

## Machine learning aplicado al análisis de un set de datos de parámetros ambientales en galpones de pollos de engorde

### Machine learning applied to the analysis of a dataset of environmental parameters in poultry farm

*Gloris Batista-Mendoza<sup>1</sup>, Edwin Juvenal Cedeño Herrera<sup>2</sup>, Gloris Cedeño-Batista<sup>3</sup>*

<sup>1</sup>Universidad de Panama, Facultad de Informática, Electrónica y Comunicación, Panamá; [gloris.batista@up.ac.pa](mailto:gloris.batista@up.ac.pa); <https://orcid.org/0000-0002-3294-6170>

<sup>2</sup>Universidad de Panama, Facultad de Informática, Electrónica y Comunicación, Panamá; [edwin.cedenoh@up.ac.pa](mailto:edwin.cedenoh@up.ac.pa); <https://orcid.org/0000-0002-5841-2759>

<sup>3</sup>Universidad de Panama, Facultad de Informática, Electrónica y Comunicación, Panamá; [gloris.cedeno@up.ac.pa](mailto:gloris.cedeno@up.ac.pa); <https://orcid.org/0000-0003-2533-0252>

*Fecha de recepción: 12 de agosto de 2023*

*Fecha de aceptación: 1 de noviembre de 2023*

DOI <https://doi.org/10.48204/j.vian.v7n2.a4566>

**Resumen:** Este artículo trata sobre el desarrollo de un modelo de Machine Learning aplicado a un conjunto de datos recopilados en una granja avícola. Su objetivo es obtener un modelo predictivo basado en variables ambientales para anticipar eventos futuros. Este modelo predictivo busca optimizar decisiones relacionadas con el bienestar ambiental de las aves y reducir costos de producción en proyectos avícolas. Para esta investigación, se obtuvo el set de datos del sistema "Smart Poultry Farm" como guía del desarrollo, se utilizó la metodología SEMMA y el lenguaje de programación Python en el entorno IDE Google Colaboratory. El modelo se construyó mediante el algoritmo de regresión logística binomial en el contexto de aprendizaje supervisado. La evaluación del modelo predictivo incluyó la matriz de confusión y métricas como: el Índice General de Calidad, Precisión, Sensibilidad, Especificidad y F1-Score. Se aplicaron distintos escenarios para anticipar la activación/desactivación de los ventiladores del galpón, en función de los parámetros ambientales: humedad, temperatura e índice de calor.

**Palabras clave:** Agromática, análisis de datos, aprendizaje supervisado, inteligencia artificial, ciencia de datos, regresión logística.

**Abstract:** This article addresses the development of a Machine Learning model applied to a dataset collected on a poultry farm. Its goal is to attain a predictive model based on environmental variables to anticipate forthcoming events. This predictive model aims to optimize decisions linked to the birds' environmental well-being and cut production costs in poultry projects. This investigation obtained information from the "Smart Poultry Farm" system, following the SEMMA methodology and utilizing the Python programming language in the Google Colaboratory IDE environment. The model was built using the binomial logistic regression algorithm in the context of supervised learning. The assessment of the predictive model encompassed the confusion matrix and metrics such as the Overall Quality Index, Accuracy, Sensitivity, Specificity, and F1-Score.

Various scenarios were employed to forecast the activation/deactivation of the poultry farm's fans, based on environmental parameters: humidity, temperature, and heat index.

**Keywords:** Agromatic, artificial intelligence, data analysis, data science, logistic regression, supervised learning.

## 1. Introducción

Hoy en día el cambio climático es una prioridad fundamental para la humanidad y los sectores, tanto económicos como sociales, quienes deben tomar medidas preventivas para que no se afecte a las distintas actividades del ser humano en los próximos años.

El Banco Interamericano de Desarrollo (BID, 2022) mencionó en su último informe mensual que para el 2030, el cambio climático puede, anualmente, empujar a 3 millones de personas a la pobreza extrema en América Latina y el Caribe. De igual forma, Racquel Moses, directora ejecutiva del Acelerador del Cambio Climático del Caribe (CCSA) en una reunión anual del BID explicó toda esta problemática y la necesidad de buscar soluciones para crear nuevas hojas de ruta climáticas en la región. Ella, junto a otros líderes intentan proponer una ambiciosa visión y convertir el cambio climático de un riesgo en una oportunidad (The Caribbean Climate-Smart Accelerator, 2022).

Dentro de la misma línea, el sector agrícola no escapa de esta realidad y debe tener en cuenta que los procesos productivos de sus proyectos están, principalmente, condicionados al comportamiento de las condiciones atmosféricas (Wicaksono et al., 2018). Además, esta industria necesita utilizar menos recursos para producir la misma cantidad de alimentos, reduciendo la cantidad de agua usada y disminuyendo el uso de productos químicos. Es por ello, que este sector requiere optimizar sus proyectos, obtener beneficios y rentabilidad, e incorporar tecnologías emergentes, mientras que el público desea alimentos más nutritivos (Ayoub Shaikh et al., 2022).

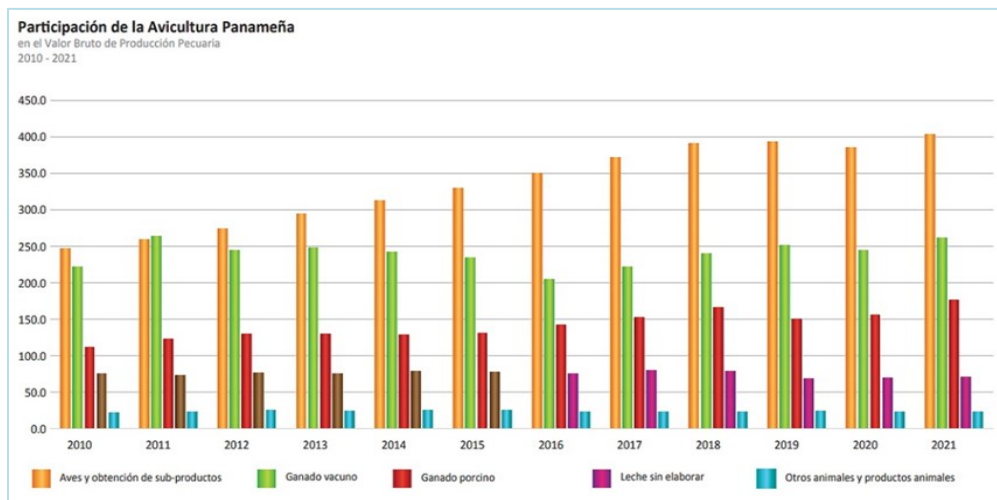
En Panamá, como en el resto del mundo, el sector avícola tiene un importante crecimiento, esto se debe a la relación costo-beneficio del rubro, la población prefiere el consumo de carne de pollo, gracias a la fuente de nutrientes que contiene, como proteínas, lípidos, vitaminas y minerales como calcio, hierro, zinc, sodio, potasio y magnesio, entre otros (*Consejo Mexicano de la Carne*, s. f.); además, de ser alimentos accesibles convirtiéndose en parte importante de la dieta de los ciudadanos.

La propuesta de esta investigación está enfocada en el sector de producción avícola de engorde en granjas con galeras en abierto, y tiene como objetivo descubrir información valiosa a través de tecnologías emergentes para ayudar a los productores avícolas a tomar decisiones correctas utilizando soluciones inteligentes en sus proyectos, pero para alcanzar los rendimientos óptimos es imprescindible gestionar un entorno adecuado que les proporcione a los pollos condiciones ambientales ideales para mejorar la productividad (Estrada et al., 2007). Los galpones de ambiente controlado (cerrados) constituyen una modalidad que recomiendan las guías de los proveedores de pollos a nivel global como Cobb (*Cobb, s. f.*) y Arbor Acres (*Aviagen, s. f.*). Sin embargo, de acuerdo al análisis de campo realizado por los investigadores, en Panamá existe un gran número de galpones operando la modalidad en abierto, debido a los altos costos de la modalidad de ambientes controlados o cerrados. Además, la mayoría de las granjas avícolas en abierto son, manualmente, operadas y monitoreadas de forma empírica por operarios, típicamente, sin formación. Aunada a la situación, no cuentan con soporte de equipos para mención de variables que influyen en los procesos productivos de los pollos.

