

REGRESIÓN MÚLTIPLE Y SEGMENTACIÓN EN EL RENDIMIENTO Y MORTALIDAD EN LA CADENA DE PRODUCCIÓN AVÍCOLA

José Eider Rivera Serna ^{1*}, José Luis Hernández Otero ^{1*}, Cristian Méndez ^{3*}, Julián Hoyos ^{3*}

¹ *Ingeniería de Sistemas, Fundación Universitaria de Popayán, Popayán, Colombia.*

² *Intelligent Management Systems (IMS), Profesor Ingeniería de Sistemas, Fundación Universitaria de Popayán, Colombia.*

³ *Profesor Ingeniería de Sistemas, Fundación Universitaria de Popayán, Colombia.*

* jose.rivera@estudiante.fup.edu.co,

* jose.hernandez@estudiante.fup.edu.co

Resumen — Este artículo analiza el rendimiento y la mortalidad como factor principal de la industria avícola, lo cual afecta directamente la rentabilidad y la sostenibilidad de la misma, es por ello, que el rendimiento en dicha industria hace referencia a criar pollos de engorde, donde influye la productividad y esta se relaciona a la vez con la cantidad de carne producida por pollo y el tiempo de crecimiento requerido para alcanzar el peso ideal. Por otro lado, la mortalidad en la cadena de producción avícola se puede generar por diversas razones, como enfermedades, problemas de crianza y condiciones de vida insalubres, lo cual pueden aumentarla, puesto que no se le da un manejo y cuidado adecuado a las aves.

En este sentido, se realizó un estudio en tres granjas avícolas productoras de carne, en el municipio de Caldon, departamento del Cauca, donde se utilizó la metodología SEMMA (Sample, Explore, Modify, Model, Assess) propuesta por uno de los principales fabricantes de inteligencia empresarial software llamado SAS Institute Inc, la cual consiste en realizar una selección, exploración y modelado aplicado a datos para identificar patrones que sirven como herramienta de apoyo para el trabajo desarrollado, por tal motivo, se tuvo en cuenta 18 variables y 756 registros u observaciones como muestra, donde se le efectuó el análisis de la regresión múltiple a partir de la matriz de coeficientes de correlación para predecir la mortalidad de las aves en cada uno de los tres proyectos (granjas), y para dar un mayor acierto en dicho estudio se creó un conjunto de datos en el que se tomaron cinco variables: “calidad_de_agua”, “ahogamiento”, “huevo_sin_vacuna”, “semillas defectuosa” y “otro”, las cuales están relacionadas directamente con la mortalidad y en el análisis de los clústeres, se obtuvo como resultado que en el proyecto 3, tiene una gran mortalidad en comparación con los otros proyectos. La razón es debido a la calidad de agua, de igual manera, se analizó el modelo de regresión lineal simple, que se centra en predecir el peso final que contiene los coeficientes del intercepto y la variable de “consumo_de_alimento_c.u”. Por tal motivo, se creó un dataset con cinco nuevos datos de la cantidad de alimento para observar cómo el peso final está relacionado con el consumo del mismo de cada ave y se logró como resultado que; a mayor cantidad de alimento consumido por los animales, mayor será el peso final de las aves. Lo cual influye en la producción y mortalidad, por ello se hizo necesario el uso del

algoritmo K-Means mediante el software R-Studio, donde se establecieron 3 grupos y cuyos resultados fueron los siguientes: mortalidad media con pesos iniciales y finales bajos, baja mortalidad y un alto peso final, baja mortalidad y un peso final medio.

De igual manera, se encontraron características en las variables predictoras que contribuyen a la prevención y a la toma de decisiones para mejorar el manejo técnico y práctico en la cadena de producción avícola. Finalmente, a partir de la regresión multi-lineal, se predice la mortalidad en relación al número de ciclos por cada proyecto productivo.

Como trabajo futuro, sería interesante analizar los planes de mejora y los procesos de toma de decisiones en cada sistema productivo, así como proponer nuevas soluciones tecnológicas que permitan visualizar de manera óptima la información que resulta del análisis de datos a los administradores de los proyectos productivos.

Palabras claves: *Producción, Regresión Múltiple, rendimiento, mortalidad, conversión, rentabilidad, Clustering*

Abstract Yield and mortality as a major factor, which directly affects the profitability and sustainability of the poultry production, are discussed in this article. In terms of yield, poultry farming focuses on raising broiler chickens, where productivity is measured in terms of the amount of meat produced per chicken and the growth time required to reach the ideal market weight. On the other hand, mortality in the poultry production chain can be caused by several reasons, such as diseases, rearing problems and unsanitary living conditions, thus, mortality can increase due to management problems and lack of care of the chickens. Hygiene and cleanliness measures and strategies are necessary to maintain healthy living conditions for the chickens. In this study we used daily data taken from three different poultry projects from 18 variables and 756 records or observations, analyzed the multiple regression from its correlation coefficient matrix, and implemented a segmentation in production and mortality from a K-Means algorithm (R-Studio software). The main result was to establish 3 groups: medium mortality with low initial and final weights, low mortality and high final weight, low mortality and medium final weight. Characteristics were found

in the variables that contribute to prevention and decision making to improve technical and practical management in the poultry production chain. Finally, from linear regression, mortality is predicted in relation to the number of cycles for each production project.

1. INTRODUCCIÓN.

La cadena avícola es un sector importante en la producción mundial de alimentos, es por ello que la integración del análisis de variables como el rendimiento y la mortalidad en la cadena de producción avícola y segmentación de datos básicos es fundamental para lograr mejorar estos factores que son vitales para la sostenibilidad financiera de esta industria [1]. Este trabajo describe algunas técnicas y metodologías que se utilizan para mejorar el rendimiento y reducir la mortalidad en la producción avícola

El sector avícola afronta grandes retos, que están relacionados a factores que pueden afectar la rentabilidad y sostenibilidad, algunos de los más comunes son: enfermedades avícolas, calidad de agua, insalubridad en los sitios de producción. Las enfermedades pueden propagarse rápidamente en los ambientes de producción avícola y puede ser devastador. Según [2] la colibacilosis es una de las principales causas de pérdidas económicas en la industria avícola mundial, y esta enfermedad puede aparecer tanto en las aves jóvenes como en las adultas.

El agua es un factor clave en la dieta de las aves, y su calidad es fundamental en la producción avícola. En [3] se describe la importancia de la calidad del agua en la producción avícola y cómo afecta la salud de las aves y la calidad del producto final. También se discuten las principales fuentes de contaminación y medidas para prevenirlas. En [4] centran su estudio en la importancia que tiene el monitoreo de la calidad del agua en la producción avícola. Se describen los parámetros fisicoquímicos que se deben medir y cómo interpretar los resultados para tomar medidas preventivas y correctivas. Por otra parte, las condiciones de vida en las granjas avícolas son fundamentales para la salud y el bienestar de las aves. Si las condiciones no son adecuadas, pueden determinar problemas de salud en las aves y afectar la calidad del producto final [5].

Los procesos de captura y transporte afectan el bienestar animal y ocasionan pérdidas de peso vivo, y hasta la muerte, lo cual impacta económicamente la actividad avícola. Además, el mantenimiento adecuado de la infraestructura es esencial para garantizar la salud y el bienestar de las aves y maximizar la producción. Los productores avícolas deben asegurarse de que las instalaciones estén adecuadamente ventiladas, iluminadas y que la calidad del agua sea alta. Además, deben considerar cuidadosamente el tamaño y la distribución de las instalaciones para garantizar que las aves tengan suficiente espacio para vivir y evitar el estrés excesivo [6-7].

Sumado a la calidad de los alimentos [7], el control de temperatura es otro factor clave que incide en la producción avícola. En [8] se evidencia que la avicultura al igual que otras industrias, tiene como objetivo principal la rentabilidad

económica y debido a la competencia del mercado es necesario buscar la eficiencia y sostenibilidad del sistema productivo. Por lo tanto, para que los pollos desarrollen al máximo el potencial productivo contenido en su genética, es indispensable manejar un entorno que les proporcione las condiciones ambientales adecuadas.

