MERU UNIVERSITY OF SCIENCE AND TECHNOLOGY

DEPARTMENT OF COMPUTER SCIENCE

BACHELOR OF SCIENCE IN DATA SCIENCE

PREDICTIVE MODEL FOR PROACTIVE LAYER FEED EFFICIENCY USING RANDOM FOREST REGRESSOR

CT2O4/106242/21: SILAS OCHIENG

A Research Project Submitted in Partial Fulfillment of the Requirements of the Bachelor of Science in Data Science of Meru University of Science and Technology

April, 2025

DECLARATION

This research project is my original work prepared with no other than the indicated sources and support, and has not been presented elsewhere for a different or similar assignment.

Name: SILAS OCHIENG'

Registration Number: CT204/106242/21

Signature______________________ Date_________________________

ACKNOWLEDGEMENT

I would want to express my sincere gratitude to my adviser, Mr. Anthony Irungu, whose knowledge and direction have greatly influenced this project. Your insightful advice and unwavering assistance have made this endeavour much better. Also deserving of my gratitude are the faculty and staff of department of computer science, whose resources and support have been invaluable.

Special thanks to my fellow classmates and friends for their continuous support, insightful discussions, and camaraderie. Your encouragement kept me motivated, especially during challenging times. Lastly, I wish to thank my family for their unconditional love, patience, and belief in me. This accomplishment would not have been possible without your constant support and sacrifices. Thank you all for being a part of this journey.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

INTRODUCTION

1.1 Background of study

One of the most important aspects of chicken farming is egg production, and getting the best results needs a better knowledge of feeding techniques. (Narváez-Solarte et al., 2006a) A thorough approach to feeding can result in greater improvements in the amount and quality of eggs laid, as proper nutrition is a greater approach to raising laying hen production.

For laying hens to produce a high volume of eggs, their dietary requirements must be met. Proteins, amino acids, calcium, phosphorus, and other vitamins and minerals are essential for meeting these needs. Because they aid in the development of the yolk and white of eggs. Crucial amino acids like lysine and methionine are also important for protein synthesis and the general quality of eggs. Specifically, methionine is necessary for the synthesis of egg proteins.(Pousga et al., 2005)

Another essential mineral for the greater robust eggshells is calcium. Compared to their non laying counterparts, laying hens need higher calcium levels, and layer meals typically contain 3.5–4.5% calcium.(Pousga et al., 2005)In order to maintain the quality of eggshells and the health of the skeleton, phosphorus and calcium must be in the right proportion for the body to effectively use nutrients. In addition, minerals like zinc and selenium, as well as vitamins A, D3, and E, are critical for sustaining general health and regulating egg and the health of the skeleton, phosphorus and calcium must be in the right proportion for the body to effectively use nutrients. In addition, minerals like zinc and selenium, as well as vitamins A, D3, and E, are critical for sustaining general health and regulating egg production. The metabolism of calcium requires vitamin D3, which emphasizes how important nutrients are in the diets of chickens.

In order to provide laying hens with the nutrition they require, the type of feed utilized is essential. Layer rations are designed especially for hens that are laying eggs; they usually have higher calcium content and a more balanced protein content. There are other kinds of these foods available, such as crumbles, mash, and pellets. Pelleted feeds are frequently used because they provide a regular nutrient intake and minimize waste.(Pousga et al., 2005)

Some farmers use supplemental feeds, like oyster shells or limestone, in addition to conventional layer foods to increase consumption of calcium.(Narváez-Solarte et al., 2006b)These supplements are especially helpful in high-production systems where chickens might need more calcium to keep the integrity of their eggshells

Egg production can only be maximized by employing effective feeding practices(King'Ori, 2011)Ad libitum feeding is a popular method in which chickens are provided with unlimited access to eat all day long. Although this approach can boost output, if it is not carefully controlled, it may also result in higher feed costs and possible health problems. On the other hand, restricted feeding lowers feed intake to increase feed conversion ratios and manage weight. This strategy can keep egg production steady and prevent obesity, but it needs to be carefully managed to make sure that nutritional requirements are still being satisfied

Egg production is significantly influenced by efficient management procedures in addition to the feed itself. Creating a regular feeding schedule promotes the general health of the chickens and aids in controlling lay patterns. (Roberts, 2004) To avoid spoiling and contamination, feed must be stored properly. Feed must be kept cold and dry to maintain its nutritious value and stop dangerous bacteria from growing. contamination, feed must be stored properly. Feed must be kept cold and dry to maintain its nutritious value and stop dangerous bacteria from growing.

The environment and genetics have a big influence on egg production as well. (Ledvinka et al., 2012) Egg yield varies throughout breeds; hybrid layers, including the ISA Brown and Lohmann Brown, are frequently selected due to their high production rates. Egg production can also be impacted by environmental factors including stress levels, housing quality, and light exposure. Enough lighting, usually 14–16 hours a day, promotes the hens' biological rhythms and aids in maintaining regular egg production.

The science of feeding chickens is always changing, with new studies looking into the effects of different feed additives and substances on egg production. (Coffey et al., 2016) Probiotics, prebiotics, and enzyme supplements are among the innovations whose potential benefits in raising egg production and increasing feed efficiency are being studied. These developments should give farmers new resources to help them improve their feeding strategies even more.

Furthermore, having access to fresh, clean water is essential for healthy digestion and general wellbeing. Changing the water in the henhouse on a regular basis keeps her hydrated and helps prevent disease, both of which are essential for best egg production(Oviedo-Rondón, 2019)

In order to modify feeding tactics as necessary, regular monitoring of the health and egg production rates of hens is essential. Through the observation of production patterns and general health, farmers are able to identify problems early on and modify feed to maximize yield. (Roberts, 2004)

1.2 Motivation for study

The increased use of data analytics in agriculture to improve farming operations is the driving force behind this study. The research attempts to optimize laying hen feeding practices, thereby enhancing egg production, by utilizing statistical models and machine learning. Using cutting-edge analytical methods to solve practical problems and improve productivity and sustainability in the poultry sector is one of the exciting opportunities that comes with integrating data science into agriculture. The research hope to provide insightful information that will help to advance agricultural innovation in the future and change feeding habits.

1.3 Problem statement

Even with greater improvements in poultry management and nutrition, many farms still face various challenges when it comes to feed efficiency in egg production. Excessive feed costs make up a significant amount of total production costs, and inefficiencies can result in higher resource consumption and worse profitability. This issue is increased by variables in feed quality, unusual hen health indicators, and a deficiency of feeding plans that are majorly designed for given production circumstances. Therefore, it is better to gain a greater understanding of the variables to improve feed efficiency which is essential to understand feed prices and need for sustainable practices.

Present-day techniques for evaluating feed efficiency frequently depend on conventional methodologies, which may miss important information hidden in complex datasets. Poultry farmers lose out on opportunities for optimization that could improve production results because of a lack of detailed analysis. This research uses data science techniques to find relationships between different types of farm data. The ultimate goal is to create prediction models that will inform more efficient feeding practices. By using data-driven insights to address these inefficiencies, egg production will become more economically viable while also advancing sustainable farming practices.

1.4 Research objectives

1.4.1 General objective

To develop a regression model to predict feed conversion ratio from past data, enabling proactive management choices.

1.4.2 Specific objectives

The objectives of this research project are:

i. To identify a historical feed data on layer hens including variables such as Total feed intake, FCR, nutrient composition
ii. To explore dataset to understand data distributions, relationships, and trends.
iii. To develop a regression model to predict feed efficiency
iv. To evaluate the regression performance
v. To deploy the model using flask web application

1.5 Significance of the study

This work, which optimizes feeding techniques using data-driven ways to overcome issues in egg production, is important for researchers and society alike. Improving the sustainability and efficiency of chicken production contributes to food security by making wholesome food more accessible. It provides an opportunity for academics to work with cutting edge data science techniques and add to the body of knowledge on poultry nutrition in academic publications. The results will inform the formulation of policies that support sustainable practices in the poultry industry, better resource management, and food safety. They will also offer evidence-based recommendations for enhancing agricultural practices and hen welfare.

1.6 Scope of the study

The scope of study will encompasses using data science tools to improve feed efficiency in egg production. In order to find trends and inefficiencies that traditional approaches frequently miss, it will include gathering and analyzing complicated datasets from chicken farms. The goal of the study is to investigate factors including feed quality, health measures, and feeding schedules designed for certain production circumstances. The goal of the project is to produce actionable insights that will assist farmers in optimizing their feeding practices, lowering feed costs, and increasing overall profitability by creating predictive models based on this investigation. In the end, the intention is to encourage the chicken sector to adopt more ecologically friendly and economically viable farming methods, thus enhancing the sustainability of egg production.

1.7 Assumptions in the study

This study assumes that reliable data on feed types, hen health, and egg production will be accessible from poultry farms. It presumes consistency in feeding practices among participants and a measurable correlation between health indicators and feed efficiency. The hens are expected to be similar breeds, minimizing genetic variation. External factors will be controlled to isolate feed efficiency impacts, and the developed predictive models will be generalizable across similar farming contexts. Lastly, it is assumed that farmers will be motivated to adopt data-driven practices for enhanced profitability and sustainability.

1.8 Limitations of study

There are various potential limitations to this study that could affect the results. Analysis depth may be hampered by a lack of high-quality data on feed kinds, hen health, and productivity. A small sample size may limit how far the results may be applied. Comparisons may become more difficult due to differences in agricultural methods, management approaches, and environmental factors between farms. Unmeasured variables may also have an impact on the complex interactions that exist between egg production, health markers, and feed efficiency. Logistical or financial limitations may also make it difficult to implement successful ideas.

CHAPTER TWO
LITERATURE REVIEW

2.1 Introduction

The optimization of poultry feed is crucial in animal husbandry, balancing economic viability, nutrition, and sustainability. Laying hens require specific nutritional formulations throughout their life cycle to enhance production and meet farmers' economic goals. Recent environmental concerns have driven research into minimizing the ecological impact of feed production. This literature review examines key studies on the formulation, optimization, and sustainability of poultry feed, highlighting advancements and future directions. The findings emphasize the importance of nutritional balance, cost-efficiency, and environmental responsibility in the poultry industry.

2.2 Automating poultry farm management with artificial intelligence

(Depuru et al., 2024)highlight the critical role of the broiler industry in meeting the growing global demand for poultry meat while emphasizing the need for effective monitoring systems to maintain optimal productivity and animal welfare. A key challenge in broiler production is the rising mortality rate, necessitating proactive disease detection strategies. The study presents a computer vision-based approach utilizing the YOLOv5s (You Only Look Once) model, which achieved a classification accuracy of 96% in identifying and categorizing broiler chickens based on age.

The model, after custom training, is converted into the ONNX (Open Neural Network Exchange) format and integrated with a centroid tracker for real-time tracking. Deployment is facilitated using the OpenCV library on a local machine, with the processed data stored in a MySQL database for further analysis. The system effectively classifies broilers into four age groups (1–4 weeks), addressing the critical period for monitoring flock health. Potential applications of this model extend to tracking temperature, weight, and flock behavior, thereby enhancing poultry management practices.

This research contributes to the broader field of automated livestock monitoring, demonstrating the potential of AI-driven solutions in improving disease prevention and overall flock management.

2.3 Predictive Modeling in Poultry Farming

(Omomule et al., 2020) introduced a fuzzy predictive model for improving the accuracy of egg production cycle predictions in poultry farming. The model was designed to handle the nonlinear and complex factors affecting egg production, such as feed quality, genetic traits, and disease prevalence. By applying fuzzy logic, the model achieved 100% prediction accuracy for Pred (30) and demonstrated superior performance compared to traditional predictive techniques. This method provides a cost-effective solution for farmers seeking to optimize egg production in real-time, offering a practical tool for addressing the uncertainties inherent in poultry farming.

2.4 Artificial intelligence and iot based monitoring of poultry health

(Singh et al., 2020) present a comprehensive review of poultry health monitoring systems utilizing Internet of Things (IoT) platforms integrated with Artificial Intelligence (AI) techniques. The study emphasizes the significance of adopting modern technologies such as sensors, video/image processing, and vocalization analysis to monitor poultry farms and assess bird health effectively. These IoT-based solutions incorporate classification algorithms and real-time data collection to enable continuous and automated surveillance of large-scale poultry operations.

The authors further note that the increasing availability of cost-effective computational resources and standardized algorithms has accelerated the implementation of such technologies. They argue that the integration of AI and IoT in poultry farm management is essential, especially considering the crucial role of poultry products as a primary source of protein. By leveraging these advanced technological solutions, farm productivity and bird welfare can be significantly improved through timely health monitoring and early detection of anomalies.

2.5 Model applications in poultry production and nutrition

(Oviedo-Rondón, 2015) discusses the development and application of mathematical models in poultry production, highlighting their significance in decision-making, research, and teaching. These models provide a structured approach to understanding complex problems related to poultry enterprise management, live production, and nutrition. By offering a systematic means of analyzing poultry production systems, mathematical models help optimize key aspects such as growth patterns, feed efficiency, and overall resource utilization. The study outlines various advancements in mathematical modeling, including models for bird growth, egg production, and enterprise management.

Despite the advantages of modeling techniques, their adoption in commercial poultry production remains limited due to challenges such as a lack of knowledge, training, and data availability. Many poultry companies and nutritionists do not utilize biological models daily, and these techniques are not widely taught in animal and poultry nutrition programs. The study emphasizes the need for greater awareness and integration of these tools in both research and commercial settings to maximize their benefits for the poultry industry. Enhancing access to training and improving data collection methods could facilitate the broader implementation of mathematical models, ultimately leading to better decision-making and efficiency in poultry production systems.

