

# NS-SVM: Bolstering Chicken Egg Harvesting Prediction with Normalization and Standardization

Aji Gautama Putrada<sup>1</sup>, Nur Alamsyah<sup>2</sup>, Mohamad Nurkamal Fauzan<sup>3</sup>, Syafrial Fachri Pane<sup>4</sup>

<sup>1,2,3,4</sup>Advanced and Creative Networks Research Center, Telkom University, Indonesia

<sup>1</sup>ajigps@telkomuniversity.ac.id, <sup>2</sup>nuralamsyah@student.telkomuniversity.ac.id,

<sup>3</sup>mnurkamalfauzan@student.telkomuniversity.ac.id,

<sup>4</sup>fachrie@student.telkomuniversity.ac.id

**Abstract** - Breeding chickens and chicken eggs are poignant, and recent studies have applied computer science to optimize this field, including chicken egg harvesting prediction. However, existing research does not emphasize the importance of data transformation to obtain optimum chicken egg harvesting prediction. This paper proposes the normalization and standardization-bolstered support vector machine (NS-SVM) method, namely normalization, and standardization, to improve the prediction of chicken egg harvest using SVM. First, we obtain the chicken egg dataset from Africa using Kaggle. The problem and solution become urgent, whereas chicken egg production can ease businesspeople to invest in chicken eggs. We adopt the normalization and standardization method from previous research. However, the notation is to differentiate the method from legacy SVM. The dataset has up to 13 features. Then we apply standard pre-processing such as label encoding and random oversampling. We also review the dataset feature using the Pearson correlation coefficient (PCC). We use two SVM kernels: radial basis function (RBF) and the 2<sup>nd</sup>-degree polynomial. Then we again apply the same model but by applying normalization and standardization. We use cross-validation with  $K = 10$  to measure the Accuracy of the compared models. The results show that normalization and standardization positively affect the prediction model of the two SVM kernels. The model with the highest performance is NS-SVM with a 2<sup>nd</sup>-degree kernel, namely *Accuracy* = 0.996. At the same time, the model with the lowest performance is SVM with RBF, namely *Accuracy* = 0.986. In addition, the results of ROC AUC analysis show that the performance of our model on the imbalanced dataset with a moderate degree is *AUC* = 0.927 to 0.993.

**Keywords:** normalization, standardization, chicken egg, prediction, support vector machine, imbalance dataset

## I. INTRODUCTION

Several studies in computer science have led to an increase in the production of chicken eggs. Reference [1] proposed a machine that can sort chicken eggs based on

their quality using light sensors, weight sensors, and fuzzy logic. Reference [2] made a chicken egg incubator connected to the Internet of things (IoT) and supplied solar energy. In machine learning, [3] predicted the price of chicken eggs using dynamic time warping (DTW) to make investing easier for chicken egg businesspeople. Reference [4] used a support vector machine (SVM) to predict the category of chicken eggs based on a combination of images and sensor metrics. However, although the results are adequate, the research is limited to a small number of datasets.

Using measurements based on various sensors sometimes results in features with ranges that have significant differences. For example, [5] used a humidity sensor ranging from 0 to 100. That range is in stark contrast to the lux sensor, which ranges from 0 to approximately 10000. Another example as in [6], although the sensor comes from a three-axis accelerometer, each gives a different axis range. The diversity of the range of feature values can result in the prediction model's performance decline [7]. To avoid the decline, pre-processing techniques such as normalization and standardization can improve the prediction model's performance [8].

This paper proposes the normalization and standardization-bolstered support vector machine (NS-SVM) method, namely normalization and standardization, to improve the prediction of chicken egg harvest using SVM. First, we get the chicken egg dataset from Africa using Kaggle. The dataset has up to 13 features. Then we apply standard pre-processing such as label encoding and random oversampling. We also review the features of the dataset using the Pearson correlation coefficient (PCC). We use two SVM kernels: radial basis function (RBF) and 2<sup>nd</sup>-degree polynomial. Then we again apply the same model but by applying normalization and standardization. We used cross-validation with  $K = 10$  to measure the accuracy of the compared models.

Many methods have implemented standardization and normalization to bolster their model's performance. However, to the best of our knowledge, there has never been a study that carried out normalization and standardization to improve the predictive performance of chicken egg harvest using SVM. Here are some of our research contributions:

1. A model for chicken egg harvesting prediction that has a better performance by adding ten features of chicken characteristics
2. A chicken egg harvesting prediction with a model bolstered by normalization and standardization, namely NS-SVM
3. A report that shows the positive effect of normalization and standardization on the model for chicken egg harvesting prediction using 2<sup>nd</sup>-degree polynomial SVM.

