Evaluating the effectiveness of PCA-based indicator reduction on ANN performance for hotel maintenance classification in Yogyakarta

Bagus Gilang Pratama[1*], *Sely* Novita Sari[2], and *Rizal* Maulana[2]

[1]Electrical Engineering Study Program, Faculty of Engineering and Planning, Institut Teknologi Nasional Yogyakarta, Indonesia
[2]Civil Engineering Study Program, Faculty of Engineering and Planning, Institut Teknologi Nasional Yogyakarta, Indonesia

Abstract. The hospitality industry in Yogyakarta faces challenges in maintaining hotel facilities efficiently, prompting the need for data-driven maintenance classification systems. This study evaluates the effectiveness of reducing eleven hotel maintenance indicators using Principal Component Analysis (PCA) and analyzing its impact on Artificial Neural Network (ANN) performance. PCA reduced the indicators to nine principal components while retaining 93% of the data variance, and these components were used to train a multilayer perceptron ANN model. The PCA-based model achieved 90.57% accuracy and a macro-F1 score of 0.9106, slightly lower than the original model using all indicators, which attained 94.34% accuracy with fewer misclassifications. This research provides the first empirical assessment of PCA-based dimensionality reduction in the hotel maintenance context, revealing that although PCA enhances computational efficiency, it can remove subtle yet important information crucial for precise classification. The results highlight that PCA alone may be insufficient for high-precision maintenance prediction and should be complemented with alternative or hybrid feature-selection methods. For hotel managers, the findings emphasize the importance of complete and informative input variables in AI-based maintenance systems, which may be further strengthened through real-time IoT data integration.

1 Introduction

The hospitality sector, as a key part of the tourism industry, requires efficient and timely facility maintenance to extend asset life, reduce operational costs, and enhance guest satisfaction. Determining maintenance needs involves numerous interrelated indicators, which can burden Artificial Neural Network (ANN) models with high-dimensional data, reducing efficiency and increasing the risk of overfitting. Principal Component Analysis (PCA) offers a solution by reducing dimensionality while preserving essential information, yet its specific impact on ANN performance in hotel maintenance contexts

* Corresponding author: bagusgilangp@itny.ac.id

remains underexplored. Previous studies integrating PCA and ANN have demonstrated improved efficiency and predictive accuracy across various domains such as intrusion detection [1], water quality prediction [2], road maintenance [3], to intelligent transportation systems [4], showing PCA's ability to streamline data inputs without compromising performance. However, these applications largely center on technical systems, with limited research addressing the service-oriented complexity of hotel facility management.

Previous studies have applied ANN to hotel and government building maintenance using full indicator sets, achieving high accuracy but lacking efficiency due to the absence of dimensionality reduction [5], [6]. While deep learning applications in industrial systems demonstrate that preprocessing enhances model performance [7], other scholars warn that PCA may discard important information if not supported by domain understanding [8]. Addressing this gap, this study integrates PCA and ANN to evaluate how reducing eleven hotel maintenance indicators into nine principal components affects classification accuracy, error rates, and model consistency across five maintenance categories. Although PCA–ANN has been widely explored in fields such as intrusion detection, water quality prediction, and transportation, its impact in service-oriented, multi-indicator hotel maintenance contexts remains unexamined. Given the complex qualitative–quantitative nature of hospitality maintenance data, this study provides a novel contribution by assessing how PCA-driven reduction influences ANN performance in hotel maintenance classification.

This study finds that while PCA enhances computational efficiency, it also raises misclassification rates, showing that each original indicator plays a crucial role in accurate hotel maintenance classification. In precision-driven sectors like hospitality, reducing inputs can compromise system reliability. The findings highlight the importance of applying PCA carefully and contextually, especially in dynamic environments like Yogyakarta. By examining the trade-off between efficiency and accuracy, this research supports the development of adaptive, lightweight AI systems that maintain decision quality in real-world hotel operations.