El objetivo de este artículo es analizar el rendimiento y la mortalidad en la cadena de producción avícola, a partir de la segmentación-clustering, aplicando el algoritmo K-Means que sirve para agrupar un conjunto de datos, permitiendo la obtención de resultados más precisos en la producción avícola y obteniendo información que contribuye al proceso de toma de decisiones [9]. Algunas de las variables consideradas fueron el peso, la edad, el sexo y la raza.

Con el análisis realizado, se encontró que es de vital importancia que en las granjas dedicadas a la producción avícola se cuente con unas condiciones adecuadas para mitigar los altos porcentajes de mortalidad y las bajas tasas de rendimiento ocasionadas, debido al mal manejo de las variables mencionadas anteriormente. Por lo cual, se reitera que mientras un proyecto avícola esté sustentado por condiciones técnicas optimas y haya una buena semilla se aumenta la probabilidad de obtener los resultados esperados.

El resto del documento esta organizado de la siguiente manera: en la sección 2 se presentan los trabajos relacionados a la investigación; la metodología es abordada en la sección 3; la sección 4 enseña los principales resultados obtenidos; y finalmente son presentadas las conclusiones.

2. TRABAJOS RELACIONADOS.

A nivel nacional se han propuesto prácticas de mejoramiento en la producción avícola, tales como: implementar prácticas de bioseguridad; lo cual es esencial para prevenir enfermedades aviares. Promover la diversificación de la producción avícola en cuanto a las razas de pollo para producción de carne [10].

A pesar de que el sector avícola viene apoyándose en metodologías que han aportado a mejorar su rendimiento y producción, todavía existe una brecha amplia en la que se puede seguir implementando y aún más cuando se hace referencia a los pequeños y medianos productores que por ende son menos industrializados [11].

A nivel Internacional se encontró que los programas de alimentación y los alimentos utilizados en la actualidad varían mucho en costo, desde excesivamente costosos hasta insuficientemente nutritivos, desde eficientes hasta derrochadores [12]. Para que sean realmente efectivos, los programas de alimentación deben estar diseñados para proporcionar al rebaño todos los nutrientes que necesita. Al pensar en avicultura, se debe de pensar en un negocio, el cual nos ofrece una mayor rentabilidad lo que se garantiza cuando las aves ganan más peso en menos tiempo y al menor costo posible.

En [13] se identificó que una de las principales preocupaciones es que se desconoce la utilización de la función de producción de la línea genética Cobb que maximice el margen bruto para los avicultores, por lo tanto, se realizó una investigación utilizando un modelo compuesto por una función de producción y una de costos que varían, para encontrar el tiempo óptimo de sacrificio para pollos de engorde de la línea genética Cobb. La función de ingresos se estructuró teniendo en cuenta un precio variable en el tiempo por gramo de pollo en pie, una función de crecimiento de un pollo promedio expresada en gramos y el número de pollos. En la determinación de la función de producción y costos que varían totales, se incluyen parámetros obtenidos mediante regresiones, utilizando datos proveídos por la empresa Cobb-Vantress [14].

En [15] se observó que la investigación se realizó en una Granja Agropecuaria con características específicas, con el fin de identificar y controlar todos los factores importantes que influyen durante la producción. El estudio tuvo una duración de ocho semanas, consiguiendo el objetivo principal correspondiente a la evaluación de los principales indicadores productivos y económicos de la crianza intensiva de pollos de las líneas Cobb y Ross de ambos sexos, además se identificó la importancia de comprender el inminente riesgo al que se exponen los resultados de la camada de pollos que son transportados cada ciclo desde las granjas hasta las plantas de faenamiento. En este trayecto, los pollos están expuestos a una amplia variedad de factores estresantes como la restricción de alimento, la captura y el transporte que afectan el bienestar animal y producen pérdidas de peso vivo hasta la muerte, lo cual impacta económicamente la actividad. Estudio que identifica las causas de falencia y evidencia la importancia de unas buenas prácticas de captura y transporte del producto final [16].

En [14] se identificó la aplicación de la metodología de segmentación de datos que tiene como objetivo estimar el valor medio y/o predecir valores particulares, basado en la información de variables independientes o predictoras. con el fin de realizar el análisis estadístico de tendencia central como la media, varianza y desviación estándar para cada etapa de cría, Estudio realizado gracias a que se evidenció que dentro de los parámetros de producción y manejo de pollos de engorde existe la necesidad de adoptar sistemas de producción adecuados a las condiciones ambientales característicos de las zonas de producción avícola específicamente de pollos de engorde de la línea Ross ya que estos determinan la rentabilidad de la crianza para poder predecir si dicha actividad es económicamente sustentable o no para el productor.

A nivel Nacional se identificó que una de las principales causas de mortalidad en las aves es el síndrome ascítico o síndrome de hipertensión pulmonar [17]. El síndrome ascítico puede considerarse como una manifestación de una insuficiencia cardíaca congestiva derecha, que provoca una hipertensión hidrostática venosa generalizada, hipertrofia cardíaca derecha y edema. Se caracteriza por afectar al pollo de engorde y a la

polla reproductora pesada, a partir de la segunda semana de vida, con máxima mortalidad a la quinta y sexta semana. Por lo cual se plantea desarrollar un protocolo para identificar y mitigar puntos de riesgo de Síndrome Ascítico mediante la implementación de buenas prácticas de manejo en pro de favorecer las variables zootécnicas de cada lote.

En [18] se encontró la construcción de un sistema de predicción durante el crecimiento del ave con relación a las variables ambientales que más influyen en este proceso, para ello se instalaron sensores (sistema IoT) en un galpón adquiriendo los datos de temperatura, humedad, velocidad del viento y el peso de las aves obteniendo datos en tiempo real. Se realizó el procesamiento de esta información con el fin de eliminar los datos atípicos y trabajar el algoritmo con los valores normalizados generando alertas en las variaciones.

En [10] se identificó la importancia de resaltar y a su vez discriminar las necesidades ambientales para la explotación del pollo de engorde en regiones tropicales con condiciones climáticas semejantes a las encontradas en Colombia, con la intención de determinar la zona termo neutral en la cual el ave lleva a cabo pequeños cambios en la producción calórica, permitiendo así que modelos de producción sean adecuados a cada tipo de zona climática.

En [14] se identificó que pueden existir efectos diferenciales no muy pronunciados con el experimento realizado mediante la alimentación pollos de engorde de la línea Ross machos y hembras como herramienta indispensable para medir el sistema de producción en términos técnicos, económicos y de desempeño del animal.

En Colombia no se han planteado estudios científicos que evalúen y determinen las zonas y los factores ambientales adecuados para el manejo de los pollos de engorde.

3. METODOLOGÍA.

3.1. Contexto de la investigación.

El estudio se desarrolló en tres granjas avícolas productoras de carne, las cuales tienen las siguientes características: i) “Villa Lucía” ubicada en la Vereda Monterilla, Caldoño Cauca, Colombia, con altitud de 1.498 metros sobre el nivel del mar (m.s.n.m), temperatura promedio de 20 °C; ii) “Pedro Salinas” ubicada en Pescador, Cauca, Colombia, con una altitud de 1.508 m.s.n.m, con temperatura promedio de 20 °C; y iii) “Rubiela Balanta” ubicada en Villarrica, Cauca, con una altura de 982 m.s.n.m, con una temperatura promedio de 24 °C, consolidando la recogida de datos a partir de 6 ciclos de producción durante todo un año (2022) considerando la mortalidad y la productividad.



Figura 1. Ubicación geográfica de los 3 proyectos avícolas (San Rafael, Monterilla, Pescador)

3.2. Conceptos a considerar.

Para el desarrollo de la investigación es importante considerar los siguientes conceptos: segmentación, clúster, regresión múltiple, Rstudio.