The composition of the remainder of this paper is as follows: Section II discusses system design. Section III reports the test results and discusses the research results with state-of-the-art papers regarding chicken egg harvesting prediction. In addition, the discussion subsection also highlights the contribution of our research. Finally, Section IV presents the conclusions of the study.

Like ours, several studies have used features related to chickens' environment. Omomule *et al.* [9] used PCC to rank several features and produced four features: chicken age, weight, quality, and quantity. The study used a fuzzy method to predict the amount of egg production based on these four features. However, they need to explain the potential increase in model performance if the number of features is added to the model. There is a research opportunity to analyze whether feature selection with PCC has a significant effect or not on the prediction performance of the model.

We propose SVM as a classification model. Some studies use other models besides SVM. Gonzalez-Mora *et al.* [10] used random forest (RF) as a classification model for predicting egg production. This study uses the air quality index (AQI) as its feature. This research results show that the value of  $r^2$  for RBF is 0.78. Comparing RF with AQI and SVM with features of the chicken environment is a research opportunity.

Several studies have observed the use of standardization and normalization to improve the predictive ability of a model. For example, Raju *et al.* [11] compared 12 different types of scaling, including normalization and standardization, to three classification models: RBF SVM, sigmoid SVM, and KNN SVM. The three models are models whose performance is

determined by the distance between the data. The case study in this study used diabetes data. The results of this study show that the model that has gone through the 12 types of scaling has better performance than the model with raw data. There is a research opportunity to also compare the 2<sup>nd</sup>-degree polynomial SVM in the case study of chicken egg harvesting prediction.

## II. METHOD

Fig. 1 shows our proposed research methodology. First, we obtain and observe the chicken egg dataset. Then we design a predictive model for chicken egg harvesting. The next step is to evaluate the model. After that, we re-iterate the process by applying normalization and standardization to the data. The next step is to evaluate the development of its predictive model. Finally, we report the findings of this research.

### A. Chicken Egg Harvesting Data and Pre-Processing

The dataset we use is an egg-producing chickens' dataset from Kaggle. The dataset contains attributes and observations of 19 egg-producing chickens, with 1000 data items in the chicken and chicken egg dataset. Researchers may use the dataset without permission for machine learning and data science research purposes. Here we use it to predict the number of eggs to be harvested.

Furthermore, we explain the features contained in the dataset. Fig. 2 shows an explanation of each feature. Thirteen features include the characteristics of chickens, such as weight, age, and color of the body parts. In addition, there are also egg characteristics such as weight and color. Then there are other explanations, such as the amount of feed and the chickens' sunlight exposure. The label is the number of eggs a chicken produces in one day. The label range is 0 to 1 egg.

In addition to normalization and standardization, we also implement several other pre-processing stages. Since we are using SVM, we need to implement a label encoder, which is a process that converts a string to a unique integer [12]. Then we apply PCC to evaluate the features present in the dataset [13]. The PCC formula (denoted by  $r$ ) is (1).

$$r = \frac{\sum(x_{1i}-\bar{x}_1)(x_{2i}-\bar{x}_2)}{\sqrt{\sum(x_{1i}-\bar{x}_1)^2 \sum(x_{2i}-\bar{x}_2)^2}}, i \in m \quad (1)$$

where  $x_{1i}$  is the first variable with data item  $i$ ,  $m$  is the dataset size,  $\bar{x}_1$  is the average of the first variable,  $x_{2i}$  is the second variable with data item  $i$  and  $\bar{x}_2$  is the average of the second variable.

