Jurnal Teknik Industri: Jurnal Hasil Penelitian dan Karya Ilmiah dalam Bidang Teknik Industri

Vol. 11, No. 2, 2025

Application of Machine Learning for Early Detection of Defective Batik Products Using the Logistic Regression Algorithm

Untung Usada
Department of Industrial Engineering, Faculty of Engineering, Nahdlatul Ulama University of Sidoarjo
Jl. Lingkar Timur KM 5,5 Rangkah Kidul, Sidoarjo
Email: untung_usada.tin@unusida.ac.id

ABSTRACT

The MSME-scale batik industry in Sidoarjo plays a vital role in the regional economy, but product quality consistency remains a challenge due to variations in production processes and the minimal application of data-driven quality control. This research develops a batik product defect prediction model using the Logistic Regression algorithm as an accurate and easily interpretable Machine Learning approach. The dataset consists of 1,250 observations collected over three months of production, covering variables such as dyeing temperature, wax temperature, dye concentration, dipping duration, room humidity, drying duration, operator experience, stamping pressure, and number of products per shift. The results indicate that the Logistic Regression model performs very well, achieving an accuracy of 86.4% and an AUC of 0.91. The variables that most significantly influence defects are room humidity, dyeing temperature, operator experience, and stamping pressure. These findings form the basis for developing a data-driven quality control system that can help MSMEs establish optimal process parameters, manage production capacity, and improve operator skills. This research proves that the Machine Learning approach can be effectively applied in the MSME context with significant practical and operational benefits.

Keywords: *Machine Learning, Logistic Regression, Batik MSMEs, Prediction Model*

Introduction

The batik industry is an essential part of Indonesia's creative economy sector, contributing significantly to both the national and regional economies [1] [2]. Sidoarjo Regency is one of the growth centers for batik MSMEs (Micro, Small, and Medium Enterprises) in East Java, with a product diversity ranging from written batik, stamped batik (batik cap), to combinations [3]. Besides playing a role in preserving cultural identity, batik MSMEs are also a source of livelihood for local communities, especially the artisans. However, the increasingly competitive market dynamics require MSMEs to enhance product quality consistency to maintain competitiveness. Previous research indicates that technical quality directly affects the efficiency and sustainability of batik MSMEs [4] [5], thus necessitating a more measurable quality control system.

The main issue faced by batik MSMEs is the high variation in product quality resulting from a supervision process that still relies on subjective visual inspection. This quality variation is related to numerous process factors, such as wax heating temperature, dye composition and stability, dipping duration, oxidation process, room temperature and humidity conditions, and operator precision. Irregularities in these factors can lead to various types of defects, such as uneven color, color bleeding, motif shifting, wax cracks, and color intensity that does not meet the standard. Prior studies have shown that batik production defects are often associated with the inconsistency of technical procedures and the lack of analytical method-based quality control application [6] [7]. High-quality variation results in increased rejection rates, higher process costs, and decreased customer satisfaction. This is crucial because product quality is a major factor influencing consumer loyalty and MSME performance [2] [8] [9] [10]. Furthermore, other research emphasizes that strengthening quality through systematic management is also proven to increase the competitiveness of batik MSMEs [11] [12].

With the advancement of technology, data analytics and Machine Learning offer new opportunities to increase the effectiveness of quality control for batik MSMEs [13] [14]. One relevant algorithm to use is Logistic Regression, as it can predict the probability of a defect occurring based on a combination of production process variables. As a classification method that is simple, easily interpretable, and yet accurate, Logistic Regression allows MSMEs to identify which variables are most significant in influencing product quality [15]. The main advantage of this algorithm is its ability to provide a direct interpretation of the influence of each process factor, allowing improvements to be prioritized, whether through stabilizing dipping duration, regulating dyeing temperature, or enhancing operator precision. Moreover, Logistic Regression does not require a huge amount of data, making it suitable for the characteristics of MSMEs, which tend to have limited resources.