En la **Figura 1**, se puede observar el comportamiento de la producción en Panamá, que realiza todos los años en concepto de producción pecuaria, siendo las aves y subproductos los de mayor aporte (*ANAVIP - La avicultura en Panamá, s. f.*).

**Figura 1**

*Participación de Avicultura Panameña en Producción Pecuaria*



Fuente: ANAVIP - La avicultura en Panamá, (s. f.)

Ciertamente, el auge de las tecnologías emergentes ha tenido un impacto relevante en la manera en que los individuos administran el conocimiento, lo que genera un cambio de rol en las personas y en el desempeño diario de la vida. Al mismo tiempo, es la responsable de la creación de grandes volúmenes de datos e información. Sin embargo, su gestión no está exento a desafíos, como es el caso del manejo y procesamiento efectivo de enormes cantidades de información, con el objetivo de maximizar los beneficios y minimizar los riesgos (Porcelli, 2020). Estos desafíos adquieren una relevancia especial en el ámbito del sector avícola, el cual constituye el foco de este estudio.

El Machine Learning se basa en algoritmos que permiten el descubrimiento de patrones ocultos en los datos, lo que ayuda a predecir eventos futuros y maximiza la toma de decisiones informadas (Vuppalapati, 2021). Además, tiene un impacto positivo en la economía y en el medio ambiente debido a que permite la creación de ecosistemas limpios e inteligentes. El propósito de esta investigación radica en la construcción de un modelo de Machine Learning basado en un conjunto de datos extraídos del sistema “Smart Poultry Farm” ubicado en una granja avícola de la ciudad de Santiago de Veraguas dedicada a la cría de pollos de engorde, con el fin de realizar predicciones asertivas basadas en los parámetros ambientales (temperatura, humedad relativa e índice de calor) de forma tal que se pueda conocer con antelación los posibles eventos dentro del galpón, contribuyendo así a la optimización de los procesos avícolas en la región.

En esta investigación se incluyen los estudios precedentes, y que representan el soporte teórico para su desarrollo, desde los relacionados a los modelos de Machine Learning, las técnicas aplicadas, algoritmos más adecuados, métodos de evaluación, a los referentes a las aplicaciones de Machine Learning en parámetros ambientales del sector agrícola y siendo parte de este, el sector avícola. A continuación, se presentan algunos ejemplos destacados de dichos estudios:

El trabajo de investigación ***“Data analytics for crop management: a big data view”*** presenta una revisión sistemática del análisis de los datos en la agricultura digital (DA), limitando el estudio al rendimiento y seguimiento de los cultivos. Los investigadores sostienen que la agricultura digital (DA) es un enfoque basado en datos que explota la

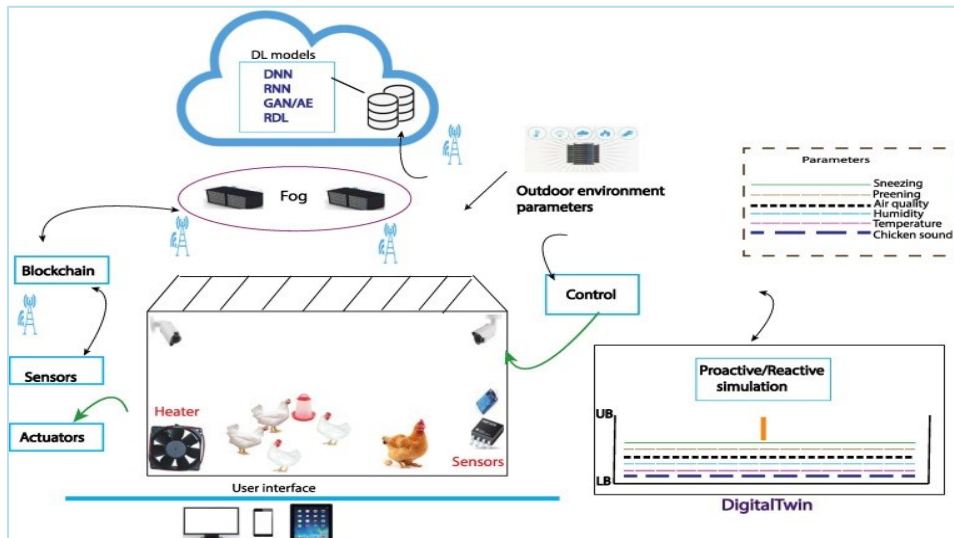
información oculta dentro de los datos recopilados para obtener nuevo conocimiento y transformar las prácticas agrícolas de una toma de decisiones intuitiva a una informada, y que de ella surgen tres dimensiones, siendo la primera dimensión la relacionada al uso de tecnologías avanzadas para mejorar las prácticas y la productividad, la segunda dimensión se refiere a los factores legales, éticos, sociales y económicos de DA, y la tercera dimensión se centra en la aplicación del análisis de big data y el machine learning (ML), para optimizar y pronosticar la producción y el uso de recursos. Esta última dimensión es el objeto de este estudio, considerando que es uno de los pilares de la agricultura debido a que con su utilización permite optimizar el coste operativo y reducir el impacto sobre el medio ambiente. Los autores presentan una metodología basada en datos derivada del proceso general de descubrimiento del conocimiento en la que establece tres fases: recopilación de datos, representación y análisis y la toma de decisiones. Los autores definieron métricas de big data (baja, media, alta) para cada una de sus dimensiones (volumen, velocidad y variedad), para medir el nivel de uso de big data en DA. Sin embargo, concluyeron que el análisis de big data en la agricultura aún se encuentra en su etapa inicial y es necesario superar muchas barreras, a pesar de la disponibilidad de datos y herramientas para analizarlos (Chergui & Kechadi, 2022).

Por su parte, la investigación titulada ***“Internet of Things and Machine Learning techniques in poultry health and welfare management: A systematic literature review”*** proporciona una revisión literaria sobre la integración de las aplicaciones IoT y el machine learning para la gestión de la salud y el bienestar de las aves de corral, brindando así una avicultura inteligente, monitoreo continuo de datos y los análisis prescriptivos con el fin de lograr un control eficiente de recursos, minimizando costos, y llevando a la toma de decisiones óptima. Esta revisión se fundamenta en 93 artículos para el estudio, publicados desde enero de 2010 hasta junio de 2022, siendo China el país con mayor número de investigaciones, seguida de Estados Unidos, Brasil y Bélgica. Entre los aportes de este estudio se obtiene una visión holística de la tecnología digital para aplicaciones de salud y bienestar avícola, técnicas de IA/ML, herramientas, datos monitoreados, tipos de datos, tipos de procesamiento, enfermedades y estado de salud. Este estudio también ofrece una

categorización de investigaciones de aplicaciones de IA/ML completo, donde se resalta que los métodos de regresión que se aplican, principalmente, en el monitoreo de aves de corral que incluye regresión lineal, logística, de lazo, SVM, ANN, RNN, DNN y de modelo de árbol, además de permitir predicciones de una variable de salida, según variables de entrada conocidas. Otro aporte del estudio es el marco inteligente de gestión de la salud y el bienestar de las aves de corral, como se puede observar en la **Figura 2**.