2.6 A framework for modelling, control and supervision of poultry farming

Lorencena et al. (2020) emphasize the importance of thermal comfort in broiler chicken production, as it directly affects food consumption and meat production efficiency. Temperature and humidity are critical factors in maintaining thermal balance, yet conventional poultry farming still relies heavily on expert observation and manual control, which can be inefficient and error-prone. Their research proposes a framework that integrates modern technologies such as sensor networks, control theory, and remote monitoring to automate climate control within poultry houses.

The study introduces a plant architecture that facilitates real-time monitoring and decision-making for maintaining optimal poultry house conditions. A controller is developed to observe environmental changes and adjust actuators accordingly while ensuring minimal restriction and compliance with operational requirements. Additionally, the framework includes a web-based supervision system that enables remote monitoring and intervention, enhancing user control over environmental parameters. By integrating automation and remote access, this study highlights the potential for improving poultry farm efficiency, reducing labor dependency, and ensuring better environmental conditions for birds.

2.7 Technological Innovations in Poultry Management

Recent technological advancements in poultry management, particularly concerning artificial lighting, have significant implications for hen welfare and productivity. (Bahuti et al., 2023) investigated the physiological responses of laying hens to varying artificial lighting intensities, developing predictive models to assess impacts on welfare and productivity. Their findings suggest that advanced modeling approaches, such as fuzzy inference systems, can significantly enhance decision-making in poultry management by predicting responses to different environmental

2.8 Research gap summary

Despite significant advancements in poultry feed formulation and sustainability, several key research gaps persist that warrant further exploration:

i. Underutilization of Data Analytics

Although studies have addressed the formulation and sustainability of poultry feed (Belkhanchi et al., 2023);(Heidari et al., 2021) there is insufficient emphasis on using data analytics to optimize feed efficiency and predict performance outcomes. Current methodologies primarily focus on static formulations without leveraging real-time data or machine learning techniques.

i. Lack of Integrated Models

Existing research has explored nutritional and genetic factors influencing egg production and quality (Luchkin et al., 2021); (Bryden et al., 2021) However, there is a lack of integrated models that incorporate environmental variables, health indicators, and feeding practices to enhance decision-making in feed management

ii. Integration of Technology and Data Science

Although studies highlight technological innovations in poultry management (Bahuti et al., 2023), there is a gap in evaluating how these technologies can be integrated with data science approaches to enhance feed efficiency and sustainability.

2.9 Summary

The optimization of poultry feed is crucial for economic viability, nutritional adequacy, and environmental sustainability in poultry farming. Research emphasizes the need for balanced formulations tailored to different growth stages. Belkhanchi et al. (2023) used linear programming to optimize feed compositions, while Heidari et al. (2021) integrated life cycle assessments into feed optimization models to minimize environmental impact while maintaining cost-efficiency. These studies highlight the balance of cost, nutrition, and sustainability in feed production, particularly given rising environmental concerns.

Advances in predictive modeling and technology present new opportunities for enhancing poultry farming outcomes. Omomule et al. (2020) demonstrated fuzzy logic models for accurate egg production predictions, while Gao et al. (2021) and Luchkin et al. (2021) explored the roles of dietary oils, protein, and vitamins in improving feed efficiency and egg quality. Probiotics also show potential for enhancing nutrient utilization and poultry health (Jha et al., 2020). However, gaps remain in integrating real-time data analytics and machine learning into feed management and incorporating environmental variables and technological innovations to improve sustainability and efficiency in poultry production.

CHAPTER THREE
RESEARCH METHODOLOGY

3.1 Introduction

This chapter describes the study approach used to use data science tools to optimize feeding practices for egg production. Creating machine learning model, creating dashboards for in-the-moment decision-making, and gathering and evaluating data from several secondary sources of chicken farm records are all part of the project. The techniques selected are essential for making precise feed efficiency estimates and guaranteeing that interested parties can put data-driven feeding plans into practice.

In order to handle the massive amounts of poultry data and find hidden patterns that impact feed efficiency, the selected techniques, such as machine learning algorithms and predictive analytics are crucial. These methodologies guarantee a methodical and impartial examination of factors impacting egg production, such as feed kind, nutrient composition, ambient circumstances, and indicators of hen health.

3.2 Research Design

A research design is an organized framework that directs a project's methodology and specifies how data collection, analysis, and interpretation are to be done. By ensuring that the research is carried out methodically, it minimizes bias and maximizes reliability.

This study is both exploratory and descriptive in nature, with the goal of describing current laying hen feeding behaviours and investigating novel approaches to data-driven optimization of these practices. Beginning with the gathering of historical data on feed kinds and egg production, the study framework is organized into several stages. The subsequent processes, which involve data analysis and the creation of predictive machine learning models, are greatly aided by this basic data. By predicting feed efficiency and production results, these models offer insightful information that can greatly improve chicken farmed feeding techniques.

The framework's last step is the development of an interactive dashboard that allows stakeholders to make decisions in real time by visualizing the analysis's findings. With the help of this method, feed efficiency can be continuously monitored, enabling managers and farmers to respond quickly to shifting operational circumstances. The system guarantees that feeding procedures may be continually refined, which eventually results in better egg production and more sustainable farming methods, by combining data gathering, analysis, and visualization.

3.3 Data Collection Methods

Data for this study is obtained from secondary sources, primarily from the dataset available on DataDryad.org ;a repository that has datasets on the intake pattern and feed preference of layer hens, including feed consumption and feed conversion ratio (FCR)1PLOS ONE .(Clark et al., 2019)This dataset encompasses historical data on hen health, feeding behavior data, average daily feed intake, total feed intake, total egg weight, feed consumed per hour, egg production rates, dietary composition and nutrient intake. The study will combine multiple datasets into a single, cohesive dataset. This merged dataset will be used for analysis and model creation. The goal is to develop insights and predictive models for optimizing feeding practices.

The study employs a technique of extensive searches of online databases, utilizing keywords related to poultry feeding and production to identify and extract pertinent datasets from trusted sources. This rigorous approach ensures that the most relevant and high-quality data are included in the analysis. To further enhance the validity of the dataset, The study cross-referenced information from multiple sources, verifying consistency and reliability across various records. This comprehensive methodology provides a robust foundation for the research, ensuring the data's integrity and relevance.

3.4 Data Preparation

The collected data undergoes a rigorous cleaning process to ensure its quality and reliability. This phase focus on removing duplicates, outliers, and erroneous entries that could skew the analysis. To facilitate this process, Python libraries such as Pandas and NumPy are employed.

These tools offer powerful functionalities for data manipulation and cleaning, enabling efficient identification and correction of inconsistencies within the dataset. By ensuring the data is accurate and consistent, the study will lay a solid foundation for subsequent analyses and modeling efforts.

In addition to general cleaning, special attention will be given to handling missing values within the dataset. Various imputation techniques will be applied based on the nature of the missing data. For numerical data, mean or median imputation will be used to fill in gaps, providing a simple yet effective way to maintain dataset integrity. For time series data, forward and backward filling methods will help preserve temporal continuity by using available neighboring data points to estimate missing values. In instances where a significant portion of data is missing and imputation is not feasible, affected instances will be removed to maintain the overall quality of the dataset. These methods will ensure that the analysis remains robust and reflective of the true feeding practices and their impacts on egg production.

To prepare the data for machine learning models, several transformation procedures are implemented to enhance its suitability and effectiveness. Normalization and standardization of numerical data will be conducted to ensure uniformity across the dataset, facilitating better model performance and convergence during training. For non-numeric variables, such as feed types or farm locations, categorical encoding techniques like one-hot encoding will be applied to convert these variables into a numerical format that machine learning algorithms can interpret. Additionally, feature engineering will be employed to derive new variables that capture essential insights, such as the feed conversion ratio (FCR), which quantifies feed efficiency. These transformation steps will not only improve the quality of the data but also enhance the predictive power of the models developed in this study.

3.5 Data Analysis Techniques

The study employs several statistical methods to gain insights into the collected data and explore relationships among key variables. Descriptive statistics provides a preliminary understanding of the dataset, highlighting critical aspects such as feed consumption patterns, egg production rates, and environmental conditions. This foundational analysis will help identify trends and anomalies within the data, laying the groundwork for more complex analyses. Additionally, correlation analysis was conducted to investigate the relationships between various factors, such as different feed types and their corresponding egg production outcomes. Understanding these correlations is crucial for informing the subsequent modeling process and optimizing feeding practices.

To develop predictive models for feed efficiency, the study employed machine learning algorithms. Regression model was employed for the continuous prediction of feed conversion ratio based on various farm data, allowing for an understanding of how change in bird behaviors in response to feed intake influence output. This approach provided insights into the factors influencing feed efficiency and help refine feeding strategies.

The implementation of these statistical methods and machine learning algorithms was supported by a variety of tools and software. Python was the primary programming language for coding and model development, utilizing libraries such as Scikit-learn for machine learning. This tool ensured a robust analysis, enhancing the study's ability to optimize feeding practices in poultry production.

3.6 Model Building

A random forest regressor was trained after performing accurate data analysis and feature engineering and removing the outliers. The dataset is splitted into training and testing sets to facilitate the assessment of model's generalizability and accuracy in real-world scenarios. A cross validation is implemented to mitigate overfitting and enhance model's robustness.

The model is then evaluated using standard regression metrics like mean squared error, mean absolute error and R^2 that provide a quantitative measure of how well the model predicts the target variable. Evaluating the model on unseen test data allows us to assess its ability to generalize and make accurate predictions on data it has not encountered during training.

3.7 Ethical Considerations

In order to maintain the integrity and accountability of the research process, the study was dedicated to adhering to a number of ethical criteria. In order to protect the private rights of anyone affected, the used dataset was obtained from a licensed source for data reuse. Throughout the study, accountability and transparency is upheld, guaranteeing that all parties involved are fully aware of the methods, results, and possible consequences. The emphasis will be on ongoing assessment and development, with procedures being updated in accordance with the ethical implications of optimizing feed to boost egg production efficiency. The study also follows ethical guidelines for animal welfare, seeking to enhance feeding procedures without endangering or stressing the animals.

3.8 Limitations

The complexity introduced by the variety of farming methods may limit the generalizability of results in other agricultural contexts. Furthermore, technological differences are important since some farms may not have access to real-time data gathering systems, which reduces the accuracy and comprehensiveness of the data collected. Data gaps are another problem; even with techniques to patch them up, smaller farms may have missing or partial data that affects forecast accuracy. Furthermore, bias may be introduced into the model's predictions by data from various farm types, feed formulations, and management techniques. Stratified sample will guarantee that all farm types are fairly represented in order to lessen this. Together, these elements highlight the difficulties in standardizing and extrapolating agricultural

3.9 Summary

This chapter has outlined the methodology used to optimize feeding practices for egg production. A combination of data collection methods, machine learning algorithms, and interactive dashboard development will ensure the study’s objectives are met. The approach, which integrates statistical and predictive analytics, ensures that key factors influencing feed efficiency are identified and that stakeholders have actionable insights for decision-making. The ethical considerations and limitations addressed ensure the study’s transparency and robustness.

CHAPTER FOUR

RESULTS AND DISCUSSION

4.1 Introduction

This chapter presents the implementation of the predictive model for proactive layer feed efficiency. It outlines the step-by-step approach taken to develop, test, and deploy the model, ensuring its practical application in optimizing poultry feeding practices. The implementation process involved several critical stages, beginning with importation of necessary libraries to facilitate data handling, data preparation and exploration, followed by model development, and finally deploying the model using a Flask web application for real-world testing.

4.2 Importing Necessary Libraries

To facilitate data analysis, visualization, and machine learning, this study utilizes several essential Python libraries, each serving a specific purpose in the implementation process.

Pandas (pd) is employed for data manipulation and analysis, offering powerful data structures such as DataFrames that enable efficient handling of structured data. Complementing this, NumPy (np**)** supports numerical computations, particularly in working with arrays, matrices, and mathematical operations, making it a fundamental tool for data processing.

For data visualization, multiple libraries are utilized. Matplotlib (plt) is used to generate static and interactive plots, while Seaborn (sns) enhances these visualizations by providing statistical graphics such as heatmaps and box plots. Additionally, Plotly (px, go, pio) enables the creation of highly interactive visualizations. Within Plotly, plotly.express (px) offers a high-level interface for straightforward plotting, plotly.graph_objects (go) allows for more customizable visualizations, and plotly.io (pio) manages input/output configurations. To maintain a consistent visualization style, the default template is set to white using pio.templates.default.

To ensure a clean output without unnecessary warning messages, the Warnings (warnings) library is employed, allowing for the suppression of system-generated warnings.

For machine learning tasks, Scikit-learn (sklearn) provides a comprehensive suite of tools for model development and evaluation. It includes performance metrics such as mean_absolute_error and r2_score for assessing regression models. Additionally, various modeling techniques are applied, including Ridge, a regularized linear regression model, and RandomForestRegressor, an ensemble-based model known for its robustness in regression tasks. The train_test_split function is also utilized to partition the dataset into training and testing sets, ensuring effective model evaluation and generalization.

By leveraging these libraries, the study efficiently processes, visualizes, and models data, ultimately supporting the development of a predictive model for proactive layer feed efficiency.

```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
import plotly.express as px
import plotly.graph_objects as go
import plotly.io as pio

pio.templates.default = "plotly_white"
warnings.filterwarnings('ignore')

from sklearn.metrics import mean_absolute_error,mean_squared_error, r2_score
from sklearn.linear_model import Ridge
from sklearn.ensemble import RandomForestRegressor
from sklearn.model_selection import train_test_split
from flask import Flask, request, jsonify
import joblib
```
Python

Table 4-1:Libraries

4.3 Loading Datasets into Pandas DataFrames

The study utilizes multiple datasets related to feed efficiency in egg production, each containing specific information that will be merged for comprehensive analysis. To facilitate this process, the datasets are first loaded into Pandas DataFrames. This step ensures that the data is structured in a tabular format, enabling efficient analysis and manipulation. By leveraging Pandas, the study can seamlessly handle data cleaning, transformation, and integration, laying a solid foundation for subsequent modeling and interpretation.

