñuela & Galván León, 2000).
- **Aprendizaje no supervisado:** El aprendizaje no supervisado está constituido únicamente por los patrones de entrada; este aprendizaje no posee la guía de una persona externa para realizar la comprobación del comportamiento de la red. La red forma su conocimiento a partir de la experiencia de los patrones de datos de preparación anteriores. (Marín, 2007).
- **Aprendizaje por refuerzo:** Este aprendizaje constituye una variación del aprendizaje supervisado en el que “no se dispone de información concreta del error cometido por la red para cada ejemplo de aprendizaje, sino que simplemente se determina si la salida producida para dicho patrón es o no adecuada”. (Isasi Viñuela & Galván León, 2000, p. 13).

#### **2.5.5.2. Fase de operación.**

En ocasiones el ajuste se enfoca mayormente en soluciones particulares de acuerdo a los datos de entrenamiento, lo que provoca un sobre entrenamiento de la red, provocando que los valores sean erróneos. Para evitar esto, es necesario utilizar datos que no fueron utilizados para el entrenamiento, a estos datos se los conoce como datos de test o prueba, y de esta forma realizar una validación de los resultados de la red neuronal, apegado a una solución generalizada del problema propuesto.

### **2.5.6. Métricas de evaluación.**

Cómo métricas de evaluación se tienen: Error Cuadrático Medio o MSE y el Valor de Regresión.

El Error Cuadrático Medio o MSE es la diferencia entre los valores obtenidos por la red neuronal y el valor del objetivo, los resultados con valores más bajos indican la precisión de la predicción, un valor de cero indica que no existe error por lo tanto el resultado sería perfecto.

El valor de Regresión R indica la correlación de los resultados de la red frente al objetivo establecido, aquí por el contrario del MSE, los valores cercanos a 1 indican mayor precisión de la red neuronal, el valor de 1 significa una correlación perfecta, un valor de 0 significa que la correlación es nula, esta métrica es una forma de mostrar cuánto es la adaptación del modelo a los datos del ensayo.

## **2.6 Trabajos relacionados**

A continuación, se presentan trabajos relacionados a esta investigación elaborados por otros autores, teniendo en común las redes neuronales artificiales.

### **2.6.1. Sistema de predicción de ruido urbano mediante redes neuronales.**

El modelo descrito por el autor se fundamenta en el análisis de las variables que se estiman tiene predominio en el ruido ambiental, realizando una recolección propia de datos para posteriormente realizar el entrenamiento de la red neuronal.

La arquitectura de la red neuronal es perceptrón multicapa, con un algoritmo de entrenamiento Levenberg-Marquart, desarrollada en Matlab

La red neuronal artificial inicialmente es creada con 25 variables independientes y como variable objetivo el nivel equivalente de ruido, posee una sola capa oculta con 7 neuronas. Realiza una reducción de variables utilizando análisis de componentes principales y varios métodos de selección para elegir las variables más idóneas; presenta resultados similares

utilizando solo 15 variables y 12 neuronas ocultas, obteniendo un valor de error que no supera el 5%. (Genaro, 2014).

### **2.6.2. Entrenamiento de redes neuronales basado en algoritmos evolutivos.**

En esta investigación el objetivo principal es entrenar a la red neuronal utilizando algoritmos evolutivos, permitiendo combinar estas dos técnicas y sustituir a los algoritmos tradicionales; al finalizar se realiza una evaluación de este algoritmo innovador frente a los algoritmos cotidianos, donde se demuestra que esta técnica presenta resultados similares al algoritmo backpropagation especialmente en la generalización del resultado.

El método no utiliza información de gradientes como lo harían los métodos tradicionales, donde en ocasiones queda atrapado en mínimos locales, efecto análogo de la detención anticipada; el método utiliza operadores que afianzan la heterogeneidad del resultado.

Una ventaja del método planteado es la disminución de los tiempos de entrenamiento, aproximadamente menor al 40% que el algoritmo backpropagación, lo que permite mejorar el rendimiento del entrenamiento, frente a los algoritmos tradicionales. (Bertona, 2005).

### **2.6.3. Modelamiento matemático y por redes neuronales artificiales del crecimiento de *Spirulina* sp. en fotobiorreactor con fuente de luz fluorescente e iluminación en estado sólido.**

Este proyecto pone de manifiesto cómo a través del uso de redes neuronales artificiales se puede evaluar los modelos matemáticos de Gompertz y logístico en el cambiante crecimiento de la spirulina en ambientes de cultivo con luz fluorescente e iluminación LED.

La estructura de la red neuronal artificial para este proyecto es del tipo hacia adelante (feedforward), con algoritmo de entrenamiento de Levenberg Marquardt, una sola capa oculta con 3 neuronas, utilizando únicamente la variable con los datos de tiempo, y como métricas de evaluación se tienen al Error Cuadrático Medio, Error Objetivo, tangente y logaritmo sigmoideal hiperbólica; la red neuronal fue desarrollada con el software Matlab.

La red neuronal mostró el coeficiente de correlación 0.993 en la prueba de la Spirulina con iluminación fluorescente, y 0.994 en la prueba con iluminación LED, demostrando una buena precisión en los resultados de la red neuronal en contraste con el modelo matemático de Gompertz. (Vásquez V, Artega G, Chanamé A, & Esquivel T, 2013).

**CAPÍTULO III.**  
**ANÁLISIS Y PREPARACIÓN DE DATOS**

### 3.1 Obtención de los datos

Los datos fueron obtenidos gracias al convenio de cooperación interinstitucional entre la Universidad Upstate de New York, Universidad Técnica de Machala, Ministerio de Salud Pública del Ecuador, Instituto Nacional de Meteorología e Hidrología (INAMHI) y Instituto Nacional de Estadísticas y Censos (INEC) a través del proyecto de investigación “Fortalecimiento de las capacidades en Ecuador: Colaboración para mejorar la vigilancia de enfermedades febriles transmitidas por vectores”, la carta de autorización se encuentra en Anexos 1, 2, 3 y 4.

Los datos recopilados para este estudio corresponden desde el mes de enero del año 2011 hasta el mes de diciembre del año 2016, con un total de 13468 registros, para esta investigación no se incluyen los datos del año 2017 porque la información del Instituto Nacional de Meteorología e Hidrología (INAMHI) no se encuentra completa.

#### 3.1.1. Datos de Control Vectorial.

Encontramos los siguientes datos proporcionados por el Departamento de Control Vectorial del Distrito de Salud Machala:

**Tabla 1.** Datos de visitas domiciliarias.

<b>Variable</b>	<b>Descripción</b>
Barrio	Nombre del barrio donde se realiza la actividad de control vectorial.
Parroquia	Corresponde al nombre de la Parroquia a la que pertenece el barrio, la ciudad de Machala tiene 8 Parroquias Urbanas (Puerto Bolívar, Jambelí, Jubones, Machala, Nueve de Mayo, La Providencia, El Cambio) y 1 Rural (El Retiro).
Código Barrio	Código asignado al barrio de intervención, está conformado por 10 dígitos, los 2 primeros indican la Provincia, los 2 siguientes el Cantón, los 2 siguientes indican la Parroquia, estos 6 dígitos son establecidos por el INEC; y los 4 últimos indican el número del barrio, este código fue asignado por la desaparecida

	institución SNEM (Servicio Nacional de Erradicación de la Malaria).
Área	Indica el área en metros cuadrados del barrio intervenido, este dato es calculado a través de la función %area en el software QGIS (Sistema de Información Geográfica) que realiza la estimación del área a través de la SRC (Sistema de Referencia de Coordenadas) del mapa.
Total Casas	Número de casas que tiene el barrio.
Habitantes	Número de habitantes que posee el barrio.
Casas Inspeccionadas	Número de viviendas que fueron inspeccionadas para determinar la presencia de criaderos de mosquitos.
Casa Cerradas	Número de viviendas que no se inspeccionaron por ausencia de los habitantes al momento de la inspección.
Casas Renuentes	Número de viviendas que no se inspeccionaron por que los habitantes no permiten realizar la inspección.
Casas Positivas	Número de viviendas donde se encontró criaderos de mosquitos <i>Aedes aegypti</i> .
Terreno Baldío	Número de terrenos o solares sin habitar.
Depósitos Inspeccionados	Número de recipientes inspeccionados como: floreros, tarrinas, llantas, cisternas, tanques, etc.
Depósitos Positivos	Número de recipientes dónde se encontró la presencia de <i>Aedes aegypti</i> .
Índice de Casa	Porcentaje de viviendas infestadas con <i>Aedes aegypti</i> / Total de Casas.
Índice de Breteau	Número de Depósitos positivos por cada 100 viviendas inspeccionadas.
Índice de Depósito	Porcentaje de recipientes infestados con <i>Aedes aegypti</i> / Total de Casas.
Larvicida	Tipo de larvicida utilizado en la visita de control vectorial.
Semana Epidemiológica	Número de la semana correspondiente al calendario epidemiológico del Ministerio de Salud Pública del Ecuador.

Fecha Fumigación Motomochila	Corresponde a la última fecha de fumigación con máquina motomochila realizada en el barrio
------------------------------	--

**Fuente:** (Distrito de Salud 07D02 - Machala)

**Elaborado por:** El autor

### 3.1.2. Datos de Clima.

Los datos fueron facilitados por el Instituto Nacional de Meteorología e Hidrología (INAMHI) en su formato original de registro, siendo este diario y cada 60 minutos:

**Tabla 2.** Datos climáticos.

Variable	Descripción
Precipitación	Cantidad de agua que ha caído en la ciudad expresada en milímetros
Temperatura max	Temperatura máxima registrada en la ciudad
Temperatura min	Temperatura mínima registrada en la ciudad
Temperatura med	Temperatura promedio registrada en la ciudad

**Fuente:** (INAMHI)

**Elaborado por:** El autor

### 3.1.3. Datos socioeconómicos.

Datos obtenidos de la página web oficial del Instituto Nacional de Estadísticas y Censos (INEC), los cuales se encuentran registrados por zonas censales.

**Tabla 3.** Datos Socioeconómicos.

Variable	Descripción
Código Zona	Código de la Zona Censal establecido por el INEC
Alcantarillado	Indicador promedio para determinar si el barrio tiene alcantarillado o alguna forma de desfogue de aguas servidas: AL1 = Inodoro y alcantarillado AL2 = Inodoro y pozo séptico AL3 = Inodoro y pozo ciego AL4 = Letrina AL5 = No tiene

Agua Potable	Indicador promedio para determinar la forma de abastecimiento de agua potable en el barrio: AP1 = Red pública AP2 = Otra fuente por tubería AP3 = Carro repartidor / triciclo AP4 = Pozo AP5 = Río vertiente o acequia AP6 = Otro
Recolección Basura	Indicador promedio para determinar la forma de eliminación de los desechos en el barrio: RB1 = Servicio Municipal RB2 = La botan a la calle, quebrada, lote RB3 = La queman RB4 = La entierran RB5 = La botan al río, acequia, canal RB6 = Otro
Nivel Socioeconómico	Indicador promedio para determinar cómo se considera el nivel socioeconómico en el barrio: NS1 = Muy pobre NS2 = Pobre NS3 = Más o menos pobre NS4 = No pobre
Nivel de Instrucción del Jefe de Hogar	Indicador promedio para determinar el nivel máximo de instrucción educativa alcanzado por el jefe de hogar en el barrio: NI0 = Menores de 5 años NI1 = Ninguno NI2 = Primaria NI3 = Secundaria NI4 = Superior

Fuente: (INEC)

Elaborado por: El autor

### 3.2 Preparación de los datos

Para el trabajo de investigación, se deben procesar los datos y obtener nuevas variables que ayudarán a la consecución del objetivo propuesto.

**Tabla 4.** Obtención de nuevas variables.

<b>Variable original</b>	<b>Nueva Variable</b>	<b>Descripción</b>
Total Casas – Casas Inspeccionadas	Casas No Inspeccionadas	Número total de viviendas que no se lograron inspeccionar, resultado de la resta del total de casas del barrio menos las casas inspeccionadas.
Fecha Fumigación Motomochila	Semana última fumigación_m	Se procedió a transformar la fecha en semana
Fecha Fumigación Motomochila	Fumigación estado	Se estable en 0 cuando la fumigación tiene un máximo de 9 días y 1 cuando es mayor a 10.
Fecha Visita	Estación	De diciembre a mayo se asigna 0, que corresponde a invierno y de junio a noviembre se asigna 1, que corresponde a verano
Fecha Visita	Mes	Obtención del número del mes a partir de la fecha de visita domiciliaria
Fecha Visita	Año	Obtención del año a partir de la fecha de visita domiciliaria
Casas Inspeccionadas	Validación	Selección o descarte del registro para su procesamiento, como condición de selección debe tener un mínimo de casas inspeccionadas del 33% del total de casas del barrio.
Alcantarillado	Alcan_1 Alcan_2 Alcan_3	Se separa esta variable para su correcta utilización, establece en 1 para los datos que presentan esta opción y 0 para los que no corresponden AL1 = Inodoro y alcantarillado AL2 = Inodoro y pozo séptico

		AL3 = Inodoro y pozo ciego
Agua Potable	Agua_1 Agua_2 Agua_3	Se separa esta variable para su correcta utilización, establece en 1 para los datos que presentan esta opción y 0 para los que no corresponden AP1 = Red pública AP2 = Otra fuente por tubería AP3 = Carro repartidor / triciclo
Recolección Basura	Recol_1 Recol_2	Se separa esta variable para su correcta utilización, establece en 1 para los datos que presentan esta opción y 0 para los que no corresponden RB1 = Servicio Municipal RB2 = La botan a la calle, quebrada, lote
Nivel Socioeconómico	Socio_2 Socio_3	Se separa esta variable para su correcta utilización, establece en 1 para los datos que presentan esta opción y 0 para los que no corresponden NS2 = Pobre NS3 = Más o menos pobre
Nivel de Instrucción del Jefe de Hogar	Instru_2 Instru_3 Instru_4	Se separa esta variable para su correcta utilización, establece en 1 para los datos que presentan esta opción y 0 para los que no corresponden NI2 = Primaria NI3 = Secundaria NI4 = Superior

Fuente: (El autor)

Elaborado por: El autor

**CAPÍTULO IV.**  
**SELECCIÓN DE TÉCNICA DE PREDICCIÓN**

#### 4.1 Selección de los datos

Una vez obtenidos y preparados los datos, continuamos con la selección de los mismos. Primeramente, se establece como variable objetivo a Casas Positivas, luego se procede a seleccionar las variables predictoras por medio de una matriz de correlación para establecer qué variables tienen más influencia con nuestra variable objetivo.