**Figura 2**

*Marco inteligente de gestión de la salud y el bienestar de las aves de corral*



Fuente: Ojo et al., (2022)

Esta investigación ofrece un compendio a otros investigadores interesados en comprender y aprovechar mejor las tecnologías avanzadas, permitiendo nuevas líneas de razonamiento que mejorarán la productividad y la rentabilidad en la industria avícola (Ojo et al., 2022).

Con respecto a la investigación **“Developing Edge AI Computer Vision for Smart Poultry Farms Using Deep Learning and HPC”**, los investigadores utilizaron una plataforma agrícola de IoT existente con una red de sensores y nodos para obtener los datos y crear los modelos de predicción de imágenes procesadas y recopiladas en granjas avícolas utilizando Faster R-CNN para la detección de pollos, y AutoML para identificar la arquitectura más adecuada para la detección y segmentación de pollos para el conjunto de datos

determinado. El desarrollo de los modelos se realizó utilizando sistemas HPC/AI fuera de línea para desarrollar modelos de predicción de aprendizaje profundo (Deep Learning) y entrenar los modelos. Los datos de entrada fueron imágenes de cámaras instaladas en las granjas avícolas, aunque las imágenes provenían de bucles en línea y fuera de línea. Para el entrenamiento de los datos, los investigadores utilizaron un aproximado de 4000 imágenes y se amplió con la herramienta RoboFlow a 9000, de las cuales se usaron para el entrenamiento 7550, para la validación 725 y para las pruebas del modelo 725. También, para el desarrollo de modelos se utilizó el lenguaje de programación Python y el paquete Detectron2 que incluye variantes de redes neuronales profundas (Cakic et al., 2023).

El estudio “***Generating action plans for poultry management using artificial neural networks***” demuestra que es posible predecir planes de acción aceptables basándose en una red de sensores que registran la gestión avícola y son determinantes como la humedad, temperatura, cantidad de ración de alimento, ventilación entre otros, luego se preprocesa mediante técnicas de Machine Learning, para el mismo se utilizó el modelo de regresión múltiple de ANN (Artificial Neural Network). Los investigadores sostienen que los grandes conjuntos (bigdata) de parámetros ambientales se combinan permitiendo toma de decisiones acertada en contra de la percepción humana que tiende a ser limitada. La validación del estudio fue basada en una comparación entre los planes de acción realizados por un experto humano (para el caso, un avicultor) contra los resultados obtenidos del modelo (Ribeiro et al., 2019).

En el artículo “***Quantile regression forests-based modeling and environmental indicators for decision support in broiler farming***”, los investigadores sostienen que para asegurar el bienestar de los animales se requiere el ajuste y control de los parámetros ambientales y de una interacción dinámica de buenas prácticas de gestión desarrollando así, un DSS. El estudio se obtuvo de un sistema cuya arquitectura consta de una serie de sensores instalados a lo largo de la galera para medir y recopilar, automáticamente, diferentes condiciones ambientales, un almacenamiento de dicha información basado en la nube y un conjunto de modelos basados en Machine Learning aprendidos a partir de datos históricos para gestionar la producción de manera inteligente y en línea. Los dispositivos de

recopilación de datos miden los parámetros ambientales (temperatura y humedad relativa) cada 15 minutos desde la llegada de las aves a la granja hasta el final de cada producción. También, este estudio consideró los pesos durante el período de crecimiento, los problemas relativos a las piernas y la mortalidad acumulada. El conjunto de datos estuvo compuesto por 20 parvadas de distintas granjas avícolas, para estimar el rendimiento de generalización se aplicó la estrategia LOOCV (validación cruzada de dejar uno fuera) del enfoque de modelado basado en bosques de regresión cuantil (Diez-Olivan et al., 2019).

Cada uno de estos proyectos de investigación proporcionan una variedad de enfoques y orientaciones respecto al uso del Machine Learning. Sin embargo, todos ellos brindan una guía fundamental para el enfoque que se pretende desarrollar. El objetivo de este estudio se centra en identificar y/o reconocer patrones relacionados con los parámetros ambientales de una granja avícola dedicada a la cría de pollos de engorde, y generar predicciones que contribuyan al avance en el sector avícola. Ante esta situación, y a las necesidades actuales surge la pregunta de investigación, ¿De qué manera el desarrollo de un modelo de Machine Learning puede predecir sobre los ventiladores (actuadores), basados en parámetros ambientales propios de un galpón en abierto, permitiendo atenuar el impacto económico de los proyectos avícolas?

## **2. Materiales y métodos**

Según Hernández-Sampieri & Mendoza Torres, (2018), la ruta cuantitativa es aplicable para fines de desarrollo de tecnología, por lo que este estudio tiene un enfoque cuantitativo, además, este enfoque utiliza la recolección y el análisis de los datos para responder a la pregunta de investigación.

Con respecto al análisis de los datos será predictivo, ya que, según su definición, es el proceso de extraer información de los datos y predecir resultados y tendencias futuras. Esta investigación busca predecir el comportamiento (activación o desactivación) de los ventiladores (actuador) a través del desarrollo de un modelo de Machine Learning basado en un bigdata de parámetros ambientales (temperatura, humedad relativa e índice de calor) dentro de una galera de pollos en abierto, permitiendo de esta manera mejoras en el



bienestar de las aves dentro del galpón y, por consiguiente, en la rentabilidad del proyecto avícola.

El tipo de muestra es probabilística debido a que según Hernández-Sampieri & Mendoza Torres, (2018), los elementos de la población tienen al inicio la misma posibilidad de ser escogidos para conformar la muestra y su principal objetivo es reducir al mínimo el error estándar. Siendo el dominio de los científicos de datos, el Machine Learning (ML) un subconjunto de la Inteligencia Artificial que utiliza técnicas estadísticas para permitir que las máquinas mejoren con la experiencia, detectando a través de sus algoritmos, patrones y aprendiendo a hacer predicciones y recomendaciones mediante el procesamiento de datos y experiencias, en lugar de recibir instrucciones de programación explícitas (Vuppalapati, 2021).

El diseño de la investigación es de carácter no experimental, debido a que no existe manipulación deliberada de las variables, solo se observan tal cual se encuentran en su entorno natural, para luego ser analizadas. Además, ya han sucedido en el tiempo (Hernández-Sampieri & Mendoza Torres, 2018).

El tipo de la investigación tiene una connotación transversal causal y se debe a que los datos se recolectaron en un tiempo y lugar determinado, para el caso de estudio, el rango de tiempo va desde 7 de mayo de 2022 hasta el 26 de septiembre de 2022 y el lugar de la extracción de los datos es del sistema “**Smart Poultry Farm**” instalado en una galera de pollos de engorde ubicado en Santiago de Veraguas, Panamá. Igualmente, se busca establecer una relación causa-efecto entre las variables independientes y la dependiente, siendo las variables independientes la temperatura, humedad relativa e índice de calor y la variable dependiente es la activación o desactivación de los ventiladores (actuador).

El sistema “**Smart Poultry Farm**” es un sistema inteligente conformado por hardware y software integrado, y basado en una arquitectura de IoT (internet de las cosas), para optimizar la eficiencia de los proyectos de pollos de engorde, permitiendo ahorros energéticos, mejoras en el factor de conversión, protección eléctrica a los equipos, alarmas configurables en situaciones de emergencia, como interrupciones de fluido eléctrico por periodos prolongados, mediante SMS o e-mails. Además, ofrece una interfaz gráfica vía

web y una App para dispositivos con recursos limitados, para el monitoreo del estado de los equipos y las variables ambientales como temperatura, humedad relativa, índice de calor en los galpones, proporcionando la visualización de los eventos en el proyecto. En la **Figura 3** se presenta el escenario general, para una mejor comprensión del estudio.

