

A Prediction Model for Estimating Egg Hatch Rate for Ghanaian Farmers Using Machine Learning

Abraham Atosona, Stephen Akobre, Mohammed Ibrahim Daabo

Department of Computer Science, Department of Cyber Security and Computer Engineering Technology, School of Computing and Information Sciences, C. K. Tedam University of Technology and Applied Sciences, Navrongo, Ghana

Email: abratososona@gmail.com, sakobre@cktutas.edu.gh, idaabo@cktutas.edu.gh

How to cite this paper: Atosona, A., Akobre, S. and Daabo, M.I. (2025) A Prediction Model for Estimating Egg Hatch Rate for Ghanaian Farmers Using Machine Learning. *Journal of Data Analysis and Information Processing*, 13, 306-313. <https://doi.org/10.4236/jdaip.2025.133019>

Received: January 10, 2025

Accepted: August 16, 2025

Published: August 19, 2025

Copyright © 2025 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0). <http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Egg loss is one of the major problems in the egg hatching industry. This study aims to support farmers in optimizing their egg hatch through the development of a prediction model. This is to enable local farmers to accurately predict the hatchability of eggs before loading them into the incubator, thereby minimizing losses and reducing time wastage. By employing advanced data analysis techniques, a predictive algorithm was developed to predict the hatchability of eggs based on various parameters, such as egg storage temperature, time of egg storage, quantity processed and other egg characteristics. This predictive model provides farmers with valuable insights into the potential outcome of the incubation process, enabling them to make informed decisions in selecting eggs for incubation. The prediction model exhibited a high level of accuracy in estimating hatchability for farmers. The practical implications of this research are significant as it helps local farmers minimize losses, reduce time wastage and improve the overall efficiency of egg hatch process. This research does not only contribute to the field of agricultural technology, but also provides practical solutions for sustainable farming practices.

Keywords

Hatchability, Algorithm, Incubation, Prediction

1. Introduction

Poultry farming in developing countries is a significant source of revenue generation for smallholder farmers. Also, raising poultry in Africa plays a vital role in providing the protein requirements for humans through its meat and eggs [1].

In modern times, farmers make use of egg incubators to hatch their eggs. This

device provides ideal temperature and humidity in an enclosed environment for egg embryo development without human intervention [2].

In poultry farming, maximizing the effectiveness of incubation, early identification of viable eggs, reducing cost on wasted/spoiled eggs and boosting overall egg hatching while minimizing resource waste are very crucial to every farmer [3]. In recent times, Machine learning is used in the poultry industry to predict egg viability and hatchability, but much has not been achieved in Ghana locally. This technology has been shown to achieve accuracy rates of over 90% in predicting hatchability, making it a useful tool for hatcheries to optimize production and reduce waste and costs [4]. The use of artificial intelligence, machine learning, among other innovative tools, has allowed us to predict early hatching rates with remarkable accuracy by improving hatchability. Electricity, time and space are saved by preventing unnecessary waiting for infertile eggs in the incubator not to hatch [5].

This study aims to develop a hatch prediction model using linear regression to help farmers select good eggs for incubation to avoid wastage of eggs, time and other resources.

2. Related Literature

Incubation is a critical aspect of poultry farming, as it allows farmers to hatch eggs under controlled conditions and ensure the best possible hatching rates. It is therefore necessary that farmers are able to predict the hatchability of the eggs to be incubated. Machine learning algorithms such as regression, decision trees or neural networks can be used to create a model that can predict the optimal results based on the data that one collected. [6] considered the online-prediction kind of standard problem of linear regression. The objective of the researchers was to estimate subsequent responses given the relevant explanatory variables and all of the prior observations. The researchers were mainly interested in predicting intervals instead of point predictions. [7] opined that the prediction distribution is a good basis for predictive inferences applied to several real-world situations. This is a distribution of the unobserved future response(s) conditional on a set of realized responses from an informative experiment. The Bayesian approach was used by the researchers to further derive the prediction distribution(s) for the multiple linear regression model. The Bayesian approach was used by the researchers to further derive the prediction distribution(s) for the multiple linear regression model. Linear regression is a popular machine learning-based prediction tool used in agriculture to predict crop yield. A study was conducted by [8] aimed to predict corn yield using linear regression. The study utilized various predictors, such as soil characteristics, weather conditions, and crop management practices, among others. The results showed that linear regression is a very powerful tool in the prediction of crop yield. The study found that soil characteristics were also significant predictors of corn yield.