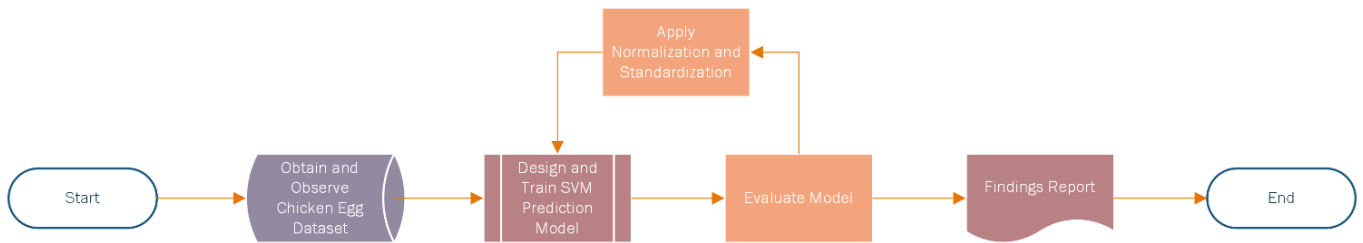


Fig. 1 Proposed research methodology

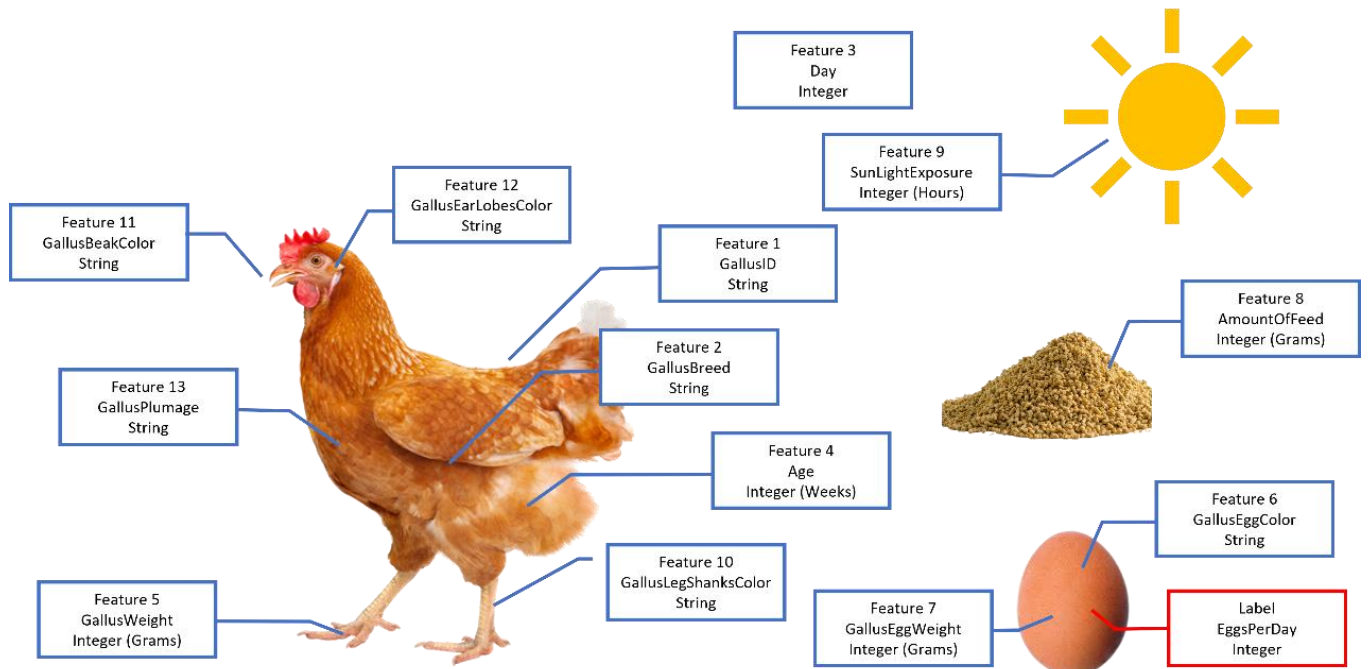


Fig. 2 Chicken egg dataset features including chicken characteristics and more.

Furthermore, we apply random oversampling [14]. The application of random oversampling is for imbalanced data. Data imbalance occurs if the composition between the output labels is usually imbalanced. There are three degrees of imbalance. Mild if the minority label has a proportion of 20 to 40%, moderate if the minority label has a proportion of 1 to 20%, and extreme if the minority label has a proportion below 1%. The way random oversampling works is to add minority class data randomly until the class proportion is no longer imbalanced. Imbalance data can affect the prediction model's performance, so it tends to choose the majority class.

**B. NS-SVM Prediction**

We use SVM to predict chicken egg harvesting. SVM is a binary classifier, where this model will put data in

one of the classes [15]. SVM creates a hyperplane that separates the two classes so that the distance between the two classes is maximized [16]. SVM has a kernel ( $K(x_1, x_2)$ ), a function that forms a hyperplane [17]. Here we compare the RBF kernel and the 2<sup>nd</sup>-degree polynomial kernel. The RBF kernel is popular because it separates data with a Gaussian distribution. It works similarly to the k-nearest neighbor (KNN) method [18]. The RBF kernel formula is (2).

$$K(x_1, x_2) = \exp\left(-\frac{\|x_1 - x_2\|^2}{2\sigma^2}\right) \quad (2)$$

where  $x_1$  is the first feature,  $x_2$  is the second feature, and  $\sigma$  is the variance.