**Figura 3**

*Escenario General del Sistema "Smart Poultry Farm"*

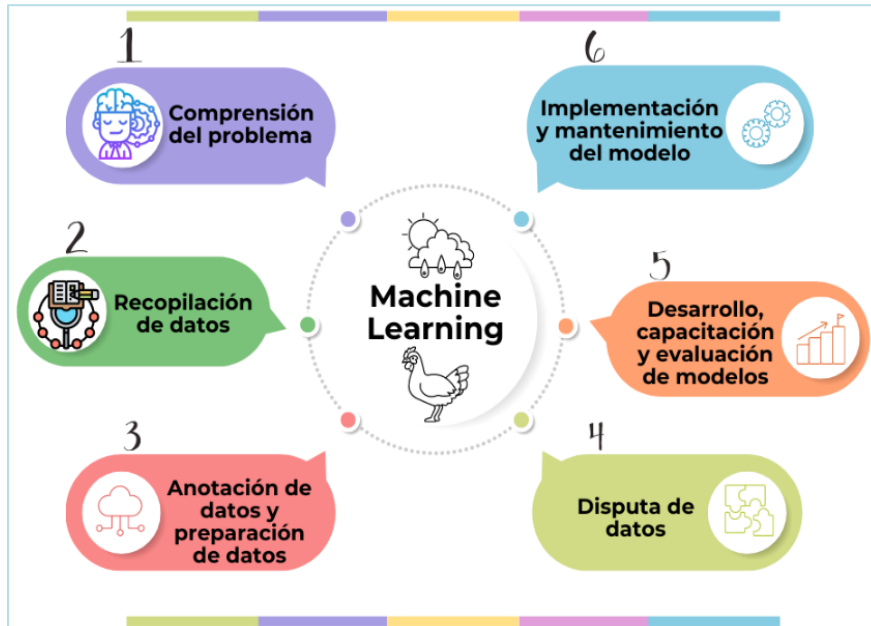


La unidad de análisis de la investigación es la muestra del conjunto de parámetros ambientales (temperatura, humedad relativa e índice de calor), para el estudio se escoge la muestra de 15,344 registros, extraídos como se mencionó, anteriormente, del sistema “Smart Poultry Farm”.

El estudio aplicará el proceso de Machine Learning extremo a extremo, y que según Pajankar & Joshi, (2022) se reduce a seis etapas descritas en el siguiente esquema (**Figura 4**) y, por consiguiente, generando la arquitectura propuesta incluido el ciclo de vida del Machine Learning propuesto como se puede observar en la **Figura 5**.

**Figura 4**

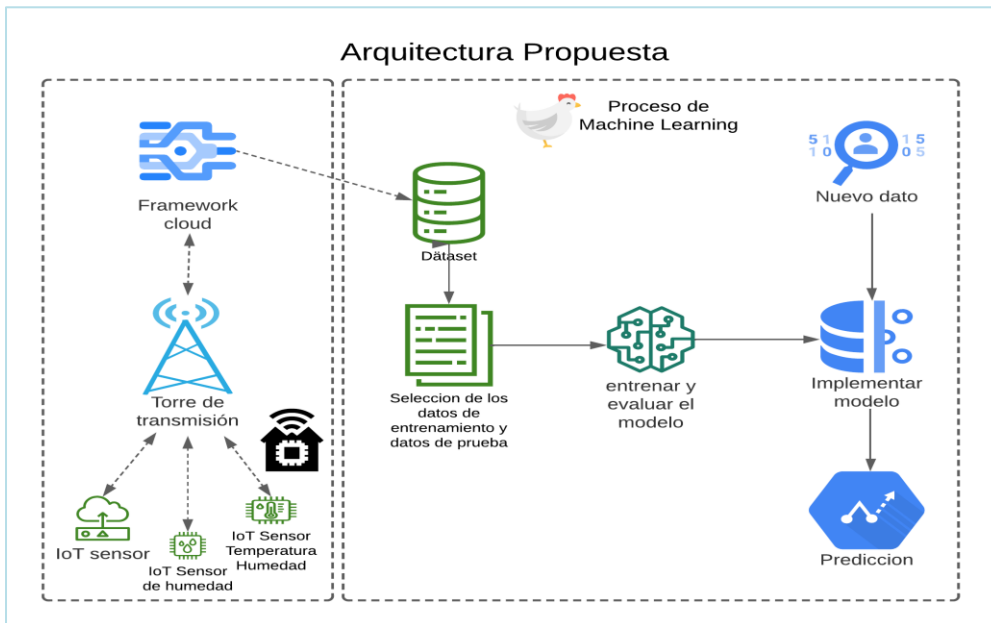
*Etapas del Machine Learning*



Fuente: Pajankar & Joshi, (2022)

**Figura 5**

*Arquitectura propuesta junto al ciclo de vida de Machine Learning*



Esta investigación se limita al tiempo y lugar de los datos recopilados por el Sistema **“Smart Poultry Farm”** como el dataset sujeto a estudio, para aplicar el Machine Learning, por tal motivo, no se profundiza en la arquitectura *per sé* del sistema de *hardware* y *software* que lo conforma, llámese tipos de sensores, cantidad de sensores, tipos de actuadores, cantidad de actuadores, arquitectura IoT, entre otras estructuras propias del sistema inteligente.

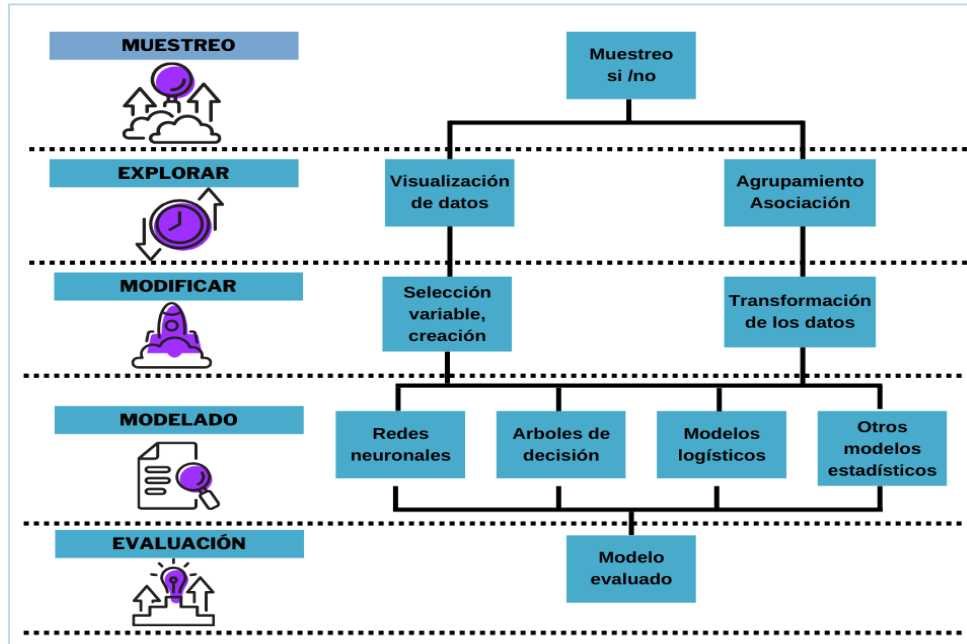
Para guiar el desarrollo de la investigación se utilizó la metodología *SEMMA*. El Instituto SAS, desarrollador de esta metodología, la define como “el proceso de seleccionar, explorar y modelar grandes cantidades de datos para descubrir patrones de negocio desconocidos” (SAS Help Center: Introduction to SEMMA, s. f.).