- **Clusters:** permite la agrupación de datos con características similares, los conjuntos de datos obtenidos son quienes generan los patrones [6, 7].
- **Segmentación:** permite separar en pequeños grupos el alto volumen de datos que se ha recolectado según el objetivo deseado [2,6,12]. Para el desarrollo de la propuesta se emplea una segmentación no dirigida, en la cual no se tiene variables implícitas en otra [6].
- **Regresión múltiple:** el objetivo es desarrollar un modelo matemático que describa como las variables predictoras están relacionadas con la variable dependiente y así poder realizar inferencia sobre cada variable independiente [6].
- **Análisis de residuos:** es una herramienta valiosa para evaluar la validez del modelo de regresión múltiple. Mediante la visualización y la interpretación de los residuos, es posible identificar problemas en el ajuste del modelo y decidir si los supuestos del modelo lineal se mantienen o si se necesita una revisión del enfoque de modelado.[1]
- **Rstudio:** es una interfaz de usuario de código libre para lenguajes de programación como R. Proporciona un ambiente de desarrollo integrado (IDE) para ayudar a los usuarios a escribir código R, ejecutarlo, visualizar los resultados y organizar sus trabajos. Está diseñado para hacer que el trabajo con R sea más fácil y más eficiente [13]. Con esta herramienta se procede

a hacer limpieza de datos, reducción de variables, apoyados en métodos y funciones. Proporciona gráficas que contribuyen hacer más comprensible el análisis de los datos.

3.2. Propuesta metodológica.

Se empleó la metodología SEMMA, desarrollada por el instituto SAS. Esta metodología se define como un proceso que permite seleccionar, explorar y modelar grandes volúmenes de datos, con el propósito de descubrir patrones de negocio desconocidos. SEMMA representa las cinco fases fundamentales del proceso: Sample (muestreo), Explore (exploración), Modify (modificación), Model (modelo) y Assess (valoración) [7] (ver Figura 2).

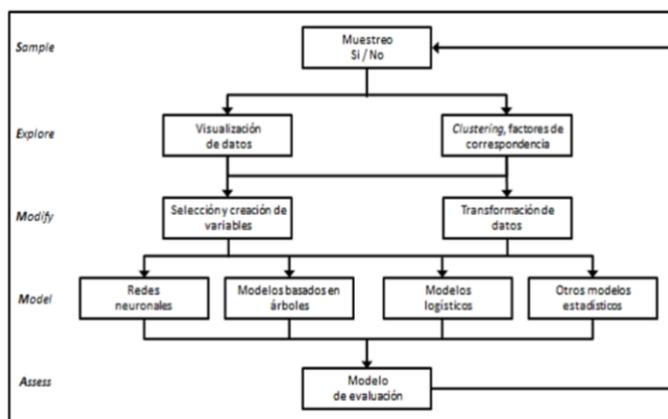


Figura 2. Propuesta metodológica – SEMMA.

3.2.1. Obtención de datos.

La obtención de la información se realizó a través de la información de una base de datos de tres proyectos avícolas (registros diarios tomados durante 6 ciclos de producción de 42 días cada uno, correspondientes a todo un año). La estructura del dataset se observa en la Tabla 1 (18 variables consideradas).

Tabla 1. Variables consideradas en el dataset, 3 proyectos avícolas.

Variable	Descripción
Condiciones climáticas	Rango de temperatura 15° - 31°
Calidad del agua	Agua turbia, agua contaminada, ph alto, ph bajo, exceso de cloro, escasez de agua, normal.
Tipo de alimento	Pre Inicio, inicio, levante, engorde, finalizador.
Consumo de alimento	47g - 640g
Peso inicial	83g - 2257g
Raza	Razas con las cuales trabajan los proyectos avícolas: cobb y ross

Ahogamiento	De 0 a 5 aves mueren diariamente por esta causa.
Semilla defectuosa	De 0 a 28 aves mueren diariamente por esta causa.
Huevo sin vacuna	De 0 a 13 aves mueren diariamente por esta causa.
Otro	De 0 a 5 aves mueren diariamente por causas relacionadas con incidentes durante el manejo.
Conversión	Cantidad de gramos que gana un ave al día según consumo de alimento y edad; está entre 8g a 92g.
Peso final	Peso diario del ave después de su primera alimentación desde el día uno (91 g) hasta el día cuarenta y dos (2374 g) que se finaliza el ciclo.
Mortalidad	De 0 a 37 aves mueren diariamente por todas las causas mencionadas anteriormente.
N° aves	Nos permite identificar el número de aver que maneja cada proyecto.
N° de proyecto	Hace referencia a los tres proyectos en los cuales nos basamos para realizar la investigación.
N° de aves recibidas	Número inicial de aves recibidas en cada proyecto.(Día uno).
N° de aves entregadas	Número de aves entregadas al final del ciclo en cada proyecto.(Día cuarenta y dos, fin del ciclo).
N° de ciclo	Permite identificar el ciclo al cual pertenecen los datos. (Se investigó en base a seis ciclos).

3.2.2. Limpieza de datos

El objetivo de la limpieza de datos es la detección de valores faltantes y/o datos anómalos con el propósito de asegurar que que estos sean confiables, precisos y consistentes para el proceso de análisis, o en términos generales, se realiza para garantizar la calidad del conjunto de datos para no afectar al proceso de análisis. En caso de detectarse valores faltantes o NA, es común reemplazarlos por la media aritmética de la respectiva variable cuantitativa o columna [20].

3.2.3. Normalización.

Este proceso encuentra el valor medio y elimina los datos que están muy separados, lo que puede afectar la ubicación del centroide. La ventaja de usar la normalización se define en los resultados de la agrupación, obteniendo trayectorias de mayor varianza para evitar errores de descripción del modelo [21].

3.2.4. Clusterización.

Esta es una técnica de aprendizaje sin control que consiste en agrupar varios objetos en subgrupos homogéneos (o clúster) dependiendo de sus similitudes o afinidades. En otras palabras, se utiliza en varias aplicaciones, como la segmentación de datos, principalmente aplicada en minería de datos. El cálculo de la matriz de distancia permite medir la similitud o la distancia entre los datos a agrupar, puede medirse de diversas maneras, dependiendo del tipo de datos y del algoritmo de clusterización utilizado [22].

Una matriz de distancia es una tabla que representa la distancia entre cada punto a través de un conjunto de datos. El cálculo de esta matriz depende del tipo de datos utilizados y del método que se quiera utilizar. Estos son algunos de los métodos más utilizados para calcular la matriz de distancia:

Es una medida común para calcular la distancia entre dos puntos cuando los datos son numéricos y en el espacio euclidiano. Se calcula como la raíz cuadrada de la suma al cuadrado de las diferencias en cada dimensión [23].

$$d(a, b) = \sqrt{\sum_{i=1}^n (a_i - b_i)^2}$$

Ec.1. Distancia Euclidiana [23].

Con esta técnica se buscan elementos en función de su proximidad para agruparlos. Usando variables previamente declaradas, recopila y almacena los resultados de los cálculos de la matriz anterior (uso de la función “dist”).

A continuación, se describen las etapas del proceso de clusterización:

3.2.5. Agrupación de clúster.

Es una técnica de análisis de datos no supervisada que se utiliza para dividir conjuntos de datos en grupos en función de la similitud de las observaciones. El objetivo es encontrar patrones o estructuras en los datos, por ende, se implementó el k-means este método asigna cada observación al grupo más cercano, donde la proximidad se define como la distancia euclidiana (u otra métrica de distancia) [24]. Luego se calcula el centroide de cada conglomerado como el promedio de todas las observaciones y se reasignan al centroide más cercano.

El algoritmo de k-means es un algoritmo iterativo y, como tal, es más eficiente para agrupar datos, más útil para el análisis que se realiza [24]. El objetivo del método del centroide es minimizar la distancia entre los objetos dentro de un clúster y maximizar la distancia entre los objetos de diferentes clústeres.

