Implementation of Naive Bayes Classifier (N.B.C.) on Chicken Egg Classification

T S Jaya[1*] and M Yusman[1]

[1] Informatics Management Study Program,
Politeknik Negeri Lampung, Indonesia

*Email: sandi@polinela.ac.id

Abstract. Egg quality has a significant impact on sales; However, many poultry farmers are still not paying attention to the effects of egg quality. This study aims to predict the quality of broiler eggs worth selling or not—classified by the Naive Bayes algorithm with the input variable (X). These factors affect the quality of chicken eggs in the form of 5 variables X and Y in chicken egg sales conditions. The data used are 215 data. The results show that the Naive Bayes model produces an accuracy of 86,7%.

1. Introduction

Eggs are known to be a rich source of nutrients. Eggs are one of the animal products consumed along with meat, fish, and dairy products. To meet consumer demand for eggs, egg production has become an important industry in many countries. The high demand for eggs must also be accompanied by the expectation and demand for good quality eggs so that egg companies compete not only for producing good quality eggs but also for being rich in nutrients like D.H.A., Omega3, etc. [1]. The fundamental nature of eggs is that they are perishable. Various types of egg damage reduce the quality of the eggs, including eggshell cracking, loss of CO2, spoilage, bacterial growth in eggs, and others. The nutritional value of a 50-gram egg is 6.3 grams of protein, 0.6 grams of carbohydrates, five grams of fat, vitamins, and minerals [2]. To determine a good hen's egg, you can look at the type, shell color, weight, and in this case, the Naive Bayesian Classifier (N.B.C.) method is used to analyze the quality and quality of the eggs being tested. They are worth selling or not.

The Naive Bayesian Classifier is a Bayesian classifier that combines the naivety of assuming that conditions between attributes are independent and predicts future opportunities based on experience [3]. One of the features of N.B.C. is that it can work with invalid attribute values, removing training data during modeling and forecasting[4]. Hence, strong N.B.C. is faced with unrelated attributes. Relational attributes can degrade the performance of the N.B.C. classifier because there are no longer any prerequisites for separating these attributes [5]. Bayes' theorem is combined with the naivety of predicting future possibilities based on experience, assuming These elements were previously used for manual calculations and analysis, making it faster, more efficient, and more accurate using the Naive Bayesian Classifier method to maximize the sale of eggs to customers.

Theoretical Basis. Naïve Bayes is a classification technique based on Bayes's theorem combined with naivety to predict future opportunities based on experience, assuming conditions between attributes are independent [7].

The characteristics of Naïve Bayes can be explained as follows [8]:

1. The Naïve Bayes method works robustly on isolated data. Naïve Bayes can handle incorrect attribute values by ignoring training data during the modeling and prediction process.
2. Strong in dealing with irrelevant attributes.
3. Attributes that have a relationship can degrade the performance of the Naïve Bayes classification because the assumption of the non-binding of the feature no longer exists.

Bayes' theorem is formulated as follows [9]:

$$P(H|X) = \frac{P(X|H)}{P(X)} \, . \, P(H) \tag{1}$$

Note :
X: Data with unknown class
H: Hypothesis data is a specific class
P(H|X): Probability of hypothesis H based on condition X (posterior probability)
P(H): Probability of hypothesis H (prior probability)
P(X|H): Probability of X based on condition on hypothesis H
P(X): Probability X

And this is the step by step how N.B.C. algorithm works [10].

Step 1 : Read the training dataset.
Step 2 : Calculate the mean and standard deviation of the predictor variables in each class.
Step 3 : Repeat
Calculate the probability of all predictor variables using the gauss density equation in each class.
Step 4 : Calculate the likelihood for each class.
Step 5 : Get the greatest likelihood.

Figure 1. Algorithm of N.B.C. Method

2. Methods

This part of the Method will explain the stages carried out in this research from beginning to end. The Method from this research is divided into some stages in Figure 2.

Figure 2. Research Method

2.1. Problem Identification

At this stage, identifying problems is carried out to determine the quality of chicken eggs worth selling. The results of interviews with farmers obtained variables used to determine egg quality. The variables used are color, size, sound, position, and thickness. The target achieved is Worth selling or not.

