Jurnal Ilmiah Matematika dan Terapan
Volume 21 Number 1 June 2024 (Pages 65 - 73)
doi: https://doi.org/10.22487/2540766X.2024.v21.i1.17214

Jurnal Ilmiah Matematika dan Terapan
ISSN : 2540 – 766X

CHICKEN EGG PRICE PREDICTION USING EXTREME GRADIENT BOOSTING METHOD

Helmy Fauzan Wahyudiputra[1*], Wina Witanti[2] and Gunawan Abdillah[3]

[1,2,3] Study Program of Informatics, Faculty of Science and Informatics
Jenderal Achmad Yani University
[1]helmy.fauzan2506@gmail.com, [2]witanti@gmail.com, [3]gunawanabdillah03@gmail.com
*(*corresponding author)*

ABSTRACT

Prediction is a way to estimate future data based on previous data patterns. This research focuses on predicting the price of chicken eggs, one of the basic needs of the community. There are various types of chicken eggs, such as omega 3 eggs, purebred chicken eggs, free-range chicken eggs, and domestic chicken eggs. This research will predict the price of purebred chicken eggs in the local market, which often experiences significant price fluctuations. In this research, to make predictions using the XGBoost (Extreme Gradient Boosting) method. XGBoost is a relatively new and advanced machine learning method, which is not widely used. This method is able to identify trends or patterns in time series data, making it suitable for predicting prices that change over time. In this study, before the data is processed by the XGBoost model, we apply normalization and data balancing techniques to improve data quality. This step is important to ensure that the model can work with cleaner and more balanced data, which can improve prediction accuracy.

Keyword : Prediction, Chicken Eggs, Extreme Gradient Boosting

I. INTRODUCTION

Information technology is currently a very useful thing for everyday life. The development of technology today can make it easier for people to get information, including information about basic necessities spread in markets. Chicken eggs are one of the ingredients for daily needs that are very important because chicken eggs are ingredients to be able to meet food needs for consumption in everyday life (Anwari and Yulita 2023). Chicken eggs are a staple that is in great demand by the community because of their affordable price. Related to this, there are several factors that make chicken eggs can increase or decrease in price, especially in local markets because the place is always crowded by people to buy something for their daily needs. The price increase in chicken eggs which is sometimes quite high occurs when approaching holidays (Puteri and Silvanie 2020). With several factors affecting the price of chicken eggs in the local market, people need information about predicting the increase or decrease in the price of chicken eggs so that people get price information for the future(Mujilahwati and Fauziah 2018).

The ups and downs of sales of staples for daily needs can increase and decrease in price suddenly due to factors that can increase prices, such as decreased production or approaching holidays. Therefore, this study aims to be able to predict the pattern of ups and downs in the price of chicken eggs in order to prepare when there is a price increase that can jump high because there are several factors for the increase in the price of chicken eggs. Because this can be a benchmark for sellers in order to prepare a sales strategy when prices are experiencing a decline in the price of chicken eggs so that there are no losses in sales.(Ghulam et al. 2022). The results of this prediction can be useful to see patterns for future chicken egg prices. By using the XGBoost or Extreme Gradient Boosting method which has a way of working to create a new model, then can pay attention to the error of the previous model. The XGBoost method can be useful for predicting future production patterns by looking at previous data patterns.(Mubarok et al. 2022). To get accurate data, the XGBoost method can use an ensemble learning approach which is useful for creating prediction models. The model provides better data accuracy than the model in the previous study.(Admirani 2018). Application of normalization and data balancing techniques to improve data quality before processing by the XGBoost model. Analytical model evaluation using various metrics such as Mean Squared Error, Root Mean Squared Error, R-squared, and Root Mean Squared Percentage Error to assess prediction accuracy. The XGBoost method can predict with high accuracy with a large amount of data. The method conducts training on the data owned in order to recognize the data pattern. The XGBoost method can produce training data that is fast enough and more accurate than the Support Vector Machine (SVM) method, the XGBoost method can make predictions in real time with accuracy reaching 79%. Therefore, in previous research, this XGBoost method can perform a prediction effectively and produce reliable predictions. Previous research conducted a prediction using the Random Forest method is enough to produce good accuracy. In previous research in 2018 predicting cell phone prices using the Random Forest method to predict with an accuracy of 81%.(Saadah and Salsabila 2021).