Moreover, if the linear kernel separates two data with a linear hyperplane, then the polynomial separates the two data with a hyperplane polynomial [19]. The 2<sup>nd</sup>-degree polynomial function is a quadratic function,

where the characteristic sign is the shape of the function in the form of a parabola. The 2<sup>nd</sup>-degree polynomial kernel formula is (3).

$$K(x_1, x_2) = (x_1^T x_2 + c)^2 \quad (3)$$

where  $c$  is the free parameter and the value  $c \geq 0$ .

The normalization function in pre-processing data changes all data ranges from 0 to 1 [20]. A model can use normalization for numerical parameters that have varying scaling [21]. The use of normalization is because the diversity of the scaling affects when the prediction model algorithm is compiling the model.

*Standardization* is a feature scaling technique that makes the average of a dataset 0 [22]. In some cases, applying standardization before performing machine learning training can improve the performance of the training model and increase the training speed [20]. Another result of standardization is that the effect of outliers also decreases. The normalization and standardization formulas is (4).

$$x_{ns} = \frac{x - x_{min}}{\sigma(x_{max} - x_{min})} - \frac{\mu}{\sigma} \quad (4)$$

where  $x$  is the feature of a dataset,  $x_{min}$  is the lowest value in the feature,  $x_{max}$  is the highest value in the feature,  $\mu$  is the mean of a feature,  $\sigma$  is the standard deviation of a feature, and finally,  $x_{ns}$  is the standard form of the feature's normal form. By applying normalization and standardization, the dataset features will have a range of -1 to 1 and an average of 0.

### C. Performance Metrics

The performance measurement of our model uses K-fold cross-validation, receiver operator curve (ROC), and the area under curve value (AUC). Sometimes a model has a generalization problem, namely the ability of a predictive model to have the same performance for each sub-sample dataset [23]. Therefore, cross-validation tests the model [24]. Fig. 3 shows the K-fold cross-validation algorithm. K-fold cross-validation tests as many as  $K$  iterations [25]. In each iteration, the algorithm divides the data into  $K$  equal parts.

The K-fold cross-validation algorithm makes one of  $K$  the validation set or  $V$ , where the rest is for training. The validation uses accuracy ( $Acc_k$ ). Here is the  $Acc_k$  formula (5).

$$Acc_k = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

where  $TP$  is a true positive,  $TN$  is a true negative,  $FP$  is a false positive, and  $FN$  is a false negative. There will be as many  $K$  values  $Acc_k$ . The final step is to aggregate these values using (6).

$$Acc = \frac{1}{K} \sum_{k=1}^K Acc_k \quad (6)$$

where  $Acc$  is aggregate accuracy.

The ROC and AUC measurement methods for imbalanced data are more precise than the accuracy score because they are scale-invariant [26]. ROC is a curve that observes the relationship between true positive rate ( $TPR$ ) and false positive rate ( $FPR$ ) [24]. When the positive prediction threshold increases,  $TP$  and  $FP$  will increase, which is observable by the ROC. AUC is the area under the ROC curve. The greater the proportion of  $TPR$  compared to  $FPR$  at each point, the better the AUC value. Here is our formula for measuring AUC (7).

$$AUC = \sum_{j=2}^J \frac{TPR_j + TPR_{j-1}}{2} (FPR_j - FPR_{j-1}) \quad (7)$$

where  $J$  is the number of  $TPR$  and  $FPR$  measurements in the probability prediction class.

## III. RESULT AND DISCUSSION

At first, we measure the PCC score of all the features in the chicken egg dataset. Fig. 4 contains the PCC score that measures the correlation between all features and the output, namely EggsPerDay. There are no features that have a strong or moderate correlation. The feature that has the PCC with the strongest correlation is Age. The category is weak negative. Two features that have a very weak correlation are AmountOfFeed and GallusWeight. The AmountOfFeed category is a weak negative, while the GallusWeight category is a weak positive. The remainder, which has a magnitude below 0.1, does not correlate with EggsPerDay.

We prepared two datasets. The first is a raw dataset, while the second is a dataset that applies normalization and standardization. Fig. 5 shows the normalization and standardization results on each feature's value distribution. Before normalization and standardization, each feature has a diverse distribution, with a maximum population of 7. The maximum range is 0 to 200. After normalization and standardization, all features have an average value of 0. Then the maximum range is -5 to 5. There is no range disparity. The maximum population size is 1.4.