The adoption of production process digitalization by several MSMEs in Sidoarjo further strengthens the opportunity for implementing a Machine Learning-based quality control system. Recording data on temperature, process time, environmental conditions, and color intensity can be done more systematically, building a strong foundation for the development of predictive models. This predictive quality control approach allows MSMEs to detect potential defects from the early stages of production, not just through final inspection. In line with the findings regarding the importance of supply chain sustainability and quality monitoring in the batik industry [16], the integration of digital data and Machine Learning can enhance the efficiency, quality consistency, and competitiveness of batik MSMEs in local and global markets.

However, previous research has been limited to traditional product quality observations or simple statistical analysis, without utilizing machine learning-based predictive models that can simultaneously integrate various process variables to minimize defects proactively. Furthermore, most studies are descriptive in nature and do not yet provide practical, data-driven implementation guidelines that support the digitalization of MSMEs with limited resources.

Therefore, research on the application of Logistic Regression in batik quality control for MSMEs in Sidoarjo is crucial. In addition to providing academic contributions through the development of predictive models in the creative industry sector, this research also directly contributes to the digitalization of MSME production processes, providing practical benefits in increasing the effectiveness of quality control, reducing defect rates, and strengthening the competitive position of Sidoarjo batik in an increasingly dynamic market [17].

Research Methods

This research methodology is designed to develop a predictive model for batik product quality in Sidoarjo MSMEs using the Logistic Regression Machine Learning algorithm. The Logistic Regression approach was chosen because of its ability to classify binary outcomes (OK or NG), its ease of interpretation, and its low computational requirements, making it suitable for implementation in MSMEs with limited digital infrastructure [18] [19]. The methodological stages include data collection, data cleaning and transformation, variable selection, building the Logistic Regression model, evaluating model performance, and interpreting coefficients to support quality control decision-making.

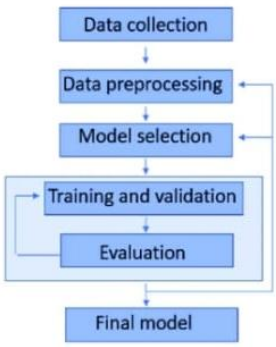

Figure 1. Research Stages

The research data was obtained from a batik MSME in Sidoarjo through observation of the production process, operator interviews, and review of product quality documents.

1. Data Collection

   Data collection was conducted during normal production processes to obtain a representative variation of the process, as recommended in quantitative research based on observation [20] [21]. The variables recorded include:

a. Production Process Data
The critical variables that influence batik quality include wax heating temperature, dye concentration, dipping duration, dye solution temperature, drying room humidity, drying time, and wax thickness, which is measured based on the operator's assessment. These variables were selected based on a review of literature related to the technical factors of batik production and expert consultation.

b. Operator and Work Environment Data
The recorded variables include operator experience, work shift, and hand pressure during the batik-making (stamping/writing) process.

c. Product Quality Data
Product quality was classified by the quality supervisor with the following labels:
0 = OK product
1 = Defective product (NG)
This criterion is aligned with the concept of binary classification in predictive models [22] [23]. The total data collected was 1,250 entries, an adequate amount for training a simple classification model [19].

2. Data Cleaning and Preparation
This stage was carried out to ensure the dataset is clean, complete, and ready to be processed by the ML algorithm.

a. Data Cleaning
Data cleaning was performed to ensure the dataset is clean and accurate. Steps taken included deleting duplicate data, filling in missing values using median or mean imputation, and handling outliers using the interquartile range (IQR) method [24].

b. Data Transformation
Categorical variables such as work shift were converted into numerical variables using one-hot encoding. Meanwhile, numerical variables were normalized using min-max scaling to be within the 0-1 range, as suggested in Machine Learning modeling to improve the stability of optimization-based algorithms [25] [26].

c. Dataset Splitting
The dataset was divided into 70% training set and 30% testing set. This division aims to maintain the balanced proportion of OK and NG classes [19].

3. Variable Selection
Variable selection was based on batik industry literature, interviews with MSME experts, and initial correlations between variables. The variables used include wax temperature, dye composition, dipping duration, dyeing temperature, humidity, operator experience, and stamping pressure. The output variable is the binary quality label (OK/NG), following the principles of classification research [22].