3.2.6. Selección del número óptimo de clusters.

La búsqueda bibliográfica reveló que el tamaño no se definió como una referencia para indicar cuántos clústeres deberían estar contenidos en el nuevo clúster resultante. Se empleó el método wss [27] para determinar el número óptimo de clústeres, en un conjunto de datos [28]. El procedimiento implica calcular la suma de las distancias al cuadrado entre cada elemento o punto en un conjunto de grupos y su centroide. Después de eso, se traza la mejor curva "wss" para el grupo mediante la búsqueda de un "codo" en la curva que apuntan a la adición de más grupos. los puntos en los que se explica la varianza significativamente.

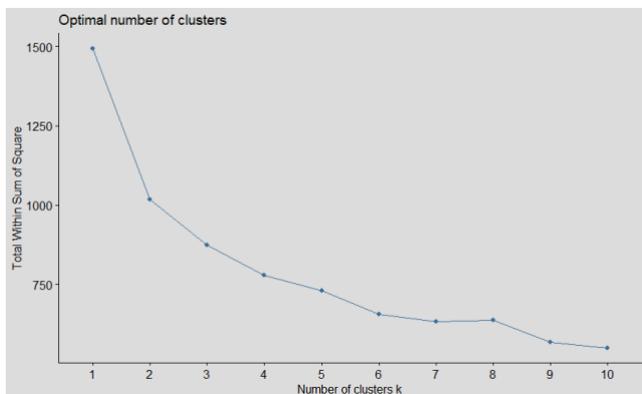


Figura 3. Cálculo número óptimo de cluster.

Según la figura anterior, la cantidad óptima de clústeres para el conjunto de datos es dos. Sin embargo, en [29] se menciona que para métodos de clustering jerárquico, se recomienda realizar un corte en el dendrograma generado, en el punto en donde se empiezan a presentar saltos bruscos. Por tal motivo, al aplicar las técnicas de clustering jerárquico con el método del centroide y el algoritmo kmeans, se tendrán en cuenta tres clústeres en este estudio.

3.2.7. Construcción de los modelos descriptivos.

Para la construcción de los modelos descriptivos, se aplicaron las técnicas de clustering jerárquico con el método del centroide, y el algoritmo kmeans. que requiere la matriz de distancia y la cantidad de clústeres por lo cual se tomaron 3 clústeres

3.2.8. Creación del modelo con el método del centroide.

Al igual que en el modelo kmeans, el método del centroide requiere de la matriz de distancia, con la diferencia de no requerir el número de clústeres durante la creación del modelo. Además, requiere especificar el método con el que se construirá el modelo, en este caso "centroid".

3.2.9. MODELOS PREDICTIVOS.

Los modelos predictivos, son comúnmente utilizados con el fin de detectar que podría ocurrir en el futuro a partir de los patrones detectados en un conjunto de datos, se utilizaron las técnicas de regresión lineal simple y regresión lineal múltiple.

La regresión lineal simple, es usada para predecir el valor de una variable conocida como variable dependiente a partir de otra variable llamada variable independiente [30]. Su descripción matemática es la siguiente:

$$y = b_0 + b_1 * x$$

Ec.2 Regresión lineal simple [30].

Antes de construir el modelo de regresión lineal simple, se generó una gráfica de la matriz de correlación de Pearson, La cual mide la relación que existe entre dos variables. Los valores de correlación pueden estar entre -1 y 1, donde un valor cercano a -1 indica una correlación negativa fuerte y 1 una correlación positiva fuerte [30]. A continuación, se presenta la matriz de correlación generada.

n_de_ciclo	-0.01	0.11	0	0.16	0	-0.29	0.2	0.2	0.09	0.28	0.02	0.02	0.2	-0.02	0	-0.01	-0.03	
n_de_aves_entregadas	0.41	0.06	0	0	0	-0.01	0.59	0.3	0.37	0.3	0.01	0.01	0.22	1	-0.8	1	1	-0.03
n_de_aves_recibidas	0.41	0.06	0	0	0	-0.02	0.61	0.33	0.38	0.32	0.01	0.01	0.28	1	-0.78	1	1	-0.01
n_de_proyecto	-0.32	-0.2	0	0.02	0	0	-0.48	-0.1	-0.3	-0.19	-0.02	-0.02	0.06	-0.79	1	-0.78	-0.8	0
n_de_aves	0.4	0.06	0.03	-0.04	-0.04	-0.01	0.6	0.32	0.38	0.31	0	-0.03	0.25	1	-0.79	1	1	-0.02
mortalidad	0.12	-0.09	0.03	0	-0.06	-0.1	0.41	0.39	0.31	0.43	-0.03	-0.08	0.25	0.06	0.28	0.22	0.2	0.2
peso_final	0.06	0.09	-0.67	0.94	1	1	0	-0.01	-0.09	-0.08	0.01	0.2	1	-0.08	-0.03	-0.02	0.01	0.01
conversion	0.04	0.07	-0.04	0.14	0.16	0	0.01	-0.01	0.01	0	0	1	0.2	-0.03	0	-0.02	0.01	0.01
cep	0.1	0.06	0	0.03	0.01	-0.35	0.45	0.39	0.15	1	1	0	0.01	0.43	0.31	-0.19	0.32	0.3
huevo_sin_vacuna	0.18	0.05	0.04	-0.01	-0.08	0.05	0.26	0.14	1	0.15	0.01	-0.08	0.31	0.38	-0.3	0.38	0.37	0.09
semilla_defectuosa	0.08	-0.03	0.06	-0.07	-0.09	-0.3	0.39	1	1	0.14	0.39	-0.01	-0.09	0.39	0.32	-0.1	0.33	0.3
atragamiento	0.23	0.07	-0.01	0	-0.02	-0.24	0.39	0.26	0.45	0.01	-0.01	0.41	0.6	-0.48	0.61	0.59	0.2	0.2
raza	0	-0.01	0	0.01	0	0	-0.24	-0.3	0.05	-0.35	0	0	-0.1	-0.01	0	-0.02	-0.01	-0.29
peso_inicial	0.05	0.06	-0.68	0.94	1	1	0	-0.02	-0.09	-0.08	0.01	0.16	1	-0.06	-0.04	0	0	0
consumo_de_alimento_c.u	0.06	0.1	-0.68	1	0.94	0.01	0	-0.07	-0.01	0.03	0.14	0.94	1	0.04	0.02	0	0	0.16
spo_de_alimento	-0.05	-0.04	-0.68	-0.68	1	1	0	-0.01	0.06	0.04	0	-0.04	-0.67	1	0.03	0.03	0	0
calidad_de_agua	0.03	-0.04	0.1	0.08	-0.01	0.07	-0.03	0.05	0.06	0.07	0.09	-0.09	0.08	-0.2	0.08	0.08	0.11	0.11
condiciones_climaticas	1	1	0.03	-0.05	0.06	0.06	0	0.23	0.08	0.18	0.1	0.04	0.06	0.12	0.4	-0.32	0.41	0.41

Figura 4. Matriz de correlación de Pearson

En la gráfica de la matriz de correlación de Pearson, se observa que las variables con una fuerte relación con la variable peso_final es: "peso_inicial" y "consumo_de_alimento_c.u".

Para esta ocasión se creó el modelo de regresión lineal simple, enfocándose en predecir el "peso_final" en base a la variable independiente "consumo_de_alimento_c.u" la cual cuenta con una valor de 0.94. donde se utilizó datos limpios que generó

una gráfica de dispersión con curva de regresión, con el propósito de verificar la relación entre estas dos variables.

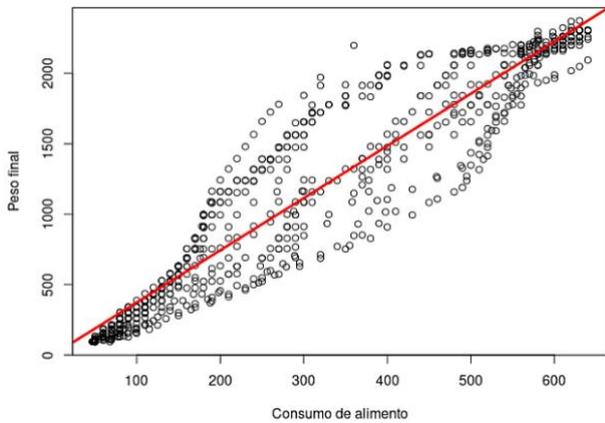


Figura 4. Gráfica de dispersión con línea de regresión.