Después del proceso preparatorio de los datos, se obtuvo 42 variables para la investigación, pero de las 19 variables que proceden de los datos del Departamento de Control Vectorial del Distrito de Salud 07D02 se descartaron 15 de la selección, porque dependen exclusivamente de la intervención de los trabajadores; es decir, dependen del factor humano, en la actividad de visitas domiciliarias. Por ejemplo, algunas variables que fueron excluidas tenemos a: Casas Inspeccionadas, Depósitos Positivos, Casas Cerradas, etc., esto para no crear un dato de predicción viciado o de sobreajuste, quedando un total de 27 variables para la matriz de correlación.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	
1- Casas_Positivas	1	0.43	0.53	0.51	0.26	-0.03	-0.24	-0.24	-0.05	-0.18	0.13	0.18	0.11	-0.07	0.08	0.00	-0.06	0.03	0.07	0.00	0.00	0.02	-0.02	0.01	-0.02	0.03	-0.04	
2- Area	0.43	1	0.82	0.82	0.43	-0.01	0.00	0.00	0.02	-0.01	0.00	0.00	0.00	-0.01	0.02	-0.04	-0.04	0.04	0.01	0.04	-0.04	0.05	-0.06	0.02	-0.12	0.11	0.02	
3- Total_Casas	0.53	0.82	1	0.99	0.41	-0.05	0.00	0.00	0.03	0.00	0.00	0.01	-0.03	-0.04	0.04	0.00	-0.06	0.03	0.07	0.00	0.00	0.04	-0.04	0.01	-0.11	0.10	0.01	
4- Habitantes	0.51	0.82	0.99	1	0.40	-0.06	0.00	0.00	0.04	-0.01	0.00	0.02	-0.03	-0.03	0.03	0.02	-0.05	0.02	0.07	-0.02	0.02	0.04	-0.05	0.03	-0.10	0.10	0.00	
5- Terreno_Baldio	0.26	0.43	0.41	0.40	1	-0.04	0.03	0.03	0.01	0.03	-0.03	-0.03	-0.02	-0.37	0.36	0.05	-0.24	0.20	0.13	-0.05	0.05	0.25	-0.20	-0.07	-0.05	0.04	0.02	
6- Anio	-0.03	-0.01	-0.05	-0.06	-0.04	1	-0.12	-0.10	-0.55	-0.07	0.14	-0.25	0.77	0.04	-0.03	-0.03	0.05	-0.04	-0.03	0.03	-0.03	-0.04	0.03	0.02	-0.03	0.02	0.06	
7- Mes	-0.24	0.00	0.00	0.00	0.03	-0.12	1	1.00	-0.08	0.65	-0.37	-0.57	-0.39	0.00	0.01	-0.02	0.00	0.01	-0.01	0.02	-0.02	-0.03	0.02	0.02	-0.02	0.01	0.02	
8- Semana	-0.24	0.00	0.00	0.00	0.03	-0.10	1.00	1	-0.09	0.66	-0.36	-0.58	-0.38	0.00	0.01	-0.02	0.00	0.01	-0.01	0.02	-0.02	-0.03	0.02	0.02	-0.02	0.01	0.02	
9- Larvicida	-0.05	0.02	0.03	0.04	0.01	-0.55	-0.08	-0.09	1	-0.07	0.01	0.38	-0.65	-0.02	0.02	0.00	0.02	-0.02	-0.01	0.00	0.00	0.00	0.01	0.02	-0.01	0.01	0.00	-0.04
10- Estacion	-0.18	-0.01	0.00	-0.01	0.03	-0.07	0.65	0.66	-0.07	1	-0.35	-0.71	-0.43	0.00	0.01	-0.02	0.00	0.01	-0.01	0.02	-0.02	-0.02	0.00	0.04	0.00	0.00	0.01	
11- Precipitacion	0.13	0.00	0.00	0.00	-0.03	0.14	-0.37	-0.36	0.01	-0.35	1	0.22	0.22	0.00	-0.02	0.04	0.01	-0.02	0.03	-0.04	0.04	-0.02	0.01	0.01	-0.01	0.01	-0.02	
12- Temperatura_max	0.18	0.00	0.01	0.02	-0.03	-0.25	-0.57	-0.58	0.38	-0.71	0.22	1	0.03	-0.01	0.00	0.00	0.01	-0.01	0.00	0.00	0.00	0.01	0.01	-0.04	0.02	-0.01	-0.04	
13- Temperatura_min	0.11	0.00	-0.03	-0.03	-0.02	0.77	-0.39	-0.38	-0.65	-0.43	0.22	0.03	1	0.03	-0.03	-0.01	0.02	-0.02	0.00	0.01	-0.01	0.00	0.00	0.00	0.00	0.02	0.01	0.04
14- alcan_1	-0.07	-0.01	-0.04	-0.03	-0.37	0.04	0.00	0.00	-0.02	0.00	0.00	-0.01	0.03	1	-0.94	-0.28	0.33	-0.21	-0.31	0.28	-0.28	-0.34	0.26	0.11	0.05	-0.06	0.04	
15- alcan_2	0.08	0.02	0.04	0.03	0.36	-0.03	0.01	0.01	0.02	0.01	-0.02	0.00	-0.03	-0.94	1	-0.07	-0.22	0.24	0.00	0.07	-0.07	0.26	-0.19	-0.11	-0.19	0.19	-0.04	
16- alcan_3	0.00	-0.04	0.00	0.02	0.05	-0.03	-0.02	-0.02	0.00	-0.02	0.04	0.00	-0.01	-0.28	-0.07	1	-0.36	-0.05	0.90	-1.00	1.00	0.28	-0.25	-0.03	0.37	-0.36	-0.01	
17- agua_1	-0.06	-0.04	-0.06	-0.05	-0.24	0.05	0.00	0.00	0.02	0.00	0.01	0.01	0.02	0.33	-0.22	-0.36	1	-0.89	-0.40	0.36	-0.36	-0.38	0.31	0.09	-0.02	0.01	0.03	
18- agua_2	0.03	0.04	0.03	0.02	0.20	-0.04	0.01	0.01	-0.02	0.01	-0.02	-0.01	-0.02	-0.21	0.24	-0.05	-0.89	1	-0.06	0.05	-0.05	0.29	-0.24	-0.08	-0.14	0.14	-0.03	
19- agua_3	0.07	0.01	0.07	0.07	0.13	-0.03	-0.01	-0.01	-0.01	-0.01	0.03	0.00	0.00	-0.31	0.00	0.90	-0.40	-0.06	1	-0.90	0.90	0.24	-0.21	-0.03	0.33	-0.31	-0.01	
20- recol_1	0.00	0.04	0.00	-0.02	-0.05	0.03	0.02	0.02	0.00	0.02	-0.04	0.00	0.01	0.28	0.07	-1.00	0.36	0.05	-0.90	1	-1.00	-0.28	0.25	0.03	-0.37	0.36	0.01	
21- recol_2	0.00	-0.04	0.00	0.02	0.05	-0.03	-0.02	-0.02	0.00	-0.02	0.04	0.00	-0.01	-0.28	-0.07	1.00	-0.36	-0.05	0.90	-1.00	1	0.28	-0.25	-0.03	0.37	-0.36	-0.01	
22- socio_2	0.02	0.05	0.04	0.04	0.25	-0.04	-0.03	-0.03	-0.01	-0.02	-0.02	0.01	0.00	-0.34	0.26	0.28	-0.38	0.29	0.24	-0.28	0.28	1	-0.88	-0.11	0.21	-0.19	-0.04	
23- socio_3	-0.02	-0.06	-0.04	-0.05	-0.20	0.03	0.02	0.02	0.00	0.01	0.00	0.26	-0.19	-0.25	0.31	-0.24	-0.21	0.25	-0.25	-0.88	1	-0.38	-0.15	0.14	0.05			
24- socio_4	0.01	0.02	0.01	0.03	-0.07	0.02	0.02	0.02	-0.01	0.04	0.01	-0.04	0.00	0.11	-0.11	-0.03	0.09	-0.08	-0.03	0.03	-0.03	-0.11	-0.38	1	-0.08	0.09	-0.02	
25- instru_2	-0.02	-0.12	-0.11	-0.10	-0.05	-0.03	-0.02	-0.02	0.01	0.00	-0.01	0.02	-0.02	0.05	-0.19	0.37	-0.02	-0.14	0.33	-0.37	0.37	0.21	-0.15	-0.08	1	-0.97	-0.03	
26- instru_3	0.03	0.11	0.10	0.10	0.04	0.02	0.01	0.01	0.00	0.00	0.01	-0.01	0.01	-0.06	0.19	-0.36	0.01	0.14	-0.31	0.36	-0.36	-0.19	0.14	0.09	-0.97	1	-0.22	
27- instru_4	-0.04	0.02	0.01	0.00	0.02	0.06	0.02	0.02	-0.04	0.01	-0.02	-0.04	0.04	0.04	-0.04	-0.01	0.03	-0.03	-0.01	0.01	-0.01	-0.04	0.05	-0.02	-0.03	-0.22	1	

Figura 11. Matriz de correlación.

Fuente: El autor

Elaborado por: El autor

Para el proceso de selección se utiliza una matriz de correlación, lo que permitirá observar qué variables predictoras tienen mayor influencia frente a la variable objetivo, como se puede

apreciar en la primera línea de la matriz de correlación, que es donde se encuentra ubicada nuestra variable objetivo frente a las variables predictoras, ninguna tiene el 100% de correlación o valor de 1, por lo tanto, estas variables presentan multicolinealidad aproximada.

Las variables con correlación positiva se ven influenciadas positivamente (relación creciente) por la variable contrastada a medida que aumenta su valor y, las variables con correlación negativa (relación decreciente) se ven influenciadas negativamente por la variable contrastada a medida que estas aumentan su valor.

**Tabla 5.** Descripción de los valores de correlación.

<b>Coefficiente</b>	<b>Interpretación</b>
0	Relación nula
0 - 0.2	Relación muy baja
0.2 - 0.4	Relación baja
0.4 - 0.6	Relación moderada
0.6 - 0.8	Relación alta
0.8 - 1	Relación muy alta
1	Relación perfecta

**Fuente:** (Cortina Januch, 2012)

**Elaborado por:** El autor

Para esta investigación se procede a realizar dos ensayos, en el que se pretende utilizar las variables que mejor resultado den a la creación de la herramienta de predicción de casas positivas con criaderos de mosquitos *Aedes aegypti*.

#### **4.1.1. Ensayo 1.**

Se tomará en consideración 11 variables que involucran datos de Control vectorial, Clima y Socioeconómicos, estas dos últimas se consideran porque existe la hipótesis que la situación socioeconómica de la población y el Clima tienen influencia en el aumento de las casas positivas con criaderos de mosquitos *Aedes aegypti*, aunque en la matriz de correlación se aprecia que tienen una relación muy baja, pero se decide realizar ensayos para aceptar o rechazar la hipótesis planteada para los datos de la investigación. Las variables seleccionadas son las siguientes:

**Tabla 6.** Variables seleccionadas de acuerdo al grado de correlación, Ensayo 1.

<b>Variable</b>	<b>Correlación</b>	<b>Interpretación</b>
Total Casas	0.53	Relación moderada
Habitantes	0.51	Relación moderada
Área	0.43	Relación moderada
Terreno Baldío	0.26	Relación baja
Temperatura max	0.18	Relación muy baja
Precipitación	0.13	Relación muy baja
Temperatura min	0.11	Relación muy baja
alcan_2	0.08	Relación muy baja
Estación	-0.18	Relación muy baja
Semana	-0.24	Relación baja
Mes	-0.24	Relación baja

**Fuente:** El autor

**Elaborado por:** El autor

#### 4.1.2. Ensayo 2.

Para este ensayo se tomará en consideración 6 variables predictoras o independientes, descartando las variables socioeconómicas y de clima, como se puede observar en la matriz de correlación estas variables presentan una relación muy baja, posteriormente se procederá a realizar pruebas para corroborar si tienen alguna inferencia en el resultado de predicción. Las variables seleccionadas para este ensayo son las siguientes:

**Tabla 7.** Variables seleccionadas de acuerdo al grado de correlación, Ensayo 2.

<b>Variable</b>	<b>Correlación</b>	<b>Interpretación</b>
Total Casas	0.53	Relación moderada
Habitantes	0.51	Relación moderada
Área	0.43	Relación moderada
Terreno Baldío	0.26	Relación baja
Semana	-0.24	Relación baja
Mes	-0.24	Relación baja

**Fuente:** El autor

**Elaborado por:** El autor

### 4.1.3. Ensayo 3.

En este ensayo se toman en consideración únicamente las variables predictoras con una relación moderada, excluyendo las variables con una baja relación, se realizarán pruebas para verificar si los resultados mejoran a partir de la inclusión de variables con relación moderada, quedando para el ensayo las siguientes variables:

**Tabla 8.** Variables seleccionadas de acuerdo al grado de correlación, Ensayo 3.

<b>Variable</b>	<b>Correlación</b>	<b>Interpretación</b>
Total Casas	0.53	Relación moderada
Habitantes	0.51	Relación moderada
Área	0.43	Relación moderada

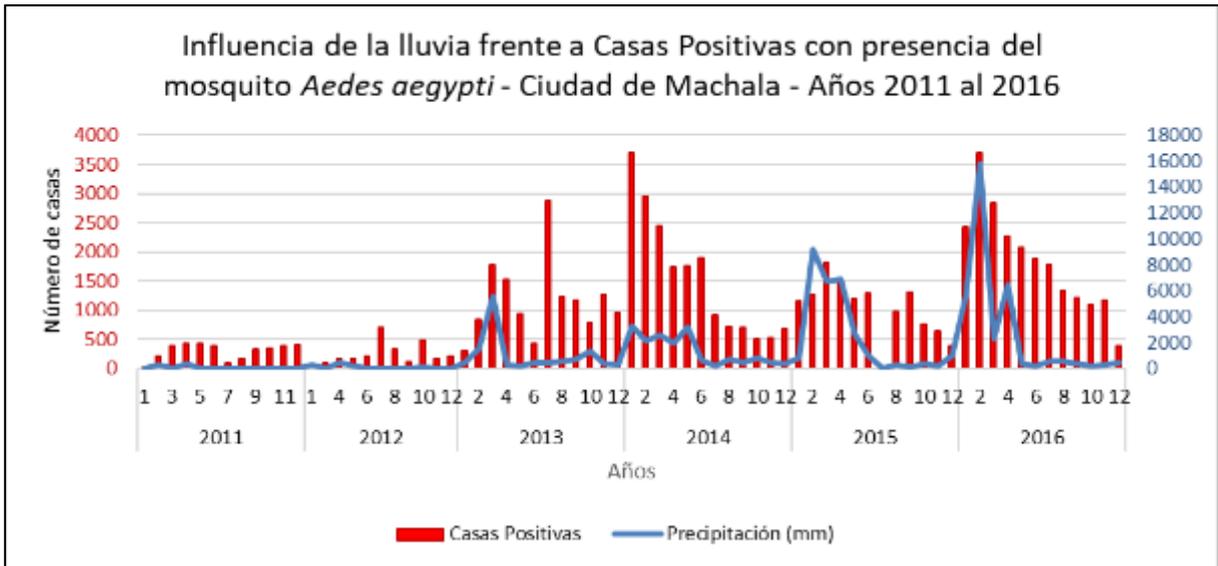
**Fuente:** El autor

**Elaborado por:** El autor

En resumen, para el ensayo 1 se cuenta con 11 variables predictoras y 1 variable objetivo; para el ensayo 2 se tiene 6 variables predictoras y 1 variable objetivo y, para el ensayo 3 se dispone de 3 variables predictoras y 1 variable objetivo, estas nos servirán para realizar las pruebas con las técnicas de predicción de nuestra investigación.