4.3.1 Dataset's overview

i. FCR data set

	Bird ID.	Total feed intake (g)	ADI (g)	Total egg wt (g)	FCR
0	R3B31	823.0	117.571429	449.0	1.832962
1	R3B47	791.0	113.000000	463.0	1.708423
2	R3B5	913.0	130.428571	488.0	1.870902
3	R3B60	818.0	116.857143	454.0	1.801762
4	R3B65	875.0	125.000000	481.0	1.819127

Table 4-2:FCR dataset

The FCR dataset provides essential information for analyzing feed efficiency in egg production. Each record in the dataset represents an individual bird, identified by a unique Bird ID (e.g., *R3B31, R3B47*), which allows for precise tracking across multiple datasets.

The dataset includes Total Feed Intake (g), representing the total amount of feed consumed by each bird in grams. This metric is fundamental in assessing feeding patterns and efficiency. Additionally, the Average Daily Intake (ADI) is calculated to reflect the bird's daily feed consumption over a given period. This value is crucial in evaluating feeding consistency and detecting variations in intake.

To measure productivity, the dataset captures the Total Egg Weight (g) produced by each bird. This helps in determining overall egg yield and its relationship with feed consumption. A key efficiency metric in the dataset is the Feed Conversion Ratio (FCR), which represents the ratio of feed intake to egg production.

ii. Feeding Behaviours Dataset

	Bird ID.	No. of feeding-bout/h	Time spent for feeding /h	No. of Still/h	No. of Rest/h	No of Head flicks/h	No. of drinking/h	No. of preening/h	No. of feeder pecking/h	No. of cage pecking	No. of Walking/h
0	R3B31	24	18	42	1	36	12	15	12	22	12
1	R3B47	84	42	72	1	18	6	16	7	23	12
2	R3B5	24	18	12	6	54	6	14	12	18	6
3	R3B60	30	18	24	18	2	6	20	6	21	12
4	R3B65	36	24	12	1	6	6	18	12	19	6

Table 4-3:Feeding behaviors dataset

The Behavioral Metrics Dataset shown above captures detailed behavioral patterns of individual birds, providing valuable insights into their feeding, drinking, and movement activities. Each record corresponds to a specific bird, identified by a unique Bird ID, ensuring precise tracking and comparison across different datasets.

A key metric in this dataset is the Number of Feeding Bouts per Hour, which measures how frequently a bird engages in feeding sessions. Complementing this, Time Spent for Feeding per Hour quantifies the duration a bird spends consuming feed, helping to assess feeding intensity and efficiency.

In addition to feeding behavior, the dataset includes metrics related to general activity levels. The Number of Still Periods per Hour and Number of Resting Periods per Hour indicate how often a bird remains inactive, which can provide insights into energy conservation and potential health issues.

Head movements are also tracked through the Number of Head Flicks per Hour, a metric that reflects alertness and potential stress levels. Number of Drinking Sessions per Hour captures hydration patterns, while Number of Preening Sessions per Hour indicates self-maintenance behaviors, which are crucial for assessing bird welfare.

The dataset also monitors pecking behaviors, distinguishing between Number of Feeder Pecking Instances per Hour, which directly relates to feed consumption, and Number of Cage Pecking Instances, which may signal stress or environmental dissatisfaction. Lastly, Number of Walking Instances per Hour provides an estimate of overall mobility and activity levels.

iii. Chemical Composition of Feeds remains

	Bird ID.	DM%	Ash%	N%	CP%	GE (kcal/kg)
0	R3B31	89.014993	11.025366	3.150	19.69	15.141612
1	R3B47	88.174224	14.232869	2.648	16.55	14.384441
2	R3B5	89.016720	11.557241	3.069	19.18	15.686275
3	R3B60	87.727906	10.741070	2.749	17.18	15.264109
4	R3B65	89.285905	15.606278	2.598	16.24	14.749675

Table 4-4:Chemical composition of feed remains

The Chemical Composition of Feed Remains Dataset provides valuable insights into the nutritional content of the uneaten feed, helping to assess feed wastage and optimize feeding strategies in egg production. Each record corresponds to an individual bird, identified by a unique Bird ID, allowing for a detailed analysis of variations in feed remains across different birds.

One of the key parameters in this dataset is DM% (Dry Matter Percentage), which represents the proportion of solid material in the leftover feed after moisture removal. A higher dry matter content in the remains could indicate selective feeding, where birds consume moisture-rich components while leaving behind drier portions of the feed.

The Ash% content measures the total mineral composition of the uneaten feed. A significant difference between the ash content of feed remains and the original feed may suggest that birds are selectively consuming certain nutrients while ignoring others, which could impact their overall mineral intake and health.

Additionally, the dataset includes N% (Nitrogen Percentage), a crucial indicator of protein retention in feed remains. From this, CP% (Crude Protein Percentage) is calculated to assess the total protein left unconsumed. If feed remains contain a high percentage of crude protein, it may suggest that birds are avoiding protein-rich components, potentially affecting their growth and egg production efficiency.

Another important metric is GE (Gross Energy, kcal/kg), which measures the total energy content of the uneaten feed. If the energy value of the feed remains is significantly high, it may indicate inefficient feed utilization, where birds consume only specific components while leaving behind more energy-dense feed. This could lead to imbalances in their diet, ultimately impacting their productivity and health.

4.4 Merging the Datasets

After loading the datasets into Pandas DataFrames, the next step is to merge them into a single dataset. This merging process is essential because each dataset contains different but related information, and combining them allows for a more comprehensive analysis. The datasets will be merged using the Bird ID column, which serves as a common key across all the datasets. By merging the datasets, the study can integrate various dimensions of data, such as feed conversion rates, bird behaviors, and chemical composition of the feed, providing a more complete picture of feed efficiency in egg production. The merge will be done using a left join strategy (how='left'), which ensures that all records from the FCR dataset are retained. Corresponding records from the Feeding Behaviours and Chemical Composition of Feeds datasets will be added where there is a matching Bird ID. If no match is found, missing values (NaN) will be introduced, indicating the absence of information for those specific birds in the secondary datasets.

The choice to use a left join ensures that no birds from the FCR dataset are excluded, maintaining the integrity of the primary dataset. It allows the FCR dataset to be enriched with additional data from the behavioral and chemical composition datasets, without losing any original records. Furthermore, by handling missing data with NaN values, this approach ensures that important records are not discarded due to missing data in the secondary datasets. This merging strategy provides a solid foundation for the analysis, enabling the study to gain deeper insights into feed efficiency and its various influencing factors.

Below is a preview of the merged dataset

	Bird ID.	Total feed intake (g)	ADI (g)	Total egg wt (g)	FCR	No. of feeding-bout/h	Time spent for feeding /h	No. of Still/h	No. of Rest/h	No of Head flicks/h	No. of drinking/h	No. of preening/h	No. of feeder pecking/h	No. of cage pecking	No. of Walking/h	DM%	Ash%	N%	CP%	GE (kcal/kg)
0	R3B31	823.0	117.571429	449.0	1.832962	24	18	42	1	36	12	15	12	22	12	89.014993	11.025366	3.150	19.69	15.141612
1	R3B47	791.0	113.000000	463.0	1.708423	84	42	72	1	18	6	16	7	23	12	88.174224	14.232869	2.648	16.55	14.384441
2	R3B5	913.0	130.428571	488.0	1.870902	24	18	12	6	54	6	14	12	18	6	89.016720	11.557241	3.069	19.18	15.686275
3	R3B60	818.0	116.857143	454.0	1.801762	30	18	24	18	2	6	20	6	21	12	87.727906	10.741070	2.749	17.18	15.264109
4	R3B65	875.0	125.000000	481.0	1.819127	36	24	12	1	6	6	18	12	19	6	89.285905	15.606278	2.598	16.24	14.749675

Table 4-5:Merged dataset

4.5 Initial DataFrame Exploration and Summary

Once the datasets have been successfully loaded and merged, it's important to perform an initial examination to understand the structure, content, and quality of the data. This is done through a series of methods to get an overview and summary of the DataFrame. In this study, several key methods have been applied: df.info(), df.shape, df.isnull().sum(), and df.describe(). These methods provide a comprehensive look into the dataset's basic properties and potential issues such as missing values.

4.5.1 Data structure overview

The study has used df.info() to provides a concise summary of the DataFrame, including column names, non-null counts, data types, and memory usage as shown below

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 68 entries, 0 to 67
Data columns (total 20 columns):
 #   Column                     Non-Null Count  Dtype
---  ------                     --------------  -----
 0   Bird ID.                   68 non-null     object
 1   Total feed intake (g)      68 non-null     float64
 2   ADI (g)                    68 non-null     float64
 3   Total egg wt (g)           68 non-null     float64
 4   FCR                        68 non-null     float64
 5   No.  of feeding- bout/h    68 non-null     int64
 6   Time spent for feeding /h  68 non-null     int64
 7   No.  of Still/h            68 non-null     int64
 8   No.  of Rest/h             68 non-null     int64
 9   No  of Head flicks/h       68 non-null     int64
 10  No.  of drinking/h         68 non-null     int64
 11  No. of preening/h          68 non-null     int64
 12  No.  of feeder pecking/h   68 non-null     int64
 13  No. of cage pecking        68 non-null     int64
 14  No.  of Walking/h          68 non-null     int64
 15  DM%                        68 non-null     float64
 16  Ash%                       68 non-null     float64
 17  N%                         68 non-null     float64
 18  CP%                        68 non-null     float64
 19  GE  (kcal/kg)              68 non-null     float64
dtypes: float64(9), int64(10), object(1)
memory usage: 10.8+ KB
```

Table 4-6:Data structure overview

The DataFrame consists of 68 entries, indexed from 0 to 67, as indicated by the RangeIndex. It contains a total of 20 columns, each representing a distinct aspect of the dataset. The **column** names and non-null count provide insight into the completeness of the data, listing each column alongside the number of non-missing values it contains. For example, the Bird ID. column has 68 non-null entries, confirming that there are no missing values in this identifier field.

The data types (Dtype) vary across columns, with numerical attributes such as Total feed intake and ADI being of type float64, while categorical or discrete numerical features like No. of feeding-bout/h are stored as int64. Additionally, the Bird ID. column is classified as object, indicating that it contains string-based identifiers.

Finally, the memory usage of the dataset is approximately 10.8 KB, reflecting the storage requirements needed to hold the DataFrame in memory.

4.5.2 Shape of the dataset

Shape shows how many feature columns and rows are there in the dataset.This command is employed print(f"Columns: {df.shape[1]}\nSamples: {df.shape[0]}") and the following output is obtained Columns: 20 ,Samples: 68 which shows that the DataFrame consists of 68 rows, each representing an individual observation, record, or sample within the dataset. Additionally, it contains 20 columns, with each column representing a distinct feature or variable that provides valuable information about the observations.

4.5.3 Checking for null values

df.isnull().sum()checks for missing values in each column of the Data Frame 'df' and returns the sum of null values for each column.