This research uses the XGBoost method to predict the price of chicken eggs in the local market based on historical data. The goal is to produce accurate predictions that can help understand and anticipate future fluctuations in chicken egg prices. By utilizing XGBoost's ability to identify data patterns and trends, this research can produce good predictions for various parties involved in the supply and sales chain of chicken eggs. The prediction results can be a useful system for planning and decision-making regarding chicken egg prices in the local market.

II. LITERATURE REVIEW

2.1. Prediction

Prediction is an attempt to estimate the time to come (Ridwansyah and Zakaria 2023) by looking at previous data which is used as a variable to be able to determine patterns that will occur in the future. Predictions can be a solution to problems regarding something that is uncertain, for example predicting the price of rice or predicting the price of eggs where the problem cannot be clearly ascertained if there is no previous data. Prediction has the main objective, namely to carry out a scientific process in which prediction functions to obtain systematic information in accordance with the existence of physical evidence.(Jange 2021). Prediction in this study focuses on recognizing patterns of price ups and downs, in order to find out future price patterns by collecting previous data which will be used as variables to be able to recognize patterns from previous data. Then the data that has been collected will be processed to be able to recognize future patterns.

2.2. Gradient Boosting

Gradient Boosting is a technique in machine learning that functions to create a predictive model. Gradient Boosting is also a development of Reccurent Neural Network (RNN), where Reccurent Neural Network can handle relationships on different time scales(Putra Nasyuli et al. 2023). The Gradient Boosting method is carried out for the evaluation of the model to be made. This method uses an ensemble technique where a model is combined iteratively to improve the performance of the model which can reduce erroneous model predictions at each iteration.(Ridwansyah and Zakaria 2023).

2.3. Extreme Gradient Boosting

The XGBoost or Extreme Gradient Boosting method is a gradient boosting development method based on the ensemble learning algorithm. The method can be used to make predictions, especially to overcome problems that have a lot of data. The XGBoost method is also known as a method that attempts to carry out training on data quickly which makes the method have good scalability. The XGBoost method is also known as a method that has an effort to train on data quickly which makes the method has good scalability. The use of the XGBoost Method has a high predictive value that can provide a prediction of the price of chicken

eggs. In the XGBoost method there is a feature that is quite helpful in data processing, which can prevent overfitting when processing data.

2.4. Boosting

Boosting was proposed by Robert Schapire which is a way to be able to collect data which then works to group weak models which are then converted into more accurate models for prediction. In this case, it can make the model that has been grouped and processed more accurate because the weak or simple model will be improved until it can become an accurate model (Adli 2021). Boosting works by calculating a model that will later be grouped into new information with previously existing data. After that, boosting combines all the previous data and the latest data to be processed so that all the data can produce better data, which makes it more accurate when making a prediction(Homepage, Suryana, and Tri Prasetio 2021).

III. METHOD

The research method discusses the methods used to predict the price of chicken eggs using the Extreme Gradient Boosting method.