---

### Algorithm 1 K-fold cross validation

---

- 1: Divide data into  $K$  equal folds
  - 2: **for**  $k$  in range(0,  $K$ ) **do**
  - 3:      $V \leftarrow Fold_k$  in data
  - 4:      $T \leftarrow data \setminus V$
  - 5:     Train  $T$
  - 6:      $Acc_k \leftarrow$  evaluate  $V$  with trained model
  - 7: **end for**
  - 8:  $Acc \leftarrow \frac{1}{K} \sum_{k=1}^K Acc_k$
- 

Fig. 3 K-fold cross-validation algorithm

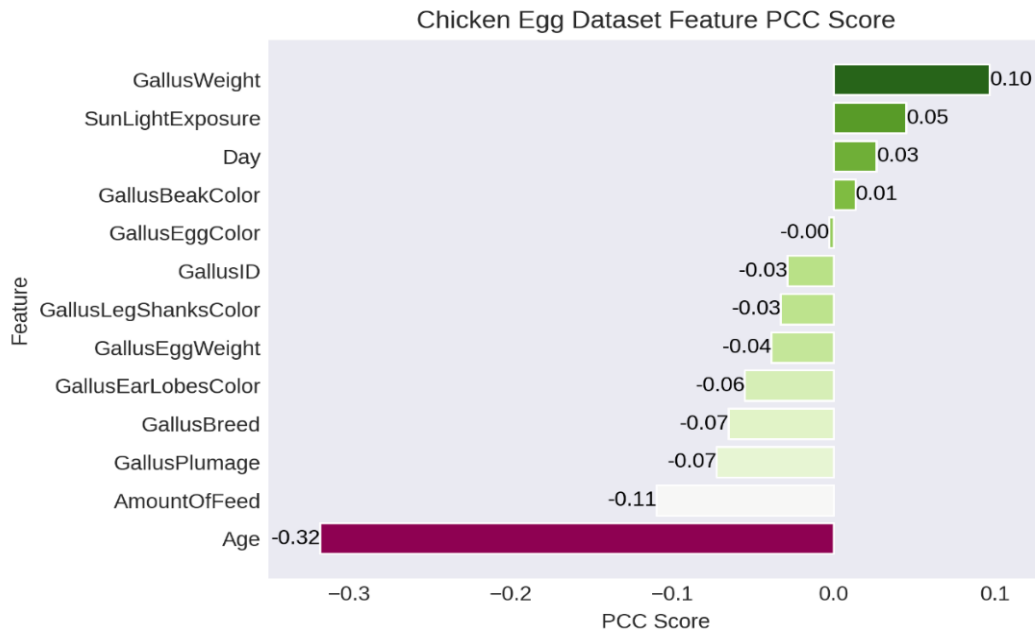


Fig. 4 Chicken egg dataset features PCC score

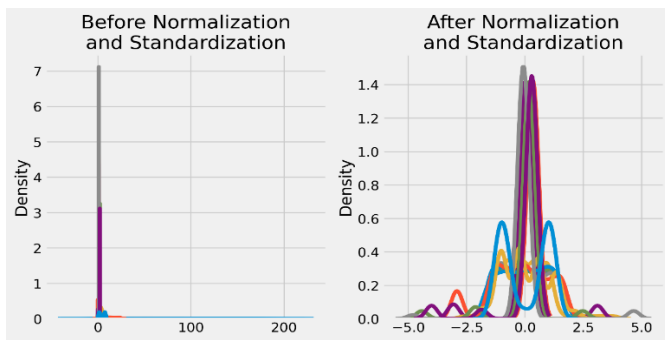


Fig. 5 Dataset distribution before and after normalization and standardization.

The next step is to prepare the dataset for training. Our train-test ratio is 50:50. We randomize the dataset by stratification. Stratification maintains the ratio label in the train and test data. Before doing the training, we observe the dataset for imbalanced data. The application of imbalanced observation data is to the training dataset. Our label, EggsPerDay, has a value range of 0 to 1, meaning there are two classes. Table I resumes the properties.

In raw data, there are 18 data with label 0, while in data with label 1, there are 482. The minority label proportion is 3.6% of the dataset, so our training dataset has a moderate imbalance degree. After we apply random oversampling, the proportion of minority labels becomes 50%, with no imbalance in the random

oversampled training data. Fig. 6 compares training datasets before and after applying random oversampling.