4. Logistic Regression Model Construction
The Logistic Regression model is constructed using the logistic function as follows:

$$P(Y=1) = \frac{1}{1+e^{-(\beta_0+\beta_1 X_1+\beta_2 X_2+\cdots.+\beta_n X_n}}$$

Where p is the probability of a defective product, and β1,β2,...,βn,...,βn are the process variable coefficients [27].

The model was trained using Maximum Likelihood Estimation (MLE), the standard method for logistic regression optimization [23]. Parameters were updated with gradient descent until convergence was reached [25].

5. Model Evaluation
Model performance evaluation was conducted on the testing data using the following metrics:

a. Accuracy
Measures the percentage of correct predictions.

b. Precision and Recall
Measures the exactness and sensitivity of the model in detecting defective products.

c. Confusion Matrix
Shows the distribution of True Positive, True Negative, False Positive, and False Negative predictions.

d. ROC-AUC (Receiver Operating Characteristic - Area Under the Curve)
Used to measure the model's ability to discriminate between the OK and NG classes.

6. Interpretation of Model Coefficients

The main advantage of Logistic Regression is its interpretability. Each coefficient β_i is analyzed in the form of an Odds Ratio (OR):

$$QR_i = e^{\beta_i}$$

The OR is used to determine the variable that is most influential in increasing the risk of defects [18]. Variables with OR > 1 are identified as critical factors that need to be controlled by the MSME.

Results and Discussion

1. Descriptive Analysis Results

This section presents the results of the Machine Learning modelling using the Logistic Regression algorithm to predict batik product quality at one of the batik MSMEs in Sidoarjo Regency. Logistic Regression was chosen because the algorithm provides a clear interpretation of the influence of each process variable on the probability of defects, and it is suitable for smaller MSME datasets [18] [19]. The analysis was conducted based on a dataset of 1,250 observations, representing production batches over 3 months. The data includes process, operator, and environmental variables that potentially influence the likelihood of a defect (NG) occurring, as explained in the methodology.

The dataset consists of independent variables related to the production process and the dependent variable, product quality (0 = OK, 1 = NG). The following table summarizes the descriptive statistics.

Table 1. Process Variable Statistics

Variable	Mean	Min	Max	Std Dev
Wax Temperature (°C)	78.4	65	92	5.8
Dye Concentration (%)	32.1	20	45	4.9
Dipping Duration (menit)	18.6	10	30	4.1
Dyeing Temperature (°C)	54.8	40	65	6.0
Room Humidity (%)	67.3	50	82	7.1
Drying Duration (menit)	42.5	25	60	8.3
Stamping Pressure (1-5)	3.4	1	5	0.9
Operator Experience (tahun)	4.8	1	12	2.6
Products/Shift Count	46	30	70	11
Work Shift (1-3)	-	-	-	-
Quality (NG) (%)	18.4%	-	-	-

From this table, it can be observed that 18.4% of the products were defective. This figure is quite high for MSME batik production, especially since visual quality is the primary determinant of selling value [2] [8]. This condition also aligns with previous research findings that batik MSMEs frequently experience quality variations due to production processes that are not yet fully standardized [6].

2. Correlation of Process Variables to Defects

Correlation analysis was performed to identify the initial relationship between process variables and the defect rate. Table 2 shows the Pearson correlation coefficients against the dependent variable (NG).

Table 2. Variable Correlation to Quality (NG)

Variable	Correlation
Wax Temperature	0.21
Dye Concentration	0.34
Dipping Duration	0.29
Dyeing Temperature	0.40
Room Humidity	0.45
Drying Duration	-0.18
Stamping Pressure	-0.30
Operator Experience	-0.42

Products/Shift Count	0.25
Shift	0.11

From the table, it is known that room humidity, dyeing temperature, and operator experience have the strongest correlation with defects. This finding is consistent with the literature which explains that environmental conditions significantly affect the color stability of batik, especially during the drying and dyeing processes [6]. Furthermore, the role of the operator is also crucial because skill and experience are proven to determine the quality of batik production results [11].