```
Bird ID.                      0
Total feed intake (g)         0
ADI (g)                       0
Total egg wt (g)              0
FCR                           0
No.  of feeding- bout/h       0
Time spent for feeding /h     0
No.  of Still/h               0
No.  of Rest/h                0
No  of Head flicks/h          0
No.  of drinking/h            0
No. of preening/h             0
No.  of feeder pecking/h      0
No. of cage pecking           0
No.  of Walking/h             0
DM%                           0
Ash%                          0
N%                            0
CP%                           0
GE  (kcal/kg)                 0
dtype: int64
```

Table 4-7:Null values

Since all columns have a count of zero it means there are no missing values in the dataset

4.5.4 Summary statistics

Summary statistics provide key insights into the distribution, central tendency, and variability of numerical features.The study has employed df.describe().T to provide summary statistics for numerical columns in the DataFrame, and transposes the result for easier readability.

	count	mean	std	min	25%	50%	75%	max
Total feed intake (g)	68.0	897.323529	89.902508	704.000000	838.500000	878.250000	960.250000	1145.000000
ADI (g)	68.0	128.235294	12.835649	100.571429	119.785714	125.964286	137.178571	163.571429
Total egg wt (g)	68.0	447.683824	44.122362	223.000000	431.875000	454.500000	471.000000	529.000000
FCR	68.0	1.992196	0.200961	1.483607	1.851469	1.934027	2.135443	2.490950
No. of feeding- bout/h	68.0	24.529412	17.553730	6.000000	12.000000	24.000000	30.000000	84.000000
Time spent for feeding /h	68.0	27.000000	13.396368	0.000000	18.000000	30.000000	36.000000	54.000000
No. of Still/h	68.0	9.794118	11.803818	1.000000	1.000000	6.000000	12.000000	72.000000
No. of Rest/h	68.0	3.838235	4.708234	1.000000	1.000000	1.000000	6.000000	24.000000
No of Head flicks/h	68.0	25.705882	19.572337	0.000000	10.500000	18.000000	36.000000	72.000000
No. of drinking/h	68.0	3.691176	3.057935	1.000000	1.000000	2.500000	6.000000	12.000000
No. of preening/h	68.0	16.367647	4.062587	10.000000	12.000000	16.000000	19.000000	26.000000
No. of feeder pecking/h	68.0	13.955882	12.001783	5.000000	6.000000	11.500000	13.000000	60.000000
No. of cage pecking	68.0	18.779412	6.737832	12.000000	12.000000	18.000000	22.000000	48.000000
No. of Walking/h	68.0	15.676471	8.739072	6.000000	12.000000	12.000000	18.000000	42.000000
DM%	68.0	88.946283	1.235518	85.239085	88.386029	89.007426	89.714506	93.095829
Ash%	68.0	18.807925	6.139027	9.137096	14.121235	17.343463	22.604969	33.503335
N%	68.0	2.593382	0.276754	1.855000	2.396000	2.596000	2.814750	3.233000
CP%	68.0	16.208529	1.729452	11.590000	14.975000	16.225000	17.595000	20.200000
GE (kcal/kg)	68.0	14.121919	0.954847	12.115804	13.419713	14.297043	14.939316	15.686275

Table 4-8:Summary statistics

The table above provides a summary of key statistical metrics for various features in the dataset. These metrics offer valuable insights into the distribution and variability of the data. The count represents the number of non-missing values, with each variable containing 68 observations. The mean denotes the average value for each feature, while the standard deviation (std) measures how spread out the values are around the mean.

The minimum (min) value indicates the lowest recorded observation, while the maximum (max) represents the highest. Additionally, the table includes quartile values: the 25th percentile (Q1) marks the value below which 25% of the data falls, the 50th percentile (median/Q2) represents the middle value that divides the dataset into two equal halves, and the 75th percentile (Q3) denotes the value below which 75% of the observations fall. These statistics help in understanding the distribution, central tendency, and spread of the dataset's features.

4.6 Exploratory data analysis

This analysis explores the dataframe, focusing on key variables within the dataset with the goal to uncover trends and patterns of features in the dataset through univariate analysis, bivariate analysis and multivariate analysis

4.6.1 Univariate analysis

This analysis involves analyzing a single variable to understand its distribution, central tendency and spread. The study has employed histogram, violin and box plot to visualize the distribution of key variables as follows

4.6.1.1 Understanding Total feed intake distribution with a histogram plot

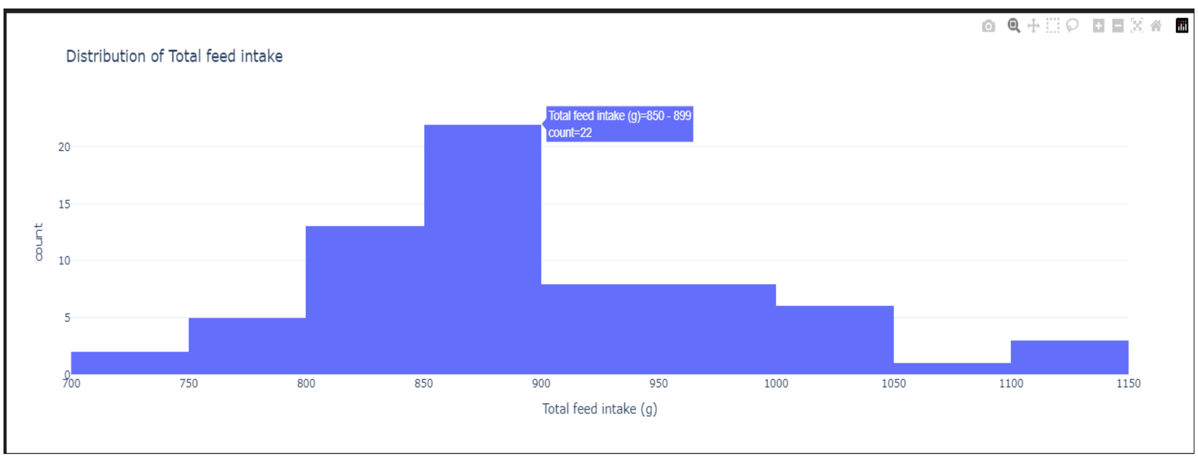

Figure 4-1:Histogram plot of total feed intake

Insights from the histogram

The X-axis represents Total Feed Intake, which ranges from 700 to 1150 and is divided into bins, such as 850-899. Each bin encompasses a specific range of total feed intake values. The Y-axis denotes the Count, representing the number of birds that fall within each corresponding feed intake range.

A key insight from the data is that the most common feed intake range among the birds is 850-899, as indicated by the tallest bar in the distribution. In this range, 22 birds are recorded, making it the most frequent total feed intake category.

4.6.1.2 violin plot of total egg weight

A violin plot of Total Egg Weight visualizes the distribution, spread, and density of egg weights, highlighting variations and potential outliers. It provides a detailed view of how egg weights are distributed among the birds, offering insights into central tendencies and variability.

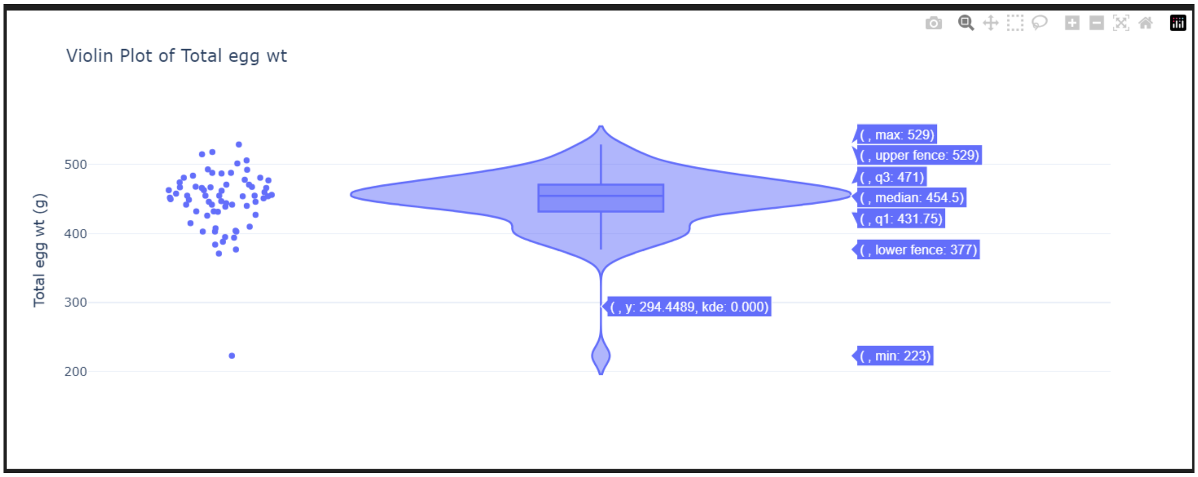

Figure 4-2:Violin plot of total egg weight

The violin plot of Total Egg Weight provides a detailed view of the distribution by combining a density plot with a box plot. The shape of the violin reflects the frequency of different egg weights, with wider sections indicating where more data points are concentrated. Notably, the thicker middle region suggests that most egg weights cluster between 450 and 470 grams.

Embedded within the violin plot is a box plot, which highlights key statistical measures. The median egg weight is approximately 454.5 grams, positioned at the center of the box. The interquartile range (IQR) spans from 431.75 grams (Q1) to 471 grams (Q3), capturing the middle 50% of the data. The whiskers extend from 377 grams to 529 grams, representing the lower and upper fences. Additionally, an outlier is observed at 223 grams, which may warrant further investigation.

The Kernel Density Estimate (KDE) curve smooths the distribution, reinforcing that the majority of eggs fall within the 450-470gram range. The overall symmetry of the plot suggests that egg weights follow a roughly normal distribution without significant skewness.

4.6.1.3 Understanding average daily intake (ADI) Distribution with a Box Plot

A box plot of Average Daily Intake (ADI) provides a clear summary of its distribution, highlighting key statistics such as the median, interquartile range, and potential outliers. This visualization helps in understanding the variability and central tendency of ADI among the birds.

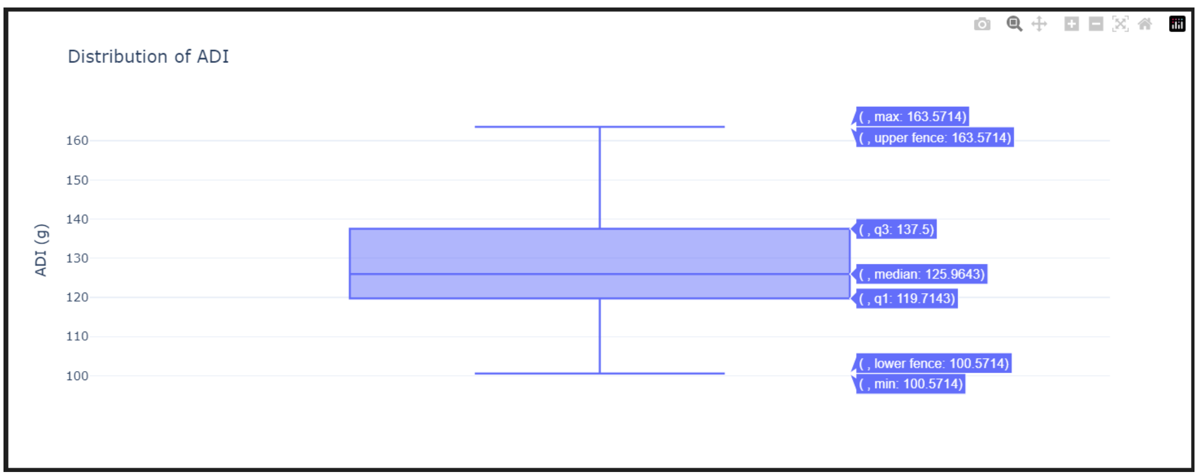

Figure 4-3:Box plot of ADI

The box plot provides a detailed visualization of the distribution of Average Daily Intake (ADI), illustrating how feed intake varies among the birds. The median ADI**,** positioned at 125.96, represents the midpoint, meaning half of the birds consume less than this amount while the other half consume more.

The middle 50% of the data falls between the first quartile (Q1) at 119.71 and the third quartile (Q3) at 137.5**,** forming the interquartile range (IQR) of 17.79 **(**137.5 - 119.71**).** This range captures the core of the dataset, excluding extreme values.

The whiskers, extending from 100.57 to 163.57, define the expected spread of ADI, covering values within 1.5 times the IQR from Q1 and Q3. Any data points beyond these whiskers are considered potential outliers**.** The minimum and maximum observed ADI values within this normal range are 100.57 and 163.57, respectively.

4.6.2 Bivariate analysis

Bivariate analysis involves analyzing two variables with the purpose of understanding whether changes in one variable affect the other. This study has made use of scatter plots, bar plots and box plot explore relationships between variables, such as total egg weight vs. total feed intake.

4.6.2.1 Scatter Plot of Total Egg Weight vs Total Feed Intake

A scatter plot is generated with an OLS trendline to reveal whether increased feed intake contributes to greater egg weight

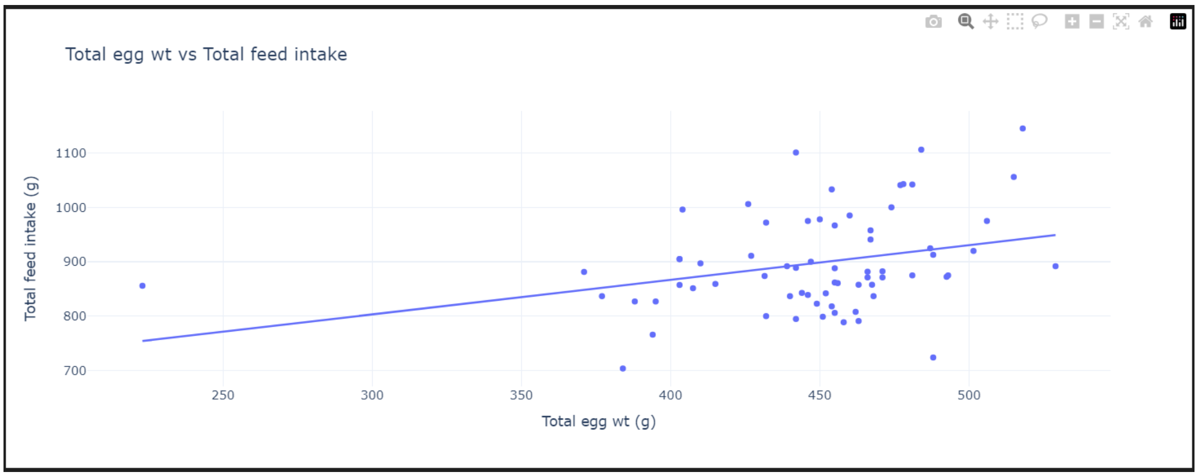

Figure 4-4:Scatter plot of total egg weight vs total feed intake

The scatter plot illustrates the relationship between Total Egg Weight and Total Feed Intake for individual birds. The X-axis represents the total weight of eggs produced by each bird, while the Y-axis indicates the total feed consumed. Each data point corresponds to a bird's total feed intake in relation to its egg weight.

To capture the overall trend, an Ordinary Least Squares (OLS) regression line is fitted to the data.

4.6.2.2 Bar plot of Total feed intake against various bird ids

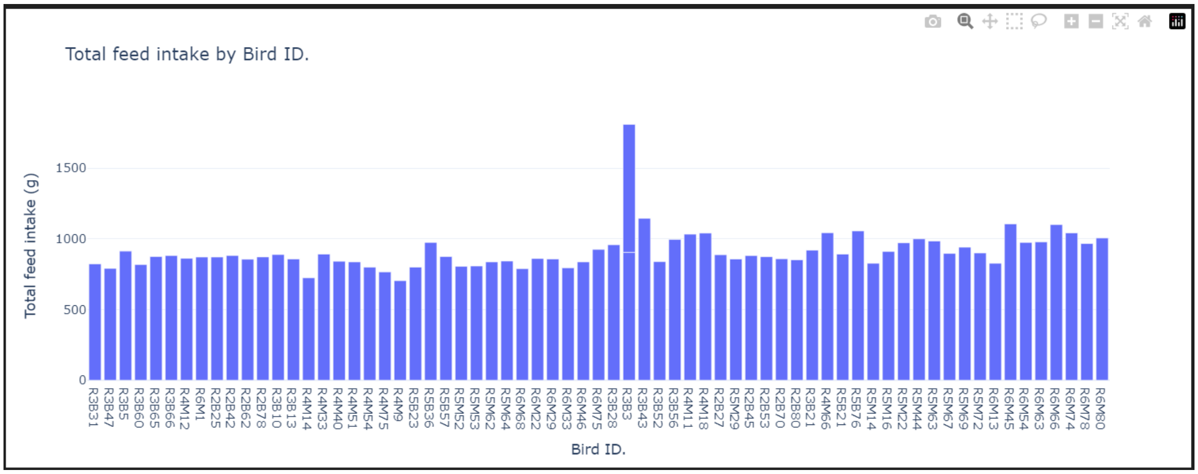

Figure 4-5: Bar plot of total feed intake against various bird ids

Each blue bar represents one bird's total feed intake. Most birds cluster around a similar range of intake, indicating consistent feeding patterns overall.

4.6.2.3 Box Plot of Total Feed Intake by Time Spent Feeding/h

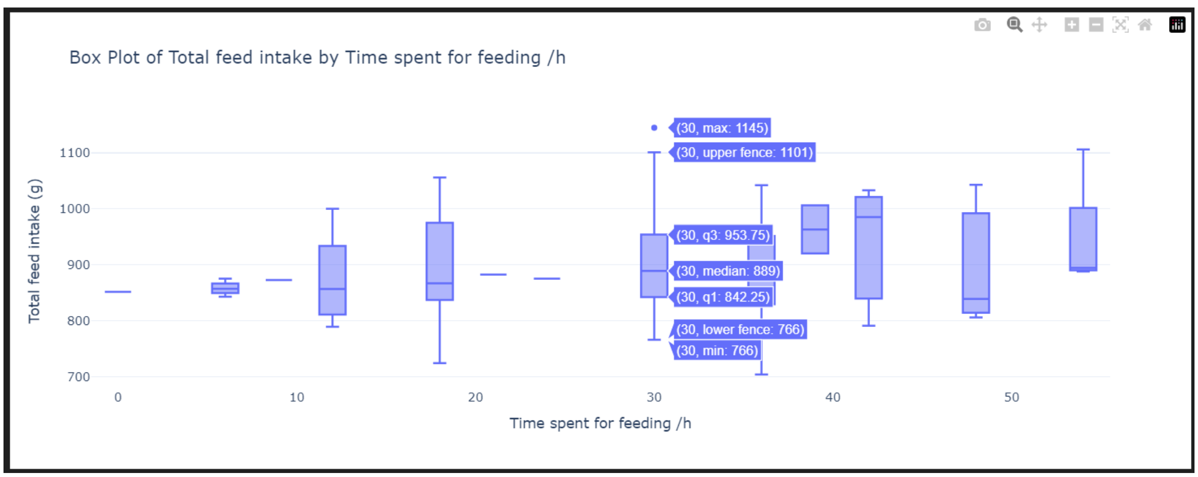

Figure 4-6: Bar plot of total feed intake by time spent feeding/h

The plot displays the relationship between Time Spent Feeding (hours) and Total Feed Intake, with the X-axis representing the amount of time each bird spent feeding (e.g., 0, 10, 20, 30 hours), and the Y-axis showing the total feed intake for those birds.

Key insights from the plot include the interquartile range (IQR)**,** which captures the middle 50% of the data. For instance, at 30 hours of feeding, the IQR spans from 842.25 to 953.75 grams. The bold line in the box represents the median, showing the middle value of feed intake, which is 889 grams for birds that spent 30 hours feeding. This helps to identify where most birds' feed intake is centered.

The whiskers on the plot, which represent the range of normal values, extend from 766 to 1101 grams for birds that fed for 30 hours. Any data points outside this range are considered outliers, and are marked as dots. For example, a bird that consumed 1145 grams at 30 hours of feeding is flagged as an outlier, indicating it ate significantly more than most of its peers.

4.6.2.4 Scatter plot total feed intake vs feed conversion ratio (FCR)

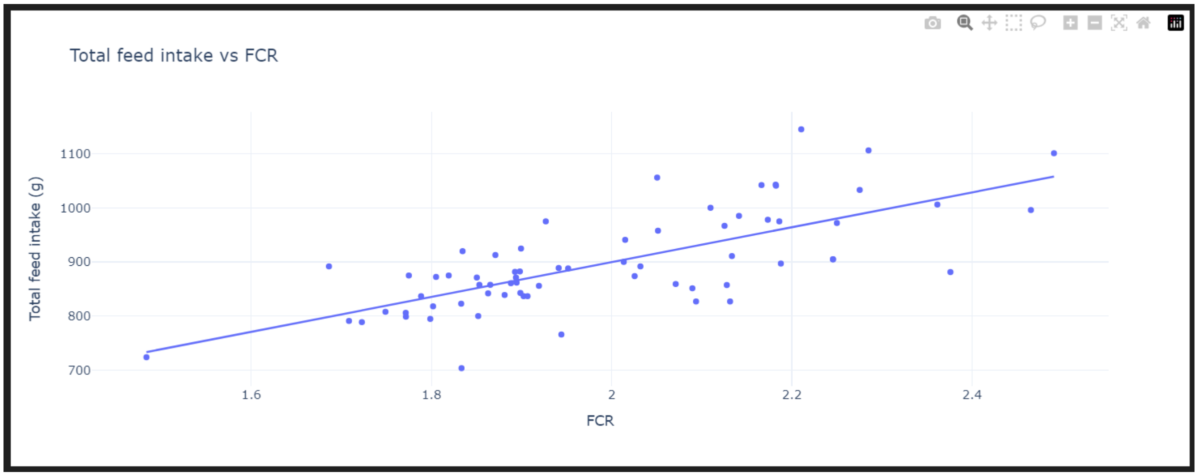

Figure 4-7:Scatter plot of total feed intake vs fcr

4.6.3 Multivariate analysis

Involves analyzing three or more variables to uncover complex relationships. A correlation matrix is created to identify relationships among multiple variables.

4.6.3.1 Correlation matrix

To explore the relationships between numerical features in the dataset, a correlation matrix is constructed. This matrix serves as a valuable tool for identifying the strength and direction of linear relationships between pairs of numerical variables.

The process begins with the selection of only the numerical features from the dataset. This ensures that the correlation analysis focuses on variables that are quantifiable, such as integers and floating-point numbers. In Python, this selection is typically performed using the select_dtypes method to filter columns of types float64 and int64.

Once the numerical features are isolated, the correlation matrix is computed. This matrix reveals the degree to which each pair of variables is linearly related, with values ranging from -1 (perfect negative correlation) to +1 (perfect positive correlation). A value close to zero indicates little to no linear relationship.

To enhance interpretability, the correlation matrix is visualized using a heatmap. This graphical representation employs color gradients to reflect the strength of the correlations, often accompanied by numerical annotations for precise interpretation. The heatmap allows for quick identification of strongly correlated features, which can be useful for feature selection, multicollinearity detection, and deeper data insights.

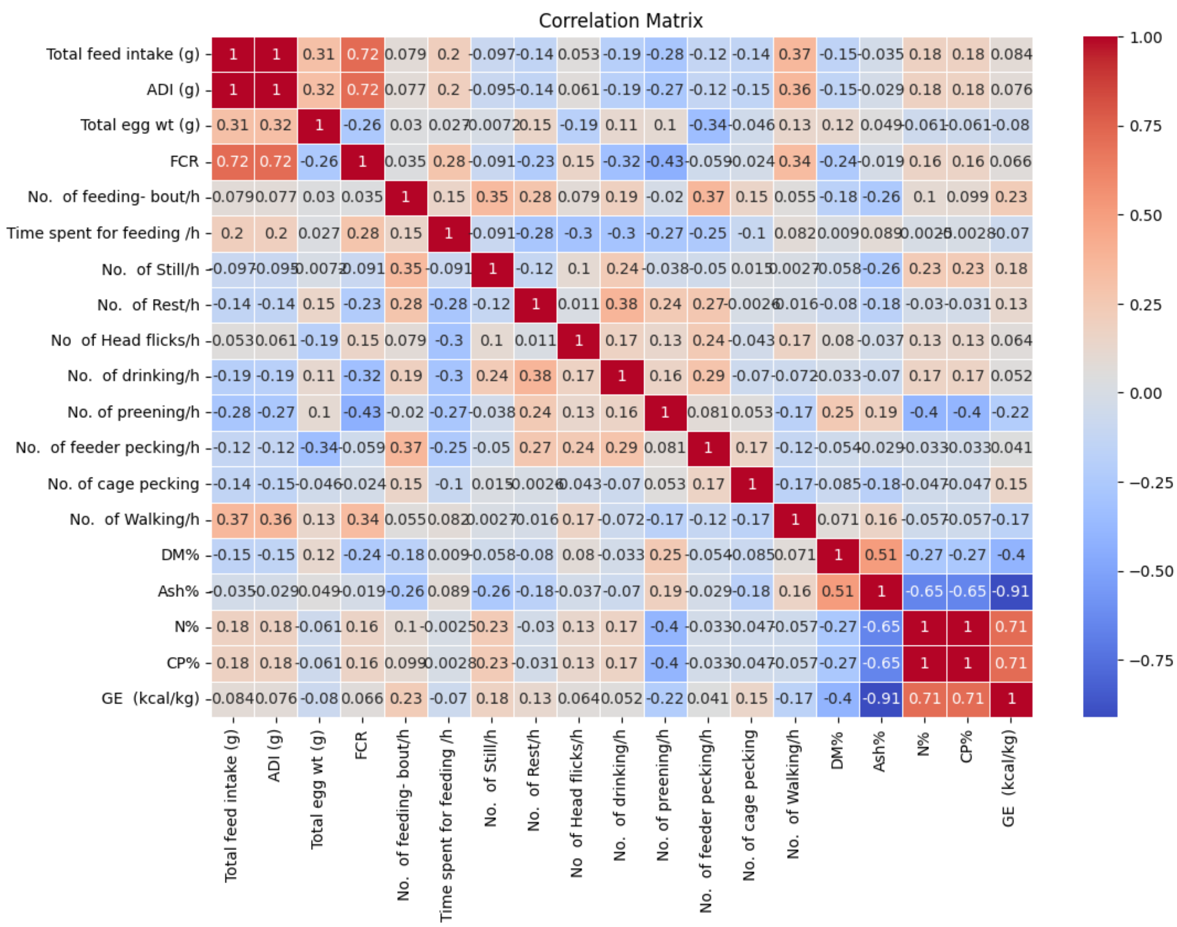

Figure 4-8: heatmap 1

4.6.3.1.1 Identify and drop Highly Correlated Features

To address the issue of multicollinearity in the dataset, the research conducted a correlation analysis to identify and remove highly correlated features. This was done to improve model performance and ensure that each feature contributes uniquely to the predictive process without redundancy.

The research began by creating a correlation matrix using only the numerical features in the dataset. To focus on unique feature relationships and avoid duplications from the symmetric matrix, extracted the upper triangle of the correlation matrix using NumPy’s trio function. This allows to isolate and evaluate only one side of the correlation matrix.

Next, identified features that had a correlation coefficient greater than 0.9 with at least one other feature. These were considered highly correlated and thus potential candidates for removal. Through this analysis, the following features were flagged: 'Ash%', 'GE (kcal/kg)', 'ADI (g)', 'CP%', and 'Total feed intake (g)'.

Then proceeded to drop these highly correlated features from the dataset to reduce redundancy. After removing them, recalculated the correlation matrix to assess whether multicollinearity had been effectively reduced.

Finally, generated an updated heatmap to visualize the revised correlation matrix. This confirmed that the remaining features exhibited lower intercorrelations, indicating improved data quality. This step was crucial in preparing a cleaner dataset for modeling, ultimately contributing to more stable and interpretable model performance.

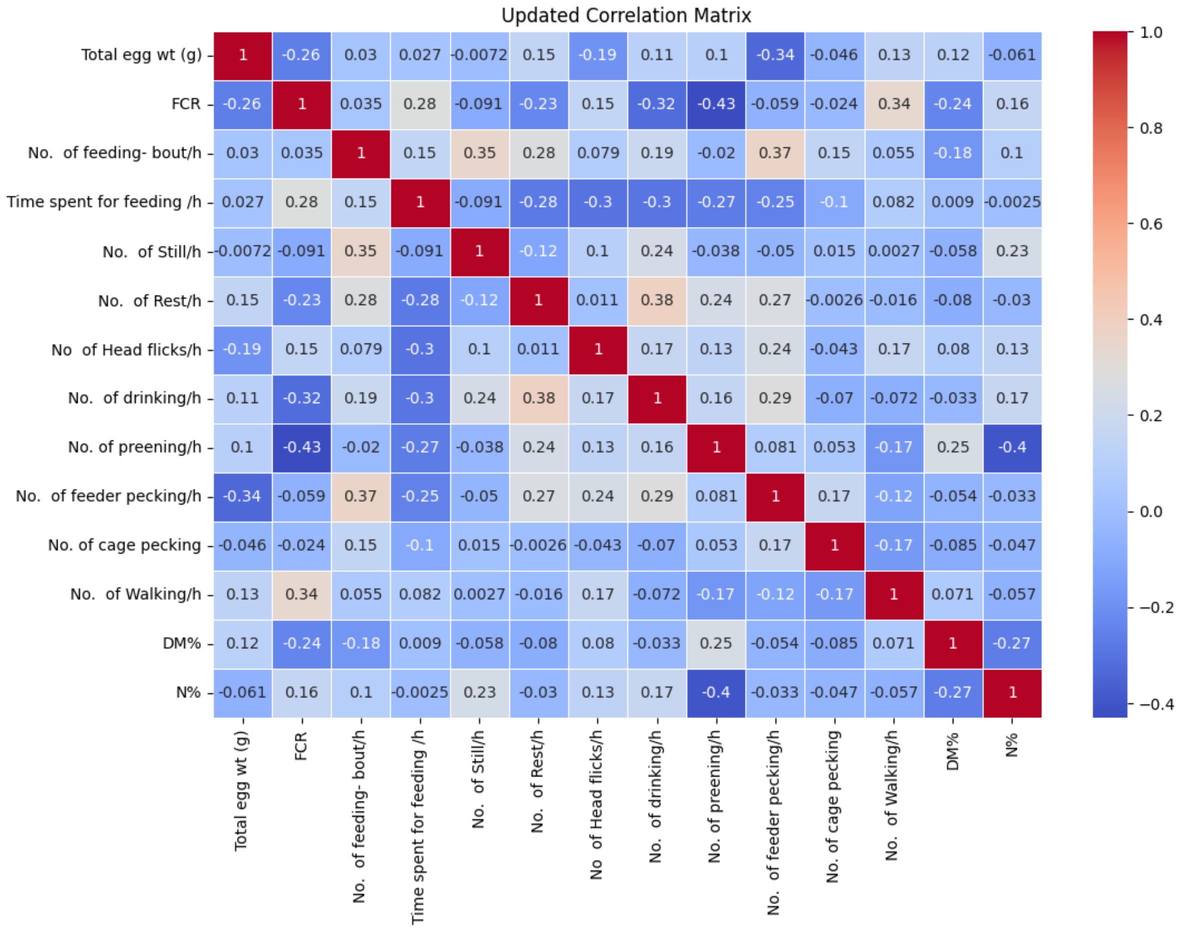

Figure 4-9:heatmap 2

4.7 Feature engineering

Feature engineering is a crucial step in the data preprocessing phase that involves transforming raw data into meaningful input features that enhance the performance of machine learning models. In this research, feature engineering was undertaken to extract, modify, and create features that better represent the underlying patterns in the data. By leveraging domain knowledge and statistical techniques, the goal was to improve model accuracy, reduce noise, and ensure that the input features align with the predictive objectives of the study. This section outlines the techniques applied to refine the dataset and optimize it for modeling tasks.

The following steps were undertaken;

4.7.1 Handling Infinite Values

The first step involved identifying and handling infinite values in the dataset. Infinite values often result from divisions by zero or invalid mathematical operations and can negatively affect both statistical analysis and machine learning algorithms. To ensure data integrity, all instances of positive and negative infinity were replaced with NaN values. This step helped standardize missing or undefined values, enabling more effective handling in later processing stages.

4.7.2 Cleaning the Data

Following the replacement of infinite values, the dataset was further cleaned by removing rows that contained zero values in key numeric columns. Specifically, entries where the 'Time spent for feeding /h' or 'Total egg wt (g)' was equal to zero were excluded. These columns are essential for computing several derived features, such as efficiency scores and behavioral ratios. Retaining such rows could lead to division by zero errors or misleading indicators, hence their removal improved the quality and reliability of the data used in subsequent analysis.

4.7.3 Creation of Derived Features

To capture deeper insights into the birds' feeding and behavioral patterns, several new features were engineered. These features were designed to enhance the dataset by reflecting efficiency, activity levels, and feeding behavior more accurately.

Drinking/Feeding Ratio:

A new feature was created to assess the frequency of drinking relative to the time spent feeding. It was computed by dividing the number of drinking occurrences per hour by the feeding time in hours. This ratio provided insight into hydration behavior during feeding sessions.

Preening/Feeding Ratio:

Similar to the drinking ratio, this feature was derived by dividing the number of preening occurrences per hour by the time spent feeding. It offered a measure of grooming activity in relation to feeding behavior.

Feed Efficiency Score:

This metric was calculated by dividing the total feed intake (in grams) by the total egg weight (also in grams). It provided a direct measure of how efficiently each bird converted feed into egg mass, which is critical for evaluating production performance.

Feeding Intensity:

To understand how feeding time related to rest time, a feeding intensity score was computed by dividing the total time spent feeding by the number of rest hours. This feature reflected the intensity or focus of feeding activity within the bird's daily routine.

Activity Ratio:

To quantify overall behavioral activity, this ratio was derived by summing several behavioral indicators—namely, the number of feeders pecking, drinking, preening, and walking actions per hour—and dividing the total by the time spent feeding. This feature represented how behaviorally active a bird was in relation to its feeding duration.

Combined Feed Intake:

Finally, a new feature was engineered by combining average daily intake (ADI) and total feed intake. This composite metric represented total feed consumption in a more comprehensive manner and was useful for understanding overall feeding trends and variations across different birds.

Each of these steps contributed significantly to enriching the dataset, improving the quality of inputs for the machine learning models, and aligning the data more closely with the study's objective of predicting and optimizing feed efficiency in egg production systems.

4.8 Outlier Detection and Removal

Following the feature engineering phase, the next critical step in the data preprocessing process was the detection and removal of outliers. Outliers are data points that significantly differ from the majority of observations and can distort statistical analyses, bias model training, and reduce predictive performance. In this research, identifying and addressing outliers was essential to ensure data consistency, improve model accuracy, and enhance the generalizability of results. This section outlines the systematic procedures undertaken to detect and eliminate extreme values from the dataset, thereby refining the quality and reliability of the data used for modeling.

4.8.1 Defining the Outlier Removal Function

A function was defined to remove outliers based on the Interquartile Range (IQR) method. The IQR is a statistical measure that identifies the range within which the central 50% of the data lies. Any data point outside this range can be considered an outlier. The function `remove_outliers` accepts three parameters:

df: The dataset from which outliers will be removed.
columns: A list of continuous columns to check for outliers.
factor: The multiplier used to determine the threshold for identifying outliers, set to a default value of 1.5. The IQR method identifies outliers as any value outside of the range defined by:

- ✓ lower bound: Q1 - (factor*IQR)
- ✓ upper bound: Q3 + (factor*IQR)

Where Q1 is the first quartile (25th percentile) and Q3 is the third quartile (75th percentile), and IQR is the difference between Q3 and Q1.
The function iterates over the specified columns, calculates the IQR for each, and filters out any values that fall outside the defined thresholds.

4.8.1.1 Implementation of Outlier Removal

After defining the function, it was applied to the dataset. The continuous columns were selected for outlier detection, and the function was executed to clean the dataset:
df_cleaned = remove_outliers(df, continuous_columns)
This step ensured that the outliers, which could potentially skew the analysis or model predictions, were removed from the dataset. By using the IQR method, the research ensured that the removal process was robust, avoiding the arbitrary removal of values while maintaining the integrity of the data distribution.

4.9 Calculation of Correlation of Feed Conversion Ratio (FCR) with Features

Following the completion of feature engineering and the removal of outliers, the next critical step in the data analysis process was to calculate the correlation between the Feed Conversion Ratio (FCR) and the selected features in the dataset. Understanding the relationship between

FCR and other features is vital for identifying which factors most influence feed efficiency in egg production systems. By evaluating these correlations, the research aims to highlight the key variables that drive feed efficiency and contribute to more informed decision-making in optimizing feeding practices.

This step involved computing the correlation coefficients between FCR and each of the relevant features after ensuring that the data was cleaned and free from any distortions caused by outliers or irrelevant data. The results of these calculations provide insights into how strongly each feature is associated with feed efficiency and inform the selection of features for model development.