TABLE I  
PROPERTIES OF THE TRAINING DATASET  
PARAMETERS

Property	Value
1 Train-test ratio	50:50
2 Stratification	Applied
3 Label	EggsPerDay
4 Label Range	0 to 1
5 Number of Labels	Two

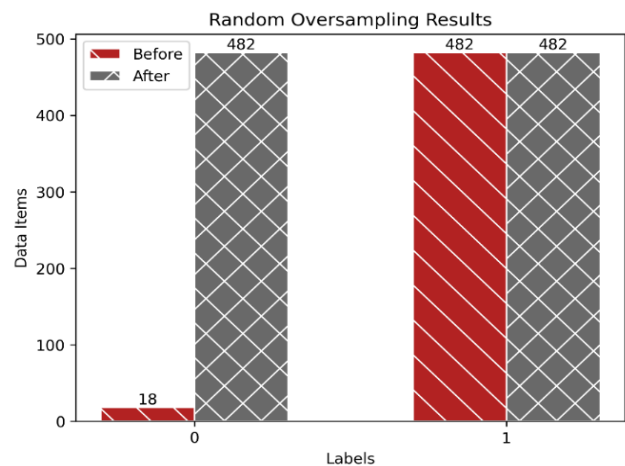


Fig. 6 Comparison of the training dataset before and after random oversampling

We trained and compared two SVM models, one with the RBF kernel and the other with the 2<sup>nd</sup>-degree polynomial kernel. Each model trains with two different datasets, namely raw and normalized and standardized datasets, the latter of which we named NS-SVM. This test compares a total of four SVM models. We used K-fold cross-validation to test all four models, with a value of  $K = 10$ . The test metric is Accuracy. Fig. 7 shows the results of the comparison. Of the four models, NS-SVM has a higher performance than SVM for each kernel. The results prove that NS-SVM improves the performance of chicken egg harvesting prediction. Then the 2<sup>nd</sup>-degree polynomial kernel has a better performance than RBF. This comparison shows the suitability of the characteristics of the dataset. The model with the highest performance is NS-SVM with a 2<sup>nd</sup>-degree kernel, *Accuracy* = 0.996. At the same time, SVM with RBF kernel is the model with the lowest performance, namely *Accuracy* = 0.986.

Finally, we conduct ROC and AUC testing. This test is because our dataset has imbalance problems to a moderate degree. Fig. 8 compares the ROC curves of the four SVM models. To interpret the curve, we use the AUC value. The curve with the highest AUC value is the 2<sup>nd</sup>-degree polynomial NS-SVM, which is 0.993. The second highest AUC value is NS-SVM RBF, with a value of 0.992. The third and fourth highest values are NS-SVM RBF and 2<sup>nd</sup>-degree polynomial SVM, with values of 0.975 and 0.927, respectively.

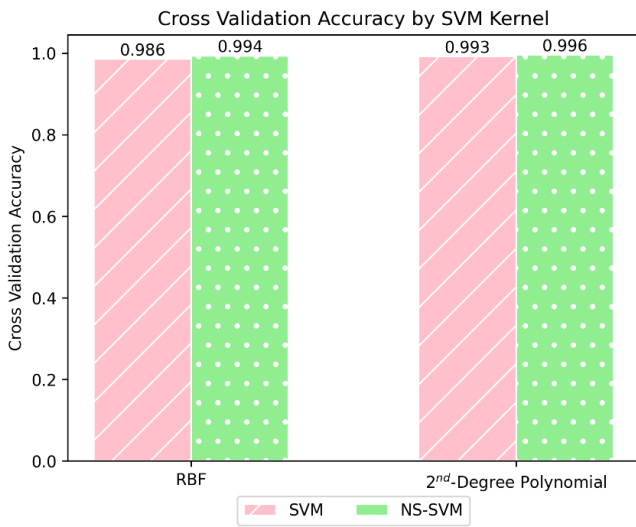


Fig. 7 Cross-validation accuracy comparison of four SVM models

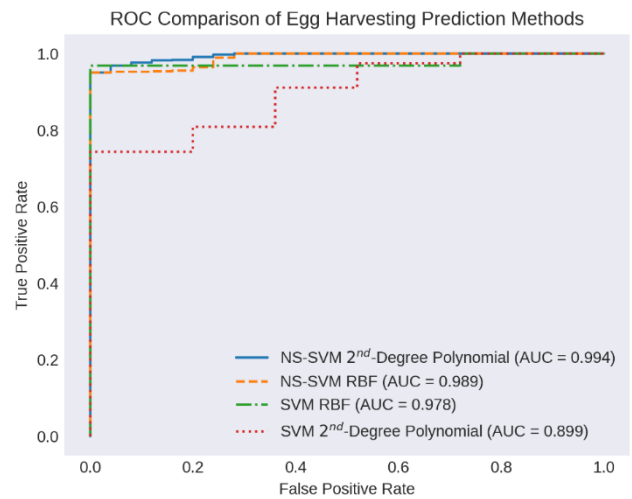


Fig. 8 ROC curve of four SVM models

We measure the PCC of 13 features in chicken egg harvesting prediction, where three have very weak correlations, namely weight, age, and feed quantity. Research [9] is also research that performs chicken egg harvesting prediction. The study stated that there were 12 features in chicken egg prediction but chose the same three features as our research. Although the study also used PCC, it did not mention other features of PCC results or did not measure them. In our research, if we add ten other features to the prediction, the performance can increase from *Accuracy* = 0.96 to *Accuracy* = 0.99. Our research contributes to the chicken egg harvesting prediction, which has a better performance by adding ten features of chicken characteristics.