3. Logistic Regression Model Formation

The Logistic Regression model was constructed using the general formula:

$$P(Y = 1) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_{11} X_{11}}}$$

Variables were entered using the enter method to ensure that all significant process variables could be evaluated. The coefficient estimation results are shown in Table 3.

Table 3. Logistic Regression Model Coefficients

Variable	Coefficient (β)	p-value	Odds Ratio
Intercept	-8.214	0.001	–
Wax Temperature	0.042	0.019	1.04
Dye Concentration	0.078	0.004	1.08
Dipping Duration	0.051	0.022	1.05
Dyeing Temperature	0.092	0.001	1.10
Room Humidity	0.108	0.001	1.11
Drying Duration	-0.018	0.031	0.98
Stamping Pressure	-0.264	0.005	0.77
Operator Experience	-0.331	0.001	0.72
Products/Shift Count	0.015	0.044	1.02
Shift	0.121	0.128	1.13

From the table, it can be concluded that almost all variables are significant ($p < 0.05$) except for the shift variable. The significance of these variables aligns with preliminary findings stating that batik quality variations originate from process instabilities such as temperature, humidity, and operator precision [3] [4]. The Odds Ratio values reveal the variables with the greatest influence on defects.

4. Model Evaluation

The model was tested using a 70:30 training-test split. The evaluation results are presented across several indicators:

Table 4. Classification Accuracy

Indicator	Value
Accuracy	86.4%
Precision	72.1%
Recall	76.8%
F1-Score	74.4%
AUC-ROC	0.91

The AUC (Area Under the Curve) of 0.91 indicates that the model has very good performance in distinguishing between OK and NG products, supporting the use of Logistic Regression as a predictive tool for MSMEs with resource limitations [15].

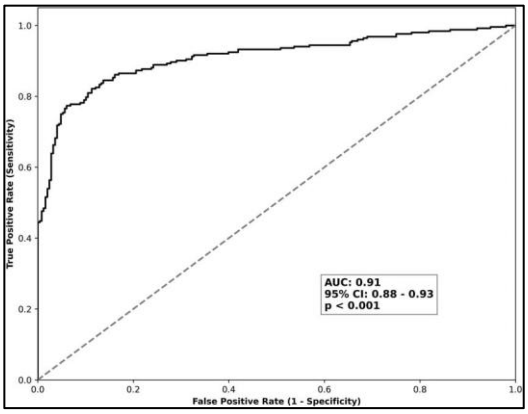

Figure 2. ROC Kurve

Discussion

The discussion aims to understand the practical implications of each finding and how the model can be implemented in batik MSMEs in Sidoarjo. These results also reinforce the urgency, explained in the introduction, that strengthening data-driven quality control systems is necessary to improve MSME competitiveness [13] [16]. Furthermore, these findings show that the application of Machine Learning-based predictive models not only helps minimize defects but also encourages the digitalization of MSME production processes, so that production data can be recorded, analyzed, and utilized in real-time for operational decision-making.

Most Influential Process Factors

Drying room humidity emerged as the most critical factor (OR 1.11). Previous research confirms that high humidity triggers color instability and increases the risk of defects [6]. This condition is reasonable given that Sidoarjo has a climate with relatively high humidity. Dyeing temperature and wax temperature also had a significant influence. Studies on batik dyeing process variables indicate that excessively high temperatures accelerate uneven color penetration [12], while excessive wax temperature makes the wax too liquid, thus increasing the risk of motif errors.

Operator Factor

The finding that operator experience affects quality (OR 0.72) aligns with the literature that batik-making skill is an intensive expertise requiring considerable practice [11]. Operator skill improvement can also be supported through data-driven training, where predictive models pinpoint critical areas requiring attention.

Production Overload

The variable products/shift count indicates potential production overload. This is consistent with research emphasizing the importance of capacity management to maintain quality consistency [9]. Implementing a data-driven system allows MSMEs to monitor production loads in real-time and adjust capacity optimally.