Para describir las fases de esta metodología se toma como referencia el artículo **“Applied Machine Learning in Social Sciences: Neural Networks and Crime Prediction”** (Reier Forradellas et al., 2020): (1) **Sample** (muestreo): En esta fase se obtiene la muestra representativa del problema de estudio. Se puede realizar la muestra creando una o más tablas de datos. Además, deben ser bastante grandes para contener la información significativa, pero pequeñas para que permitan ser procesadas. (2) **Explore** (explorar): Luego de obtener la muestra, hay que explorar la información obtenida para ubicar relaciones, tendencias o anomalías para simplificar dichos datos y optimizar el modelo a crear. (3) **Modify** (modificar): En esta fase se modifican los datos de manera que las variables sean creadas, seleccionadas e incluso sufran pequeñas transformaciones enfocándolas al proceso de selección del modelo. (4) **Model** (modelado): El propósito de esta fase es fijar una relación entre las variables explicativas y las variables objeto de estudio, permitiendo discurrir el valor de estas con un determinado nivel de confianza. En este punto se pueden utilizar técnicas que incluyen métodos estadísticos tradicionales o técnicas basadas en datos como por ejemplo las redes neuronales, árboles de decisión, entre otras. (5) **Assess** (Evaluación): Esta fase evalúa el funcionamiento del proceso de ML considerando la utilidad y confiabilidad. Estos resultados se evalúan mediante el análisis de la bondad del modelo o modelos, contrastados con otros métodos estadísticos o con nuevas poblaciones

muestrales. A continuación, en la **Figura 6**, se presentan las fases mencionadas, anteriormente.

**Figura 6**

*Fases de la metodología SEMMA*



Fuente: SAS Help Center: Introduction to SEMMA, (s. f.)

El aprendizaje supervisado es el tipo de aprendizaje seleccionado, debido a que se basa en entrenar a un modelo por medio de diferentes datos, para predecir una variable partiendo de estos mismos conjuntos de datos (Larose & Larose, 2014), es decir, transforma los datos o descubre patrones sin la necesidad de un conjunto de etiquetas de capacitación explícitas (Pajankar & Joshi, 2022). Se utilizó el algoritmo de regresión logística binomial, porque es un algoritmo de clasificación de Machine Learning que se utiliza para predecir la probabilidad de una variable dependiente categórica, es decir, que permite describir los datos y explicar la relación entre la variable dependiente cualitativa dicotómica y el resto de las variables independientes. En esta investigación, la variable predictiva y dependiente binaria es el ventilador (actuador) y las variables predictoras e independientes son: temperatura, humedad relativa e índice de calor (Vuppalapati, 2021).

Para evaluar la calidad del modelo de Machine Learning, se generará la matriz de confusión, la cual presenta cuatro cuadrantes que se interpretan de la siguiente manera: verdadero positivo (cuadrante inferior derecho), verdadero negativo (cuadrante superior izquierdo), falso negativo (cuadrante inferior izquierdo) y falso positivo (cuadrante superior derecho). Al referirse a verdadero, su interpretación es que los valores se predijeron con precisión, y falso hubo un error o una predicción incorrecta. El eje Y representa los datos reales y el eje X representa la predicción (Zheng et al., 2022). Y, con respecto a las métricas aplicadas para evaluar algoritmos de Machine Learning fueron: Índice general de calidad (Accuracy), Precisión, Exhaustividad (Recall), Especificidad (Specifity) y F1-Score (Brownlee, 2020).

Además, para el desarrollo del estudio se usó la *interfaz IDE Google Colaboratory*, ya que, es un entorno de codificación que crea un *Jupyter Notebook* para trabajar, algo similar a un cuaderno en línea que permite escribir y ejecutar código arbitrario de Python, y es recomendado para tareas de Machine Learning, análisis de datos y educación (*Google Colaboratory*, s. f.). Las librerías empleadas en *lenguaje Python* fueron: Pandas y Numpy para el manejo de altos volúmenes de datos, Matplotlib y Seaborn para representación gráfica de los datos, NumPy se utilizó para la manipulación de elementos de datos de matriz numérica, SciPy se usó para cálculos numéricos como es el caso del cálculo de la correlación de Pearson, Sklearn es una librería que unifica los principales algoritmos y funciones en un solo marco, lo que facilita, enormemente, todas las etapas de preprocesamiento, entrenamiento, optimización y verificación del modelo de predicción. La librería Scikit-learn de Python se utilizó para generar las métricas de evaluación. La misma implementa todas las métricas: `metrics.confusion_matrix`, `metrics.accuracy_score`, `metrics.precision_score`, `metrics.recall_score`, y `metrics.f1_score`.

### **3. Resultados**

En esta sección se presentan los resultados obtenidos del desarrollo del modelo de Machine Learning guiado por la metodología SEMMA. En la primera fase de la metodología, denominada *muestreo*, se realiza un análisis de la información inicial obtenida sobre los

datos, y se determina el volumen apropiado para la construcción del modelo, seleccionando el rango de fecha que va desde el 7 de mayo de 2022 hasta el 26 de septiembre de 2022, recogiendo un total de 15,344 registros. El detalle de la información extraída del set de datos (bigdata) se resume en la **Tabla 1** y que influyen en el proceso y clasificación para las categorías de preparación y validación.

**Tabla 1**

*Conjunto de datos previo a la selección de la muestra*

Campo	Tipo de Dato	Descripción
ISO Time Stamp	Date	Fecha y hora
Temperatura	Numérico	Índice de temperatura
Humedad	Numérico	Índice de humedad
Indexh	Numérico	Índice de calor
Fan	Binario	Indica el encendido o apagado de los ventiladores
Nebulizador	Binario	Indica el encendido o apagado de los nebulizadores

Seguidamente, en la fase de *modificación* de la metodología SEMMA, se limpian los datos, para que pasen a la fase de modelado. Para el caso de esta investigación, se procedió a eliminar atributos dentro del conjunto de datos y comprobar que no existan datos nulos, quedando como muestra la siguiente **Tabla 2**.

**Tabla 2**

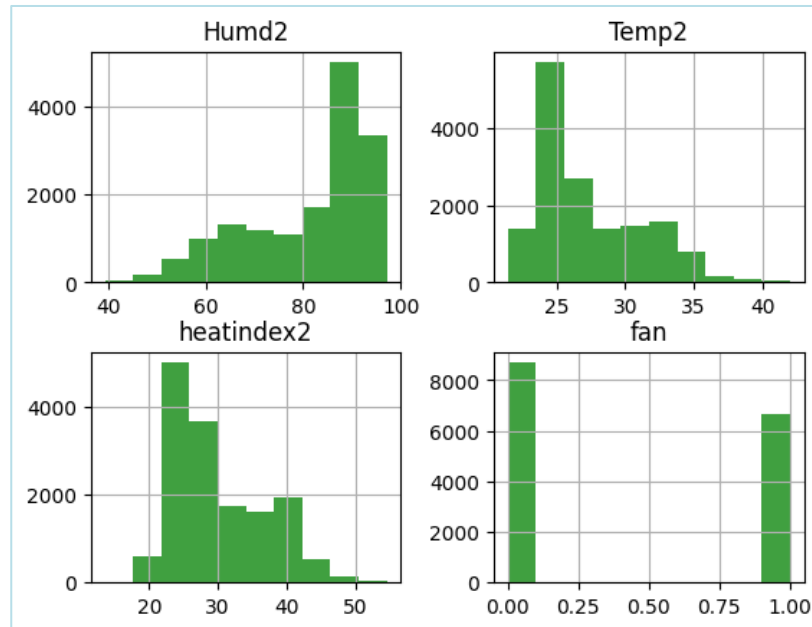
*Conjunto de datos apropiado para la selección del modelo*

Campo	Tipo de Dato	Descripción
<b>Temp</b>	Numérico	Índice de temperatura
<b>Humd</b>	Numérico	Índice de humedad
<b>Heatindex</b>	Numérico	Índice de calor
<b>Fan</b>	Binario	Indica el encendido o apagado de los ventiladores

En la **Figura 7** se muestra el conjunto de datos apropiado para la selección del modelo, sus mínimos y máximos, y en qué intervalos se concentra la mayoría de los registros.