## 4.2 Series Temporales

A continuación, se realizan varias figuras de Serie de tiempo, para representar la relación existente entre las variables predictoras con la variable objetivo y así continuar con la búsqueda de la predicción de las Casas Positivas para criaderos de *Aedes aegypti*.

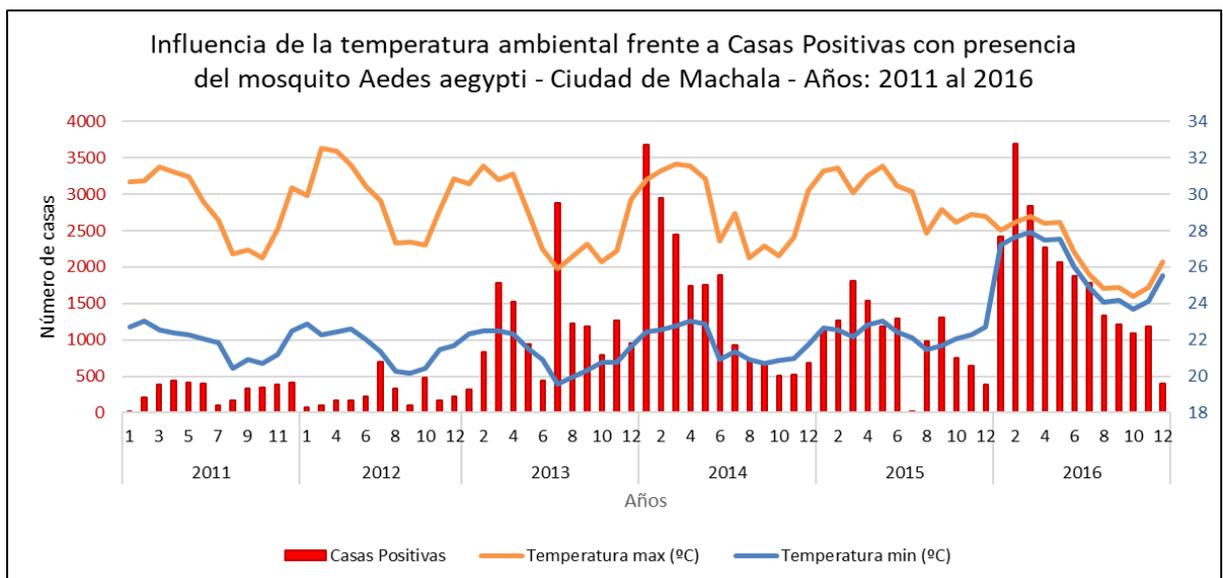


**Figura 12.** Influencia de la variable Precipitación frente a Casas Positivas con presencia del mosquito *Aedes aegypti*.

Fuente: El autor

Elaborado por: El autor

Como se puede apreciar, existe una relación no muy definida entre la presencia de lluvia y las casas positivas con criaderos de mosquitos *Aedes aegypti*, la cantidad de milímetros cúbicos de lluvia no tiene una relación marcada frente a la presencia de casas positivas, especialmente en los meses que corresponden a la estación de invierno.



**Figura 13.** Influencia de las variables Temperatura máxima y mínima frente a Casas Positivas con presencia del mosquito *Aedes aegypti*.

Fuente: El autor

Elaborado por: El autor

En la figura se puede observar que la temperatura tampoco tiene una relación definida frente a los valores de casas positivas de nuestro estudio, ya que en el ciclo de vida del mosquito *Aedes aegypti*, la fase acuática del mosquito se ve afectada por diversos cambios medioambientales, y uno de esos factores es la temperatura, ya que se sospecha que entre más alta es la temperatura, más corto es el periodo de la fase acuática del mosquito y pasa a la fase adulta o aérea más rápido que cuando la temperatura es más baja.

Después de analizar varias técnicas de predicción, para esta investigación se ha seleccionado las Redes Neuronales Artificiales, porque es una técnica muy popular, de buenos resultados en la predicción, cuando los datos de ingreso presentan ruido y tienen gran complejidad; los datos preparados para esta investigación no presentan una relación causa – efecto bien definida como para tratar de utilizar redes bayesianas.

#### **4.3 Selección de la herramienta**

Se ha seleccionado el software matemático MATLAB con entorno de desarrollo integrado (IDE), el cual es un software del tipo propietario desarrollado por la empresa MathWorks, el cual es un software multiplataforma y posee un lenguaje de programación propio. Las funciones principales de Matlab son: creación de matrices, representación y análisis de datos y funciones, implementación de algoritmos, modelado simulación y prueba de prototipos, análisis de datos, entre otros. (<https://la.mathworks.com>).

#### **4.4 Crear Red Neuronal en Matlab**

MatLab cuenta con varias opciones al momento de crear Redes Neuronales, de estas tenemos: Neural Net Clustering, Neural Net Fitting, Neural Net Pattern Recognition, Neural Net Times Series. Para el ensayo utilizaremos la Neural Net Fitting porque permite resolver el problema de adaptación de datos, y por el objetivo que se quiere alcanzar y la naturaleza de los datos de estudio es la más adecuada.

La Neural Net Fitting o Red Neuronal Ajustada utiliza la técnica feed-forward o alimentación hacia adelante de dos capas ocultas, donde las conexiones de las neuronas van en una sola dirección y no forman ningún ciclo; posee tres algoritmos de entrenamiento a elegir. El algoritmo Levenberg-Marquardt (`trainlm`), Bayesian Regularization (`trainbr`) y, Scaled Conjugate Gradient (`trainscg`), estos algoritmos se analizarán en la sección de entrenamiento de la red neuronal artificial.

En el entorno de trabajo de Matlab, se procede a importar la fuente de datos para la creación de la red neuronal, se cuenta con un total de 42 variables y 9260 registros en total.

Se establece la variable objetivo (Casas Positivas) y las 11 variables independiente o predictoras seleccionadas para el ensayo 1, 6 variables predictoras para el ensayo 2 y, 3 variables predictoras para el ensayo 3, con los siguientes comandos:

```
y = [t.Casas_Positivas];
```

### **Ensayo 1**

```
x = [t.Area, t.Total_Casas, t.Habitantes, t.Terreno_Baldio, t.Mes, t.Semana, t.Estacion,  
t.Precipitacion, t.Temperatura_max, t.Temperatura_min, t.alcan_2];
```

### **Ensayo 2**

```
x = [t.Area, t.Total_Casas, t.Habitantes, t.Terreno_Baldio, t.Mes, t.Semana];
```

### **Ensayo 3**

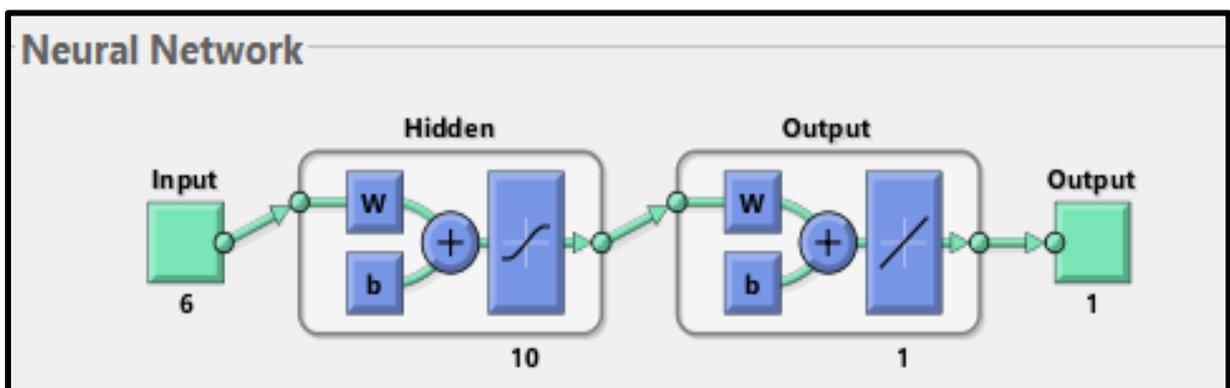
```
x = [t.Area, t.Total_Casas, t.Habitantes];
```

Una vez determinada la variable objetivo ( $y$ ) y las variables predictoras ( $x$ ), cabe mencionar que la variable  $x$  se repite para los tres ensayos, pero esto no implica que exista algún tipo de inconveniente porque las pruebas se realizan en ambientes separados; luego se utiliza la herramienta `nftool` para crear una Red Neuronal Ajustada, y por la organización de los datos establecidos se selecciona la opción matriz de filas.

Matlab utiliza la técnica de detención anticipada para mejorar la generalización en todas las funciones de creación de las redes neuronales supervisadas, esta técnica divide la totalidad de los datos en tres grupos, el primer grupo de datos corresponde al de entrenamiento, el segundo grupo de datos los utiliza para la validación y el tercer grupo de datos los utiliza para las pruebas.

Se establece los datos de prueba y validación, de los 9260 registros disponible para la investigación, se asigna aleatoriamente 6482 registros para el entrenamiento que corresponden al 70%, 1389 registros para la validación que equivale al 15% y 1389 datos para la prueba que es el 15% restante. Además, para el entrenamiento de la red neuronal se establece un total de 300 épocas para el ensayo 1 y 2, y 1000 épocas para el ensayo 3.

Se realizó pruebas con diferentes cantidades de neuronas ocultas, llegando a probar hasta con 20 neuronas ocultas, con lo que la red se vuelve más flexible pero el problema a resolver se puede tornar muy particular y la solución general se podría ver afectada. Aun así, no se obtuvo ningún cambio significativo, quedando establecida la arquitectura de la red neuronal con 10 neuronas ocultas en el ensayo 1 al igual que en el ensayo 2 y 3, como se observa en la siguiente figura:



**Figura 14.** Arquitectura de la red neuronal – Ensayo 2.

**Fuente:** El autor

**Elaborado por:** El autor

#### 4.5 Entrenamiento de la Red Neuronal

El siguiente paso es entrenar a la red neuronal, para lo cual Matlab cuenta con tres algoritmos de entrenamiento para la Red Neuronal Ajustada: Levenberg-Marquardt, Bayesian Regularization y, Scaled Conjugate Gradient.

Como lo indica la documentación de Matlab, el algoritmo Levenberg-Marquardt es recomendado para la mayoría de los problemas, utiliza cálculos jacobianos y para su ejecución requiere más memoria, pero utiliza menos tiempo para el entrenamiento deteniéndose cuando la generalización ya no mejora, de acuerdo al error cuadrático medio de las muestras de validación.

El algoritmo Bayesian Regularization para su funcionamiento utiliza más tiempo, pero puede dar mejores resultados de generalización para datos pequeños o ruidosos, ya que no utiliza datos de validación para el entrenamiento, su ejecución finaliza con la minimización adaptativa de los pesos.

Por último, el algoritmo Scaled Conjugate Gradient es recomendado para problemas muy grandes, utiliza cálculos de gradientes por lo que ocupa menos memoria para su funcionamiento y se basa en direcciones conjugadas, pero no realiza búsqueda de líneas en cada iteración.

Para el entrenamiento de la Red Neuronal Ajustada en esta investigación, se utilizó el algoritmo Bayesian Regularization, debido a la naturaleza de los datos disponibles, porque en conjunto estos no presentan un comportamiento estacional y se presentan en una forma de ruido, cabe indicar que se realizaron pruebas con los otros dos algoritmos, pero los resultados no fueron satisfactorios.