En la gráfica 4 se observa como los datos tratan de mantenerse cerca de la línea roja, indicando un fuerte relación y concentración de los datos. Además, a medida que el consumo de alimento aumenta, también lo hace el peso final de las aves de corral.

La regresión lineal múltiple se busca predecir el valor de una variable dependiente a partir de dos o más variables independiente [31]. Su descripción matemática es la siguiente:

$$y = b_0 + b_1 * x_1 + b_2 * x_2 + b_3 * x_3 + b_n * x_n$$

Ec.3 Regresión lineal múltiple [31].

Para la construcción de los modelos de regresión lineal múltiple, en un inicio se crearon tres subconjuntos de datos, según el número de cada uno de los tres proyectos y se eliminó la variable proyecto de cada uno de estos, al no ser necesaria, con el propósito de realizar predicciones para cada uno de los proyectos.

Después de haber creado los subconjuntos de datos, se inició la creación de los modelos de regresión lineal múltiple por cada uno de los proyectos.

La construcción de los modelos se enfocó en predecir la mortalidad. En un inicio a partir de todas las variables, pero al evaluar los modelos mediante distintas métricas de evaluación, se encontró que estos no eran los óptimos para predecir, por lo cual, se crearon nuevos modelos, teniendo en cuenta estas métricas, obteniendo como resultado, modelos ideales para cada uno de los proyectos 1, 2, 3, se puede observar se tienen en cuenta una gran cantidad de variables, a diferencia del modelo de regresión lineal simple.

Para realizar predicciones con los modelos de regresión lineal simple y múltiple, se utilizó la función predict de R. que requiere un dataset con nuevos datos que siguen la estructura

de los modelos de regresión lineal construidos, para realizar predicciones en base a ellos.

4. RESULTADOS

En esta etapa se presentan los resultados obtenidos al aplicar las técnicas antes mencionadas.

4.1. Resultados del preprocesamiento

4.2. Limpieza de datos

A continuación, se presenta el proceso de limpieza de datos, empezando por la detección de valores faltantes, utilizando la función plot_intro. A continuación, se presenta el resultado.

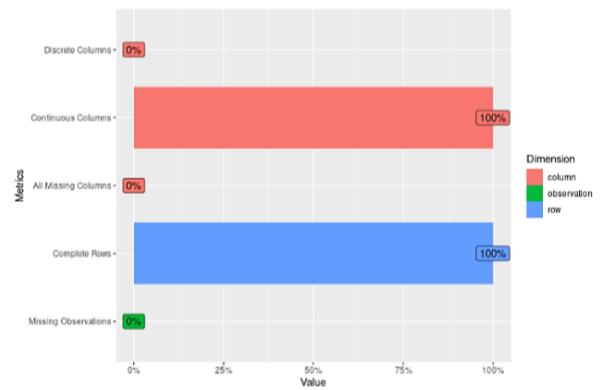


Figura 5. Distribución de las variables

En la figura se logra observar que el 100% de los datos son de naturaleza continua o numérica. Además, las barras de “missing observations” y “all missing columns” son inexistentes, lo que indica que los datos están completos.

Realizada la detección de valores faltantes, se procede con la detección de valores anómalos, para ello se utilizaron diagramas de cajas y bigotes, los cuales facilitan esta detección de forma visual, facilitando dicha tarea. A continuación, se presentan algunas de las variables en las que se detectaron datos anómalos.

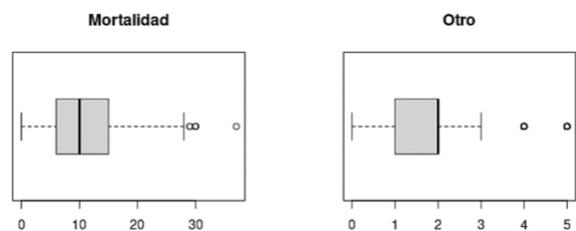


Figura 6. Detección de datos anómalos

En la figura se observan dos de las variables en las que se encontraron datos anómalos, siendo aquellos que se observan en ambos diagramas.

Para tratar los datos anómalos, estos son reemplazados por NA y posteriormente se reemplazan por la media aritmética de cada columna, como ya se había mencionado anteriormente. A continuación, se visualiza el resultado del proceso.

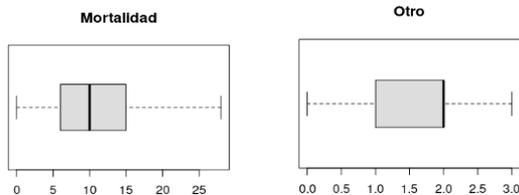


Figura 7. Detección de datos anómalos

4.3. Normalización

Para la normalización de los datos limpios se utilizó uno de los métodos más comunes, conocido como normalización lineal uniforme, el cual escala los datos a un rango previamente definido por el usuario empleando la función `data normalizada - data` [32].

4.4. Resultados modelos descriptivos

Modelo jerárquico con método del centroide agrupando en tres clústeres.

C	condici	calid	tipo_	consum	pes	r	aho	semill	huevo	o	co	pe	mo	n°_	n°_d	n°_de_	n°_de_	n°_de_
ou	ones_c	ad_d	de_al	o_de_ali	o_i	a	ga	a_def	o_sin	t	nve	so	rtal	de_	e_pr	aves_r	aves_e	de_
st	limatic	e_ag	iment	mento_c	nici	z	mie	ectu	_vacu	r	rsi	_fi	ida	ave	oyec	ecibid	ntregad	as
er	as	ua	o	/u	al	a	nto	sa	na	o	ón	nal	d	s	to	as	as	ciclo
1	268	3.5	2.59	2.22	2.3	2	1.91	1.03	1.43	2	2.6	2.3	1.9	1.8	2.51	1.8	1.8	2.5
					5						6	2	5	3				
2	5	3.12	0	4.66	4.7	5	2	0.45	5	3	0.3	4.5	2.3	4.0	0	4.4	4.25	2
					7						6	7	2	5				
3	2.19	3.12	0	3.65	4.3	0	5	2.27	0.62	5	3.8	4.3	2.5	4.7	0	5	5	3
					7						7	1		7				

. Resultado método del centroide

4.5. Clúster 1: Compuesto por aves con baja mortalidad y peso final. Provistos de una calidad de agua un poco más baja, que los otros clústeres y un bajo número de aves recibidas y entregadas.

4.6. Clúster 2: Conformado por aves con un peso final muy bueno, y una mortalidad media, respecto a los otros clústeres. Criados en una condición climática con temperaturas altas, alimentados con gran cantidad de alimento y recibidas con un alto peso inicial. En cuanto a la mortalidad, las principales

Se creó un dendrograma del método del vecino más cercano con fines demostrativos.

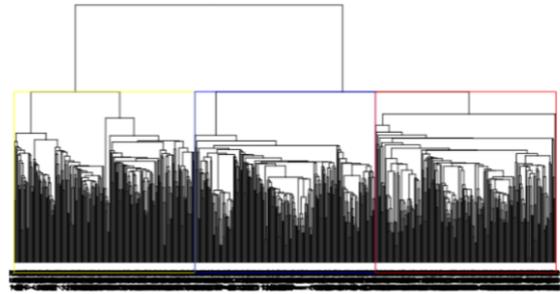


Figura 8. Dendrograma vecino más cercano

En la figura 9 se observa cómo se agrupan los datos en tres clústeres identificados por los rectángulos de colores amarillo, azul y rojo, según la distancia entre ellos.