```
Top features correlated with FCR:
FCR                            1.000000
Total feed intake (g)          0.725955
ADI (g)                        0.721912
Feed Efficiency Score          0.626399
No.  of Walking/h              0.332554
Time spent for feeding /h      0.307114
Feeding Intensity              0.298183
N%                             0.160797
CP%                            0.160365
No  of Head flicks/h           0.160320
GE  (kcal/kg)                  0.066921
No.  of feeding- bout/h        0.043505
Ash%                          -0.020002
No. of cage pecking           -0.037417
No.  of feeder pecking/h      -0.061153
No.  of Still/h               -0.088828
DM%                           -0.233439
No.  of Rest/h                -0.234685
Activity Ratio                -0.238541
Total egg wt (g)              -0.253501
No.  of drinking/h            -0.317935
Drinking/Feeding Ratio        -0.328842
Preening/Feeding Ratio        -0.355169
No. of preening/h             -0.427503
Name: FCR, dtype: float64
```

4.10 Feature selection

After performing correlation analysis, a careful selection of features was made to ensure that only the most relevant variables are included in the predictive model for Feed Conversion Ratio (FCR). The selected features were chosen based on their significant correlation with FCR, as well as their potential to contribute meaningful insights into feed efficiency and bird behavior. These features reflect various aspects of the birds' feeding habits, activity levels, and feed characteristics, all of which play a role in influencing feed efficiency.

The following features were selected:

Feeding Intensity:

This feature represents the ratio of time spent feeding to the number of rest hours per day. It is an important measure of how actively the birds are consuming feed in relation to their resting time.

Total Feed Intake (g):

This feature captures the total amount of feed consumed by the birds. It directly impacts feed efficiency and is an essential predictor of FCR.

No. of Feeding Bouts/h:

This variable reflects the frequency of feeding events per hour. It provides insight into the birds' feeding behavior and their tendency to engage in multiple feeding sessions.

No. of Head Flicks/h:

Head flicks, indicative of birds' feeding behavior, can signal feeding efficiency and the birds' interaction with the feed. This feature could provide valuable information on the birds' feeding efficiency and habits.

No. of Drinking Events/h:

This feature tracks the frequency of drinking events per hour. While not directly related to feeding, it may indicate the birds' overall health and well-being, which can influence feed consumption and efficiency.

No. of Preening Events/h:

Preening behavior, though not directly tied to feeding, could provide additional context for understanding the birds' overall activity levels and health, potentially influencing feed intake and efficiency.

No. of Feeder Pecking Events/h:

This feature represents the frequency of pecking at the feeder. Frequent pecking events may correlate with higher levels of feed intake, affecting the overall feed conversion efficiency.

No. of Walking Events/h:

Walking behavior is a good indicator of activity levels. It is included as it may relate to the birds' overall health and energy expenditure, which can indirectly affect feed intake and conversion.

GE (kcal/kg):

Gross energy content of the feed is a crucial factor affecting feed efficiency. Higher energy content may lead to better feed conversion and improved overall FCR.

N%:

Nitrogen percentage in the feed is an important nutritional component. It directly influences the growth and productivity of the birds, thereby impacting feed efficiency.

4.11 Model building

The selected features are used to create the feature matrix (`X`) and the target variable (`y`) for the model:

```
X = df[selected_features]
y = df['FCR']
```

4.11.1 Train-test split

To properly evaluate the performance of the model, the dataset is divided into two subsets: training and testing sets. This division allows the model to learn patterns and relationships from one portion of the data (the training set) and then be evaluated on a separate, unseen portion (the testing set). The testing set provides an opportunity to assess the model's generalization ability, ensuring that it can make accurate predictions on new, previously unseen data, rather than just memorizing the training data.

The dataset is typically split using an 80/20 ratio, where 80% of the data is used for training the model, and 20% is reserved for testing. This ratio is commonly used because it provides a sufficient amount of data for both training and evaluation. To ensure consistency and reproducibility of the results, a fixed random state is used during the split process.

Implementation

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

Here; X_train and y_train contain the training data and target variable, respectively, X_test and y_test contain the testing data and target variable, respectively, test_size=0.2 indicates that 20% of the data is allocated for testing and random_state=42 ensures that the split is reproducible, so the results can be consistently reproduced across different runs.

This data split sets the stage for training the model on the training set and evaluating its performance on the testing set.

4.11.2 Model training

The next step involves training a Random Forest Regressor model using the selected features. Random Forest is a powerful ensemble learning method that builds multiple decision trees and combines their predictions to enhance accuracy and reduce the likelihood of overfitting. This method leverages the strength of various decision trees to provide a more robust and reliable prediction, especially in complex datasets.

To begin, the model is initialized with 100 decision trees, setting the parameter `n_estimators=100`. This means the model will use 100 individual trees to make its predictions. Additionally, a fixed `random_state=40` is specified to ensure that the results are reproducible across multiple runs. This ensures that the model's training process can be consistently replicated for evaluation purposes.

The model is then trained on the training data, using the feature matrix **(`X_train`)** and the target variable (`y_train`) to learn the patterns and relationships between the inputs and the output.

rf_selected = RandomForestRegressor(n_estimators=100, random_state=40)`

rf_selected.fit(X_train, y_train)`

4.12 Model evaluation

After training the Random Forest Regressor model, the next critical step is to evaluate its performance. This is done using standard regression metrics that provide a quantitative measure of how well the model predicts the target variable, in this case, the Feed Conversion Ratio (FCR). Evaluating the model on unseen test data allows us to assess its ability to generalize and make accurate predictions on data it has not encountered during training.

The use of regression metrics helps to identify areas of improvement and ensures that the model is not overfitting or underfitting. These metrics provide valuable insights into the model's accuracy, precision, and overall predictive capability. By examining the results of these metrics, we can determine the model's effectiveness in predicting FCR and its suitability for deployment in real-world scenarios.

The following results was obtained:

MAE (selected features): 0.05559074654857137

MSE (selected features): 0.004667683883349603

R^2 Score (selected features):0.8568807955125473

Mean Absolute Error (MAE):

The MAE for the model was calculated as **0.0556**. This metric indicates the average magnitude of errors in the model's predictions. A lower MAE is preferred, and in this case, the relatively low MAE suggests that the model's predictions are, on average, very close to the actual **FCR** values, with only minor deviations.

Mean Squared Error (MSE):

The MSE was found to be **0.0047**. This metric measures the average squared difference between the predicted and actual values, with larger errors being penalized more heavily. The low MSE indicates that the model has effectively minimized larger errors, contributing to its overall accuracy.

R^2Score:

The R^2 score for the model was **0.8569**, meaning that approximately 86% of the variability in FCR is explained by the model's selected features. An R^2 value of this magnitude reflects a strong ability of the model to capture the relationship between the input features and the target variable, indicating good model performance.

4.13 Feature importance

Understanding which features contribute most to the model's predictions is essential for both interpretability and improving the model. The Random Forest Regressor offers a simple approach to extract feature importance values after training. These values highlight the relative significance of each feature in predicting FCR. By examining these importance scores, we can gain valuable insights into which features have the greatest impact on the model's predictions. This information not only helps in interpreting how the model makes its decisions but also allows for better refinement and optimization by focusing on the most influential variables while considering the potential removal of less impactful ones.

	Feature	Importance
0	Total feed intake (g)	0.553434
1	No. of preening/h	0.103428
2	No. of cage pecking	0.077873
3	GE (kcal/kg)	0.053116
4	Feeding Intensity	0.041136
5	No. of Walking/h	0.037072
6	No of Head flicks/h	0.033751
7	No. of feeding- bout/h	0.026175
8	No. of drinking/h	0.026026
9	N%	0.024668
10	No. of feeder pecking/h	0.023320

Table 4-9: Feature importance

4.14 Visualizing the results

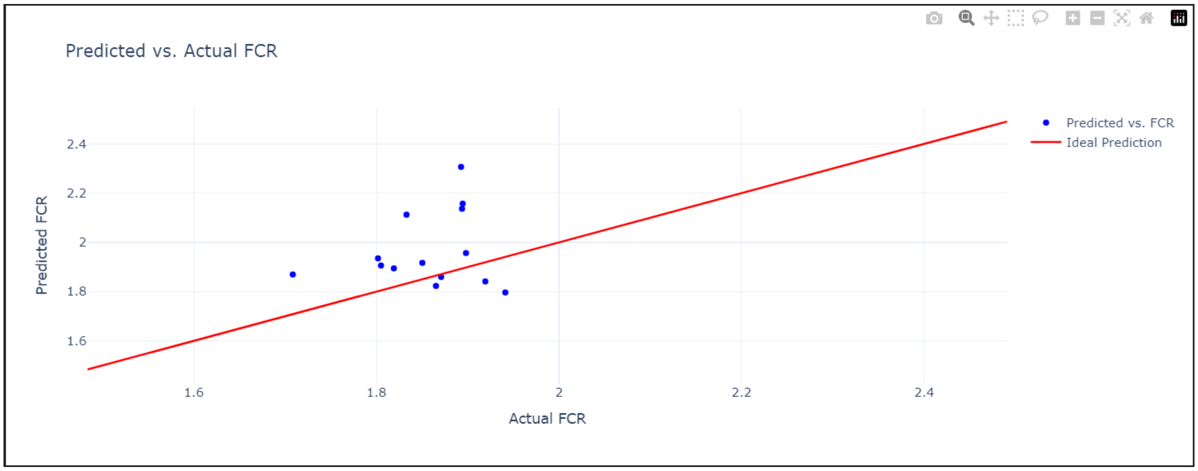

Figure 4-10: Visualizing the results

The blue scatter points represent the predicted FCR values against the actual ones. Ideally, these points should cluster along the diagonal line. The red line represents the ideal prediction scenario (`y = x`), where predictions exactly match the actual values. The closer the points are to this line, the better the model performed. The title and axis labels ensure clear communication of what the plot represents.

4.15 Save the trained model for flask use

The trained model is saved for future use in a Flask application.