2 Method

This study uses a quantitative experimental approach to evaluate the effectiveness of Principal Component Analysis (PCA) in reducing indicators for an Artificial Neural Network (ANN) model used in hotel maintenance classification. The ANN was rebuilt using PCA-transformed data, excluding the original 11 indicators previously tested. Data were collected from 175 respondents across 48 hotels in Yogyakarta, selected via purposive sampling. The validated dataset includes 11 indicators related to hotel maintenance, such as architecture, structure, MEP systems, outdoor areas, and housekeeping. Maintenance eligibility scores were classified into five levels from Highly Unfit to Highly Fit for Preservation. Data processing employed Python with Scikit-Learn for PCA and TensorFlow/Keras for ANN development. The study also provides a mathematical overview of PCA and ANN to support the technical implementation of indicator reduction and classification modeling [5], [9].

Table 1. PCA and ANN Method Algorithm [5]

PCA		ANN
Steps / Components	**Formula/Description**	
Data Normalization	$x' = \frac{x - min(x)}{max(x) - min(x)}$ (Min-Max Scaler)	$Accuracy = \frac{Number\ of\ Correct\ Predictions}{Total\ Predictions} = \frac{TP + TN}{TP + TN + FP + FN}$
Matriks Kovarians	$C = \frac{1}{n-1}(X^T.X)$	$Precision = \frac{TP}{TP + FP}$
Eigenvalue and Eigenvector	$C.v = \lambda.v$	$Recall = \frac{TP}{TP + FN}$
Key Component Selection	Select k eigenvector with the largest eigenvalue to form the transform matrix P	$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$
Data Transformation	$Z = X.P$	Confusion Matrix

This study was conducted to evaluate the effectiveness of reducing hotel maintenance indicators using PCA on the performance of an ANN classification model. Eleven validated maintenance indicators were first preprocessed through Min-Max normalization and outlier removal before PCA was applied, yielding nine principal components that preserved 93% of the data variance an optimal threshold commonly used to maintain essential information while avoiding unnecessary model complexity. These components served as inputs to a multilayer perceptron ANN with a 9-28-44-5 architecture, selected through preliminary hyperparameter screening and learning-curve analysis to ensure stable convergence and prevent underfitting. The model employed ReLU activation in hidden layers, Softmax in the output layer, and Stochastic Gradient Descent (SGD) optimization to enhance training stability on PCA-transformed features. Performance was evaluated using accuracy, precision, recall, F1-score, and confusion matrix metrics to determine how indicator reduction influenced predictive accuracy and reliability.

3 Results and Discussion

The results of a series of analysis stages carried out in the research, starting from the dimension reduction process using PCA to the development and evaluation of the ANN model. Each stage is analyzed to find out the extent to which the reduction of indicators affects the model's performance in classifying the level of hotel maintenance and maintenance needs. The presentation of results focused on the effectiveness of data transformation through PCA, the classification performance of the ANN model, and comparison with the results of previous studies using data without reduction. This analysis serves as the basis for assessing the efficiency and accuracy of the proposed approach.

3.1 Dimensional Reduction Results with PCA

The dimension reduction process in this study was carried out using the PCA method on eleven indicators of hotel maintenance and maintenance implementation that had been determined in the previous study. The main goal of implementing PCA is to simplify the number of input variables without eliminating the important information contained in the data. The results of the analysis showed that the nine main components produced were able to maintain a cumulative variance of 93% of the total information in the eleven initial indicators. This shows that despite the reduction of the two variables, most of the data patterns and structures are still significantly represented in the components of the PCA transformation. Visualization of the results of the dimension reduction process using PCA is presented in Figure 1.

3.2 ANN Model Development and Training Results (After PCA Reduction)

After the dimension reduction process using PCA, the Artificial Neural Network (ANN) model was developed using a perceptron multilayer architecture. The number of neurons in the input layer was adjusted for the nine main components of the PCA transformation. Meanwhile, the output layer consists of five neurons representing five categories of hotel maintenance feasibility levels, namely Highly Unfit for Preservation (HUP), Unfit for Preservation (UP), Moderately Fit for Preservation (MFP), Fit for Preservation (FP) and Highly Fit for Preservation (HFP).