A continuación, se resumen los parámetros para el entrenamiento de la Red Neuronal Ajustada aplicados a los ensayos 1, 2 y 3:

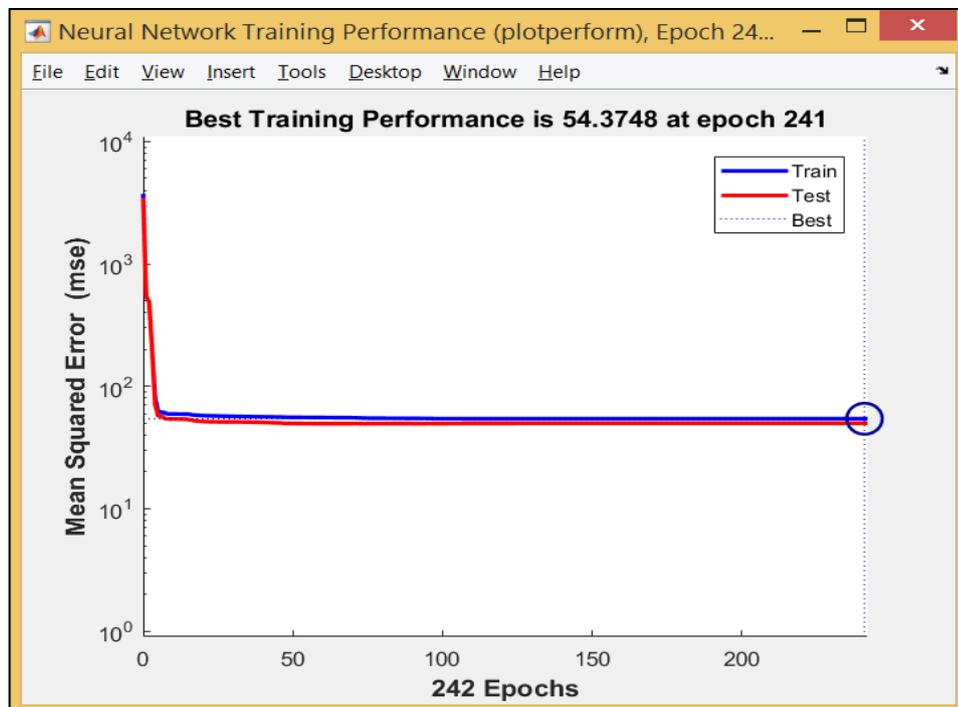
**Tabla 9.** Parámetros de entrenamiento de la Red Neuronal Ajustada

Grupo de datos	Porcentaje	Valor	Capas ocultas	Neuronas ocultas	Épocas
Entrenamiento	70%	6482	2	10	Ensayo 1 y 2 (300) Ensayo 3 (1000)
Validación	15%	1389			
Prueba	15%	1389			
Total	100%	9260	2	10	-

**Fuente:** El autor

**Elaborado por:** El autor

El mejor rendimiento del grupo de datos de entrenamiento de acuerdo al valor de MSE (error cuadrático medio) es de 50.12 en la época 69 para el ensayo 1, 54.37 en la época 241 para el ensayo 2 y, 68.36 en la época 915, de aquí en adelante no se produce ningún sobreajuste significativo en el entrenamiento de la red neuronal, con lo que queda justificado el valor de entrenamiento de 300 épocas para los ensayos 1 y 2, y 1000 épocas para el ensayo 3, como lo podemos observar en la siguiente figura:



**Figura 15.** Mejor rendimiento del grupo de datos de entrenamiento de la red neuronal – Ensayo 2.

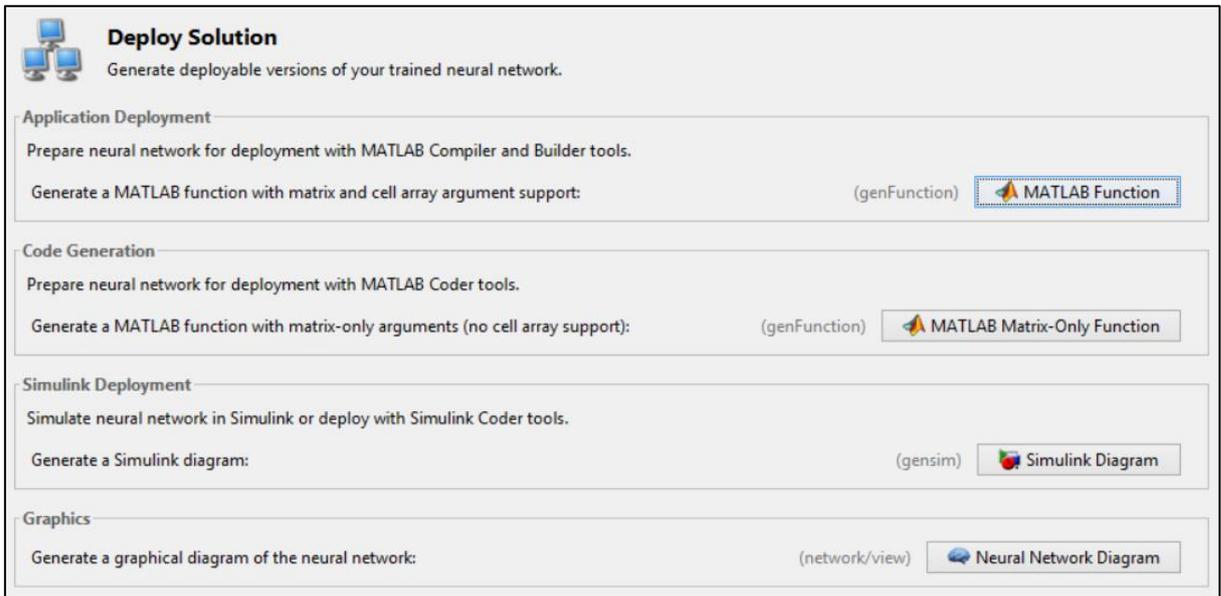
**Fuente:** El autor

**Elaborado por:** El autor

**CAPÍTULO V.  
IMPLEMENTACIÓN DE LA TÉCNICA**

## 5.1 Creación del scrip de la Red Neuronal Entrenada

Una vez realizado el entrenamiento de la red neuronal ajustada y luego de una serie de repeticiones de entrenamiento, hasta que la red tenga convergencia, se procede a generar el código seleccionando la opción MATLAB Function, este scrip permitirá predecir las Casas Positivas con presencia de criaderos de mosquitos *Aedes aegypti*, a partir de los nuevos datos que se ingresen a la red neuronal entrenada.



**Figura 16.** Generación del scrip de la red neuronal entrenada.

**Fuente:** El autor

**Elaborado por:** El autor

Se presenta un resumen del código resultante de la red neuronal entrenada a la cual se la adjudicó con el nombre de RedNeuronalEntrenad.m:

```

function [Y,Xf,Af] = RedNeuronalEntrenada(X,~,~)

% ===== MODULE FUNCTIONS =====

% Map Minimum and Maximum Input Processing Function
function y = mapminmax_apply(x,settings)
y = bsxfun(@minus,x,settings.xoffset);
y = bsxfun(@times,y,settings.gain);
y = bsxfun(@plus,y,settings.ymin);
end

% Sigmoid Symmetric Transfer Function
function a = tansig_apply(n,~)
a = 2 ./ (1 + exp(-2*n)) - 1;
end

% Map Minimum and Maximum Output Reverse-Processing Function
function x = mapminmax_reverse(y,settings)
x = bsxfun(@minus,y,settings.ymin);
x = bsxfun(@rdivide,x,settings.gain);
x = bsxfun(@plus,x,settings.xoffset);
end

```

**Figura 17.** Resumen del scrip de la red neuronal entrenada.

**Fuente:** El autor

**Elaborado por:** El autor

El código completo del scrip de la Red Neuronal Ajustada se encuentra en la sección de Anexos.

## 5.2 Ingreso de datos y simulación en la Red Neuronal Entrenada

El siguiente paso es ingresar los datos que se desea predecir en la red neuronal entrenada, proporcionando los datos seleccionados en los ensayos 1, 2 y 3 (11 variables predictoras para el ensayo 1, 6 variables predictoras para el ensayo 2 y, 3 variables predictoras para el ensayo 3) para la creación de la red neuronal, cabe recordar que los datos que se obtienen están en base a la granularidad de semanas, excepto en el ensayo 3 que ya no se considera la variable semana por tener una relación baja.

Para la simulación de la red neuronal se realizarán 2 pruebas en semanas distintas para verificar si el resultado se próxima a lo realizado en la actividad real de control vectorial, se utilizarán datos para la predicción de la semana 20 y 49 correspondiente al mes de mayo y diciembre respectivamente del año 2017 y se comparará con la información reportada en las mismas semanas, por el personal del Departamento de Control Vectorial del Distrito de Salud 07D02 Machala.

En el software Matlab se inserta las variables independientes o predictoras con el nombre x1, los datos deben ser acorde a la semana que se desea predecir, para esto se ejecuta el siguiente código:

### Ensayo 1

*x1=[t1.Area,t1.Total\_Casas,t1.Habitantes,t1.Terreno\_Baldio,t1.Mes,t1.Semana,t1.Estacion,t1.Precipitacion,t1.Temperatura\_max,t1.Temperatura\_min,t1.alcan\_2];*

### Ensayo 2

*x1=[t1.Area,t1.Total\_Casas,t1.Habitantes,t1.Terreno\_Baldio,t1.Mes,t1.Semana];*

### Ensayo 3

*x1=[t1.Area,t1.Total\_Casas,t1.Habitantes];*

A continuación, una muestra de los datos que son requeridos para el ingreso y asignación de la variable x1, se tiene una variable adicional que es Código de Barrio, pero esta es solo para referenciar y ubicar los datos de acuerdo a los barrios de la ciudad de Machala:

**Tabla 10.** Datos para el ingreso en la red neuronal entrenada para la predicción – Ensayo 1.

Codigo_Barrío	Area	Total_Casas	Habitantes	Terreno_Baldio	Mes	Semana	Estacion	Precipitacion	Temperatura_max	Temperatura_min	alcan_2
0701060052	28540.37	48	246	7	12	49	0	0.8	26.0	25.2	0
0701060050	81062.99	275	1,409	21	12	49	0	0.8	26.0	25.2	0
0701060029	26413.68	85	436	8	12	49	0	0.8	26.0	25.2	0
0701070012	43022.68	241	1,235	16	12	49	0	0.8	26.0	25.2	0
0701070024	55490.87	80	410	7	12	49	0	0.8	26.0	25.2	0
0701010008	131465.60	268	1,374	3	12	49	0	0.8	26.0	25.2	0
0701030021	15868.40	67	284	1	12	49	0	0.8	26.0	25.2	1
0701060030	10936.98	34	174	4	12	49	0	0.8	26.0	25.2	0
0701060021	56186.81	69	354	11	12	49	0	0.8	26.0	25.2	0
0701060042	91501.98	101	434	9	12	49	0	0.8	26.0	25.2	0

**Fuente:** (El autor)

**Elaborado por:** El autor

**Tabla 11.** Datos para el ingreso en la red neuronal entrenada para la predicción – Ensayo 2.

Codigo_Barrío	Area	Total_Casas	Habitantes	Terreno_Baldío	Mes	Semana
0701060052	28540.37	48	246	7	12	49
0701060050	81062.99	275	1,409	21	12	49
0701060029	26413.68	85	436	8	12	49
0701070012	43022.68	241	1,235	16	12	49
0701070024	55490.87	80	410	7	12	49
0701010008	131465.60	268	1,374	3	12	49
0701030021	15868.40	67	284	1	12	49
0701060030	10936.98	34	174	4	12	49
0701060021	56186.81	69	354	11	12	49
0701060042	91501.98	101	434	9	12	49

**Fuente:** El autor

**Elaborado por:** El autor

**Tabla 12.** Datos para el ingreso en la red neuronal entrenada para la predicción – Ensayo 3.

Codigo_Barrío	Area	Total_Casas	Habitantes
0701060052	28540.37	48	246
0701060050	81062.99	275	1,409
0701060029	26413.68	85	436
0701070012	43022.68	241	1,235
0701070024	55490.87	80	410
0701010008	131465.60	268	1,374
0701030021	15868.40	67	284
0701060030	10936.98	34	174
0701060021	56186.81	69	354
0701060042	91501.98	101	434

**Fuente:** El autor

**Elaborado por:** El autor

Una vez que se tienen lista las variables predictoras con el nombre  $x_1$ , se ejecuta el código que permite realizar la simulación de la Red Neuronal y obtener los datos de predicción de Casas Positivas como lo muestra la siguiente línea de código para cada ensayo:

### Ensayo 1

$$Y = \text{RedNeuronalEntrenadaE1}(x_1)$$

### Ensayo 2

$$Y = \text{RedNeuronalEntrenadaE2}(x_1)$$

### Ensayo 2

$$Y = \text{RedNeuronalEntrenadaE3}(x_1)$$

Donde:

Y = Lista de valores de salida de predicción de la red neuronal

x1 = Lista de variables predictoras

RedNeuronalEntrenadaE1 = Nombre de la Red Neuronal – Ensayo 1

RedNeuronalEntrenadaE2 = Nombre de la Red Neuronal – Ensayo 2

RedNeuronalEntrenadaE3 = Nombre de la Red Neuronal – Ensayo 3

Como se mencionó anteriormente se ha establecido el nombre de la variable x1 tanto para el ensayo 1, 2 y 3, esto no representa ningún problema ya que las pruebas se han realizado en ambientes diferentes, por lo que no existe la posibilidad de alguna confusión o error en el procedimiento.

### **5.3 Obtención de los resultados**

Una vez conseguidos los resultados de la red neuronal, se procede a calcular el valor absoluto, porque la diferencia entre el valor de predicción y el objetivo puede ser negativo en algunos casos.