**Figura 7**

*Visualización general de los datos de entrada*



Para la fase de *modelado* de la metodología SEMMA, se fijaron los parámetros ambientales y su relación de forma que permita el nivel de confianza. En este punto se utilizan técnicas que incluyen métodos estadísticos tradicionales o técnicas basadas en datos para comprobar, cuantitativamente, dicha relación, como se detalla a continuación:

- **Análisis estadístico de los parámetros ambientales**

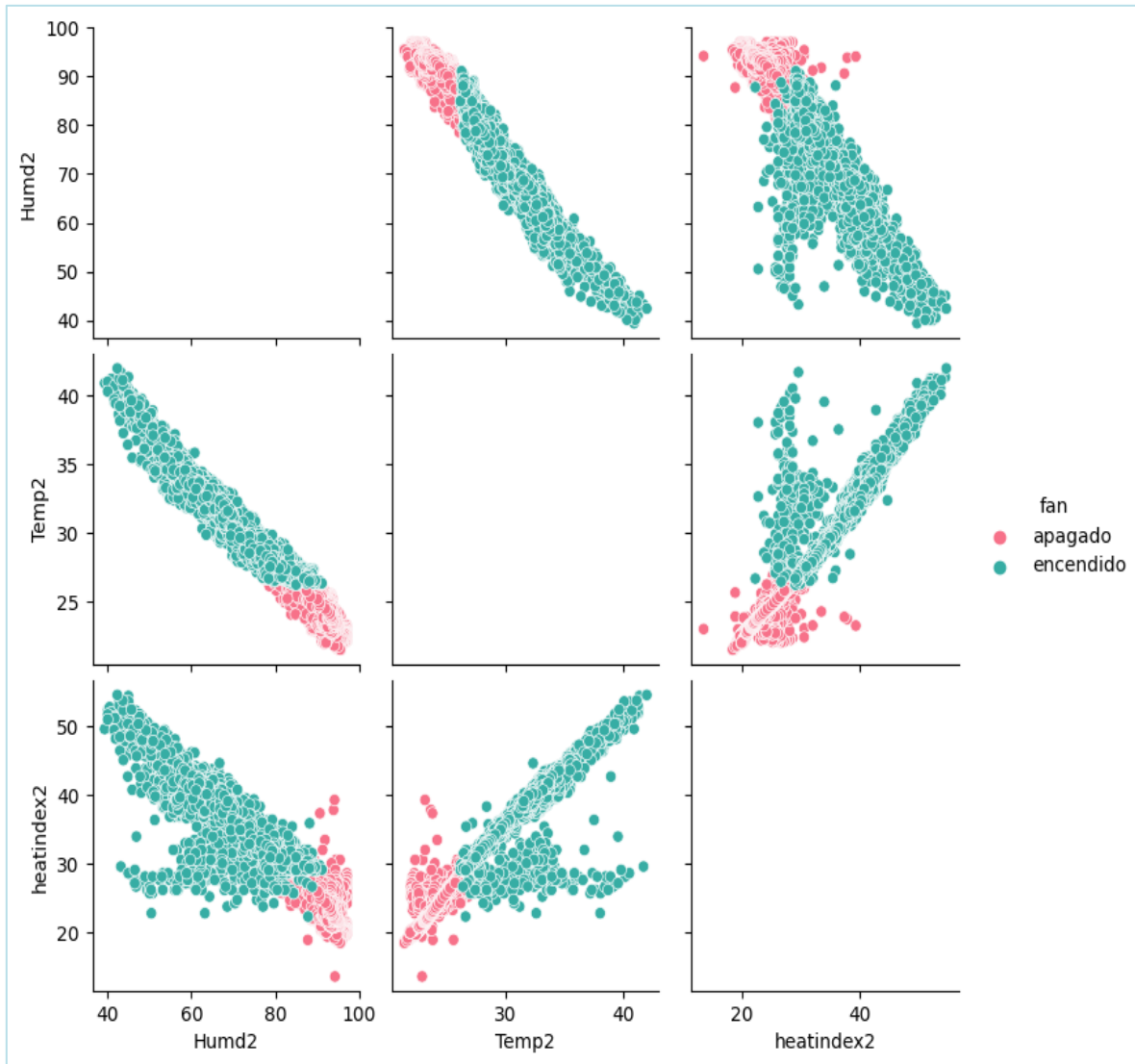
Del análisis exploratorio preliminar realizado de los datos a través del uso de herramientas estadísticas se obtuvo una mejor imagen de lo que representan los datos.

En la **Figura 8** ~~Error! No se encuentra el origen de la referencia.~~ se observa la interrelación de los datos usando gráfico de pares y cómo se concentran, linealmente, las salidas por colores: encendido del ventilador (actuador) en verde, y apagado en color rojo en relación con el comportamiento de las variables ambientales: temperatura, humedad relativa e índice de calor.



**Figura 8**

*Gráfico de pares entre el ventilador (actuador) y las variables humedad, temperatura e índice de calor*



- **Selección de datos de entrenamiento y datos de prueba o test**

Para determinar el *modelado* es necesario entrenar un grupo de datos que serán los que encontrarán un patrón de comportamiento determinado, y con los datos de prueba (test) se podrá predecir nuevos casos (vea **Figura 5**).

Las variables predictoras son: Temperatura, Humedad e Índice de Calor, como se puede observar en la **Tabla 3**.

**Tabla 3**

*Variables Predictoras del modelo*

<b>Campo</b>	<b>Descripción</b>
Temp	Índice de temperatura
Humd	Índice de humedad
Heatindex	Índice de calor

En la **Tabla 4**, se observa la variable predictiva, la cual es el encendido o apagado del ventilador (actuador).

**Tabla 4**

*Variable Predictiva del modelo*

<b>Campo</b>	<b>Descripción</b>
Fan	Indica el encendido o apagado de los ventiladores

En la **Tabla 5** se muestran los datos de entrenamiento que corresponden a 12,275 registros del conjunto de datos ambientales, equivalentes al 80% del total de los datos de muestra.

**Tabla 5**

*Datos para entrenamiento del modelo*

	<b>Humd</b>	<b>Temp</b>	<b>Heatindex</b>	<b>Fan</b>
6856	90.05	24.19	24.05	0
8057	92.6	23.45	22.34	0
13137	89.07	25.55	26.29	0
13223	96.53	22.53	20.1	0
...	...	...	...	...
12725	59.99	33.72	42.35	1
8333	90.7	24.58	24.78	0
11698	90.35	22.96	21.77	0
2910	91.2	25.7	27.46	0

*Nota.* 12275 filas × 4 columnas.

Para los datos de prueba se obtuvieron 3,069 registros como se puede observar en la **Tabla 6**, y equivale al 20% del total de los datos de muestra.

**Tabla 6**

*Datos para prueba del modelo*

	<i>Humd</i>	<i>Temp</i>	<i>Heatindex</i>	<i>Fan</i>
0	91.14	24.23	24.02	0
13	92.62	24.06	23.55	0
21	93.74	23.89	23.1	0
25	94.16	23.89	23.06	0
...	...	...	...	...
15308	70.06	31.11	37.6	1
15321	69.74	30.3	36.76	1
15328	69.36	29.8	34.86	1
15341	78.82	28.08	32.28	1

*Nota.* 3069 filas × 4 columnas.