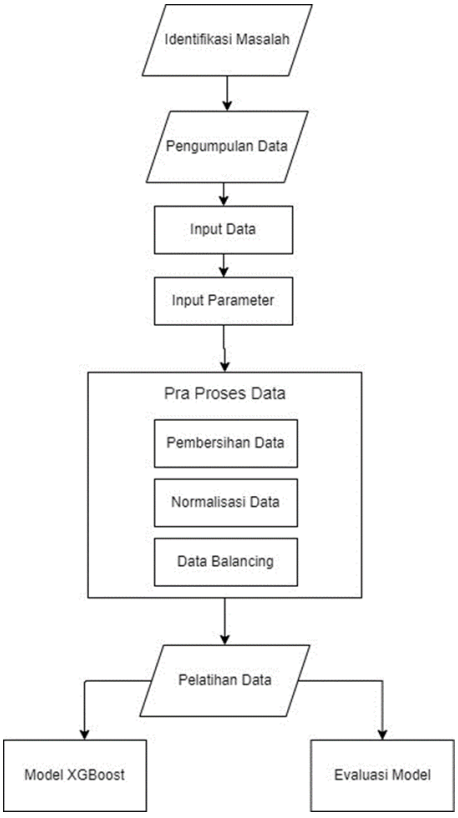

Figure 1 : Routine Diagram

In the flow chart analysis, the initial stage is to identify the problem where the identification of this problem is required to understand the case related to the factors of chicken egg prices that experience ups and downs, and identify variables regarding the price of chicken eggs. Then determine the problem in the related case in the form of data and methods that will be used. The second stage was to collect the data obtained to carry out this research, namely chicken egg price data in CSV format, where the data amounted to 4,259 data. The third stage of entering the Dataset This is the stage where the existing dataset is entered into the system, which then the model will be trained with the aim of making a prediction. The model is trained to provide prediction results with high accuracy. The fourth stage introduces the XGBoost parameters into the system, which are intended as a reference when making efforts to control the performance of the XGBoost model. The following are the parameters of the XGBoost model: Colsamp_bytree: Features used at each iteration to prevent overfitting, Maxdepth : This parameter indicates the maximum depth of the tree, Min_child_weight : This parameter indicates the minimum weight of the node, Subsample: This parameter indicates the data to be trained to prevent overfitting, Objective: This parameter indicates the main goal the model is trying to achieve, Gamma: This parameter determines the truncation of the tree, Eval_metric: This parameter indicates the evaluation metric used when training the split data. Then at the next stage perform data cleaning in the data pre-processing process, data cleaning involving data that has been collected will go through the process of identifying data that has invalid values. Invalid data will then be treated or deleted so that the data obtained becomes more valid. Normalization The next data stage is data normalization which aims to transform data that is able to determine the variables in the dataset have the same scale. The following equation is used in the normalization process.

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad \text{.....(1)}$$

The explanation of the formula above is as follows:

X_norm = Normalized value
X_min = Minimum value in the dataset
X_max = Maximum value in the dataset
X = Original value

Then the next stage performs data balancing which aims to overcome the imbalance of classes in a dataset. This data balancing usually handles data that has a larger or smaller number of samples than other samples. Then the last stage carried out is evaluating the accuracy results obtained on data that has been grouped and trained, in order to get the results of predictive values that have high accuracy. In this case, the evaluation used is RMSE (Root Mean Square Error) which is a square root of MSE (Mean Square Error) which can affect the dataset. There is an RMSE value that can be used using an equation formula.

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(a_i - y_i)^2}{n}} \quad \text{.....(2)}$$

The explanation of the above equation is as follows:

a_i = actual data value

y_i = predicted value of the data

n = the number of data

IV. TESTING

System testing is the stage of testing the software to determine the success of prediction and determine the success rate of chicken egg prediction using Extreme Gradient Boosting technique.

Table 1 : Test Result

Fault Metrics	Value
Mean Squared Error	21954515.65896017
Root Mean Squared Error	4685.564604074963
R-squared	0.6866364473210181
Root Mean Squared Percentage Error	15.475194254599147%

The results of the above tests regarding the evaluation of this model provide the Mean Squared Error (MSE) is 21954515.65896017 which measures the average square of the error between the actual value and the predicted value. The Root Mean Squared Error (RMSE) is 4.685.564604074963 which is the square root of the MSE and provides a measure of the average prediction error. The R-squared value is 0.687 which indicates that about 68.7% of the variation in the target variable can be explained by this model. The Root Mean Squared Percentage Error (RMSPE) is 15.48%, which indicates that the average prediction error is about 15.48% of the actual value. This percentage error provides a more intuitive understanding of the prediction accuracy relative to the actual value scale.

4.1. Extreme Gradient Boosting Algorithm Testing

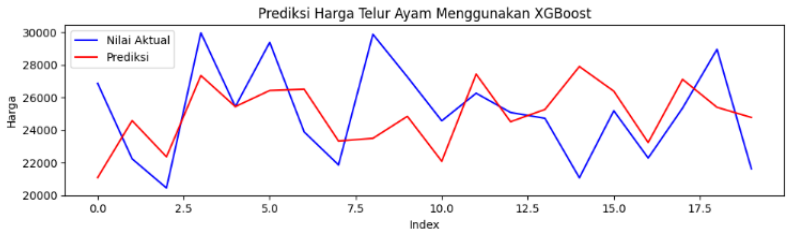

Figure 2 : Index Chart

This graph explains the comparison between the actual price and the predicted price of chicken eggs using the XGBoost method. There are significant fluctuations in the actual price of chicken eggs throughout the period shown. The XGBoost model used is able to capture the general trend of price movements, although it is not always accurate in predicting extreme peaks and valleys. At some points the predictions were quite close to the actual values, but at