Machine learning techniques have also been applied to poultry productivity

prediction. [9] conducted a review of various machine learning methods, including linear regression, neural networks and support vector machines for this purpose. Their study found that linear regression models are effective in predicting poultry productivity when appropriate features are selected and model parameters are optimized. Another literature review on this topic was conducted by [10], who discussed the various types of linear regression models that have been used for predicting poultry performance. They analyzed the strengths and limitations of different models, such as simple linear regression, multiple linear regression, and polynomial regression, and concluded that the choice of model depends on the specific prediction task and data characteristics. Furthermore, [11] examined the application of linear regression models in predicting egg hatchability. Their study revealed that linear regression models can effectively predict egg hatchability by incorporating various factors, such as incubation temperature, humidity, and egg weight. They also highlighted the importance of using appropriate feature selection techniques and sample sizes to ensure accurate predictions. A study by [12] investigated the effectiveness of linear regression models in predicting egg hatchability in poultry farming. The authors analyzed various factors that impact egg hatchability, including egg weight, egg shape index, and incubation temperature. They developed a multiple linear regression model to predict hatchability and compared its performance with that of a simple linear regression model. The study found that the multiple linear regression model achieved a higher accuracy in predicting egg hatchability than the simple linear regression model. However, the authors also noted that the predictive power of the model was affected by the quality and quantity of the input data, as well as the model parameters.

After a critical review of related studies, it has been established that Linear Regression is a crucial tool in machine learning as it has been used in numerous prediction studies. Overall, machine learning has the potential to revolutionize the poultry farming industry by providing insights into the production process, predicting outcomes with high accuracy, improving efficiency, disease management, and animal welfare. While there are challenges associated with implementing machine learning in poultry farming, such as data availability and model interpretability, the benefits are significant and warrant further investigation.

3. Methodology

This study aimed to predict egg hatch rate using linear regression. The study collected data on egg viability/hatchability history from two hatchery centers in Tamale in Northern Region and Kumasi in the Ashanti Region of Ghana. The study combines data from Tamale (extreme weather) and Kumasi (favorable weather), making it robust and representative for conclusions across Ghana.

One thousand and fifty (1050) dataset was collected from the two centers. From the data collected, several independent variables were extracted, including quantity of eggs processed, breed of bird, age of birds, breeding nature (free range/intensive), egg storage temperature, egg duration of storage, transport distance, average size of egg collected, fertility and hatch rate, with hatch rate as the dependent var-

iable. These features were considered crucial by the researcher as they are major factors that drive egg hatch rate in the incubation. **Table 1** below shows the various descriptions of the variables that were used in this study.

Table 1. Description of data collected.

Independent Variable	Description
Quantity of eggs processed	Number of eggs collected and loaded into the incubator at a season.
Breed of bird	This variable captures the specific breed or species of birds under consideration. Different breeds may exhibit variations in reproductive performance and hatchability.
Age of birds	This indicates age of the birds involved in egg production.
Breeding nature	Describes whether the birds are on free range breeding nature or intensive breeding nature.
Egg storage temperature	Temperature at which eggs are stored before incubation.
Egg duration of storage	Amount of time (days) the eggs stored before being loaded into the incubator.
Transport distance	Distance or length of egg transportation.
Average size of egg collected	Average physical size or weight of collected eggs.
Fertility	Reproductive fertility of the birds.

3.1. Data Preprocessing

Collected data was preprocessed before being used for the model training. All variables considered were used for the model. The preprocessing steps included removing missing or null values, removing duplicate entries and encoding categorical variables as numerical variables. The data was then divided into training dataset and test dataset with an 80:20 split ratio, respectively. The 80% of the data was used for the training and 20% used for testing the model.