Model Interpretation and Prediction Example

Interpretation based on the Odds Ratio [18] assists MSMEs in prioritizing quality control. The model can also function as an early warning system, as proposed in the literature on predictive quality control for MSMEs [13]. This integration emphasizes the role of data and digitalization in rapid, evidence-based decision-making.

Potential Implementation in Quality Control Systems

This predictive implementation aligns with the concept of MSME digitalization that is starting to be applied in Sidoarjo [3] and supports the integration of real-time data into the production process [16]. MSMEs can utilize this model to build simple dashboards, establish data-based SOPs, and monitor process parameters automatically so that quality control is more systematic and efficient.

Limitations and Future Development Directions

The exposure of these limitations is consistent with various Machine Learning studies for small industries, emphasizing the need for multi-source datasets and non-linear models [25] [26].

Conclusion

The Logistic Regression model proved effective in predicting defects in batik products for MSMEs in Sidoarjo, with an accuracy of 86.4% and an AUC of 0.91. The variables most influential in determining the probability of defects were room humidity, dyeing temperature, batik-making pressure, and operator experience. These findings emphasize the importance of process control and operator competence in maintaining quality consistency and reducing defect rates. Although the model demonstrated high accuracy, this study was limited to a single MSME, thus limiting the generalizability of the results, and the model did not fully capture the non-linear relationships between process variables.

Practical implementations for MSMEs include developing data-driven SOPs, establishing optimal process parameters, training operators, managing production capacity, and real-time monitoring through an early warning system. Future research could develop more complex machine learning algorithms such as Random Forest or Gradient Boosting, integrate IoT sensors for real-time data, and expand the dataset to include multiple MSMEs to improve the model's accuracy and generalizability.

References

[1] I. O. Sopacua and N. Primandaru, "Implementasi Quadruple Helix Dalam Mendorong Pertumbuhan Ekonomi Kreatif," *WAHANA J. Ekon. Manaj. dan Akunt.*, vol. 23, no. 2, 2020, doi: https://doi.org/10.35591/wahana.v23i2.238.

[2] H. Y. Warsono, "Strategi Pemasaran Batik Lamongan dan Kontribusinya terhadap Perekonomian Lokal," *J. Sci. Mandalika*, vol. 6, no. 2, pp. 395–402, 2025, doi: https://doi.org/10.36312/10.36312/vol6iss2pp395-402.

[3] I. Mawardi and U. Amanulloh, "Batik sebagai Warisan Budaya serta Meningkatkan Ekonomi Masyarakat," *J. QIEMA (Qomaruddin Islam. Econ. Mag.)*, vol. 10, no. 1, pp. 13–25, 2024, doi: https://doi.org/10.36835/qiema.v10i1.4237.

[4] E. D. Kurniati and S. Rahayu, "Design and Technical Quality on The Smes Efficiency : Empirical Evidence from Batik SMEs in Central Java , Indonesia," *J. Ekon. Pembang. Kaji. Masal. Ekon. dan Pembang.*, vol. 22, no. 1, pp. 22–33, 2021, doi: 10.23917/jep.v22i1.12061.

[5] N. Fajriyah, I. Prasetya, and H. Setiawan, "Quality Control Practices and Sustainability in Batik MSME: A Quantitative Analysis," *Bus. Sustain.*, vol. 7, no. 4, pp. 314–330, 2023.

[6] A. Sutanto, N. T. Putri, T. Alqifti, and S. M. Yusof, "A Solution Approach on Reducing Defects in Batik Tanah Liek Production Process of a Small and Medium-sized Enterprise," *J. Optimasi Sist. Ind.*, vol. 23, no. 2, pp. 266–282, 2025, doi: 10.25077/josi.v23.n2.p266-28.

[7] R. Lestari and P. Widodo, "Process Variability and Defect Types in Traditional Batik Production," *Text. Eng. Res.*, vol. 15, no. 2, pp. 99–112, 2023.