other points there were considerable differences. The model provides smoother predictions compared to sharp fluctuations in the actual values. Prediction accuracy results varied, overall the XGBoost model performed quite well in predicting the general trend of chicken egg prices. But there is still room for improvement, especially in handling extreme fluctuations in actual prices.

4.2. Normalization Testing

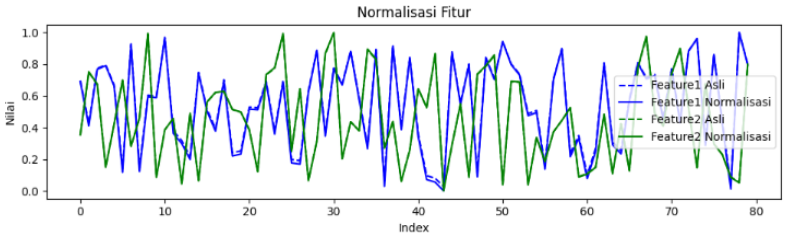

Figure 3 : Normalization Graph

This graph describes the normalization process of two different features, which are shown with blue lines for Feature1 and green for Feature2. The X-axis represents the data index, while the Y-axis shows the normalized feature values between 0 and 1. The two features have different fluctuation patterns before and after normalization. Feature1 tends to have smoother variations, while Feature2 shows sharper and more frequent fluctuations. The normalization process successfully scales both features into the same range, allowing for easier comparison between the two. But the unique characteristics of each feature are still preserved, as seen from the different fluctuation patterns that still exist after normalization.

V. RESULTS AND DISCUSSION

The results of using Extreme Gradient Boosting on egg price prediction produce stable price predictions. Can be seen in the following figure 4.

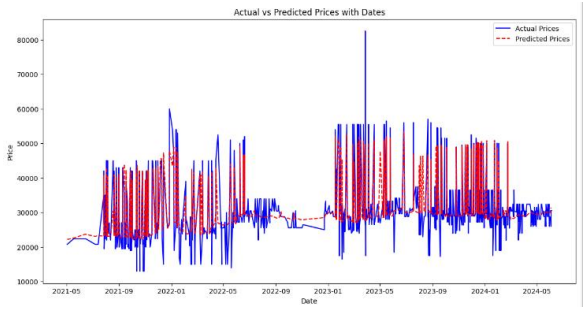

Figure 4 : Price Prediction Chart

The figure above shows the prediction graph of chicken egg prices using the Extreme Gradient Boosting method produces a graph that has a stable price. The model prediction (red line) follows the actual data trend (blue line) well. This shows that the model is able to capture basic patterns in price

data. However, there are some periods where the model's predictions deviate quite far from the actual prices, especially at very high or very low price peaks. In such cases it can be concluded that the model may have difficulty predicting extremes or outliers in the data.

The graph shows that the actual price has considerable fluctuations and there are often volatile price changes in a short period of time. This can be a challenge for predictive models, as high volatility is difficult to predict with high accuracy. Around mid-2022 to early 2023 there were many sharp price increases and decreases in actual prices that could not be fully followed by the model predictions. This graph shows that the XGBoost model is quite effective in capturing the general trend of prices, but there is room for improvement especially in predicting very extreme price fluctuations. The previous evaluation of error metrics such as MSE and R-squared can provide a more quantitative picture of the quality of the model at work and where it falls short.

VI. CONCLUSION

This research predicts the price of chicken eggs in local markets in Padang city, West Sumatra province using the Extreme Gradient Boosting method. Based on the test results, the XGBoost method is trained using data that has been processed. Model evaluation is done using metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared. The test results show that the MSE value is 21,954,515.65896017, RMSE is 4,685.564604074963, and R-squared is 0.6866364473210181. The Root Mean Squared Percentage Error (RMSPE) obtained is 15.475194254599147%. Analysis of the test results shows that the XGBoost model is able to explain 68.7% of the variation in the target variable, namely chicken egg prices. The prediction graph shows that the model follows the trend of the actual data well, although there are some periods where the prediction deviates quite far from the actual price. This shows that the model still struggles to predict extreme price fluctuations or outliers in the data.