Se obtiene una lista de datos correspondiente a la variable objetivo de la red neuronal con el nombre Y, estos datos son valores de predicción de Casas Positivas de los tres ensayos, para este ejemplo las pruebas son realizadas con datos de la semana 49 del año 2017, la cual se lo anexa a los datos trabajados por el personal de Control Vectorial para realizar la respectiva comparación, como lo muestra la siguiente tabla:

**Tabla 13.** Comparación de datos trabajados frente a datos de predicción – Ensayo 1, 2 y 3.

Parroquia	Barrio	Codigo_Barrío	Trabajado	Predicción		
			Distrito 07D02	Ensayo 1	Ensayo 2	Ensayo 3
La Providencia	Urseza 2 Sector 1	0701010090	16	17	23	34
La Providencia	Alborada 2	0701010011	13	9	9	11
La Providencia	Urseza 2 sector 2	0701010091	12	13	23	28
Nueve de Mayo	Venezuela	0701040026	12	7	17	25
Jubones	Rodrigo Murillo	0701070023	8	5	7	9
La Providencia	Alborada 1	0701010010	6	10	8	12
Jambeli	25 de Diciembre	0701060007	5	4	6	10
Puerto Bolívar	4 de Abril	0701030007	5	3	8	12
Nueve de Mayo	Nuevo Pilo	0701040022	5	8	17	21
Puerto Bolívar	Acapulco	0701030012	4	1	2	3
Jubones	Cristo del Consuelo	0701070005	4	7	9	12
La Providencia	La Kattya	0701010044	4	3	3	6
Puerto Bolívar	Venecia del Mar	0701030044	4	2	5	6
La Providencia	10 de Agosto	0701010001	3	2	2	5
La Providencia	20 de Junio	0701010006	3	7	4	7
Jubones	Corina Parrales	0701070004	3	4	8	8
Puerto Bolívar	Rumiñahui	0701030039	3	2	2	4
Puerto Bolívar	24 de Diciembre	0701030004	2	1	2	2
Nueve de Mayo	Jaime Roldos	0701040014	2	2	6	8
Jubones	Las Americas	0701070018	2	2	3	5
Nueve de Mayo	Los Girasoles	0701040017	2	2	6	9
La Providencia	Mario Minuche	0701010067	2	0	1	1
Puerto Bolívar	24 de Julio	0701030005	1	3	3	5
Puerto Bolívar	8 de Septiembre	0701030010	1	1	3	4
Puerto Bolívar	Centenario	0701030019	1	0	2	2
La Providencia	Cuba Libre	0701010020	1	1	2	2
Puerto Bolívar	Olmedo	0701030032	1	2	3	3
La Providencia	Teniente Hugo Ortiz	0701010087	1	1	1	0
Jambeli	9 de Octubre	0701060010	0	4	4	6
Jambeli	Abdon Calderon	0701060011	0	3	6	10
Jambeli	Brisas del Mar	0701060016	0	11	16	18
Jambeli	Cdla. El Chofer	0701060018	0	2	3	4
Jambeli	Los Pinos	0701060037	0	2	3	5
La Providencia	Machala Libre	0701010064	0	24	19	34
Jambeli	Patria Nueva	0701060041	0	6	8	8
Puerto Bolívar	Vicente Rocafuerte	0701030045	0	1	2	2

**Fuente:** El autor

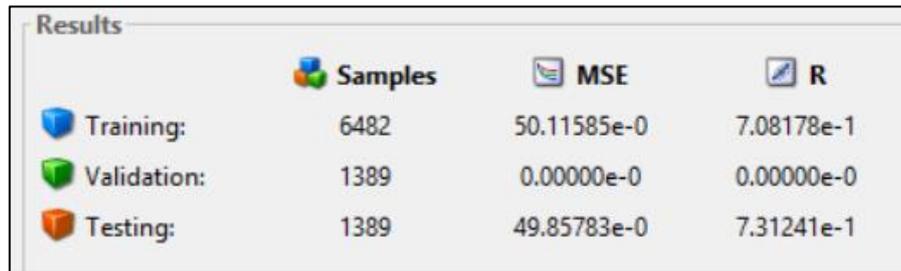
**Elaborado por:** El autor

**CAPÍTULO VI.**  
**EVALUACIÓN DE LA TÉCNICA**

## 6.1 Métricas de evaluación

Para la evaluación de la técnica de predicción seleccionada se utilizará como métricas: el MSE Mean Square Error o Error Cuadrático Medio, traducción al español, y los valores de Regresión R.

Es así que después de lograr la convergencia de la red neuronal obtenemos como resultado 7.31241 como valor de R (análisis de regresión) que corresponde al 73% de precisión y 49.86 como valor de MSE (error cuadrático medio) en el grupo de datos de Prueba para el ensayo 1, un total de 7.14622 como valor de R que corresponde al 71% de precisión y 49.90 como valor de MSE en el grupo de datos de Prueba para el ensayo 2 y, para el ensayo 3 el valor de 61% de precisión y 53.9 como valor de MSE en el grupo de datos de Prueba, tal como lo muestran la siguientes figuras:

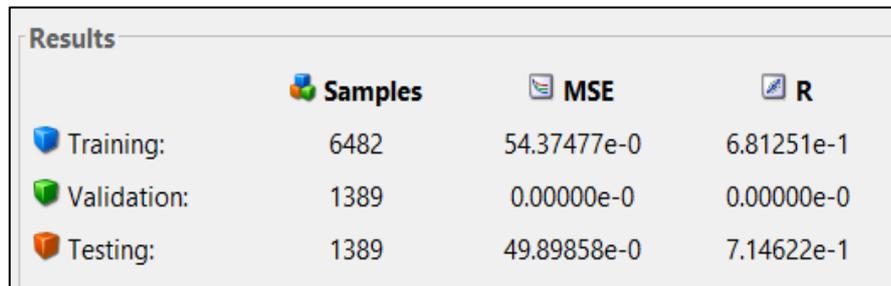


	Samples	MSE	R
Training:	6482	50.11585e-0	7.08178e-1
Validation:	1389	0.00000e-0	0.00000e-0
Testing:	1389	49.85783e-0	7.31241e-1

**Figura 18.** Resultado del entrenamiento de la red neuronal artificial – Ensayo 1.

**Fuente:** El autor

**Elaborado por:** El autor



	Samples	MSE	R
Training:	6482	54.37477e-0	6.81251e-1
Validation:	1389	0.00000e-0	0.00000e-0
Testing:	1389	49.89858e-0	7.14622e-1

**Figura 19.** Resultado del entrenamiento de la red neuronal artificial – Ensayo 2.

**Fuente:** El autor

**Elaborado por:** El autor

Results			
	 Samples	 MSE	 R
 Training:	6482	68.35746e-0	5.86405e-1
 Validation:	1389	0.00000e-0	0.00000e-0
 Testing:	1389	53.91770e-0	6.14566e-1

**Figura 20.** Resultado del entrenamiento de la red neuronal artificial – Ensayo 3.

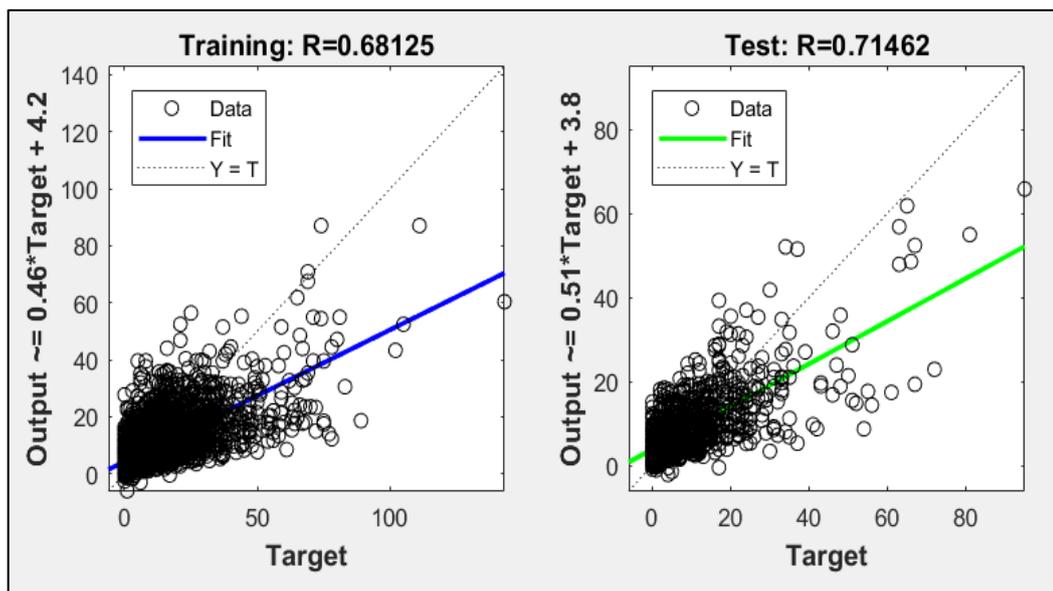
**Fuente:** El autor

**Elaborado por:** El autor

Por lo general, el rendimiento de una red neuronal ajustada ya entrenada puede medirse por el MSE o error cuadrático medio, en los grupos de entrenamiento y prueba, pero es de mucha ayuda utilizar el análisis de regresión proporcionado por Matlab con la respuesta de la red y el objetivo establecido.

El análisis de regresión presenta los resultados de la red frente al objetivo de entrenamiento; para un ajuste perfecto, los datos deben caer a lo largo de la línea punteada de 45 grados siendo el valor de  $R = 1$ , si el valor de  $R = 0$  el ajuste es nulo lo que indica que no existe ninguna relación entre la salida de datos de la red neuronal y el objetivo. La línea continua muestra la mejor relación entre los datos de salida y el objetivo, el color azul representa los datos de entrenamiento y el color verde representa los datos de prueba.

El resultado del entrenamiento de la red neuronal ajustada de la investigación, como se puede observar en la figura del análisis de regresión podemos distinguir que en los datos de pruebas tenemos un valor de  $R = 0.73124$ , que equivale al 73% de precisión para el ensayo 1, un valor de  $R = 0.71462$  que representa el 71% de precisión para el ensayo 2 y, para el ensayo 3 un valor de  $R = 0.614566$  que representa el 61% de precisión en la predicción de los datos estudiados.



**Figura 21.** Análisis de regresión de la red neuronal – Ensayo 2.

**Fuente:** El autor

**Elaborado por:** El autor

En la siguiente tabla se presenta un resumen de los valores obtenidos en las métricas de evaluación de la Red Neuronal Ajustada para los ensayos 1, 2 y 3.

**Tabla 14.** Resumen de resultados de las métricas de evaluación

Descripción	Ensayo 1	Ensayo 2	Ensayo 3
Número de Variables predictoras o independientes	11	6	3
MSE (error cuadrático medio)	49.86	49.90	53.91
Valor de Regresión	0.73124	0.71462	0.61457
Precisión de la Regresión	73%	71%	61%

**Fuente:** El autor

**Elaborado por:** El autor

El resultado de las métricas utilizadas nos indica que, a pesar del valor de MSE o error cuadrático medio es elevado, alrededor del 50% para los ensayos 1 y 2, lo que indica que no es muy preciso, pero el valor de Regresión si nos muestra un resultado favorable del 73% de precisión para el ensayo 1 y del 71% para el ensayo 2 en la predicción de los valores de Casas Positivas con criaderos de mosquitos *Aedes aegypti*, como lo demuestra (Cortina Januch, 2012) en su Tesis Doctoral “Aplicación de técnicas de inteligencia artificial a la predicción de contaminantes atmosféricos”. El ensayo 3 presenta un valor de 54 en MSE más elevado que los dos primeros ensayos, lo que indica que existe un margen de error más amplio y un valor

de Regresión del 61% lo que también indica que los resultados son menos precisión que los dos primeros ensayos.

Una vez realizado las pruebas pertinentes a los tres ensayos, se decide utilizar la Red Neuronal Ajustada del ensayo 2, por ser la que utiliza la menor cantidad de variables predictoras o independientes, optimizando el proceso de predicción de casas positivas con criaderos de mosquitos *Aedes aegypti*. El ensayo 1 no es seleccionado porque se requiere más variables predictoras disponibles para el ingreso de datos a la Red Neuronal Ajustada y el resultado es muy similar al ensayo 2 y, el ensayo 3 también no es seleccionado porque su precisión es la más baja de todos los ensayos, y al no contar con la variable semana el resultado obtenido es el mismo para todas las simulaciones realizadas, lo que se convierte en un resultado constante, ver resultados en el Anexo 9.

Los resultados de la Red Neuronal Ajustada del ensayo 2 se verán reflejados en los mapas temáticos de la ciudad de Machala.

## **6.2 Mapa temático**

Representamos los valores de predicción de la red neuronal, así como también los datos trabajados por el personal del Departamento de Control Vectorial, en un mapa temático de la ciudad de Machala, utilizando como herramienta GIS (Sistema de Información Geográfica) el software QGIS anteriormente Quantum GIS, siendo este un software libre y opera bajo la licencia GNU GPL.

Los mapas temáticos se basan en un mapa topográfico para representar información específica la cual puede ser cualitativa o cuantitativa, en el caso de esta investigación se grafican utilizando la Estratificación de Riesgo Entomológico para Índice Aédicos que utiliza el Departamento de Control Vectorial del Distrito de Salud 07D02 Machala, el cual se establece en la siguiente tabla:

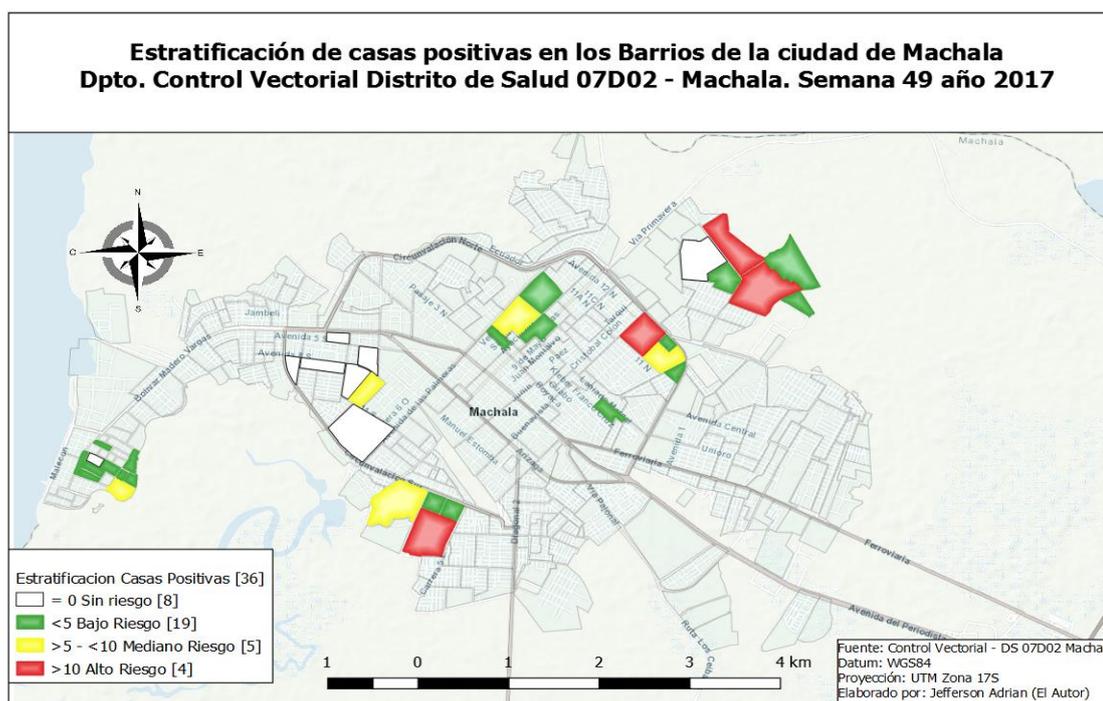
**Tabla 15.** Estratificación de Riesgo Entomológico para Índices aélicos.