```
joblib.dump(rf_selected, 'model.pkl')
print("Model saved as model.pkl")
```

4.16 Model deployment

This section outlines the process of deploying the predictive model for feed conversion ratio (FCR) forecasting in a Flask web application. The application serves as an interface to interact with the trained machine learning model, perform real-time predictions, and optimize feed management for poultry farmers. It also provides a dashboard for visualizing historical FCR data and receiving recommendations on optimal feed schedules based on predicted FCR values.

4.16.1 Framework and Libraries Used

The deployment is powered by Flask, a lightweight web framework, which is used to define routes and handle HTTP requests. Additionally, Dash is used to create an interactive web-based dashboard for visualizing FCR data and optimizing feeding schedules. The model is loaded using joblib, and the dataset is read using Pandas for handling the data.

4.16.2 Loading the Model and Dataset

The trained model is loaded using joblib, and the dataset is read from a CSV file containing bird data, including features used for prediction. The expected features for model prediction are defined and extracted from the dataset to ensure consistent data formatting during inference.

4.17 Frontend

The HTML frontend provides a user-friendly interface for the Greenest Poultry Feed Efficiency Prediction. It consists of several key components that work together to collect input data, display prediction results, and show historical predictions.

4.17.1 Interface architecture

The frontend interface is built using HTML5 for semantic structure and content organization, while Tailwind CSS was employed for responsive styling and layout design. Bootstrap components were selectively integrated to enhance UI elements, particularly for modal dialogs and form controls. Vanilla JavaScript was implemented to handle all application logic, form validations, and API communications without relying on external frameworks, ensuring lightweight performance and straightforward maintenance. This technology stack was chosen to create a clean, efficient user interface that delivered robust functionality while maintaining excellent performance across devices.

4.17.2 Main components

4.17.2.1 Navigation bars

Successfully created a fixed navigation bar at the top of the interface that displayed the Greenest Poultry logo on the left and included navigation links to important sections such as Home, About, Predict, History, and Contact on the right. The navigation bar featured a consistent green color palette across all pages to strengthen brand identity, incorporating hover effects on the links to enhance user interaction. By utilizing a flexbox layout, achieved appropriate spacing and alignment of elements, ensuring the navigation remained accessible and visually attractive on various screen sizes while remaining fixed at the top of the viewport for easy access to main sections.

Figure 4-11:Navigation bars

4.17.2.2 Feature section

Organized the input form into clearly defined feature sections, each with appropriately labeled input fields. For Intensity Features, a Feeding Intensity input field with a range of 0-100 units. The Amount of Feed Consumed section included a required Total Feed Intake field accepting values between 0-500g. For Behavior Metrics, created six distinct input fields covering feeding-bout frequency (0-50/h), Head flicks (0-100/h), drinking (0-50/h), preening (0-60/h), feeder pecking (0-200/h), and Walking (0-100/h). Finally, the Feed Remains Composition section was developed with two specialized inputs: GE (0-5000 kcal/kg) and N% (0-100). Each section was visually separated and clearly labeled to ensure intuitive data entry, with range validations to maintain data integrity.

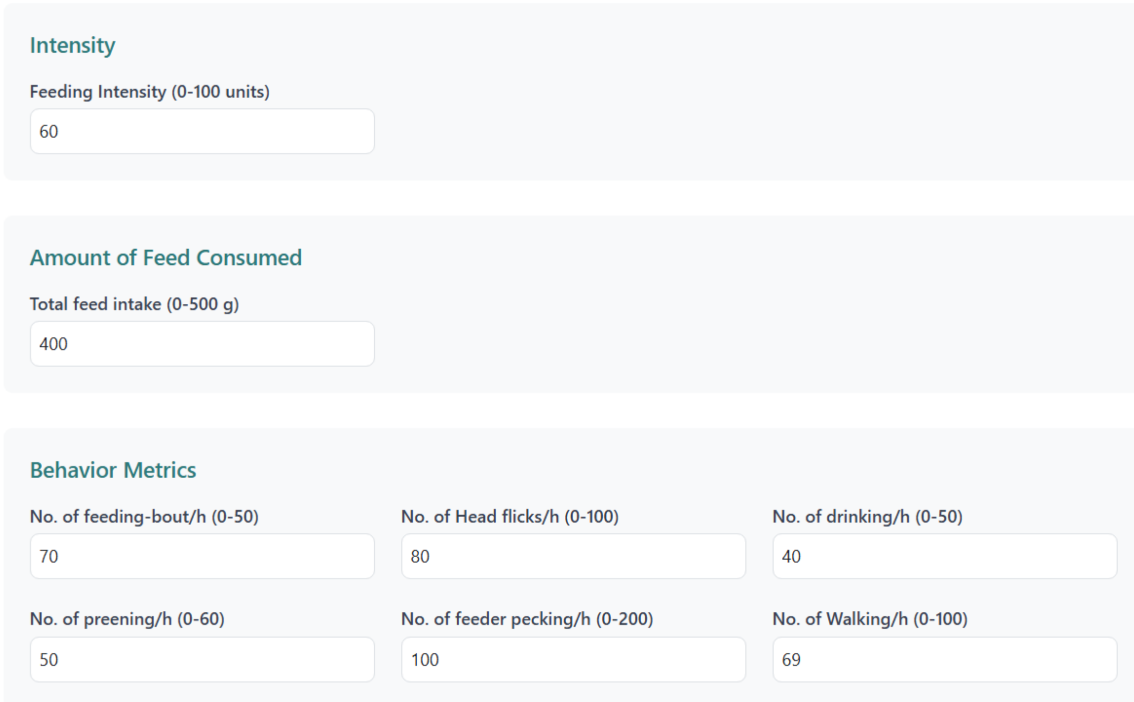

Figure 4-12:Feature section

4.17.2.3 Validation

Performed data validation to ensure data integrity and improve user experience. This is achieved through real-time validation when users moved focus away from input fields (on blur event), immediately checking for proper value ranges and formats. For invalid entries, clear visual warnings below each respective input field, highlighting the specific issue. Additionally, created a summary validation panel at the top of the form that consolidated all warnings for easy reference. To emphasize mandatory fields, specifically marked both the total feed intake and feeding intensity inputs as required, with visual indicators that alerted users if these critical fields were left blank. This multi-layered validation approach effectively guides users to provide complete and accurate data while maintaining an intuitive form-filling experience.

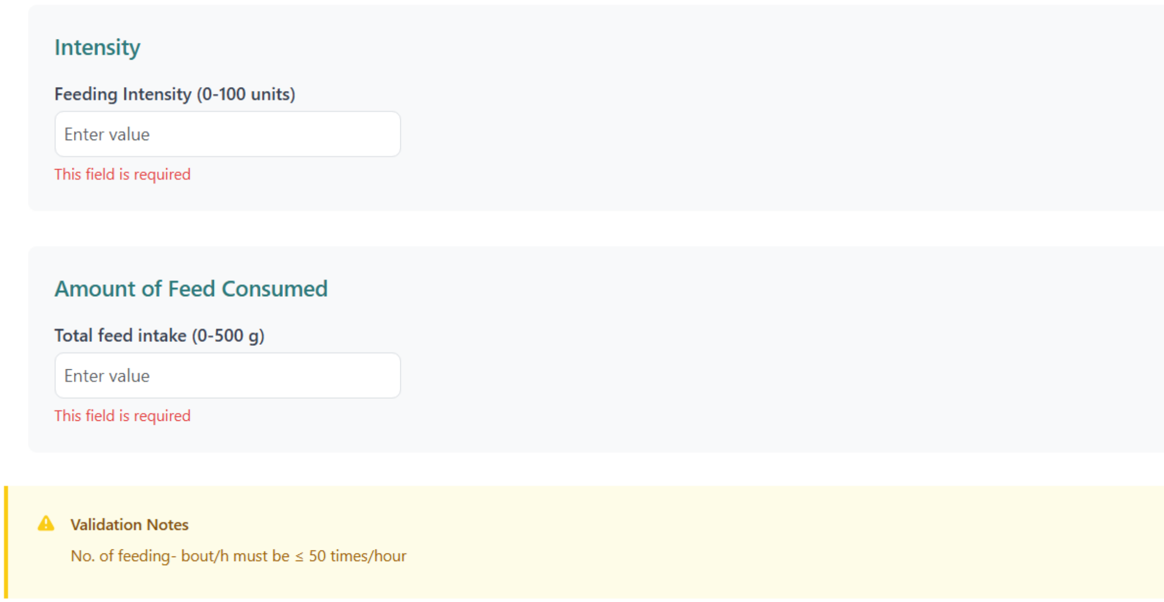

Figure 4-13:Feature validation

4.17.2.4 Javascript function

The JavaScript functions efficiently handled all application operations. The `collectFormData()` gathered form inputs into a JSON object while `resetForm()` cleared all fields. For validation, `validateField()` verified individual inputs against rules and `showValidationWarnings()` displayed backend notes. API interactions were managed through `predict()` for FCR calculations, `optimize()` for feed planning, and `loadHistory()` for retrieving past results. UI functions controlled modal displays, dynamic content refreshes, and error messaging, creating a responsive user experience.

4.17.2.5 Action Buttons

The interface has two key functional buttons; the "Predict Efficiency" button that submits validated form data to the backend for processing, and the "Reset Form" button that cleared all input fields while preserving the validation structure.

Figure 4-14:Action buttons

4.18 Backend

This API predicts and optimizes poultry feed conversion ratio (FCR) using behavioral and nutritional metrics. It maps form fields like feeding intensity to the API's feature mapping, validates inputs, tracks prediction history, and generates optimal feed plans. Which runs on base URL.