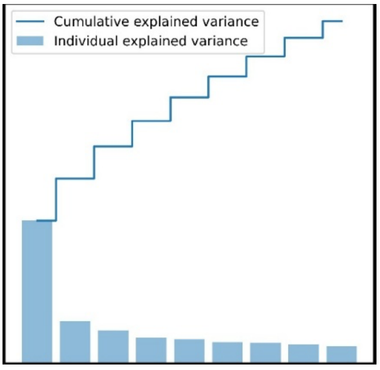

Fig 1. Dimensional Reduction Results with PCA

The ANN model in this study was developed using a multilayer perceptron architecture with two hidden layers of 28 and 44 neurons, employing ReLU activation for non-linearity and Softmax in the output layer for multi-class probability distribution. The model was trained using Stochastic Gradient Descent (SGD) and Sparse Categorical Crossentropy over 400 epochs with a batch size of 4, a configuration selected to balance model complexity with the dataset size and classification objectives. This approach produced strong performance, achieving 90.57% accuracy, a Macro-F1 Score of 0.9106, and a Weighted-F1 Score of 0.9053. However, five misclassifications were recorded among the 53 test samples, primarily in categories with closely related maintenance eligibility scores.

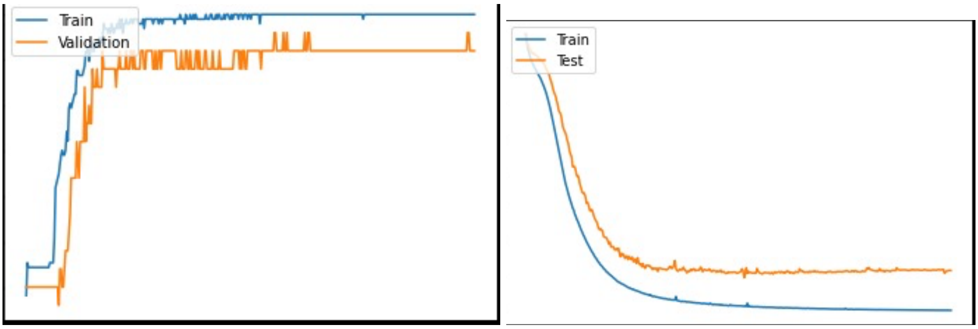

Fig 2. ANN Model Accuracy and Loss Graphs During the Training and Validation Process

The confusion matrix shows 5 misclassifications, most of which occur in the intermediate categories such as Moderately Fit for Preservation (MFP) and Unfit for Preservation (UP). This indicates that after the deduction of the indicator, the model has little difficulty distinguishing between adjacent classes in terms of score value. The results of the comparison between the ANN model using the original 11 indicators and the ANN model resulting from the reduction of indicators through PCA showed a significant difference in performance. In the previous study, the use of all indicators resulted in a classification accuracy of 94.34%, with only three misclassifications and a macro-F1 score value of 0.9464.

Training Set						
TARGET / OUTPUT	HUP	UP	MFP	FP	HFP	SUM
HUP	9 16.98%	3 5.66%	0 0.00%	0 0.00%	0 0.00%	12 75.00% 25.00%
UP	0 0.00%	10 18.87%	0 0.00%	0 0.00%	0 0.00%	10 100.00% 0.00%
MFP	0 0.00%	0 0.00%	15 28.30%	1 1.89%	0 0.00%	16 93.75% 6.25%
FP	0 0.00%	0 0.00%	1 1.89%	8 15.09%	0 0.00%	9 88.89% 11.11%
HFP	0 0.00%	0 0.00%	0 0.00%	0 0.00%	6 11.32%	6 100.00% 0.00%
SUM	9 100.00% 0.00%	13 76.92% 23.08%	16 93.75% 6.25%	9 88.89% 11.11%	6 100.00% 0.00%	48 / 53 90.57% 9.43%