3.2. Training the Model

The regression model was trained using the preprocessed dataset. The model was trained to predict hatch rate using the independent variables as predictors and performance of the model was evaluated on the test set using commonly used regression metrics such as mean squared error (MSE) and R-squared (R^2).

3.3. Linear Regression as a Prediction

Linear regression can be used as a prediction tool by fitting a linear equation to a set of data points. This equation can then be used to predict the value of the depend-

ent variable for new values of the independent variables.

The linear regression model assumes that there is a linear relationship between the dependent variable and the independent variables. This means that the change in the dependent variable is proportional to the change in the independent variables.

Linear regression is used as a prediction tool in this study as the dependent variable which is the hatch rate is continuous in nature and there is no existence of multicollinearity between the independent variables.

In developing an efficient model, data collected undergoes wrangling and subsequently trained using Linear Regression to ensure accurate model generation. Independent variables for training the model are: quantity of eggs, breed of bird, age of egg, mode of breeding, storage temperature, duration of storage, egg transport distance, average size of egg and egg fertility. These were subjected to heatmap correlation analysis and Pair Plot, which yielded the positive results, appropriate for training the model.

3.4. Hatch Rate Prediction Model

The linear regression equation used to predict hatch rate of eggs is:

$$Y = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4 + b_5X_5 + b_6X_6 + b_7X_7 + b_8X_8 + b_9X_9 \quad (1)$$

where:

Y is dependent variable (hatch rate of eggs) and $X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_9$ are the independent variables (Quantity of eggs, Breed, Age of birds, Free Range, Storage temperature, Time of storage, Distance, Average Size, Fertility), respectively.

$b_0, b_1, b_2, b_3, b_4,$ and b_5 are regression coefficients estimated using ordinary least squares method.

4. Results and Discussions

The Hatch Rate Prediction Model is a critical component of our Indigenous Smart Incubator, which uses machine learning algorithms to analyze and predict the viability of eggs in real-time. The goal of our model is to improve hatch rates, reduce production costs, and increase productivity in the poultry industry.

To establish a reliable model, the data collected were wrangled and subsequently trained using Linear Regression to ensure accurate model generation. The independent variables, which include quantity of eggs, breed of bird, age of egg, mode of breeding, storage temperature, duration of storage, egg transport distance, average size of egg and egg fertility, were subjected to heatmap correlation analysis, which yielded the following result.

Figure 1 indicates that maximum correlation of 0.41. This indicates there is no strong correlation between the independent variables and also, this suggests that linear regression can be applied to the dataset as there is no multicollinearity between the independent variables.

	QUANTITY	BREED	AGE	FREE RANGE	STORAGE TEMPERATURE	TIME OF STORAGE	TRANSPORT DISTANCE	AVERAGE SIZE OF EGG	FERTILITY
QUANTITY	1	-0.0012	-0.042	-0.045	0.17	0.25	0.091	-0.17	-0.012
BREED	-0.0012	1	0.26	-0.12	0.24	0.13	-0.15	0.11	-0.022
AGE	-0.042	0.26	1	-0.11	-0.012	0.14	-0.038	-0.065	-0.41
FREE RANGE	-0.045	-0.12	-0.11	1	0.086	-0.14	0.26	-0.011	0.13
STORAGE TEMP	0.17	0.24	-0.012	0.086	1	0.14	-0.066	0.065	-0.13
TIME OF STORAGE	0.25	0.13	0.14	-0.14	0.14	1	-0.16	-0.097	-0.037
TRANS DISTANCE	0.091	-0.15	-0.038	0.26	-0.066	-0.16	1	0.21	-0.15
AVG SIZE EGG	-0.17	0.11	-0.065	-0.011	0.065	-0.097	0.21	1	-0.24
FERTILITY	-0.012	-0.022	-0.41	0.13	-0.13	-0.037	-0.15	-0.24	1

Figure 1. Heatmap correlation analysis.