REFERENCES

[1]. Adli, Danung Nur. 2021. “Prediksi Harga Jagung Menggunakan Metode Fuzzy Time Series Dengan Atau Tanpa Menggunakan Markov Chain.” *Jurnal Nutrisi Ternak Tropis* 4(1): 49–54.

[2]. Admirani, Ica. 2018. “Penerapan Metode Fuzzy Time Series Untuk Prediksi Laba Pada Perusahaan.” *JUPITER (Jurnal Penelitian Ilmu dan Teknologi Komputer)* 10(1): 19–31.

[3]. Anwari, Ihda, and Intan Nurma Yulita. 2023. “Penggunaan Machine Learning Untuk Prediksi Harga Telur Ayam Ras Di Kota Bandung.” *Community Development Journal : Jurnal Pengabdian Masyarakat* 4(3): 6001–5. http://journal.universitaspahlawan.ac.id/index.php/cdj/article/view/17689.

[4]. Ghulam, Brillian, Ash Shidiq, Muhammad Tanzil Furqon, and Lailil Muflikhah. 2022. “Prediksi Harga Beras Menggunakan Metode Least Square.” *Pengembangan Teknologi*

Informasi dan Ilmu Komputer 6(3): 1149–54.

[5]. Homepage, Journal, Nana Suryana, and Rizki Tri Prasetio. 2021. "Penanganan Ketidakseimbangan Data Pada Prediksi Customer Churn Menggunakan Kombinasi SMOTE Dan Boosting." *IJCIT (Indonesian Journal on Computer and Information Technology)* 6(1): 31–37. https://creativecommons.org/licenses/by-sa/4.0/.

[6]. Jange, B. 2021. "Prediksi Harga Saham Bank BCA Menggunakan Prophet." *Journal of Trends Economics and Accounting …* 2(1): 1–5. https://journal.fkpt.org/index.php/jtear/article/view/168.

[7]. Mubarok, Muhammad Rizky et al. 2022. "HYPER-PARAMETER TUNING PADA XGBOOST UNTUK PREDIKSI KEBERLANGSUNGAN HIDUP PASIEN GAGAL." 09(02): 391–401.

[8]. Mujilahwati, Siti, and Suci Nur Fauziah. 2018. "Pemodelan Ooad Aplikasi Prediksi Harga Sembako Berbasis Android." *Antivirus : Jurnal Ilmiah Teknik Informatika* 12(1): 1–7.

[9]. Puteri, Kandari, and Astried Silvanie. 2020. "Machine Learning Untuk Model Prediksi Harga Sembako." *Jurnal Nasional Informatika* 1(2): 82–94.

[10]. Putra Nasyuli, Leriansyah et al. 2023. "Penerapan Model Machine Learning Algoritma Gradient Boosting Dan Linear Regression Melakukan Prediksi Harga Kendaraan Bekas Application Of Machine Learning Models and Gradient Boosting Algorithms Doing Linear Regression Vehicle Price Prediction Used." *Penerapan Model Machine Learning Algoritma Gradient Boosting dan Linear Regression Melakukan Prediksi Harga Kendaraan Bekas* 2(2): 299–310. https://jurnal.unity-academy.sch.id/index.php/jirsi/index%0Ahttp://creativecommons.org/licenses/by-sa/4.0/.

[11]. Ridwansyah, Muhammad, and Hadi Zakaria. 2023. "Implementasi Algortima Gradient Boosting Pada Aplikasi Hutang Piutang Perorangan Secara Berbasis Web Untuk Meningkatan Akurasi Prediksi Pelunasan Hutang (Studi Kasus : PT Naila Kreasi Mandiri)." 1(4): 440–51.

[12]. Saadah, Siti, and Haifa Salsabila. 2021. "Prediksi Harga Bitcoin Menggunakan Metode Random Forest." *Jurnal Komputer Terapan* 7(1): 24–32.