Fig 3. Confusion Matrix Model ANN

The performance of the original ANN model was higher than that of the PCA-based model, which achieved 90.57% accuracy with five misclassifications. This decrease in accuracy indicates that although PCA retains 93% of the data variance, the reduction process may remove fine-grained and nonlinear relationships among the original indicators that are essential for accurate classification. Several indicators such as architectural conditions, MEP reliability, and housekeeping quality contain subtle variations that play an important role in distinguishing hotel maintenance categories. Because PCA is a linear transformation, these nuanced patterns are not fully preserved in the principal components, causing the ANN to lose access to critical cues needed for precise differentiation. As a result, misclassifications predominantly occurred in adjacent categories like MFP and UP, where the model relies heavily on the detailed interactions among the full set of eleven indicators. Details of the metric comparison between models are presented in Table 2.

Table 2. Evaluation Metrics Of The ANN Classifier For Road Maintenance Priority Classification

Accuracy	Misclassification Rate	Macro-F1	Weighted-F1
0.9057	0.0943	0.9106	0.9053

The results showed that the application of PCA succeeded in reducing eleven indicators of hotel maintenance implementation to nine main components, while maintaining 93% of the total data variance. This reduction results in efficiency in terms of the number of inputs used by the ANN model, which could theoretically speed up the training process and reduce computational complexity. However, the model's performance has decreased slightly in terms of accuracy. The ANN model based on the original 11 indicators previously recorded an accuracy of 94.34%, while the model with the input of the PCA transformation results only reached 90.57%. In addition, the number of misclassification errors increased from three to five cases, which indicates the possibility of missing important information in the reduction process, especially nonlinear patterns or minor details relevant to precise classification. To provide a more detailed overview, the performance of the classification by each eligibility category is presented in Table 3.

Table 3. Class-Wise Performance Metrics Of The ANN Model For Road Maintenance Priority Classification

Class name	Precision	1-Precision	Recall	False Negative Rate	F1 Score	Specificity (TNR)	False Positive Rate (FPR)
HUP	1.0000	0.0000	0.7500	0.2500	0.8571	1.0000	0.0000
UP	0.7692	0.2308	1.0000	0.0000	0.8696	0.9302	0.0698
MFP	0.9375	0.0625	0.9375	0.0625	0.9375	0.9730	0.0270
FP	0.8889	0.1111	0.8889	0.1111	0.8889	0.9773	0.0227
HFP	1.0000	0.0000	1.0000	0.0000	1.0000	1.0000	0.0000

This phenomenon mirrors the findings of [1] who stated that although PCA can improve the computing efficiency of ANNs, this method also risks omitting minor features that are actually significant to the classification results. The linear characteristics of PCA are one of its main limitations, PCA has limitations in detecting nonlinear relationships between variables in predictive systems. In the context of hotel maintenance indicators, which are often complex and interrelated, PCA-based reduction risks removing important correlations that ANN models rely on to accurately recognize patterns.

The findings of this study are consistent with prior research indicating that PCA, when used alone, may overlook important information needed for accurate classification. Studies in environmental and infrastructure systems show that aggressive component reduction can lower model performance, while feature selection methods such as mutual information or correlation-based filtering often produce more stable outcomes than linear techniques like PCA [2], [10]. This highlights the need to consider semantic relevance rather than relying solely on variance retention. Alternatives such as autoencoders and LASSO, which preserve nonlinear relationships and suppress less influential variables, have also been recommended [5]. In hotel maintenance decision-making, high accuracy is essential, as ignoring the relative importance of individual indicators increases the risk of misclassification. Evidence from models like LSTM and CNN further demonstrates the value of architectures capable of capturing temporal and spatial dependencies, especially when integrated with real-time IoT data [7], [11]. In this study, three of the five misclassifications occurred between closely related categories, underscoring how even minor information loss can affect outcomes. Therefore, while PCA improves efficiency, its use should be complemented with techniques that preserve critical data characteristics to develop robust and context-sensitive predictive models.