Table 2 indicates that there is no strong positive or negative correlation between the independent variables. The maximum correlation 0.2615 exists between the free range and transport distance, which is still minimal and hence we can affirm that there is no multicollinearity between the independent variables. No multicollinearity between the independent variables implies that Linear Regression can be applied to the dataset.

Table 2. Correlation values for the independent variables.

X.corr()	QUANTITY	BREED	AGE	FREE RANGE	STORAGE TEMPERATURE	TIME OF STORAGE	TRANSPORT DISTANCE	AVERAGESIZE OF EGG	FERTILITY
QUANTITY	1.00000	-0.00122	0.042461	0.045087	0.172494	0.248807	0.091236	-0.17239	-0.01187
BREED	-0.001223	1	0.261389	-0.1239	0.237366	0.130912	-0.150602	0.109286	-0.02229
AGE	0.042461	0.261389	1	-0.11159	-0.012272	0.144933	-0.038297	-0.064789	-0.41339
FREE RANGE	0.045087	-0.1239	-0.11159	1	0.085716	-0.13672	0.261487	-0.010552	0.134942
STORAGE TEMPERATURE	0.172494	0.237366	-0.01227	0.085716	1	0.139141	-0.065788	0.065435	-0.13178
TIME OF STORAGE	0.248807	0.130912	0.144933	-0.13672	0.139141	1	-0.163669	-0.096745	-0.0372
TRANSPORT DISTANCE	0.091236	-0.1506	-0.0383	0.261487	-0.065788	-0.16367	1	0.211022	-0.15251
AVERAGE SIZE OF EGG	-0.17239	0.109286	-0.06479	-0.01055	0.065435	-0.09675	0.211022	1	-0.23979
FERTILITY	-0.011873	-0.02229	-0.41339	0.134942	-0.131784	-0.0372	-0.15251	-0.239793	1

4.1. Test of Prediction Model and Its Accuracy

To verify the correctness of the model, 20% of the data was reserved and used for the model testing and the result of the testing was measured using the coefficient of determination (R^2) regression formula, which yielded 0.6781.

The high R-squared value of 67.81 in linear regression implies that the model explains 67.81% of the total variance in the dependent variable based on the inde-

pendent variables included in the model.

This value can be interpreted as the proportion of the variation in the dependent variable that can be accounted for by the independent variables in the model.

4.2. Mean Squared Error (MSE)

The MSE takes into account the squared differences between the actual hatch rates and the predicted values, providing a measure of how spread out these errors are across the dataset. The mean squared error (MSE) was computed is 14.37. A lower MSE generally indicates better model performance, suggesting that the model's predictions are closer to the true values.

It can also be said that on average, the predictions made by the linear regression model are off by approximately 14.37 units in the context of hatch rate.

4.3. Linear Regression Equation of the Model

Intercept and coefficient of the independent variables, the linear regression equation of the trained model can be expressed as follows in Equation (2):

$$Y = 169.1382 - 0.0111X_1 - 3.9850X_2 - 4.9637X_3 + 8.9641X_4 - 3.0556X_5 - 2.2422X_6 + 11.9571X_7 + 13.14785X_8 + 0.0038X_9 \quad (2)$$

where:

X_1 = Quantity of eggs, X_2 = Breed, X_3 = Age of birds, X_4 = Free range, X_5 = Storage temperature, X_6 = Time of storage, X_7 = Distance, X_8 = Average size, X_9 = Fertility.

5. Conclusions

This study aimed to develop a prediction model for estimating the egg hatch rate for local farmers using a linear regression model in machine learning. Through the collection and analysis of relevant data, including various independent variables such as the quantity of eggs processed, breed of bird, age of birds, breeding nature, egg storage temperature, egg duration of storage, transport distance, average size of eggs collected, fertility, and hatch rate, we were able to derive valuable insights.