- **Entrenamiento del modelo**

El *entrenamiento del modelo* se realizó por medio del algoritmo de regresión logística binomial, y se ajustó el modelo a los datos de entrenamiento. Se aplicaron distintos escenarios a predecir, es decir, se le entregó al modelo entrenado, los datos de prueba para comprobar si el modelo era capaz de predecir sobre la decisión de encendido o apagado del ventilador (actuador) con respecto a los parámetros ambientales (temperatura, humedad e índice de calor). En la siguiente

**Tabla 7**, se muestran las predicciones generadas producto de los distintos escenarios basados en los nuevos datos.

**Tabla 7**

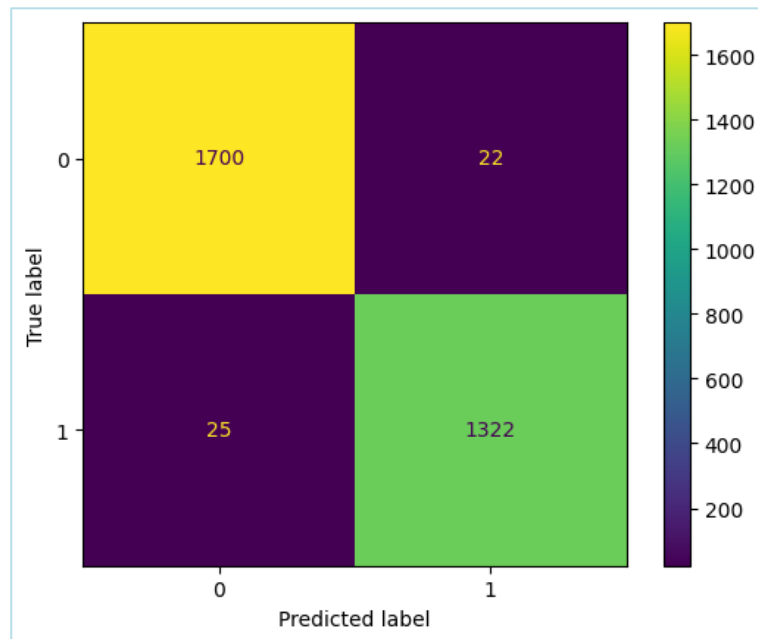
*Escenarios de prueba con las predicciones generadas*

	<i>Humd</i>	<i>Temp</i>	<i>Heatindex</i>	<i>Predicción</i>
Escenario 1	91.58	24.26	24.07	0
Escenario 2	73.15	29.72	35.04	1
Escenario 3	53.01	37.80	28.12	1
Escenario 4	82.06	25.46	27.93	0
Escenario 5	91.83	24.40	24.32	0
Escenario 6	59.19	34.44	43.42	1

La última etapa de la metodología SEMMA, se evalúa la *calidad del modelo* de Machine Learning. Esto se logra generando la matriz de confusión del modelo entrenado (**Figura 9**), y su respectivo análisis.

**Figura 9**

*Matriz de confusión*



Los resultados métricos aplicados al modelo se pueden visualizar en el informe de clasificación que se muestra en la **Tabla 8**.

**Tabla 8***Informe de Clasificación*

	<i>Precisión</i>	<i>Exhaustividad (Recall)</i>	<i>Especificidad (Specifity)</i>	<i>F1-Score.</i>	<i>Soporte</i>
apagado	0.99	0.99		0.99	1722
encendido	0.98	0.98		0.98	1347
			0.98		3069
	Índice general de calidad (Accuracy)			0.98	3069
macro avg	0.98	0.98		0.98	3069
weighted avg	0.98	0.98		0.98	3069

**4. Discusión**

En el análisis exploratorio preliminar realizado a los datos a través del uso de herramientas estadísticas se resaltan tres histogramas a los cuales se les generó el *coeficiente de correlación de Pearson*, ellos son: el histograma de la correlación entre en Índice de calor y la Temperatura, el histograma de correlación entre en Índice de calor y la Humedad, y el histograma de la correlación entre la Temperatura y la Humedad (ver **Figura 8**). Para el histograma de la correlación entre en Índice de calor y la Temperatura, el coeficiente de correlación de Pearson fue de 0.9654, lo que se interpreta como una correlación positiva casi perfecta y que las variables se correlacionan, directamente. Con respecto al histograma de correlación entre en Índice de calor y la Humedad, el coeficiente de correlación de Pearson arrojó -0.9421, lo que se interpreta como una correlación negativa, es decir, que las variables se relacionan inversamente. Y para el histograma de la correlación entre en la Temperatura y la Humedad, el coeficiente de correlación de Pearson fue de -0.9822, lo que se interpreta como una correlación negativa, es decir, que las variables se relacionan inversamente.

Para la selección de datos de entrenamiento y datos de prueba o test, se determinó que el 80% del conjunto de datos son para datos de entrenamiento y para los datos de prueba el porcentaje restante (20%) del conjunto de datos, este último subconjunto es de

gran importancia, ya que, se podrá comprobar si el modelo que se ha generado de los datos de entrenamiento es funcional (Ver **Tabla 5** y **Tabla 6**).

Las pruebas aplicadas al conjunto de datos de prueba demostraron que el modelo de Machine Learning realizó las predicciones correctamente, basadas en los parámetros ambientales (temperatura, humedad e índice de calor), demostrado a través de las métricas de evaluación para algoritmos de Machine Learning, aplicados. La **Figura 9** muestra el resultado de la matriz de confusión generada del modelo entrenado, donde los valores de la diagonal principal Verdaderos Negativos (VN) = 1700 y Verdaderos Positivos (VP) = 1322 corresponden a los valores estimados de forma correcta por el modelo. Por su parte, la otra diagonal, representa los casos en los que el modelo muestra error o una predicción incorrecta, generando Falso Negativo (FN) = 25 y Falso Positivo (FP) = 22. La evaluación del modelo logístico con la matriz de confusión obtuvo una alta precisión y un porcentaje de las predicciones correctas de medio hacia alto.

Con respecto a las métricas aplicadas para evaluar algoritmos de Machine Learning, la interpretación de los resultados son los siguientes: el *índice general de Calidad (Accuracy)* fue del 98.46%, lo que indica que el modelo es capaz de predecir, correctamente, en la mayoría de las veces, el resultado de la *métrica de Precisión* fue del 98.36%, es decir, que el modelo acertará en la predicción del encendido de los ventiladores (actuador) en un 98.36% y se equivocará en un 1.64% de las veces que prediga. Esta métrica nos da la calidad de la predicción del modelo, la *Exhaustividad (Recall)* del modelo fue del 98.14%, lo que significa que el modelo identifica, correctamente, el encendido de los ventiladores (actuador) tomando en cuenta los parámetros ambientales, la *Especificidad (Specifity)* mide qué tan bien el modelo predice el apagado de los ventiladores(actuador), el mismo arroja un resultado del 98% de asertividad. Esta métrica es exclusiva de la clasificación binaria, y el resultado de *F1-Score* de un 98%, confirmando que el modelo cumple con el objetivo de esta investigación, debido a que la puntuación F1-Score es la media armónica entre la precisión y la exhaustividad. Por consiguiente, el modelo es lo, suficientemente, bueno para usarse en escenarios de la vida real.