4 Conclusion

The reduction of eleven hotel maintenance indicators to nine principal components using PCA successfully retained 93% of the data variance, indicating strong structural efficiency. However, when applied to the ANN classification model, this reduction led to a performance decline, with accuracy decreasing from 94.34% to 90.57% and misclassifications increasing from three to five cases. These findings suggest that although PCA simplifies input variables, it may also remove essential information needed for distinguishing categories with closely related characteristics. Therefore, PCA alone is less suitable for high-precision classification systems, and its use should be complemented with feature selection techniques such as correlation analysis, mutual information, or autoencoders to balance efficiency and predictive integrity. For hotel managers, the results emphasize the need to preserve informative indicators within AI-based maintenance systems, particularly as future applications may integrate real-time IoT data for more adaptive decision-making. This study also acknowledges several limitations, including a dataset confined to 48 hotels in Yogyakarta, which may limit generalizability, and the use of purposive sampling that may introduce bias. Additionally, the PCA–ANN outcomes may differ in sensor-driven or dynamically changing environments. Future research should incorporate larger and more diverse datasets, cross-regional validation, and hybrid dimensionality reduction approaches that better preserve nonlinear indicator interactions.

References

1. R. P. KiranVarma and V.V. Kumari, Feature optimization and performance improvement of a multiclass intrusion detection system using PCA and ANN. Int J Comput Appl. **44**(13), 4–9 (2012). https://doi.org/10.5120/6321-8668
2. Y. Meng, S. N. Qasem, M. Shokri, Dimension reduction of machine learning-based forecasting models employing principal component analysis. Mathematics. **8**(8), 1233 (2020). https://doi.org/10.3390/math8081233
3. K. Gharehbaghi, Artificial neural network for transportation infrastructure systems, MATEC Web of Conferences. **81**, 05001 (2016). https://doi.org/10.1051/matecconf/20168105001.

4. D. A. Gomes, da Silva, M. Â. L. Moreira, J. H. d. Miranda, L. P. Fávero and M. dos. Santos, Selection of the most relevant indicators to improve data monitoring in a state-owned passenger transportation using PCA. 327–342 (2024). https://doi.org/10.1007/978-981-99-8324-7_28
5. B. G. Pratama, S. N. Sari and O. Yuliani, Classification based on artificial neural network for regency road maintenance priority. Aviation Electronics, Information Technology, Telecommunications, Electricals, and Controls (AVITEC). **7(**3), (2025)
6. D. I. Nastac and P. D. Cristea, An ANN - PCA adaptive forecasting model, in International Conference on Systems, Signals and Image Processing, https://ieeexplore.ieee.org/document/6208191/ (2012) 514–517.
7. T. Abbasi, K. H . Lim, and K. S. Yam, Predictive maintenance of oil and gas equipment using recurrent neural network, in IOP Conf Ser Mater Sci Eng. **495**, (2019) 012067. https://doi.org/10.1088/1757-899X/495/1/012067
8. L. da F. Costa, P. R. V. Boas, F. N. Silva, and F. A. Rodrigues, A pattern recognition approach to complex networks. Journal of Statistical Mechanics: Theory and Experiment. **2010** (11), P11015 (2010). https://doi.org/10.1088/1742-5468/2010/11/P11015
9. N. Navaroli, D. Turner, A. I. Concepcion, and R. S. Lynch, Performance Comparison of ADRS and PCA as a Preprocessor to ANN for Data Mining, in 2008 Eighth International Conference on Intelligent Systems Design and Applications, IEEE, (2008) 47–52. https://doi.org/10.1109/ISDA.2008.133
10. H. Sharma, G. Suri, and V. Savara, An Approach Combining DEA and ANN for hotel performance evaluation. International Journal of E-Adoption. **12**(1), 15–29 (2020). https://doi.org/10.4018/IJEA.2020010102
11. F. Mesa, G.C. Velez, and P.P.C. Alzate, Analysis of economic and social indicators through the principal components analysis. Contemporary Engineering Sciences. **11**(16), 763–769 (2018). https://doi.org/10.12988/ces.2018.8120