Riesgo	Escala de índices	Color
Sin riesgo	0	Blanco
Bajo	< 5	Verde
Mediano	> 5 - < 10	Amarillo
Alto	> 10	Rojo

Fuente: El autor

Elaborado por: El autor

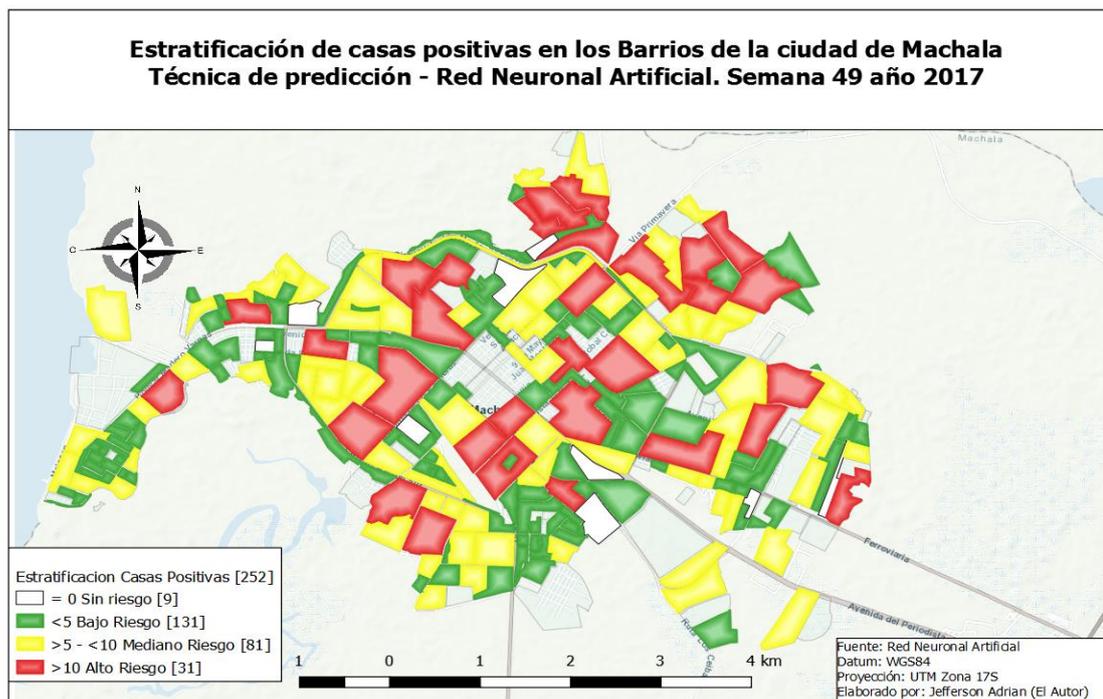
Aunque los resultados de predicción obtenidos por la Red Neuronal Artificial del ensayo 2 se basan en el número de Casas Positivas y no en Índice de Casas, se utiliza la Estratificación de Riesgo Entomológico en los barrios que se ven afectados con la presencia de este indicador para una mejor representación y comprensión de los resultados, como lo demuestran las siguientes Figuras:



**Figura 22.** Mapa temático de la ciudad de Machala con las casas positivas, trabajadas por el personal del Departamento de Control Vectorial del Distrito de Salud 07D02, semana 49 del mes de diciembre de 2017.

Fuente: El autor

Elaborado por: El autor



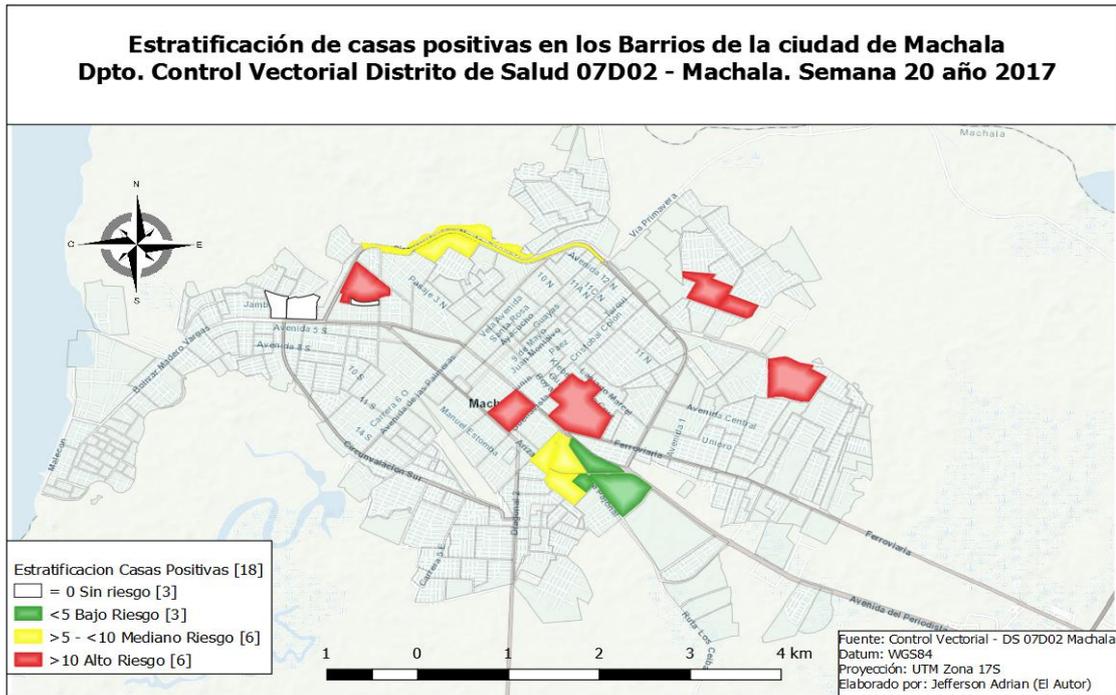
**Figura 23.** Mapa temático de la ciudad de Machala con las casas positivas utilizando red neuronal artificial, semana 49 del mes de diciembre de 2017.

**Fuente:** El autor

**Elaborado por:** El autor

Como se puede observar y al realizar una comparación, en el primer mapa se muestra los barrios que presentan Casas Positivas con criaderos de mosquitos, trabajados en la semana 49 del año 2017 por el Departamento de Control Vectorial; mientras que, en el segundo mapa, se muestran los barrios con Casas Positivas en todos los barrios de la ciudad de Machala utilizando datos de predicción para la misma semana 49, en lo que se puede observar que la clasificación del riesgo de las casas positivas tiene muy buena aproximación de los resultados, lo que indica que los datos representados son un aporte importante para la localización de las casas positivas con criaderos de mosquitos *Aedes aegypti*.

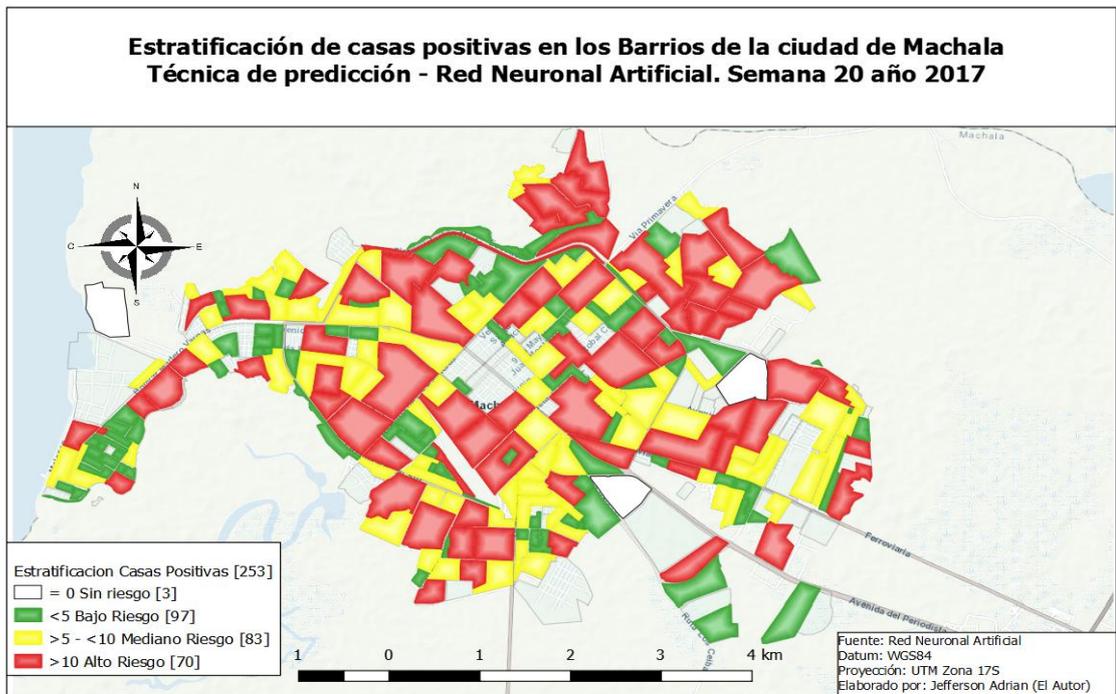
Ahora se presentará otro ejemplo con los valores de casas positivas en los mapas temáticos, utilizando datos de la semana 20 del mes de mayo del año 2017.



**Figura 24.** Mapa temático de la ciudad de Machala con las casas positivas trabajadas por el personal del Departamento de Control Vectorial del Distrito de Salud 07D02, semana 20 del mes de mayo de 2017.

**Fuente:** El autor

**Elaborado por:** El autor



**Figura 25.** Mapa temático de la ciudad de Machala con las casas positivas utilizando red neuronal artificial, semana 20 del mes de mayo de 2017.

**Fuente:** El autor

**Elaborado por:** El autor

Al igual que en el ejemplo anterior, se puede realizar una comparación entre los dos mapas donde se muestran los barrios que presentan Casas Positivas con presencia de criaderos de mosquitos trabajadas en la semana 20 del año 2017 por el Departamento de Control Vectorial, y los datos de predicción para la misma semana 20, teniendo igualmente una aproximación muy buena de resultados trabajados frente a los resultados de predicción, los mismos que se encuentran clasificados utilizando la Estratificación de riesgo entomológico en ambos mapas.

A continuación, se resumen los resultados de predicción de cada uno de los ensayos realizados en esta investigación, utilizando la Estratificación de riesgo entomológico representados en los Mapas Temáticos, al igual que los resultados presentados por el Departamento de Control Vectorial del Distrito de Salud 07D02 Machala correspondiente a la semana 49 del año 2017.

**Tabla 16.** Comparación de resultados de predicción - semana 49 del año 2017.

<b>Riesgo</b>	<b>Distrito 07D02</b>	<b>Ensayo 1</b>	<b>Ensayo 2</b>	<b>Ensayo 3</b>
Bajo	19	15	14	8
Mediano	5	2	4	1
Alto	4	2	3	4
Total	28	19	21	13
Porcentaje	100%	68%	75%	46%

**Fuente:** El autor

**Elaborado por:** El autor

Al revisar el cuadro comparativo con los resultados de los tres ensayos frente a lo realizado por el Departamento de Control Vectorial, se tiene como resultado en el ensayo 1 el 68% de aciertos en la predicción de barrios que presentan Casas Positivas con criaderos de *Aedes aegypti* en relación a la estratificación del riesgo entomológico, en el ensayo 2 el 75% de aciertos y para el ensayo 3 el 46% de precisión, lo que demuestra que el ensayo 2 presenta buenos resultados con solo 6 variables predictoras o independientes.

## CONCLUSIONES

Una vez finalizado el trabajo de investigación se obtienen las siguientes conclusiones:

- Se logra cumplir con el objetivo de desarrollar una herramienta que permita predecir los criaderos de mosquitos *Aedes aegypti*, lo que permitirá tomar medidas de prevención y control de estos vectores, protegiendo a la población de la ciudad de Machala de posibles problemas de salud, a causa de enfermedades como Dengue, Zika y Chikungunya.
- Se evidenció que el Departamento de Control Vectorial del Distrito de Salud 07D02 Machala no cuenta con un software informático a medida para el ingreso y control de datos, la información recopilada por las diferentes actividades es reportada en múltiples archivos de Excel.
- Se constató que la elaboración de la programación semana de los barrios a intervenir en las visitas domiciliarias del Departamento de Control Vectorial la realizan de tres formas detalladas a continuación: 1) seleccionando los barrios con antecedentes de índices aédicos altos, 2) Eligiendo a los barrios que no han sido intervención por un período considerable de tiempo porque se los considera de bajo riesgo, y 3) al transcurrir aproximadamente dos meses y medio vuelven a repetir los barrios seleccionados inicialmente, con lo que se cumple un ciclo de intervención.
- La herramienta de predicción se basa en el uso de Redes Neuronales Artificiales, la misma que fue desarrollada con el software propietario Matlab, utilizando la opción Neural Net Fitting y como algoritmo de entrenamiento Bayesian Regularization; la arquitectura de la Red Neuronal queda conformada por 6 variables predictoras como entrada, 2 capas ocultas con 10 neuronas ocultas para la primera capa y 1 neurona oculta para la segunda y, como salida 1 variable objetivo.
- Los mapas temáticos se desarrollaron con el software QGIS que es un Sistema de Información Geográfica, que cuenta con licencia GNU GPL, el cual permitió graficar los resultados de la herramienta de predicción; para esta actividad se procedió a realizar polígonos en referencia a los barrios de Machala y codificarlos para establecer la relación de enlace de datos, posteriormente se creó los estilos de clasificación

basado en reglas, para la estratificación de casas positivas utilizadas por el Distrito de Salud 07D02 Machala.

- Se demuestra que los datos de clima y socioeconómicos, no ejercen en los datos de este estudio una influencia significativa en los resultados para la variable objetivo Casas Positivas, como inicialmente se deducía que dichas variables tenían una influencia bien definida en la presencia de los criaderos de mosquitos *Aedes aegypti*.
- La evaluación de la herramienta utilizando la técnica de predicción Red Neuronal Artificial, se la realizó utilizando las métricas de Error Cuadrático Medio o MSE y el Valor de Regresión, obteniendo como resultado 49.90 y 71% de precisión respectivamente, valores alcanzados en el ensayo 2. Basándose en el valor de Regresión, los resultados de predicción de la Red Neuronal Artificial son alentadores para nuestra investigación.