Table 1. Data Variable

Variable	Attribute	Information
X1	Color	Pale
		Bright
X2	Size	Small
		Normal
		Large
X3	Sound	Muted
		Voiced
X4	Position	Sinking
		Middle
		Floating
X5	Thickness	Thin
		Thick

2.2. Data Collecting

At this stage, the data that will be used is collected. The data used is data taken in one of the laying hens in Bandar Lampung. The data used are 215 samples. The data will be grouped into training data as many as 150 samples and test data as many as 65 samples.

2.3. Training

At this stage, the training data will be trained using the Naive Bayes method. Before being classified, it is preprocessing is carried out. Preprocessed data, a color attribute with pale value becomes 0; the bright value becomes 1. Size attribute with small value becomes 0; the average value becomes 1, the significant value becomes 2. Voice attribute with muted value becomes 0, and voiced value becomes 1. Attribute the position with the sinking value 0, the floating value 1, and the floating value 2. The thickness attribute with the thin value being 0 and the thick value being 1. After the preprocessing process, the training dataset sample will be trained using the Naive Bayes method.

2.4. Testing

Testing is done to determine whether the algorithm is running, assess the performance of the algorithm that has been implemented, and test calculations in training data by using data testing. During this testing phase, analysis of precision, recall, and accuracy is also carried out to test the accuracy of the application being built.

A Precision calculates the estimated proportion of actual positive cases [11].

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

A Recall calculates the estimated proportion of positive cases that are correctly identified [11].

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

An Accuracy is a calculation of the proportion of the total number of correct predictions [11].

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (4)$$

Where
T.P.: True Positive
TN: True Negative
F.P.: False Positive
F.N.: False Negative

3. Result and Discussion

This section contains the results of the test and analysis and discussion of the difficulties that have been carried out. Testing the data for saleable eggs uses a dataset of 215 data. The dataset was taken from one of the laying hens in Bandar Lampung. Before being classified, preprocessing is carried out by changing the categorical data type into a numerical data type. Preprocessed data, a color attribute with pale value becomes 0; the bright value becomes 1. Size attribute with small value becomes 0; the average value becomes 1, the significant value becomes 2. Voice attribute with muted value becomes 0, and voiced value becomes 1—attribute position with sinking value to 0, floating value to 1, and floating value to 2. The thickness attribute with a thin value becomes 0, and the thick value becomes 1. The results of the test can be seen in table 2.

Table 2. Sample Dataset

Colour	Size	Voice	Position	Thickness	Worth
1	0	0	1	0	Yes
1	2	0	0	0	Yes
1	2	1	2	0	No
0	1	1	2	1	No
0	0	1	1	0	No
1	1	0	0	1	Yes
0	1	0	0	1	No
…	…	…	…	…	…
1	1	1	0	1	Yes
1	2	1	1	0	No
0	0	0	2	0	No
0	0	1	1	1	Yes
1	2	0	0	0	Yes
1	0	0	1	0	Yes
1	1	1	2	0	No
0	1	0	1	1	Yes

The sample data was processed using Orange Data Mining Software. The workflow can be seen in figure 2.

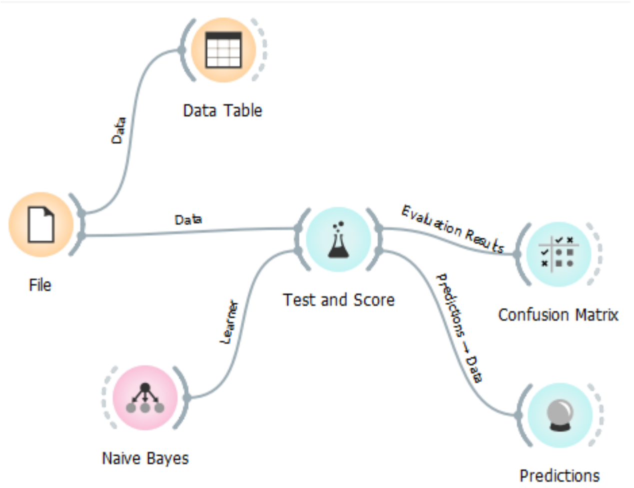

Figure 3 Processing Sample data Workflow

The Result of the Confusion Matrix can be Seen in Figure 4.