## 5. Conclusiones

- Esta investigación ofreció la oportunidad al sistema inteligente “**Smart Poultry Farm**”, de expandir su marco de trabajo, debido a que el conjunto de datos inéditos recolectados y leídos de los sensores ubicados en la galera de pollo en abierto fueron sometidos a algoritmos de Machine Learning para obtener predicciones de resultados precisas y basadas en los datos.
- Este estudio abre nuevas posibilidades en el empleo del Machine Learning para análisis de decisiones basadas en el comportamiento de las variables ambientales dentro de estos galpones avícolas, ofreciendo mejoras a la toma de decisiones, maximizando la eficiencia en la producción avícola y mitigando los efectos del cambio climático.
- La utilización de la metodología SEMMA facilitó el manejo del dataset por ser una herramienta intuitiva. Por su parte, el entorno de Google Colaboratory facilitó la construcción de los algoritmos de Machine Learning en Python permitiendo el procesamiento de los datos de forma ágil.
- El algoritmo de regresión logística binomial permitió predecir la probabilidad de que la variable dependiente cualitativa dicotómica tenga un valor específico en función de las variables independientes.
- El modelo construido es lo, suficientemente, bueno para usarse en escenarios de la vida real, y con su implementación permitirá conocer con antelación los posibles eventos dentro del galpón y contribuirá así a la optimización de los procesos avícolas en la región.
- Por último, los resultados obtenidos en este estudio sientan las bases para futuras investigaciones en el área, además, será la base para solventar problemas inherentes a decisiones ambientales futuras.

## Agradecimientos

Al Sistema Nacional de Investigación (SNI) de la Secretaría Nacional de Ciencia, Tecnología e Innovación (SENACYT - Panamá) y a los investigadores del proyecto “**Smart Poultry Farm**” de la convocatoria pública DINE-INNOVATEC 2021 de la Secretaria Nacional de Ciencia, Tecnología e Innovación (SENACYT, Panamá).



## Referencias bibliográficas

- ANAVIP - *La avicultura en Panamá*. (s. f.). <https://www.anavip.org/index.php/la-avicultura-en-panama/>
- Aviagen. (s. f.). <https://aviagen.com/la/tech-center/>
- Ayoub Shaikh, T., Rasool, T., & Rasheed Lone, F. (2022). Towards leveraging the role of machine learning and artificial intelligence in precision agriculture and smart farming. *Computers and Electronics in Agriculture*, 198, 107119. <https://doi.org/10.1016/J.COMPAG.2022.107119>
- Banco Interamericano de Desarrollo. (2022). Clima, capital y un nuevo comienzo para los jóvenes. *Boletín mensual*. <https://www.linkedin.com/pulse/clima-capital-y-un-nuevo-comienzo-para-/>
- Brownlee, J. (2020). *Data Preparation for Machine Learning: Data Cleaning, Feature Selection, and Data Transforms in Python* (Machine Learning Mastery, Ed.). [https://books.google.com.pa/books?id=uAPuDwAAQBAJ&printsec=frontcover&source=gbs\\_ge\\_summary\\_r&cad=0#v=onepage&q&f=false](https://books.google.com.pa/books?id=uAPuDwAAQBAJ&printsec=frontcover&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false)
- Cakic, S., Popovic, T., Krco, S., Nedic, D., Babic, D., & Jovovic, I. (2023). Developing Edge AI Computer Vision for Smart Poultry Farms Using Deep Learning and HPC. *Sensors* 2023, 23(6), 3002. <https://doi.org/10.3390/S23063002>
- Chergui, N., & Kechadi, M. T. (2022). Data analytics for crop management: a big data view. *Journal of Big Data* 9(1), 1-37. <https://doi.org/10.1186/S40537-022-00668-2>
- Cobb. (s. f.). <https://www.cobb-vantress.com/resource/featured>
- Consejo Mexicano de la Carne. (s. f.). <https://comecarne.org/>
- Diez-Olivan, A., Averós, X., Sanz, R., Sierra, B., & Estevez, I. (2019). Quantile regression forests-based modeling and environmental indicators for decision support in broiler farming. *Computers and Electronics in Agriculture*, 161, 141-150. <https://doi.org/10.1016/J.COMPAG.2018.03.025>
- Estrada, M. M. E., Márquez, S., & Restrepo, L. (2007). Efecto de la temperatura y la humedad relativa en los parámetros productivos y la transferencia de calor en pollos de engorde. *Revista Colombiana de Ciencias Pecuarias*, 20(3), 288-303. <https://doi.org/10.17533/UDEA.RCCP>
- Google Colaboratory. (s. f.). [https://colab.research.google.com/?utm\\_source=scs-index](https://colab.research.google.com/?utm_source=scs-index)

- Hernández-Sampieri, R., y Mendoza Torres, C. P. (2018). *Metodología de la Investigación: Las rutas Cuantitativa Cualitativa y Mixta*. McGraw-Hil. <http://repositorio.uasb.edu.bo:8080/bitstream/54000/1292/1/Hernández- Metodología de la investigación.pdf>
- Larose, D. T., & Larose, C. D. (2014). *Discovering Knowledge in Data: An Introduction to Data Mining*. 2<sup>da</sup> Ed. Wiley. <https://www.wiley.com/en-us/Discovering+Knowledge+in+%20Data%3A+An+Introduction+to+Data+Mining%2C+2nd+Edition-p-9780470908747>
- Ojo, R. O., Ajayi, A. O., Owolabi, H. A., Oyedele, L. O., & Akanbi, L. A. (2022). Internet of Things and Machine Learning techniques in poultry health and welfare management: A systematic literature review. *Computers and Electronics in Agriculture*, 200, 107266. <https://doi.org/10.1016/J.COMPAG.2022.107266>
- Pajankar, A., & Joshi, A. (2022). *Hands-on Machine Learning with Python: Implement Neural Network Solutions with Scikit-learn and PyTorch*. Apress. <https://doi.org/10.1007/978-1-4842-7921-2>
- Porcelli, A. M. (2020). Inteligencia Artificial y la Robótica: sus dilemas sociales, éticos y jurídicos. *Derecho Global. Estudios sobre Derecho y Justicia*, 6(16), 49-105. <https://doi.org/10.32870/DGEDJ.V6I16.286>
- Reier Forradellas, R., Nández Alonso, S., Jorge-Vazquez, J., & Rodriguez, M. (2020). Applied Machine Learning in Social Sciences: Neural Networks and Crime Prediction. *Social Sciences*, 10(1), 4. <https://doi.org/10.3390/SOCSCI10010004>
- Ribeiro, R., Casanova, D., Teixeira, M., Wirth, A., Gomes, H. M., Borges, A. P., & Enembreck, F. (2019). Generating action plans for poultry management using artificial neural networks. *Computers and Electronics in Agriculture*, 161, 131-140. <https://doi.org/10.1016/J.COMPAG.2018.02.017>
- SAS Help Center: *Introduction to SEMMA*. (s. f.). <https://documentation.sas.com/doc/en/emref/14.3/n061bzurmej4j3n1jni8bbj1a2.htm>
- The Caribbean Climate-Smart Accelerator. (2022). *Una Hoja De Ruta Climática Para América Latina y El Caribe*. Una Hoja De Ruta Climática Para América Latina y El Caribe. <https://vimeo.com/695183333>
- Vuppalapati, C. (2021). *Machine Learning and Artificial Intelligence for Agricultural Economics*. Springer International Publishing. <https://doi.org/10.1007/978-3-030-77485-1>

- Wicaksono, D., Perdana, D., & Mayasari, R. (2018). Design and analysis automatic temperature control in the broiler poultry farm based on wireless sensor network. *Proceedings - 2017 2nd International Conferences on Information Technology, Information Systems and Electrical Engineering, ICITISEE 2017, 2018-Janua*, 450-455. <https://doi.org/10.1109/ICITISEE.2017.8285549>
- Zheng, M., Wang, F., Hu, X., Miao, Y., Cao, H., & Tang, M. (2022). A Method for Analyzing the Performance Impact of Imbalanced Binary Data on Machine Learning Models. *Axioms*, 11(11). <https://doi.org/10.3390/AXIOMS11110607>