A continuación, se muestra la tabla obtenida a partir del método de clustering jerárquico con el método del centroide. Es importante aclarar que en esta se utilizó una nomenclatura que tiene en cuenta el orden de las variables del dataset, tal como se puede observar en la tabla 2, obtención de datos.

causas fueron otros y la calidad del agua. En cuanto al peso final, la cantidad de alimento consumida es de gran importancia.

4.7. Clúster 3: conformado por aves con una alta mortalidad y con el mejor peso final. Una de las principales causas de la mortalidad es el ahogamiento, semilla defectuosa, huevo sin vacuna y en cierta medida la calidad del agua. En tanto al peso final, el consumo de alimento tuvo gran influencia.

En resumen, la alta mortalidad está ligada al ahogamiento, a la semilla defectuosa, al huevo sin vacuna y la calidad del agua, por lo que es necesario encontrar medidas para evitar esto. En cuanto al peso final, es importante tener en cuenta el consumo de alimento de estos animales para tener un buen rendimiento.

4.8. Modelo descriptivo kmeans

El análisis del modelo descriptivo con el algoritmo kmeans, se realizó apoyado en un mapa de calor, el cual determinó el análisis de los clústeres, al estar conformado por celdas de colores, lo cual indica su valor. En esta ocasión el color rojo indica valores altos, mientras que el color azul indica valores bajos.

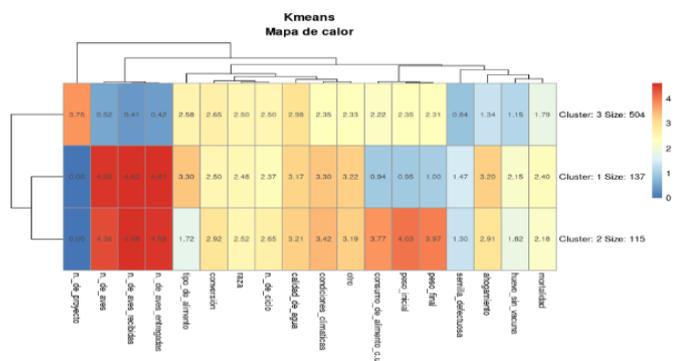


Figura 9. Mapa de calor kmeans

En la figura 10. Se puede observar que al igual que el método del centroide, para el algoritmo kmeans, se agruparon los datos en tres clústeres.

4.9. Clúster 1: conformado por 137 datos. Se observa una mortalidad media con pesos iniciales y finales bajos. El peso final bajo está relacionado con un bajo consumo de alimento. Además, las aves fueron provistas por agua de una calidad media y unas condiciones climáticas un poco altas.

4.9.1. Clúster 2: conformado por 115 datos. Se logra observar una baja mortalidad y un alto peso final debido al alto consumo de alimento que se logra evidenciar. Las aves fueron provistas de una calidad de agua un poco más alta que el clúster 1. Lo mismo ocurre con las condiciones climáticas.

4.9.2. Clúster 3: Compuesto por 504 datos. Se observa una baja mortalidad y un peso final medio. Con una condición climática y calidad de agua bajos, respecto a los otros clústeres.

En resumen, se observan diferencias significativas en la mortalidad y el peso final de las aves en función de la calidad de agua, el consumo de alimento y las condiciones climáticas

en las que fueron criados los animales para los tres clústeres identificados con el algoritmo kmeans.

4.9.3. Resultados modelos predictivos

4.9.4. Modelo de regresión lineal simple

El modelo de regresión lineal simple, el cual fue usado para predecir el peso final de las aves en base al consumo de alimento es el siguiente:

$$y = 2.22992 + 3.70628 * x$$

Ec.4 Regresión lineal simple tal como se observa en la **Tabla 3.**

Coeficientes	
(Intercept)	2.22992
consumo_de_alimento_c.u	3.70628

La figura anterior contiene los coeficientes del intercepto y la variable de “consumo_de_alimento_c.u”, necesarios para la ecuación de regresión lineal simple.

Para probar el modelo, se creó un dataset con cinco nuevos datos de la cantidad de alimento, para observar cómo el peso final, está relacionado con el consumo de alimento de cada ave de corral, como se muestra en la **Tabla 4.**

Tabla 5. proyectos 1

Consumo_de_alimento_c.u	Predicción_peso_final
1	2596.63
2	1855.37
3	558.17
4	2781.94
5	2967.25

En la figura anterior se observan los datos del dataset creado para poner a prueba el modelo de regresión lineal simple.

Al cargar los nuevos datos al modelo de regresión lineal, se obtuvo como resultado que, a mayor cantidad de alimento consumido por los animales, mayor será el peso final de las aves, tal y como se esperaba según los análisis de los clústeres.

4.9.5. Modelo de regresión lineal múltiple

Fueron usados para predecir la mortalidad de las aves con base a distintas variables independientes por cada uno de los tres proyectos. Debido a la extensión de los modelos y del pequeño valor de los coeficientes, estos se omitirán y se presentan directamente los resultados obtenidos por estos modelos.

Para probar el modelo de los proyectos 1 y 2, se creó un dataset con datos observados en las tablas 5 y 6 que a continuación se relacionan. En el cual se tomaron cinco datos aleatorios del conjunto de

Item	condiciones climáticas	calidad de agua	tipo de alimento	consumo de alimento_c/u	peso inicial	raza	ahogamiento	semilla defectuosa	huevo sin vacuna	otro	conversión	peso final	n° de aves recibidas	n° de ciclo
1	25	normal	levante	540	1913	cobbers	2	3	1	3	70	1983	8400	CICLO_UNO
2	26	normal	engorde	570	1983	ross	0	1	9	0	73	1995	8500	CICLO_DOS
3	18	ph_elevado	levante	560	1913	ross	3	7	1	3	70	1983	8400	CICLO_TRES
4	22	normal	engorde	600	2159	cobbers	1	13	1	1	14	1964	8400	CICLO_CUATRO
5	22	normal	engorde	600	2173	cobbers	3	6	1	2	25	1989	8400	CICLO_CUATRO

Tabla 6 . proyectos 2

condiciones climáticas	calidad de agua	tipo de alimento	consumo de alimento_c/u	peso inicial	raza	ahogamiento	semilla defectuosa	huevo sin vacuna	otro	conversión	peso final	n° de aves recibidas	n° de ciclo
29	normal	engorde	600	2143	ross	5	5	13	5	16	2162	14500	CICLO_DOS
23	normal	engorde	560	2133	ross	2	5	1	5	10	2147	14500	CICLO_TRES
20	normal	engorde	560	2143	cobbers	1	3	4	5	16	2160	15000	CICLO_CUATRO
31	normal	engorde	490	2133	ross	0	1	1	0	10	2146	15000	CICLO_CINCO
22	normal	engorde	500	2143	ross	0	0	0	0	16	2162	15000	CICLO_CINCO
29	normal	engorde	520	2143	cobbers	5	2	4	3	16	2161	15500	CICLO_SEIS

Debido a la extensión del nombre de las variables, se utilizó una nomenclatura siguiendo el orden de las siguientes variables: condiciones_climaticas, calidad_de_agua, tipo_de_alimento, consumo_de_alimento_c.u, peso_inicial, raza, ahogamiento, semilla_defectuosa, huevo_sin_vacuna, otro, conversión, peso_final, n._de_aves_recibidas y n._de_ciclo

Tabla 7

. proyectos 3

Item	condiciones climáticas	calidad de agua	tipo de alimento	consumo de alimento_c/u	peso inicial	raza	ahogamiento	semilla defectuosa	huevo sin vacuna	otro	conversión	peso final	n° de aves recibidas	n° de ciclo
1	26	normal	engorde	598	2173	cobbers	1	1	1	1	25	2263	7100	CICLO_UNO
2	24	normal	engorde	598	2198	cobbers	2	1	1	1	29	2292	7100	CICLO_UNO

3	17	normal	engorde	640	2257	ro ss	0	0	12	0	46	2304	7100	CICLO _DOS
4	26	normal	engorde	610	2227	ro ss	1	2	1	2	30	2263	7200	CICLO _TRES
5	20	normal	engorde	630	2257	ro ss	2	1	2	1	46	2303	7100	CICLO _CINCO

Para la nomenclatura de los nuevos datos del proyecto 3, se siguió el siguiente orden: condiciones climáticas, calidad de agua, tipo de alimento, consumo de alimento c.u, raza, ahogamiento, semilla defectuosa, huevo sin vacuna, otro, conversión, n. de aves recibidas, n. de aves entregadas y n. de ciclo.