Several other studies, such as [10], used an AQI dataset from chicken coops and used RF to predict chicken egg production, but the prediction performance was not as good as ours. Our research contributes to chicken egg harvesting prediction with an improved model, NS-SVM.

Several studies, such as [11], have shown that standardization and normalization influence predictive models that use the distance between data as the training algorithm. These prediction models include RBF SVM and KNN. In this study, apart from RBF SVM, we also show that normalization and standardization also positively influence the prediction model of 2<sup>nd</sup>-degree polynomial SVM. Our research contributes to a report that shows the positive effect of normalization and standardization on the chicken egg harvesting prediction using 2<sup>nd</sup>-degree polynomial SVM.

#### IV. CONCLUSION

We have successfully implemented a proposed novel model called NS-SVM, which can improve the

performance of chicken egg harvesting prediction. We use a Kaggle dataset called Egg Producing Chickens Dataset, which consists of 13 features and one label. We compare two SVM kernels, the RBF and the 2nd-degree polynomial kernel. K-fold cross-validation and ROC AUC curve analysis are the test metrics we use. The results showed that normalization and standardization positively affect the prediction model of the two SVM kernels. The model with the highest performance is NS-SVM with a 2nd-degree kernel, where the *Accuracy* = 0.996. At the same time, the model with the lowest performance is SVM with RBF kernel, with an *Accuracy* of 0.986. In addition, the ROC AUC analysis results show that our model's performance on the imbalanced dataset with a moderate degree is *AUC* = 0.927 to 0.993.

#### ACKNOWLEDGEMENT

We thank Phuzo Soko for sharing their Egg Producing Chickens dataset on Kaggle. Thanks to the dataset, we have made significant findings in our research on chicken eggs. Hopefully, in the future, this paper can provide the same benefits.