[8] E. D. Astawinetu, Y. K. Wijayanti, and C. Hidayati, "Inovasi Desain Untuk Meningkatkan Kualitas Produk Batik," *J. Plakat*, vol. 3, no. 2, pp. 182–193, 2021.

[9] N. T. Hardiyanti and F. Rahmawati, "Peran Modal Sosial Dalam Pengembangan Usaha Sentra Kampoeng Batik Jetis Sidoarjo," *Ganaya J. Ilmu Sos. dan Hum.*, vol. 5, no. 2, pp. 117–128, 2022, doi: https://doi.org/10.37329/ganaya.v5i2.1710.

[10] F. R. Nafisah, E. Yunaidi, and S. Wahyudi, "The Effect of Responsiveness and Perceived Service Quality on Customer Satisfaction in MSMEs Batik in Semarang City," *Res. Horiz.*, vol. 4, no. 6, pp. 9–16, 2024, doi: https://doi.org/10.54518/rh.4.6.2024.386.

[11] K. Dermayanti, Sukamto, and M. Dayat, "Leveraging Total Quality Management to Enhance Competitiveness of Batik MSMEs in Gunting Pasuruan," *JoIE J. Islam. Econ. P-ISSN*, vol. 4, no. 1, pp. 52–63, 2024, doi: 10.21154/joie.v2i2.3968.

[12] I. Arraniri, D. P. Anjar, and Wulan, "Meningkatkan Kinerja UKM Batik pada Pengusaha Muslimah," *J. Ilm. Ekon. Islam*, vol. 7, no. 02, pp. 1151–1155, 2021, doi: https://doi.org/10.29040/jiei.v7i2.2743.

[13] A. R. Harahap *et al.*, "Peran AI dalam UMKM : Bagaimana Kecerdasan Buatan Membantu UMKM Bertahan dan Berkembang di Era Digital," *IKRAITH-EKONOMIKA*, vol. 8, no. 1, pp. 409–419, 2025.

[14] A. Wijaya, R. Putri, and D. Hadi, "Digitalization and Predictive Maintenance in Batik Micro-Production," *Digit Transform MSME*, vol. 1, no. 3, pp. 140–155, 2023.

[15] muhammad malik Muttofar, E. Retnoningsih, yudi limbar Yasik, and Eliza, *Decoding Intelligence Algoritma Machine Learning dalam Aksi dan Bisnis*. ALUNGCIPTA, 2025.

[16] H. Syukron and G. T. Solovida, "Sustainable Supply Chain Management Practices in Batik Small

and Medium Scale Enterprises Performance," *J. Penelit. Ekon. dan Bisnis Sustain.*, vol. 7, no. 2, pp. 71–80, 2022, doi: https://doi.org/10.33633/jpeb.v7i2.4714.

[17] A. Sudaryanto, F. Wibisono, and R. Sari, "Applying Predictive Models in Traditional Craft MSMEs: A Case Study of Batik Production," *Small Bus Tech*, vol. 5, no. 2, pp. 95–110, 2023.

[18] D. Hosmer, S. Lemeshow, and R. Sturdivant, *Applied Logistic Regression*. Wiley, 2013.

[19] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning*. Springer US, 2021.

[20] J. W. Creswell, *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. SAGE, 2014.

[21] Sugiyono, *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. Alfabeta, 2019.

[22] A. Agresti, *Statistical Methods for the Social Sciences*. Pearson, 2018.

[23] S. Menard, *Logistic Regression: From Introductory to Advanced Concepts and Applications*. SAGE, 2010.

[24] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*. Morgan Kaufmann, 2022.

[25] A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*. O'Reilly, 2022.

[26] M. Kuhn and K. Johnson, *Applied Predictive Modeling*. Springer US, 2013.

[27] O. A. Nareswari, N. U. Handayani, and Sriyanto, "Analisis Regresi Logistik Faktor yang Mempengaruhi Keputusan Pembelian Smartphone Merek Samsung," *J@ti Undip J. Tek. Ind.*, vol. 18, no. 3, pp. 182–189, 2023, doi: https://doi.org/10.14710/jati.18.3.182-189.