Resultados predicciones de mortalidad con regresión lineal múltiple

Tabla 8. predicciones de mortalidad con regresión lineal múltiple

	calidad de _agua	ahogami ento	huevo sin vacuna	semillas def ectuosa	ot ro
Proye cto1	10	13	12	9	15
Proye cto2	6	8	7	7	5
Proye cto3	19	15	27	20	15

Para el análisis de los resultados se hará énfasis en las variables calidad de agua, “ahogamiento”, “huevo sin vacuna”, “semillas defectuosa” y “otro” las cuales están relacionadas con la mortalidad según el análisis de los clústeres.

En general, se logra observar que el proyecto 3, tiene una gran mortalidad en comparación con los otros proyectos. La razón puede ser debida a la calidad de agua, ya que si se observa esta variable (x2) de los nuevos datos de cada proyecto. Es el proyecto 3 quien contiene una menor calidad de esta. Lo mismo ocurre en la variable ahogamiento (x7 para proyectos 1 y 2 – x6 para proyecto 3), en donde los valores de estas variables son más altos para el proyecto 3.

Continuando con lo anterior, la variable en la que se observa una gran diferencia en sus valores es semilla defectuosa (x8 para proyectos 1 y 2 – x7 para proyecto 3), nuevamente para el proyecto 3, superando por mucho los valores de esta variable en los otros proyectos. Esto sugiere, la necesidad de buscar posibles soluciones para este problema, puesto que es la variable que más influyó en la mortalidad del proyecto 3, junto con la calidad del agua con la que fueron provistos los animales.

4.9.6. Resultados análisis de residuos regresión múltiple modelo 1 y 2.

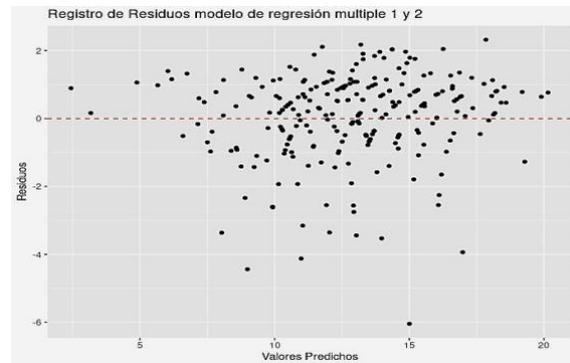


Figura 10. Grafica análisis de residuos regresión múltiple modelo 1 y 2

Los residuos de los modelos 1 y 2 están dispersos uniformemente alrededor de cero en un gráfico sin patrones discernibles, por lo tanto, se considera que se cumple una buena indicación que el modelo de regresión se ajusta bien.

4.9.6. Resultados análisis de residuos regresión múltiple modelo 3

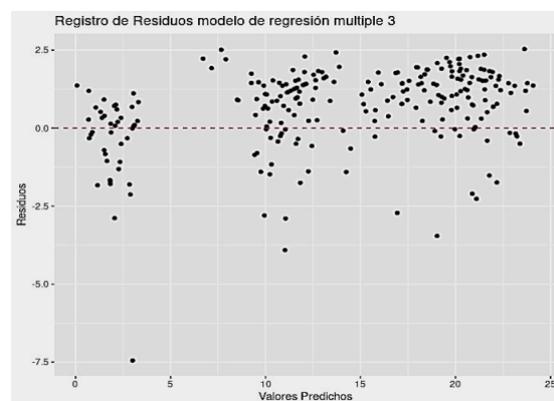


Figura 11. Grafica análisis de residuos regresión múltiple modelo 3

Los residuos del modelo 3 están organizados de forma aleatoria por lo tanto el modelo de regresión múltiple se ajusta perfectamente a la predicción de la respuesta calculada.

Con respecto a los trabajos [2] y [3] existe una similitud donde en el primero se realizó un análisis del consumo de alimento por ave perteneciente a la línea Ross y la conversión alimenticia

de estos, así como el porcentaje de mortalidad mediante el software R Studio utilizando una estadística descriptiva para lo cual se emplean variables similares como: conversión, alimento y edad, pero aplicadas con un enfoque diferente. Como también existe un grado de similitud entre nuestra investigación con el segundo trabajo citado al aplicar el algoritmo de regresión múltiple donde el objetivo de estimar el valor promedio y/o estimar un determinado valor y , a partir de información de k variables independientes o predictor x . Resulta la correlación de las variables “consumo de alimento” vs “conversión alimenticia”, “consumo de agua” vs “temperatura”, “ganancia de peso” vs “conversión alimenticia”, “consumo de agua” vs “consumo de alimento”, esto significa que la relación múltiple entre las dos variables es más fuerte cuanto más cercano a uno esté el valor del coeficiente de correlación en ambas direcciones. Cuanto más cercano a cero sea el coeficiente de correlación, más débil será la conexión entre las dos variables.

CONCLUSIONES

En esta investigación, se emplearon diversas técnicas analíticas y herramientas estadísticas en RStudio para analizar el rendimiento y la mortalidad en la cadena de producción avícola. El objetivo principal fue identificar patrones, factores y posibles soluciones que puedan mejorar la eficiencia y reducir la mortalidad en la industria avícola.

En el proyecto avícola, se emplearon diversas técnicas de regresión para modelar los datos. Después de un análisis exhaustivo, se llegó a la conclusión de que la regresión múltiple se ajusta mejor a los datos recopilados. Esto significa que, al considerar múltiples variables predictoras en lugar de una sola, el modelo de regresión múltiple proporcionó una representación más precisa de la relación entre estas variables y la variable de respuesta en comparación con otras técnicas de regresión.

A través del análisis de regresión múltiple, se identificaron factores determinantes del rendimiento avícola como la calidad del alimento, la calidad de agua, la genética y el manejo sanitario tienen una influencia significativa en el rendimiento avícola. La correcta optimización y balance de estos factores pueden conducir a un aumento en la productividad y la rentabilidad de la cadena de producción avícola.

A través de la aplicación del análisis de clustering, se identificaron factores importantes que afectan la producción y el rendimiento avícola. En relación al peso final de las aves, se encontró que la calidad de la alimentación es un factor importante a tener en cuenta. En lo que concierne a la tasa de mortalidad de estos animales, la aplicación de estas técnicas de análisis de datos, posibilitó el descubrimiento de que los

factores relacionados con lo antes mencionado son la calidad del agua, semilla defectuosa, huevo sin vacuna y ahogamiento.

Mediante el empleo del análisis de regresión lineal simple, se logró identificar la notable influencia que la alimentación tiene sobre el peso final de los pollos. Este hallazgo sugiere que una alimentación de mayor calidad para los pollos resulta en un incremento proporcional en su peso final, lo cual conlleva a un rendimiento superior en términos de producción.

Aunado a lo anteriormente expuesto, la realización de un análisis de regresión lineal múltiple, permitió identificar como los factores asociados a una mayor tasa de mortalidad influyen en estas aves. Los resultados sugieren la necesidad de buscar soluciones con el propósito de prevenir estos problemas, y lograr optimizar el balance de dichos factores, es posible predecir un aumento tanto en la productividad como en la rentabilidad de la cadena de producción avícola.

Mediante el análisis de datos relacionados con la mortalidad, los factores de riesgo asociados con tasas más altas de mortalidad, como las enfermedades contagiosas y las condiciones ambientales desfavorables. La detección temprana de estos factores y la implementación de medidas preventivas adecuadas son fundamentales para reducir la mortalidad y, por ende, mejorar la eficiencia de la producción avícola.