#### REFERENCES

- [1] I. Mahesa, A. G. Putrada, and M. Abdurohman, "Egg quality detection system using fuzzy logic method," *Kinet. Game Technol. Inf. Syst. Comput. Netw. Comput. Electron. Control*, pp. 207–216, 2019.
- [2] F. Peprah, S. Gyamfi, M. Amo-Boateng, E. Buadi, and M. Obeng, "Design and Construction of Smart Solar Powered Egg Incubator Based on GSM/IoT," *Sci. Afr.*, p. e01326, 2022.
- [3] C. Liu, L. Zhou, L. Höschle, and X. Yu, "Food price dynamics and regional clusters: machine learning analysis of egg prices in China," *China Agric. Econ. Rev.*, no. ahead-of-print, 2022.
- [4] I. Y. Purbasari, F. T. Anggraeny, and N. R. Harianto, "Classification of broiler chicken eggs using support vector machine (svm) and feature selection algorithm," in *Proceedings*, 2018, vol. 1, no. 1, pp. 505–511.
- [5] S. Karuniawati, A. G. Putrada, and A. Rakhmatsyah, "Optimization of grow lights control in IoT-based aeroponic systems with sensor fusion and random forest classification," in *2021 International Symposium on Electronics and Smart Devices (ISESD)*, 2021, pp. 1–6.
- [6] B. A. Fadillah, A. G. Putrada, and M. Abdurohman, "A Wearable Device for Enhancing Basketball Shooting Correctness with MPU6050 Sensors and Support Vector Machine Classification," *Kinet. Game Technol. Inf. Syst. Comput. Netw. Comput. Electron. Control*, 2022.
- [7] S. Rao, P. Poojary, J. Somaiya, and P. Mahajan, "A COMPARATIVE STUDY BETWEEN VARIOUS PREPROCESSING TECHNIQUES FOR MACHINE LEARNING," *Int. J. Eng. Appl. Sci. Technol.*, vol. 5, no. 3, pp. 2455–2143, 2020.
- [8] X. Larriva-Novo, V. A. Villagr a, M. Vega-Barbas, D. Rivera, and M. Sanz Rodrigo, "An IoT-focused intrusion detection system approach based on preprocessing characterization for cybersecurity datasets," *Sensors*, vol. 21, no. 2, p. 656, 2021.
- [9] T. G. Omomule, O. O. Ajayi, and A. O. Orogun, "Fuzzy prediction and pattern analysis of poultry egg production," *Comput. Electron. Agric.*, vol. 171, p. 105301, 2020.
- [10] A. F. Gonzalez-Mora, A. N. Rousseau, A. D. Larios, S. Godbout, and S. Fournel, "Assessing environmental control strategies in cage-free aviary housing systems: Egg production analysis and Random Forest modeling," *Comput. Electron. Agric.*, vol. 196, p. 106854, 2022.
- [11] V. G. Raju, K. P. Lakshmi, V. M. Jain, A. Kalidindi, and V. Padma, "Study the influence of normalization/transformation process on the accuracy of supervised classification," in *2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT)*, 2020, pp. 729–735.
- [12] S. Das, S. Kumar Mahata, A. Das, and K. Deb, "Disease Prediction from Drug Information using Machine Learning," *Am. J. Electron. Commun.*, vol. 1, no. 4, pp. 16–21, 2021.
- [13] M. B. Satrio, A. G. Putrada, and M. Abdurohman, "Evaluation of Face Detection and Recognition Methods in Smart Mirror Implementation," in *Proceedings of Sixth International Congress on Information and Communication Technology*, 2022, pp. 449–457.
- [14] A. G. Putrada, I. D. Wijaya, and D. Oktaria, "Overcoming Data Imbalance Problems in Sexual Harassment Classification with SMOTE," *Int. J. Inf. Commun. Technol. IJoICT*, vol. 8, no. 1, pp. 20–29, 2022.
- [15] A. G. Putrada, M. Abdurohman, D. Perdana, and H. H. Nuha, "Machine Learning Methods in Smart Lighting Toward Achieving User Comfort: A Survey," *IEEE Access*, vol. 10, pp. 45137–45178, 2022, doi: 10.1109/ACCESS.2022.3169765.
- [16] M. Ameliasari, A. G. Putrada, and R. R. Pahlevi, "An evaluation of SVM in hand gesture detection using IMU-based smartwatches for smart lighting control," *J. Infotel*, vol. 13, no. 2, pp. 47–53, 2021.
- [17] B. H. Farizan, A. G. Putrada, and R. R. Pahlevi, "Analysis of Support Vector Regression Performance in Prediction of Lettuce Growth for Aeroponic IoT Systems," in *2021 International Conference Advancement in Data Science, E-learning and Information Systems (ICADEIS)*, 2021, pp. 1–6.
- [18] J. Gardner, G. Pleiss, K. Q. Weinberger, D. Bindel, and A. G. Wilson, "Gpytorch: Blackbox matrix-matrix

- gaussian process inference with gpu acceleration,” *Adv. Neural Inf. Process. Syst.*, vol. 31, 2018.
- [19] I. S. Al-Mejibli, D. H. Abd, J. K. Alwan, and A. J. Rabash, “Performance evaluation of kernels in support vector machine,” in *2018 1st Annual International Conference on Information and Sciences (AiCIS)*, 2018, pp. 96–101.
- [20] A. Ambarwari, Q. J. Adrian, and Y. Herdiyeni, “Analysis of the effect of data scaling on the performance of the machine learning algorithm for plant identification,” *J. RESTI Rekayasa Sist. Dan Teknol. Inf.*, vol. 4, no. 1, pp. 117–122, 2020.
- [21] D. Borkin, A. Némethová, G. Michal’čonok, and K. Maiorov, “Impact of data normalization on classification model accuracy,” *Res. Pap. Fac. Mater. Sci. Technol. Slovak Univ. Technol.*, vol. 27, no. 45, pp. 79–84, 2019.
- [22] T. P. Van, H. N. Thanh, and T. M. Thanh, “Improving Phonetic Recognition with Sequence-length Standardized MFCC Features and Deep Bi-Directional LSTM,” in *2018 5th NAFOSTED Conference on Information and Computer Science (NICS)*, 2018, pp. 322–325.
- [23] D. Berrar, “Cross-Validation.” 2019.
- [24] I. D. Wijaya, A. G. Putrada, and D. Oktaria, “Penggunaan Metode K-fold Untuk Data Imbalance Pada Klasifikasi Hwe Dan Qpq Dalam Kejahatan Tweet Pelecehan Seksual,” *EProceedings Eng.*, vol. 8, no. 5, 2021.
- [25] “CIMA: A Novel Classification-Integrated Moving Average Model for Smart Lighting Intelligent Control Based on Human Presence,” *Complexity*, vol. 2022, pp. 1–19, Sep. 2022, doi: 10.1155/2022/4989344.
- [26] J. Huo, Y. Gao, Y. Shi, and H. Yin, “Cross-modal metric learning for AUC optimization,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 10, pp. 4844–4856, 2018.