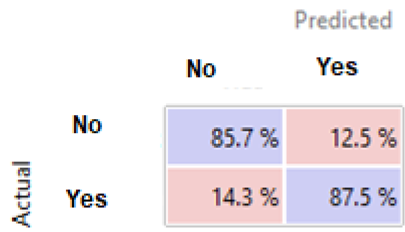

Figure 4 Confusion Matrix

The result of prediction using the Naïve Bayes Classifier model can be seen in table 3.

Table 3. Comparison of Actual Data with Predictions

Actual	Predictions
Yes	Yes
No	No
Yes	Yes
No	iNo
Yes	Yes
Yes	Yes
Yes	Yes
No	No
…	…
Yes	Yes
No	No
No	Yes
Yes	Yes
No	No
Yes	Yes
No	No

From Table 3, the dataset test with the Naive Bayes classifier model has an accuracy of 86.7%, precision of 86.7%, and recall 86.7%.

4. Conclusion

Based on the study results using the Naive Bayes classifier on the dataset that was tested with actual egg data from laying hens breeders, it resulted in 86.7% accuracy, 86.7% precision, and 86.7% recall. Of all the attributes tested were proven to affect the quality of the eggs produced. Therefore, using a naive Bayes classifier can help lay hens breeders determine the quality of the eggs produced so that egg sales increase.

References

[1] Dong J, Dong X, Li Y, Zhang B, Zhao L, Chao K and Tang X 2020 Prediction of infertile chicken eggs before hatching by the Naïve-Bayes Method combined with visible near infrared transmission spectroscopy *Spectrosc. Lett.* **53** 327–36

[2] Maimunah M and Whidhiasih R 2017 Identifikasi Mutu Telur Ayam Berdasarkan Kebersihan Kerabang Menggunakan Jaringan Syaraf Tiruan *Informatics Educ. Prof.* **2** 234362

[3] Osisanwo F ., J.E.T A, O A, J. O H, O O and J A 2017 Supervised Machine Learning Algorithms: Classification and Comparison *Int. J. Comput. Trends Technol.* **48** 128–38

[4] Vijay R, Vangara B, Thirupathur K, and Vangara S P 2020 Opinion Mining Classification using Naive Bayes Algorithm *Int. J. Innov. Technol. Explor. Eng.* **9** 495–8

[5] Ali Z M, Hassoon N H, Ahmed W S and Abed H N 2020 The Application of Data Mining for Predicting Academic Performance Using K-means Clustering and Naïve Bayes Classification *Int. J. Psychosoc. Rehabil.* **24** 2143–51

[6] Xu F, Pan Z and Xia R 2020 E-commerce product review sentiment classification based on a naïve Bayes continuous learning framework *Inf. Process. Manag.* **57** 102221

[7] Meiriza A, Lestari E, Putra P, Monaputri A and Lestari D A 2020 Prediction Graduate Student Use Naive Bayes Classifier *Sriwijaya International Conference on InformationTechnology* vol 172 (Atlantis Press) pp 370–5

[8] Zhang H, Jiang L, and Yu L 2021 Attribute and instance weighted naive Bayes *Pattern Recognit.* **111**

[9] Zhang H, Jiang L and Yu L 2020 Class-specific attribute value weighting for Naive Bayes *Inf. Sci. (Ny).* **508** 260–74

[10] Marcos De Moraes R, Soares E A D M G and MacHado L D S 2020 A double weighted fuzzy gamma naive bayes classifier *J. Intell. Fuzzy Syst.* **38** 577–88

[11] Santoso H A, Rachmawanto E H, Nugraha A, Nugroho A A, Setiadi D R I M and Basuki R S 2020 Hoax classification and sentiment analysis of Indonesian news using Naive Bayes optimization *Telkomnika (Telecommunication Comput. Electron. Control.* **18** 799–806