Además, se debe prestar especial atención a los factores determinantes del rendimiento y a las medidas preventivas para reducir la mortalidad, con el objetivo de mejorar la productividad y la sostenibilidad de la cadena de producción avícola.

Colocar aporte al avicultor Con base en la experiencia se confirma que el presente estudio está acorde con la práctica, ya que las variables identificadas en este artículo como influyentes en la mortalidad coinciden en su gran mayoría con las causas identificadas empíricamente. Como lo son: “calidad_de_agua”, “ahogamiento”, “huevo_sin_vacuna”, “semillas_defectuosa” y “otro”. Así como también las variables influyentes en el rendimiento; como lo son: “peso_inicial”, “consumo_de_alimento_c.u”, “peso_final”. Siendo de gran importancia al predecir los resultados del ciclo mediante un estudio.

En última instancia, los resultados obtenidos en esta investigación ofrecen valiosas perspectivas y pautas para la toma de decisiones informadas, buscando la optimización del rendimiento y la reducción de la mortalidad para garantizar una producción sostenible y rentable en el futuro. Lo que fomentará una producción avícola cada vez más eficiente y sostenible.

AGRADECIMIENTOS

En primera instancia agradecer a los tres proyectos avícolas quienes compartieron la información y recolección de los datos para realizar este estudio de investigación, de igual manera a la Fundación Universitaria de Popayán quien fue nuestro eje fundamental aportando las herramientas suficientes para el resultado del mismo así mismo al grupo IMS quien nos brindó las asesorías y acompañamiento para mostrar este gran resultado.

REFERENCIAS

- [1] B. Aponte, A. González, y A. González, «Fases de la cadena de suministro de las empresas avícolas.», *Revista Venezolana de Gerencia*, vol. 18, n.o 64, 2014, doi: 10.31876/revista.v18i64.11178
- [2] E. Ruiz, «Presentación: Minería de Datos Inteligencia Artificial», *Revista Iberoamericana de Inteligencia Artificial*, vol. 10, pp. 7-9, 2006, [En línea]. Disponible en: <http://www.redalyc.org/articulo.oa?id=92502901>
- [3] E. Olazábal, Y. García, y H. Z. Serrano, «TENDENCIAS DE LA MORTALIDAD POR ENFERMEDADES ENTERICAS BAJO LA INFLUENCIA DE LAS MEDIDAS PREVENTIVAS EN UNA».
- [4] R. Quiroga Martínez y United Nations. Economic Commission for Latin America and the Caribbean., *Indicadores ambientales y de desarrollo sostenible : avances y perspectivas para América Latina y el Caribe*. CEPAL, 2007. [21]
- [5] «2calidad_agua_higienizacion_avelina_bellostas_texto_46_symp_aeca».
- [6] «Proyecto Final-Characterización Sistema de Producción Animal».
- [7] C. P. Torres-Vinueza, L. J. Ron-Garrido, y J. E. Grijalva-Olmedo, «Evaluación de factores de riesgo que afectan la mortalidad en pollos de engorde durante el proceso de traslado granja-planta de faenamiento en el centro norte de la región interandina», *Siembra*, vol. 8, n.o 1, abr. 2021, doi: 10.29166/siembra.v8i1.2559.
- [8] B. Jonathan Joel Meca Avalo, «UNIVERSIDAD NACIONAL DE PIURA “IMPLEMENTACION DE UN PLAN HACCP PARA LA EMPRESA AVIMEC SAC DEDICADA A LA CRIANZA Y FAENADO DE AVES UBICADA DE INVESTIGACION INOCUIDAD ALIMENTARIA», 2022.
- [9] «Vista de Análisis de parámetros productivos de pollos de engorde en una avícola comercial - municipio de Cáqueza – Cundinamarca».
- [10] Y. Cuarta, S. Sexta, P. De Producción En, E. De, y L. A. De, «INCIDENCIA Y MORTALIDAD CAUSADA POR SÍNDROME ASCÍTICO ENTRE LA».
- [11] «2020_Tesis_Triana_Ardila_Santiago».
- [12] A. Lenin y P. Arguello, «ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL», 2007. [27]
- [13] E. Oscar Daniel Maldonado Pinto Asesores Rommel Reconocido M. Luis Sandoval, A. Tanya Müller García, A. M. Maier Acosta, R. Soto, y H. Zavala Membreño, «Escuela Agrícola Panamericana, Zamorano».
- R. Productivo et al., «UNIVERSIDAD NACIONAL DEL CENTRO DEL PERÚ FACULTAD DE ZOOTECNIA TESIS PRESENTADA POR».
- C. P. Torres-Vinueza, L. J. Ron-Garrido, y J. E. Grijalva-Olmedo, «Evaluación de factores de riesgo que afectan la mortalidad en pollos de engorde durante el proceso de traslado granja-planta de faenamiento en el centro norte de la región interandina», *Siembra*, vol. 8, n.o 1, abr. 2021, doi: 10.29166/siembra.v8i1.2559.
- M. Q. Quispe, «UNIVERSIDAD MAYOR DE SAN ANDRES, LA LINEA ROSS 308 EN LA ESTACION EXPERIMENTAL DE COTA COTA», 2009.
- «3Interacción de los factores ambientales nacional».
- M. J. Barreto Beltrán, Y. Fierro Rojas, D. M. Torres Novoa, y D. E. Cortes Martínez, «Análisis de parámetros productivos de pollos de engorde en una avícola comercial - municipio de Cáqueza – Cundinamarca», *Agricolae & Habitat*, vol. 2, n.o 1, nov. 2019, doi: 10.22490/26653176.3519.
- M. Antonio y E. Mina, «RETOS Y PERSPECTIVAS DE LAS TECNOLOGÍAS DE INFORMACIÓN CAPITULO III. CRISP-DM: CONOCIMIENTO Y COMUNICACIÓN DE UNA METODOLOGÍA PARA MINERÍA DE DATOS. Autor».
- N. Stefany, P. Saavedra, I. Alejandra, y Z. Gallo, «Tesis para optar el Título de Ingeniero Industrial y de Sistemas».
- «UNIVERSIDADE FEDERAL FLUMINENSE ESCOLA DE ENGENHARIA DEPARTAMENTO DE ENGENHARIA QUÍMICA E DE PETRÓLEO YAGO MIRANDA BENITES ANÁLISE EXPLORATÓRIA DE DADOS DO PROCESSO DE REFORMA SECA DO METANO NITERÓI 2/2022».
- J. Luis y V. Villardón, «INTRODUCCION AL ANÁLISIS DE CLUSTER».
- F. M. Hernández Arellano Millward Brown México, «El Concepto de Distancia y su Aplicación en Estadística Multivariada».
- R. J. Juárez, «Ganadería de precisión, una revisión a los avances dentro de la avicultura enfocados a la crianza de pollos de engorde.», *Prisma Tecnológico*, vol. 14, n.o 1, pp. 38-48, feb. 2023, doi: 10.33412/pri.v14.1.3652.
- S. Valero Orea, A. Salvador Vargas, y M. García Alonso, «Minería de datos: predicción de la deserción escolar mediante el algoritmo de árboles de decisión y el algoritmo de los k vecinos más cercanos».
- A. Centeno, M.-R. Directora, M. Jesús, y G. Abad, «TRABAJO FIN DE CARRERA Big Data. Técnicas de machine learning para la creación de modelos predictivos para empresas».
- «Regresion_lineal_multiple (1)».
- T. Chamba, A. Cristina, F. Carrera, y E. Rodrigo, «Desarrollo de un modelo de gestión de datos, que permita la posibilidad de

disminuir la concentración de clientes, estudio de caso: PROING CIA LTDA», 2021.

- [29] «47 - Machine learning essentials Practical Guide in R».
- [30] «Capítulo 6 Análisis de Regresión Simple».
- [31] J. M. Rojo Abuín, «Regresión lineal múltiple».
- [32] «Documento_completo.pdf-PDFA.pdf (1)».