## RECOMENDACIONES

- Empezar futuras investigaciones en lo referente a la predicción de criaderos de mosquitos *Aedes aegypti*, esperando que, con el trabajo de investigación realizado, se abra la brecha para que se incentive la investigación de predicción en el campo de control vectorial.
- Sugerir al Ministerio de Salud para que aborden un sistema de control e ingreso de datos de control vectorial que permitan agilizar y regularizar la información, para así tener a disposición datos que permitan realizar investigaciones futuras. También, recomendar la inclusión de variables como tipo de depósitos, depósitos cubiertos, tipo de viviendas, etc., en el registro de datos general; estas variables servirán para incluirlas en futuras investigaciones y saber si su influencia ayuda a mejorar la precisión de la predicción.
- Probar diferente software y nuevas técnicas de predicción para los criaderos de mosquitos. Además, solicitar al Distrito de Salud 07D02 Machala proporcione muchos más datos de control vectorial y así poder mejorar la precisión de la predicción.
- Realizar investigaciones de predicción sobre otra clase de mosquitos como *Anopheles* responsable de transmitir paludismo o malaria; y *Culex*, aunque este mosquito no reporta en el Ecuador transmisión de enfermedad alguna, pero resulta un problema de molestia pública y es el que se lo encuentra en mayor cantidad.
- Proponer al organismo rector del control vectorial, la implementación de un plan piloto con el uso de esta herramienta de predicción de criaderos de mosquitos *Aedes aegypti*, en la ciudad de Machala, con el fin de tener un instrumento de apoyo que se sume a las actividades técnicas que realizan, para priorizar y optimizar recursos y evitar así un incremento de criaderos de mosquitos.

## BIBLIOGRAFÍA

- Aluja, T. (2001). La Minería de Datos, entre la Estadística y la Inteligencia Artificial. *Qüestió*, 25(3), 479–498.
- Bertona, L. F. (2005). *Entrenamiento de redes neuronales basado en algoritmos evolutivos. Grado, Tesis de Ingeniería Informática, UNIVERSIDAD DE BUENOS AIRES.*
- Corso, C. L., & Gibellini, F. (2011). Aplicación de redes bayesianas usando Weka., 939–948.
- Cortina Januch, M. G. (2012). *Aplicación de Técnica de Inteligencia Artificial a la Predicción de Contaminantes Atmosférico.* Universidad Politécnica de Madrid.
- Dueñas-Reyes, M. X. (2009). Minería de datos espaciales en búsqueda de la verdadera información. *Ingeniería Universidad de Bogotá*, 13(1), 137–156.
- Erandi, R., Martínez, B., Ramírez, N. C., Gabriel, H., Mesa, A., Suárez, I. R., ... Morales, S. L. B. (2009). Árboles De Decisión Como Herramienta En El Diagnóstico Médico.
- Felgaer, P. (2004). *Optimización de Redes Bayesianas basado en Técnicas de Aprendizaje por Inducción. Reportes Técnicos en Ingeniería del Software.*
- Genaro, N. (2014). Sistema de predicción de ruido urbano mediante redes neuronales, 212. Retrieved from <http://digibug.ugr.es/bitstream/10481/35111/1/24184639.pdf>
- Gras, J. A. (2001). *Diseños de series temporales: técnicas de análisis.* Universitat de Barcelona. Retrieved from [https://books.google.com.ec/books?id=IGptN\\_0cXMwC](https://books.google.com.ec/books?id=IGptN_0cXMwC)
- Isasi Viñuela, P., & Galván León, I. M. (2000). *Redes de neuronas artificiales: un enfoque Práctico.* Madrid.
- Jiawei, H., Kamber, M., Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques.* San Francisco, CA, itd: Morgan Kaufmann. <https://doi.org/10.1016/B978-0-12-381479-1.00001-0>
- Jirón Araya, I. F. (2015). Redes Neuronales Artificiales Aplicadas a la Seguridad de la Información, 1–28.
- López, C. P. (2007). *Minería de datos: técnicas y herramientas.* Paraninfo Cengage Learning. Retrieved from [https://books.google.es/books?id=wz-D\\_8uPFCEC](https://books.google.es/books?id=wz-D_8uPFCEC)
- Marín, J. M. (2007). Introducción a las redes neuronales aplicadas. Manual Data Mining, 1–31. Retrieved from <http://halweb.uc3m.es/esp/Personal/personas/jmmarin/esp/DM/tema3dm.pdf>
- Martínez F.J, Ordieres J., Pernía A., A. F. (2005). Minería de datos en series temporales para la búsqueda de conocimiento oculto en histórico de procesos industriales. *Research Fund for Coal and Steel*, 31–38. Retrieved from <http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:No+Title#0>
- Matich, D. J. (2001). Redes Neuronales: Conceptos Básicos y Aplicaciones. *Historia*, 55. Retrieved from <ftp://decsai.ugr.es/pub/usuarios/castro/Material-Redes->

Neuronales/Libros/matich-redesneuronales.pdf

- Moreno, M., Miguel, L., García, F., & Polo, J. (2008). Aplicación de técnicas de minería de datos para la evaluación del rendimiento académico y la deserción estudiantil. *Iiis.Org*, 14. Retrieved from [http://www.iiis.org/CDs2010/CD2010CSC/CISCI\\_2010/PapersPdf/CA156FK.pdf](http://www.iiis.org/CDs2010/CD2010CSC/CISCI_2010/PapersPdf/CA156FK.pdf)
- Norvig, P., & Russell, S. (2014). *Inteligencia artificial, Un Enfoque Moderno. Elsevier Brasil*. (Vol. 1). <https://doi.org/M-26913-2004>
- Rodríguez, D., & Dolado, J. (2007). Redes bayesianas en la ingeniería del software. *Cc.Uah.Es*, 1–21. Retrieved from <http://www.cc.uah.es/drg/b/RodriguezDolado.BBN.2007.pdf>
- Rubio, S., & Rubio, B. (2012). Análisis de decisión y tratamiento de la incertidumbre, 1–24.
- Vásquez V, V., Artega G, P., Chanamé A, K., & Esquivel T, A. (2013). Modelamiento matemático y por redes neuronales artificiales del crecimiento de *Spirulina* sp. en fotobiorreactor con fuente de luz fluorescente e iluminación en estado sólido. *Scientia Agropecuaria*, 4(3), 199–209. Retrieved from <http://www.redalyc.org/articulo.oa?id=357633706006>

## **ANEXOS**

## Anexo 1: Solicitud de aprobación del Convenio

# UPSTATE

MEDICAL UNIVERSITY

Center for Global Health &  
Translational Science

Quito, 13 de Agosto 2013

Dr. Santiago Escalante  
Director Nacional de Vigilancia - MSP

Asunto: Aprobación proyecto *Fortalecimiento de las Capacidades en Ecuador: Colaboración para mejorar la vigilancia de enfermedades transmitidas por vectores.*

En su despacho:

Por medio de la presente tengo el agrado de informar que el proyecto *Fortalecimiento de las Capacidades en Ecuador: Colaboración para mejorar la vigilancia de enfermedades transmitidas por vectores.* Presentado por el Dr. Timothy Endy y la Dra. Anna Stewart Ibarra del Center for Global Health and Translational Science de la Upstate Medical University (CGHATS UMU) ha cumplido con todas las observaciones realizadas en la Evaluación de protocolos de investigación, se adjunta el Informe Final respectivo, por lo que solicito muy comedidamente la Aprobación oficial por parte de su dirección y poder seguir en esta importante iniciativa que involucra a diferentes instituciones nacionales e internacionales en un tema de prioridad nacional.

En espera de su favorable respuesta a la presente.

Atentamente,



Mercy J. Borbor Cordova, Ph.D.  
09066777786

Coordinadora nacional de proyectos de investigación del Center for Global Health & Translational Science - Upstate Medical University – State University of New York.



Cc: Dra. Alexa Zambrano  
Directora del INSPI, MSP.

Ec. Ruth Lucio Romero  
Inteligencia de la Salud, MSP.

Dr. Tim Endy  
Director del Proyecto, CGHATS UMU

Dra. Anna Stewart Ibarra  
Co-directora del Proyecto, CGHATS UMU

## Anexo 2: Aprobación del Convenio



Quito, 13 de agosto de 2013  
Of. MSP-CGDES-DIS-036-2013

Señora  
**Mery J. Borbor Cordova, Ph.D**  
Coordinadora Nacional de Proyectos de Investigación del Center for Global  
Health & Translational Science – Upstate Medical University - State University of  
New York  
Presente

**Asunto: “Proyecto de Fortalecimiento de las capacidades en Ecuador:  
Colaboración para mejorar la vigilancia de enfermedades transmitidas por  
vectores”.**

La Dirección de Inteligencia de la Salud de la Coordinación General de Desarrollo  
Estratégico en Salud, informa que fueron salvadas las observaciones realizadas al  
proyecto en referencia, por tal motivo informo que el estudio se encuentra  
**APROBADO** para su ejecución.

Atentamente,

**Dr. Santiago Escalante**  
**DIRECTOR NACIONAL DE INTELIGENCIA DE LA SALUD**  
Ministerio de Salud Pública



### Anexo 3: Nombramiento de Coordinador Provincial



5 de junio del 2017

A quien corresponda:

Por medio de la presente, la suscrita Anna M. Stewart Ibarra, PhD MPA, Directora del Programa de Investigación de América Latina, Centro de Salud Global y Ciencia Translacional Universidad Estatal de Nueva York (SUNY) Upstate Medical University certifico que el Dr. Félix Efraín Beltrán Ayala, Mgs, ecuatoriano con cédula de identidad numero 0701045502, profesor de la Universidad Técnica de Machala se ha desempeñado como Coordinador Provincial y Co-investigador del Proyecto de Investigación **"Fortalecimiento de las capacidades en Ecuador: Colaboración para mejorar la vigilancia de enfermedades febriles transmitidas por vectores"**, el mismo que se ejecuta en la provincia de El Oro - Ecuador, desde el mes de septiembre de 2013 hasta la presente fecha.

El Dr. Beltrán durante su tiempo de colaboración en el proyecto ha demostrado voluntad, creatividad, y profesionalismo.

Esta certificación se extiende en base a los archivos que reposan en mi oficina.

Atentamente,

Anna M. Stewart Ibarra, PhD, MPA

Assistant Professor, Department of Medicine and Public Health and Preventative  
Medicine  
Director, Latin America Research Program  
Center for Global Health & Translational Science  
SUNY Upstate Medical University  
4100 Institute for Human Performance  
505 Irving Ave., Syracuse, NY 13210 USA  
Phone: +1 315 464 6489 (office), +1 315 657 8270 (cell)

## Anexo 4: Autorización de uso de datos



Machala, 27 de Enero de 2017

A quien corresponda:

Yo, Dr. Félix Efraín Beltrán Ayala, Mgs, en calidad de Coordinador Provincial y Co-investigador del Proyecto de Investigación "Fortalecimiento de las capacidades en Ecuador: Colaboración para mejorar la vigilancia de enfermedades febriles transmitidas por vectores", autorizo al Sr. Jefferson Eduardo Adrián Delgado con cédula de identidad No: 0703183756, el mismo que es un colaborador y trabajador de este proyecto; hacer uso de los datos e información recabados a través del convenio de cooperación interinstitucional, solo con fines educativos e investigativos para su tesis de grado para poderse incorporar como Ingeniero en Sistema en la Universidad Técnica Particular de Loja, en la cual no podrá divulgar total o parcialmente los datos de ninguna institución colaboradora donde consten nombres de personas o su identificación, únicamente le está permitido presentar los resultados obtenidos por medio de su investigación.

Particular que comunico para los fines pertinentes.

Atentamente,

Dr. Félix Efraín Beltrán Ayala, Mgs.  
**COORDINADOR PROVINCIAL E INVESTIGADOR DEL PROYECTO  
PROFESOR TITULAR AUXILIAR NIVEL 1 GRADO 1 - UTMACH**



## Anexo 5: Scrip de la Red Neuronal Entrenada – Ensayo 1

```
function [Y,Xf,Af] = RedNeuronalEntrenada(X,~,~)

% Input 1
x1_step1.xoffset =
[0;0;2463.81;8;34;0;1;1;0;0;23.5434523809524;19.2285714285714;0];
x1_step1.gain = [0.00544959128065395;2;4.73129654013768e-
06;...;0.2198233562316;2];
x1_step1.ymin = -1;

% Layer 1
b1 = [-0.038111232499287731;2.7706921789563239;...;2.533002375666308];
IW1_1 = [-1.5149651492539042 -0.71578683519701181 ...-1.1391611659352341];

% Layer 2
b2 = 0.44355175834256966;
LW2_1 = [0.11410008100085892 2.107552557926851 ...-2.4693892932104005];

% Output 1
y1_step1.ymin = -1;
y1_step1.gain = 0.013986013986014;
y1_step1.xoffset = 0;

% ===== SIMULATION =====

% Format Input Arguments
isCellX = iscell(X);
if ~isCellX
    X = {X};
end

% Dimensions
TS = size(X,2); % timesteps
if ~isempty(X)
    Q = size(X{1},1); % samples/series
```

```

else
    Q = 0;
end

% Allocate Outputs
Y = cell(1,TS);

% Time loop
for ts=1:TS

    % Input 1
    X{1,ts} = X{1,ts}';
    Xp1 = mapminmax_apply(X{1,ts},x1_step1);

    % Layer 1
    a1 = tansig_apply(repmat(b1,1,Q) + IW1_1*Xp1);

    % Layer 2
    a2 = repmat(b2,1,Q) + LW2_1*a1;

    % Output 1
    Y{1,ts} = mapminmax_reverse(a2,y1_step1);
    Y{1,ts} = Y{1,ts}';
end

% Final Delay States
Xf = cell(1,0);
Af = cell(2,0);

% Format Output Arguments
if ~isCellX
    Y = cell2mat(Y);
end
end

```

```
% ===== MODULE FUNCTIONS =====
```

```
% Map Minimum and Maximum Input Processing Function
```

```
function y = mapminmax_apply(x,settings)
```

```
y = bsxfun(@minus,x,settings.xoffset);
```

```
y = bsxfun(@times,y,settings.gain);
```

```
y = bsxfun(@plus,y,settings.ymin);
```

```
end
```

```
% Sigmoid Symmetric Transfer Function
```

```
function a = tansig_apply(n,~)
```

```
a = 2 ./ (1 + exp(-2*n)) - 1;
```

```
end
```

```
% Map Minimum and Maximum Output Reverse-Processing Function
```

```
function x = mapminmax_reverse(y,settings)
```

```
x = bsxfun(@minus,y,settings.ymin);
```

```
x = bsxfun(@rdivide,x,settings.gain);
```

```
x = bsxfun(@plus,x,settings.xoffset);
```

```
end
```