The correlation analysis, as demonstrated by the heatmap and pair plot, revealed a lack of multicollinearity among the independent variables, indicating that they contributed unique information to the prediction model. The R^2 value of 0.6781 indicated that approximately 67.81% of the variability in the egg hatch rate could be explained by the selected independent variables. Furthermore, the mean squared error of 14.37 indicated that the model's predictions were reasonably accurate. This study ultimately makes a significant contribution to its scope as it helps farmers in Ghana appropriately predict egg hatch rates, thereby minimizing egg losses in the process of incubation.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Attia, Y.A., Rahman, M.T., Hossain, M.J., Basiouni, S., Khafaga, A.F., Shehata, A.A., et al. (2022) Poultry Production and Sustainability in Developing Countries under the COVID-19 Crisis: Lessons Learned. *Animals*, **12**, Article 644. <https://doi.org/10.3390/ani12050644>
- [2] Radhakrishnan, K., Jose, N., Sanjay, S.G., Cherian, T. and Vishnu, K.R. (2014) Design and Implementation of a Fully Automated Egg Incubator. *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering*, **3**, 7666-7672. https://www.ijareeie.com/upload/2014/february/54_Radhakrishnan.pdf
- [3] Bolzan, A., Machado, R. and Piaia, J. (2008) Egg Hatchability Prediction by Multiple Linear Regression and Artificial Neural Networks. *Revista Brasileira de Ciência Avícola*, **10**, 97-102. <https://doi.org/10.1590/s1516-635x2008000200004>
- [4] Ayoub Shaikh, T., Rasool, T. and Rasheed Lone, F. (2022) Towards Leveraging the Role of Machine Learning and Artificial Intelligence in Precision Agriculture and Smart Farming. *Computers and Electronics in Agriculture*, **198**, Article ID: 107119. <https://doi.org/10.1016/j.compag.2022.107119>
- [5] Çevik, K.K., Koçer, H.E. and Boğa, M. (2022) Deep Learning Based Egg Fertility Detection. *Veterinary Sciences*, **9**, Article 574. <https://doi.org/10.3390/vetsci9100574>
- [6] Vovk, V., Nouretdinov, I. and Gammerman, A. (2009) On-line Predictive Linear Regression. *The Annals of Statistics*, **37**, 1566-1590. <https://doi.org/10.1214/08-aos622>
- [7] Rahman, A. and Khan, S. (2007) Prediction Distribution for Linear Regression Model with Multivariate Student-T Errors under the Bayesian Approach. <https://www.researchgate.net/publication/257939560>
- [8] Cubillas, J.J., Ramos, M.I., Jurado, J.M. and Feito, F.R. (2022) A Machine Learning Model for Early Prediction of Crop Yield, Nested in a Web Application in the Cloud: A Case Study in an Olive Grove in Southern Spain. *Agriculture*, **12**, Article 1345. <https://doi.org/10.3390/agriculture12091345>
- [9] Nosratabadi, S., Ardabili, S., Lakner, Z., Mako, C. and Mosavi, A. (2021) Prediction of Food Production Using Machine Learning Algorithms of Multilayer Perceptron and ANFIS. *Agriculture*, **11**, Article 408. <https://doi.org/10.3390/agriculture11050408>
- [10] Fontana, I., Tullio, E., Fernández, A.P. and Berckmans, D. (2023) Frequency Analysis of Vocalisation in Relation to Growth in Broiler Chicken. *7th European Conference on Precision Livestock Farming*, Milan, September 2015, 3. https://www.researchgate.net/publication/281836267_Frequency_analysis_of_vocalisation_in_relation_to_growth_in_broiler_chicken
- [11] Klein, D.R., Vale, M.M.d., Silva, M.F.R.d., Kuhn, M.F., Branco, T. and Santos, M.P.d. (2020) Data Mining as a Hatchery Process Evaluation Tool. *Scientia Agricola*, **77**, e2018007. <https://doi.org/10.1590/1678-992x-2018-0074>
- [12] Oleforuh-Okoleh, V.U. (2016) Hatchability Prediction in Chickens Using Some External Egg Quality Traits. *Asian Journal of Animal Sciences*, **10**, 159-164. <https://doi.org/10.3923/ajas.2016.159.164>