```
PS C:\PROJECT4.2> & "C:/Users/Silas Ochieng/AppData/Local/Programs/Python/Python313/python.exe" "c:/PROJECT4.2/Python files/app.py"
 * Serving Flask app 'app'
 * Debug mode: on
WARNING: This is a development server. Do not use it in a production deployment. Use a production WSGI server instead.
 * Running on http://127.0.0.1:5000
Press CTRL+C to quit
```

Table 4-10:App.py

Model Prediction Route (`/predict`):

Accepts POST requests and feature data from the frontend. Processes the incoming data, performs feature extraction, and feeds it into a pre-trained machine learning model for predicting the feed conversion ratio (FCR) and then returns the predicted FCR value and a corresponding alert message for further user guidance.

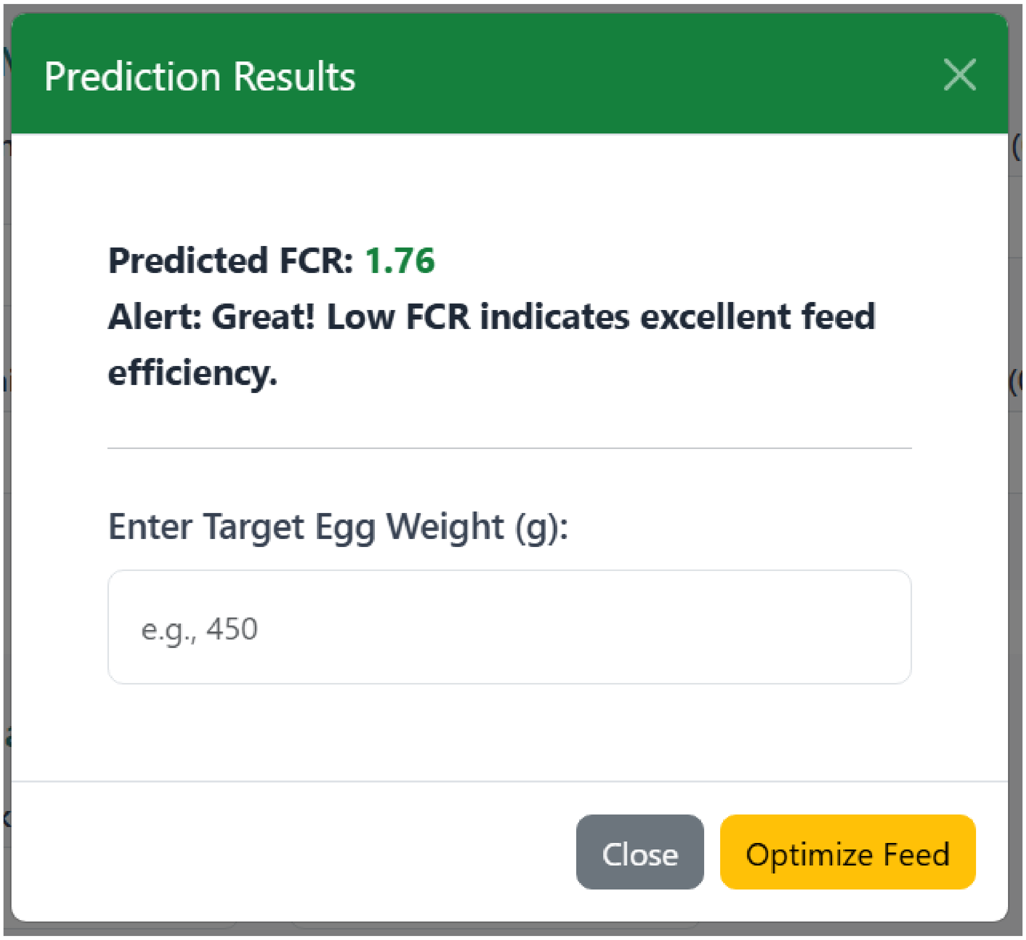

Figure 4-15:Predicted results

Feed Optimization Route (`/optimize`):

Accepts POST requests with target egg weight data. Calculates optimal feed plan based on target egg weight, leveraging the optimization logic from optimization.py. A successful response shows

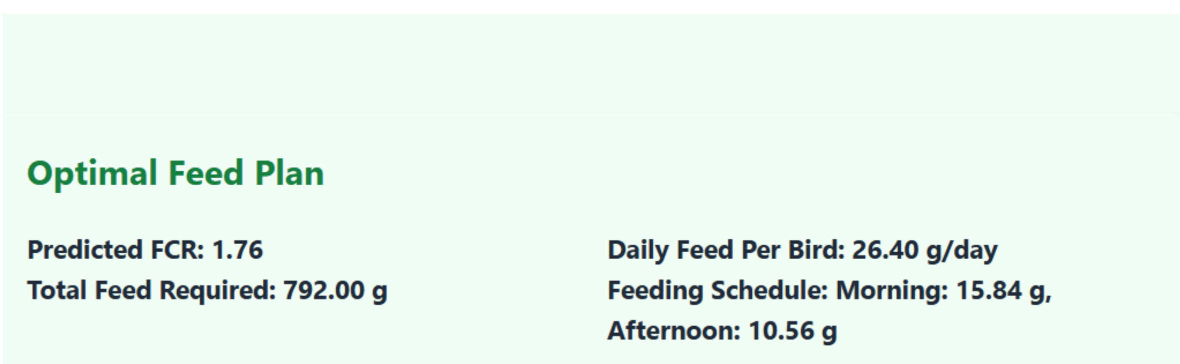

Figure 4-16:Feed plan

4.19 Summary

The deployment of this model provides an efficient tool for optimizing feed efficiency in poultry farming. By combining machine learning predictions with feed optimization algorithms, the system offers valuable insights for farmers, improving sustainability and egg production efficiency. Future iterations will include better scalability, real-time data handling, and more dynamic user inputs for continuous optimization.

CHAPTER FIVE
CONCLUSION AND RECOMMENDATIONS

5.1 Conclusion

This project aimed to develop a data-driven solution for optimizing feed efficiency in poultry farming through predictive modelling and web deployment. The project followed a structured data science pipeline, beginning with data acquisition and ending with model deployment and application testing.

The process began with the identification and collection of relevant datasets, which included key features such as feed intake, egg production metrics, and behavioural indicators. These datasets were thoroughly pre-processed and merged to form a comprehensive dataset suitable for analysis and model building.

Extensive exploratory data analysis (EDA) was conducted to uncover underlying patterns and relationships within the data. This included; univariate analysis to understand the distribution of individual variables, bivariate analysis to examine interactions between pairs of variables and multivariate analysis to explore complex relationships involving multiple features.

Based on insights gained from the EDA, appropriate features were selected for model training. A Random Forest Regressor was chosen due to its robustness and ability to handle non-linear relationships. The model was trained, tuned, and evaluated for its ability to predict key targets such as Feed Conversion Ratio (FCR) and use the predicted feed conversion ratio and the desired egg weight to compute the optimal feed amount and plan

To make the solution accessible and interactive, the trained model was deployed using the Flask web framework. A user-friendly interface (index.html) was developed to allow users to input bird-specific data, receive predictions, and optimize feeding strategies in real time.

The application was thoroughly tested to ensure it functions as intended, providing accurate predictions and optimized feed schedules based on user inputs. The model’s deployment successfully bridges the gap between data science and practical application in the poultry domain.

Overall, this project demonstrates the effective integration of data science techniques and web technologies to deliver a real-world solution for feed optimization and efficiency in egg production.

5.2 Recommendations

To enhance the performance and applicability of the deployed feed efficiency prediction system, several key recommendations are proposed for future development. These enhancements focus on improving scalability, integrating real-time data, and enabling continuous model improvement through user feedback.

Firstly, while the current model performs effectively for small- to medium-scale poultry setups, scaling up to large commercial operations will require significant upgrades. As the volume of data increases, the system must be capable of handling larger datasets without compromising prediction speed or accuracy. This can be achieved by optimizing data processing pipelines and incorporating distributed computing tools such as Apache Spark or Dask. Additionally, deploying the model and application on cloud platforms like AWS, Google Cloud, or Microsoft Azure will enable elastic scalability and better resource management, ensuring the system remains responsive and efficient under high-demand scenarios.

Secondly, the current model relies on static data inputs. For real-time decision-making and dynamic optimization, it is essential to integrate real-time data streams from sensors placed in the poultry environment. These sensors can provide continuous data on variables such as feed intake, water consumption, bird activity, and environmental conditions. By leveraging Internet of Things (IoT) technologies and stream processing tools like Apache Kafka or MQTT protocols, the system can adapt its predictions on the fly, leading to more precise feed adjustments and improved operational efficiency.

Lastly, incorporating a user feedback loop is crucial for the system's long-term effectiveness. Allowing users—such as farm managers or technicians—to provide feedback on prediction accuracy and real-world outcomes can help identify areas where the model may need refinement. This feedback can be collected through the application's interface and stored for future use in retraining or fine-tuning the model. Employing techniques such as active learning, where the system selectively learns from the most informative new data, will enable the model to evolve and maintain high performance over time.

In summary, enhancing scalability, enabling real-time integration, and establishing a user feedback mechanism will significantly improve the robustness, adaptability, and value of the deployed system. These improvements will help transition the model from a proof-of-concept stage to a fully scalable, intelligent tool for modern poultry farm management.

REFERENCES

Bahuti, M., Junior, T. Y., de Lima, R. R., Fassani, É. J., Ribeiro, B. P. V. B., Campos, A. T., & Abreu, L. H. P. (2023). Statistical and fuzzy modeling for accurate prediction of feed intake and surface temperature of laying hens subjected to light challenges. *Computers and Electronics in Agriculture*, *211*, 108050.

Belkhanchi, H., Ziat, Y., Hammi, M., & Ifguis, O. (2023). Formulation, optimization of a poultry feed and analysis of spectrometry, biochemical composition and energy facts. *South African Journal of Chemical Engineering*, *44*(1), 31–41.

Bryden, W. L., Li, X., Ruhnke, I., Zhang, D., & Shini, S. (2021). Nutrition, feeding and laying hen welfare. *Animal Production Science*, *61*(10), 893–914.

Clark, C. E. F., Akter, Y., Hungerford, A., Thomson, P., Islam, M. R., Groves, P. J., & O'Shea, C. J. (2019). The intake pattern and feed preference of layer hens selected for high or low feed conversion ratio. *PloS One*, *14*(9), e0222304.

Coffey, D., Dawson, K., Ferket, P., & Connolly, Ajj. (2016). Review of the feed industry from a historical perspective and implications for its future. *Journal of Applied Animal Nutrition*, *4*, e3.

Depuru, B. K., Putsala, S., & Mishra, P. (2024). Automating poultry farm management with artificial intelligence: Real-time detection and tracking of broiler chickens for enhanced and efficient health monitoring. *Tropical Animal Health and Production*, *56*(2), 75.

Heidari, M. D., Gandasasmita, S., Li, E., & Pelletier, N. (2021). Proposing a framework for sustainable feed formulation for laying hens: A systematic review of recent developments and future directions. *Journal of Cleaner Production*, *288*, 125585.

King'Ori, A. M. (2011). Review of the factors that influence egg fertility and hatchability in poultry. *International Journal of Poultry Science, 10*(6), 483–492.

Ledvinka, Z., Zita, L., & Klesalová, L. (2012). Egg quality and some factors influencing it: a review. *Scientia Agriculturae Bohemica*, *43*(1), 46–52.

Luchkin, A. G., Lukasheva, O. L., Novikova, N. E., Zyatkova, A. V, & Yarotskaya, E. V. (2021). Feasibility study of the influence of the diet on the quality characteristics of poultry production. *IOP Conference Series: Earth and Environmental Science*, *640*(3), 032041.

Narváez-Solarte, W., Rostagno, H. S., Soares, P. R., Uribe-Velasquez, L. F., & Silva, M. A. (2006a). Nutritional requirement of calcium in white laying hens from 46 to 62 wk of age. *International Journal of Poultry Science*, *5*(2), 181–184.

Narváez-Solarte, W., Rostagno, H. S., Soares, P. R., Uribe-Velasquez, L. F., & Silva, M. A. (2006b). Nutritional requirement of calcium in white laying hens from 46 to 62 wk of age. *International Journal of Poultry Science*, *5*(2), 181–184.

Omomule, T. G., Ajayi, O. O., & Orogun, A. O. (2020). Fuzzy prediction and pattern analysis of poultry egg production. *Computers and Electronics in Agriculture*, *171*, 105301.

Oviedo-Rondón, E. O. (2015). Model applications in poultry production and nutrition. In *Nutritional modelling for pigs and poultry* (pp. 125–140). CABI Wallingford UK.

Oviedo-Rondón, E. O. (2019). Holistic view of intestinal health in poultry. *Animal Feed Science and Technology, 250*, 1–8.

Pousga, S., Boly, H., & Ogle, B. (2005). Choice feeding of poultry: a review. *Livestock Research for Rural Development*, *17*(4), 45–46.

Roberts, J. R. (2004). Factors affecting egg internal quality and egg shell quality in laying hens. *The Journal of Poultry Science*, *41*(3), 161–177.

Singh, M., Kumar, R., Tandon, D., Sood, P., & Sharma, M. (2020). Artificial intelligence and iot based monitoring of poultry health: A review. *2020 IEEE International Conference on Communication, Networks and Satellite (Comnetsat)*, 50–54.

APPENDICES

Budget

#	ITEM	DESCRIPTION	COST
1	Equipment	Hp laptop	40000/=
2	Internet	Data bundles	2000/=
3	Miscellaneous	Fare	0.00/=
4	Communication	Phone bills	0.00/=
5	printing	Ream papers	1000/=
		TOTAL	**KSH.43000**

Work plan

TASK/WEEKS	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8	Week 9	Week 10	Week 11	Week 12
Data Identification												
Data pre-processing												
Feature engineering												
Model development												
Model training												
Model testing												
Model deployment												
Report writing												