## Anexo 6: Scrip de la Red Neuronal Entrenada – Ensayo 2

```
function [Y,Xf,Af] = myNeuralNetworkFunction(X,~,~)

% Input 1
x1_step1.xoffset = [2463.81;8;34;0;1;1];
x1_step1.gain = [4.73129654013768e-
06;0.00238663484486874;0.0005249343832021;0.0208333333333333;0.181818181818182
;0.0384615384615385];
x1_step1.ymin = -1;

% Layer 1
b1 = [-1.7529108275410035;-2.2196223966673623;-
0.023053755120628835;1.5490975523153303;-1.9481104368634019;-
1.6338954935166381;-0.56795291341599952;0.67225635218283331;-
0.51695820441180018;0.37693216965826487];
IW1_1 = [-0.3927215286916611 2.8502504556746544 -2.3549157763226867
0.92955821996328492 0.47716576796205129 -1.0074696970679666;-
1.8964229338650731 2.4427267006950477 -0.6444614147742449 -
0.004672345548854295 -0.93581325676502858 0.71284446590599038;-
0.66933505948346217 0.73621114474191285 -0.16161279264344428 -
0.024341092542806055 -0.35268970557865292
1.4405192798535207;1.1483865055744076 -2.307439082074521 1.0941744825974098 -
0.22915416485669129 1.0297026798575832 0.64007332993422772;-
0.96031386741163649 2.264583414955355 -1.1344138272932609 0.17943459524722874
-0.57212558962360904 -2.1831368090619083;-1.4869112802796498
2.3016473757114557 -0.88579068373704928 -0.011039430664256361 -
1.3909583744389065 0.54558402413449503;-0.3302290090485509 -
0.16353235967597157 0.065534235882259373 -0.31804607425222386
0.20361145712033576 0.39773422469443936;0.84737866868173661
0.33213319959952603 -0.58709759818991858 0.80266742342347175
0.33464983097971313 -0.40336697175676511;-1.155364023159718 -
0.08978315415669369 0.64596008712295394 -0.85659325875997172 -
0.98888893633177988 0.85749148218412596;0.07921510832933161
0.32294703823628135 0.55225366687748323 -0.16125079190262978
0.38264003571704253 -0.78030498034060203];

% Layer 2
b2 = 0.27381697083001416;
LW2_1 = [2.335471787569591 -2.5922177303976111 -0.52310553446543817
2.7221778662386251 0.88939530841055048 3.2182245990468621 1.633863108596318
2.0808107910704798 1.3608243755915288 0.42951728272420703];

% Output 1
y1_step1.ymin = -1;
y1_step1.gain = 0.013986013986014;
y1_step1.xoffset = 0;

% ===== SIMULATION =====

% Format Input Arguments
isCellX = iscell(X);
if ~isCellX
```

```

    X = {X};
end

% Dimensions
TS = size(X,2); % timesteps
if ~isempty(X)
    Q = size(X{1},1); % samples/series
else
    Q = 0;
end

% Allocate Outputs
Y = cell(1,TS);

% Time loop
for ts=1:TS

    % Input 1
    X{1,ts} = X{1,ts}';
    Xp1 = mapminmax_apply(X{1,ts},x1_step1);

    % Layer 1
    a1 = tansig_apply(repmat(b1,1,Q) + IW1_1*Xp1);

    % Layer 2
    a2 = repmat(b2,1,Q) + LW2_1*a1;

    % Output 1
    Y{1,ts} = mapminmax_reverse(a2,y1_step1);
    Y{1,ts} = Y{1,ts}';
end

% Final Delay States
Xf = cell(1,0);
Af = cell(2,0);

% Format Output Arguments
if ~isCellX
    Y = cell2mat(Y);
end
end

% ===== MODULE FUNCTIONS =====

% Map Minimum and Maximum Input Processing Function
function y = mapminmax_apply(x,settings)
y = bsxfun(@minus,x,settings.xoffset);
y = bsxfun(@times,y,settings.gain);
y = bsxfun(@plus,y,settings.ymin);
end

% Sigmoid Symmetric Transfer Function
function a = tansig_apply(n,~)
a = 2 ./ (1 + exp(-2*n)) - 1;
end

```

```
% Map Minimum and Maximum Output Reverse-Processing Function
function x = mapminmax_reverse(y,settings)
x = bsxfun(@minus,y,settings.ymin);
x = bsxfun(@rdivide,x,settings.gain);
x = bsxfun(@plus,x,settings.xoffset);
end
```

## Anexo 7: Scrip de la Red Neuronal Entrenada – Ensayo 3

```
function [Y,Xf,Af] = myNeuralNetworkFunction(X,~,~)

% Input 1
x1_step1.xoffset = [2463.81;8;34];
x1_step1.gain = [4.73129654013768e-06;0.002386634844486874;0.0005249343832021];
x1_step1.ymin = -1;

% Layer 1
b1 = [0.72310723609597494;-0.19852492770654295;1.0854742011799208;-
1.3023567770526276;-1.3698186533299666;-0.85233171043135347;-
2.2355822665811083;-1.9043239878835521;3.1086261681586178;-
0.49744633944688865];
IW1_1 = [1.6005386447122329 -0.46650392408532337 -0.071424376803267062;-
2.2265124384265707 1.6375053676573934 -0.15963407556452763;1.7340640178341036
-1.2746554035897431 -3.0022806886369309;-2.4837785543052751 1.7595138481084289
1.8138808703268716;-1.7708524750234982 1.6266717882283095 4.5155114809193986;-
4.2421717634708429 0.4828724120484611 1.0812026312973757;-0.78707935584334077
-0.87462985224367962 -0.85281377980615702;-0.77938516427988835
0.94684425996871568 1.5522750465329658;0.28733067543080992 1.0821293957907401
2.1990584185891384;-4.0963270411409338 1.8750016902764943
0.090273786666387093];

% Layer 2
b2 = -1.778289002859043;
LW2_1 = [3.159365954016601 2.2903086398555055 2.9290915909189259
1.8845196913386255 1.3944515902599866 1.2396610994928436 1.6640956563121756 -
1.1299398042750153 0.6813535085273752 -1.5750482646351234];

% Output 1
y1_step1.ymin = -1;
y1_step1.gain = 0.013986013986014;
y1_step1.xoffset = 0;

% ===== SIMULATION =====

% Format Input Arguments
isCellX = iscell(X);
if ~isCellX
    X = {X};
end

% Dimensions
TS = size(X,2); % timesteps
if ~isempty(X)
    Q = size(X{1},1); % samples/series
else
    Q = 0;
end

% Allocate Outputs
Y = cell(1,TS);
```

```

% Time loop
for ts=1:TS

    % Input 1
    X{1,ts} = X{1,ts}';
    Xp1 = mapminmax_apply(X{1,ts},x1_step1);

    % Layer 1
    a1 = tansig_apply(repmat(b1,1,Q) + IW1_1*Xp1);

    % Layer 2
    a2 = repmat(b2,1,Q) + LW2_1*a1;

    % Output 1
    Y{1,ts} = mapminmax_reverse(a2,y1_step1);
    Y{1,ts} = Y{1,ts}';
end

% Final Delay States
Xf = cell(1,0);
Af = cell(2,0);

% Format Output Arguments
if ~isCellX
    Y = cell2mat(Y);
end
end

% ===== MODULE FUNCTIONS =====

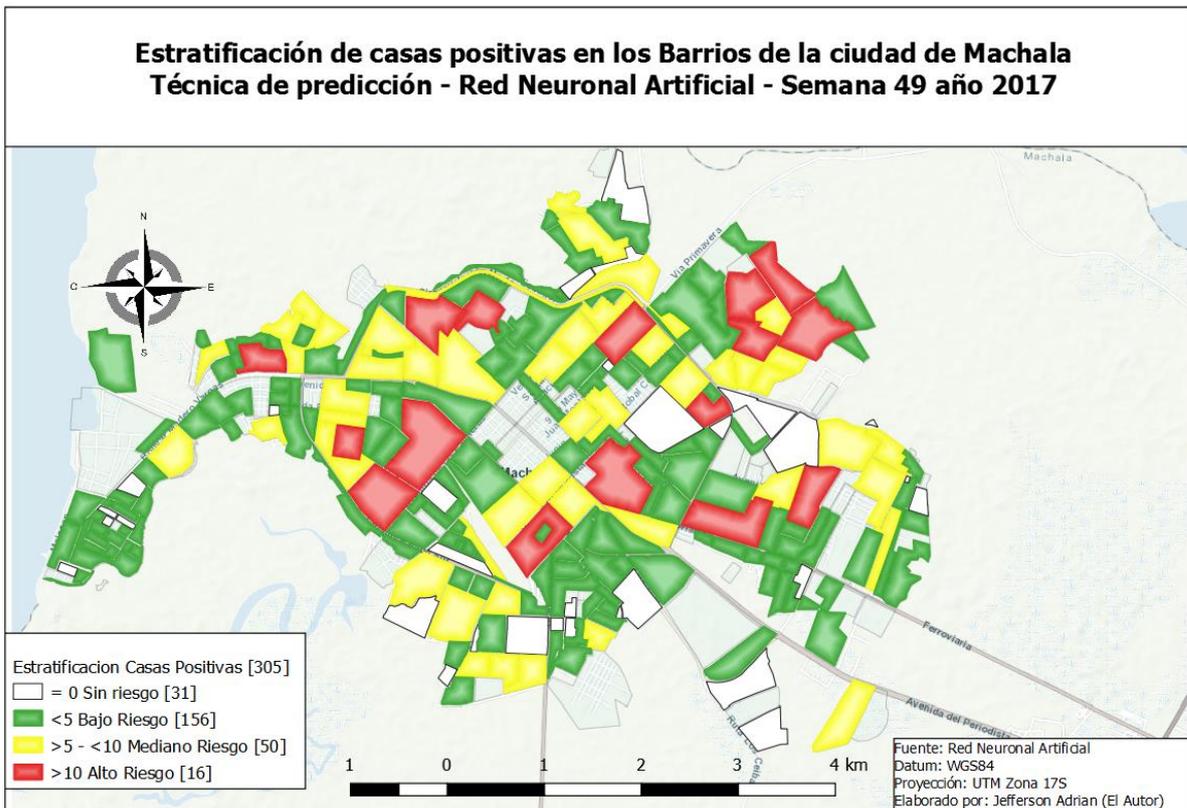
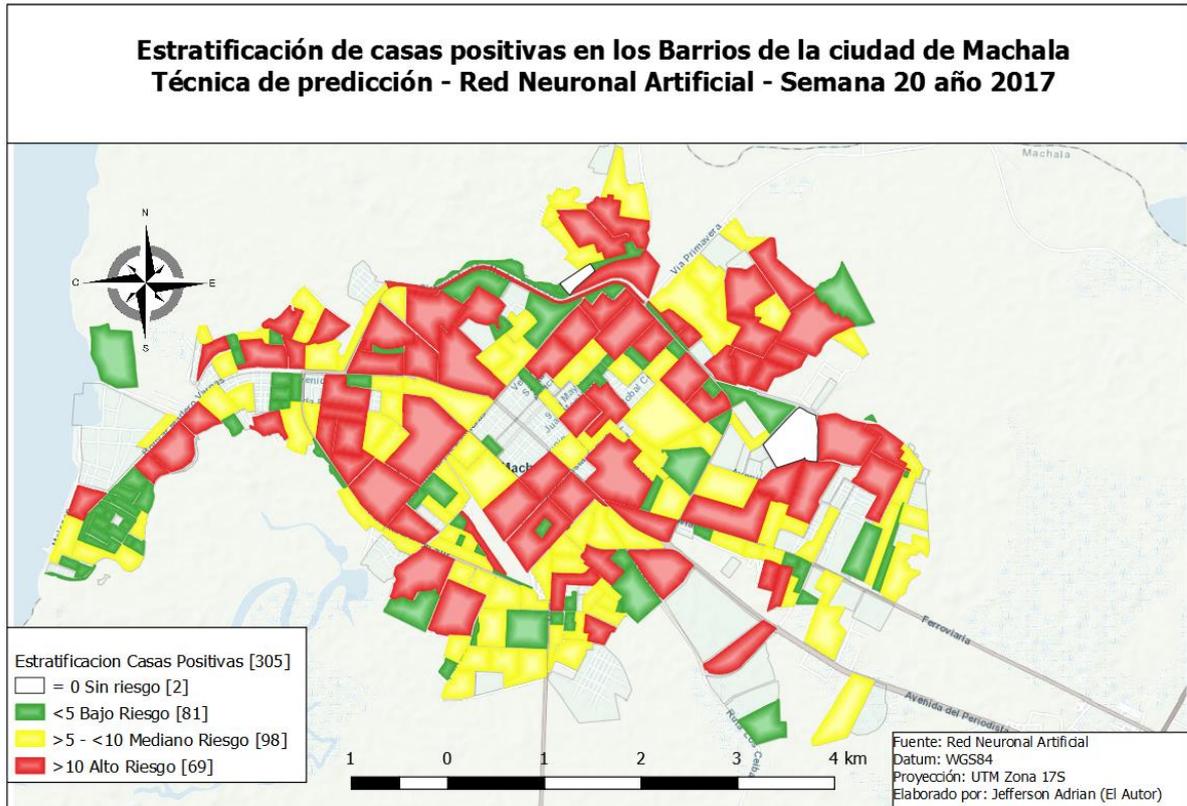
% Map Minimum and Maximum Input Processing Function
function y = mapminmax_apply(x,settings)
y = bsxfun(@minus,x,settings.xoffset);
y = bsxfun(@times,y,settings.gain);
y = bsxfun(@plus,y,settings.ymin);
end

% Sigmoid Symmetric Transfer Function
function a = tansig_apply(n,~)
a = 2 ./ (1 + exp(-2*n)) - 1;
end

% Map Minimum and Maximum Output Reverse-Processing Function
function x = mapminmax_reverse(y,settings)
x = bsxfun(@minus,y,settings.ymin);
x = bsxfun(@rdivide,x,settings.gain);
x = bsxfun(@plus,x,settings.xoffset);
end

```

**Anexo 8: Mapas de predicción de casas positivas correspondiente a la semana 20 y 49**  
**– Ensayo 1**



**Anexo 9: Mapas de predicción de casas positivas correspondiente a la semana 20 y 49  
– Ensayo 3**

